МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ПОЛТАВСЬКИЙ ІНСТИТУТ ЕКОНОМІКИ І ПРАВА
ВИЩОГО НАВЧАЛЬНОГО ЗАКЛАДУ
«ВІДКРИТИЙ МІЖНАРОДНИЙ УНІВЕРСИТЕТ РОЗВИТКУ ЛЮДИНИ «УКРАЇНА»

А. Т. Опря

СТАТИСТИКА

(модульний варіант з програмованою формою контролю знань)

НАВЧАЛЬНИЙ ПОСІБНИК

Рекомендовано
Міністерством освіти і науки України
для студентів вищих навчальних закладів

Київ
«Центр учбової літератури»
2012

УДК 311(075.8)
ББК 60.6я73
О–62

Гриф надано
Міністерством освіти і науки України
(Лист № 1/11–6044 від 24.07.2009 р.)

Рецензенти:
Рогоза М. Є. – доктор економічних наук, професор;
Савчук В. К. – доктор економічних наук, професор.

Опря А. Т.
О–62 Статистика (модульний варіант з програмованою формою контролю знань). Навч. посіб. – К.: Центр учбової літератури, 2012. – 448 с.

ISBN 978-611-01-0266-7

Розглядаються теми Освітньо-професійної програми (2010 р.) навчального курсу «Статистика» економічних факультетів вищих закладів освіти. Висвітлено питання математичної статистики і загальної теорії статистики в модульному варіанті з програмованою формою контролю знань: методологічні засади статистики; статистичне спостереження; зведення і групування статистичних даних; узагальнюючі статистичні показники; аналіз рідів розподілу; аналіз подібності розподілів; статистичні методи вимірювання взаємозв'язків; аналіз інтенсивності динаміки; аналіз тенденцій розвитку та коливань; індексний метод; вибірковий метод; подання статистичних даних: таблиці, графіки, карти. Додатково розглянуто науково-пізнавальні теми: перевірка статистичних гіпотез; методи багатомірного статистичного аналізу.

Для викладачів і студентів вищих навчальних закладів.

Може бути корисним аспірантам і науковцям, а також працівникам економічних служб.

УДК 311(075.8)
ББК 60.6я73

ISBN 978-611-01-0266-7

ВСТУП

Зрослі вимоги до статистики як фундаментальної навчальної дисципліни (поряд з математикою та інформатикою), а також нагальна потреба в підвищенні її наукового рівня зумовлюють необхідність перебудови і самого навчального курсу. А саме, викладання статистики повинне забезпечувати створення надійної методологічної бази поглибленого економічного аналізу з комплексним використанням традиційних статистичних і сучасних математико-статистичних методів.

Оволодіння матеріалом відповідних тем має дати студенту перші навички самостійної роботи з масовими статистичними даними і забезпечити створення теоретичної бази для наукового пошуку в курсовому та дипломному проектуванні, а з часом – і у майбутній практичній чи науковій діяльності.

При підготовці підручника враховано досвід викладання навчальних курсів «Статистика» та «Методика викладання статистики» в Національному аграрному університеті та інших вузах, наявність раніше опублікованих автором підручників, навчальних посібників і інших навчально-методичних і нормативних матеріалів.

Враховуючи відчутні психологічні перешкоди щодо широкого використання методів математичної статистики у вирішенні різного роду аналітичних задач, окремі теми підручника викладено у максимально можливому наближенні висвітлення їх прикладного значення, а при необхідності – їх математичної природи (а не математичних доведень). Особливого значення надано економічній інтерпретації результатів математичних розрахунків, що повинно сприяти творчому засвоєнню статистичних прийомів обробки економічної інформації.

Зважаючи на теоретичну важливість деяких тем і їх практичну значущість в аналізі соціально-економічних явищ, а також те, що для підготовки магістрів і аспірантів потрібно більш глибокий розгляд питань теорії, методології та практики статистичних досліджень, окремі розділи роботи викладено більш поглиблено і ширше, з висвітленням ряду питань за результатами власних досліджень автора.

При написанні підручника враховувався той факт, що галузь економічних досліджень, в якій передбачається застосування математико-статистичних методів, специфічна, а саме те, що незалежність одиниць спостереження і нормальний характер розподілу тут не є правилом. Вказана особливість значно ускладнює методи статистичної оцінки. Такі обставини зумовили детальніше викладення окремих тем навчальної програми. Серед них: вибірковий метод спостереження, дисперсійний і кореляційно-регресійний аналіз.

Підручник розрахований на програмований контроль засвоєних знань студентами в розрізі окремих тем. Автор застосував лінійну програму контролю знань, а саме той її варіант, коли студент повинен знайти правильну відповідь із кількох відповідей і перейти до наступного кадру інформації.

Форма подання програмованого матеріалу і концентрованість його змісту мають такі переваги перед посібниками та підручниками з традиційною формою викладення: 1) сприятимуть перевірці засвоєння знань по результатах відповідей (кількість правильних відповідей), що свідчить про успіхи студента в процесі вивчення кожної теми; 2) власний безперервний контроль служитиме зворотним зв'язком щодо розуміння студентом матеріалу; 3) пристосованість для самостійного вивчення курсу; 4) можливість уточнення найбільш істотних питань; 5) можливість звести широке коло питань до основних (вузлових).

Автор виражає вдячність колегам Петренко Ж.А. і Старокожко В.І. за допомогу при підготовці рукопису підручника до видання.

МОДУЛЬ 1

ТЕМА 1. МЕТОДОЛОГІЧНІ ЗАСАДИ СТАТИСТИКИ

§ 1.1. Загальне поняття статистики, її галузі

Термін «статистика» походить від латинського «status», що означає положення, стан явищ. Від кореня цього слова виникли слова «stato» (держава), «statista» (статистик, знавець держави), «statistiks» (статистика – певна сума знань, зведень про державу). Цей термін існує століття, хоч зміст його неодноразово змінювався.

У науковій літературі слово «статистика» вживають із XVIII століття за змістом як державознавство. Проте статистика почала свій розвиток значно раніше – у середині XVII століття.

Нині термін «**статистика**» вживають у кількох значеннях:

1) це – дані, які характеризують масові суспільні явища;

2) процес збирання, зберігання і обробки даних про масові суспільні явища, тобто галузь практичної діяльності, спрямованої на одержання, обробку, аналіз і видання масових даних про явища і процеси суспільного життя;

3) це – наука, яка вивчає величину, розміри і кількісну сторону масових суспільних явищ у нерозривному зв'язку з якісною стороною цих явищ, з їх соціально-економічним змістом.

Наукова система статистики складається із статистичної теорії, статистичної методології та зведених результатів статистичних досліджень.

Статистична теорія являє собою загальне вчення про розміри суспільних явищ і статистичних показників, які їх характеризують. Вона включає також вивчення зв'язків між статистичними показниками, розвитку, змін змісту і форми статистичних показників.

Статистична методологія – це сукупність статистичних методів дослідження. Вона розробляє питання збирання зведень про розміри суспільних явищ, вивчення зв'язків між величинами та динаміки, принципів і прийомів аналізу статистичних даних.

Статистична наука являє собою нерозривну єдність статистичної теорії і статистичної методології.

Зведені результати статистичних досліджень – це сукупність конкретних науково обґрунтованих статистичних даних (наприклад, показники кількості тварин за їх видами на певну дату, показники обсягу продукції тваринництва за певний рік і т. д.).

Найпоширенішим у статистичній літературі є визначення статистики як науки, сформульоване у 1945 р. на науковій нараді з питань статистики: «Статистика – самостійна суспільна наука. Вона вивчає кількісну сторону масових суспільних явищ у нерозривному зв'язку з їх якісною стороною. Статистика вивчає кількісну сторону суспільного виробництва в єдності виробничих сил і виробничих відносин, а також явищ культурного і політичного життя суспільства. Вона вивчає вплив природних і технічних факторів на кількісні зміни суспільного життя та вплив розвитку суспільного виробництва на природні умови життя суспільства».

У науковій літературі можна зустріти й іншу думку щодо визначення статистики як науки. Окремі автори вважають, що статистика не має свого предмета дослідження, а є лише своєрідним методом пізнання. Ми не погоджуємось з такою думкою, адже наук, які не мають свого предмета дослідження, не може бути. Тобто це, по суті, заперечує існування статистики як самостійної науки.

Відповідно до наукової дисципліни (статистичної науки) статистикою називають навчальну дисципліну у вищих і середніх спеціальних навчальних закладах.

Терміном «статистика» також називають сукупність цифрових зведень, які характеризують ті чи інші явища суспільного життя або їх сукупність (наприклад, статистика посівних площ, статистика урожайності тощо).

Статистикою в математичній статистиці називають деякий параметр *f*, який залежить від $x_1, x_2, x_3, ..., x_n$. Значення параметра є одним з ряду можливих. Отже, кожна статистика в цьому розумінні має свій розподіл імовірностей.

Приклад. Припустимо, x_1 – урожайність вівса у *i*-му господарстві району. Середню урожайність по всіх господарствах району можна в такому разі назвати статистикою. Якщо урожайність господарств порівняти між собою, то максимальна чи мінімальна урожайність теж будуть статистиками. Особливістю змісту терміна «статистика» в цьому розумінні є те, що тут підкреслюється наявність чіткого математичного правила (алгоритму) одержання величини параметра із сукупності спостережень.

Статистика як самостійна наука пройшла складний шлях свого становлення. У процесі розвитку в її складі виділилися: математична статистика, загальна теорія статистики, соціально-економічна статистика (у більш вузькому розумінні – соціальна і економічна статистика), галузеві статистики.

Існує градація статистик і за сферами людської діяльності, сферами обігу, диференціацією економічних наук.

Якщо розглядати перелічені вище види як систему наукових статистичних дисциплін, слід назвати і деякі нові розділи статистики, створені під впливом інтеграційних зв'язків з іншими науками, це – статистичне моделювання, статистичне прогнозування та ін.

Розглянемо коротко зміст кожного з видів статистики за традиційною схемою їх поділу.

Математична статистика – це галузь математичних знань. Вона розробляє раціональні прийоми (способи) систематизації, обробки і аналізу даних статистичних спостережень масових явищ з метою встановлення характерних для них статистичних закономірностей, використання для наукових і практичних висновків.

У математичній статистиці більшість методів обробки статистичних даних ґрунтується на імовірнісній природі цих даних. Галузь застосування таких статистичних методів обмежується вимогами, щоб явища, які досліджуються, були підпорядковані достатньо визначеним імовірнісним закономірностям. Математична статистика абстрагується від матеріального змісту масових явищ, які вона характеризує, озброює дослідника математичним апаратом.

Найважливіші розділи математичної статистики такі: статистичні ряди розподілу, оцінка параметрів розподілу, закони розподілу вибіркових характеристик, перевірка статистичних гіпотез, дисперсійний, кореляційно-регресійний, коваріаційний аналіз. Останнім часом знаходять поширене використання методи багатомірного статистичного аналізу – факторний аналіз, метод головних компонент, кластерний аналіз тощо. Названі методи математичної статистики дещо складніші від зазначених вище і потребують для їх опанування певних математичних знань. Вони також вивчаються як окремі розділи математичної статистики.

Математична статистика виконує роль основи для застосування власне математичних методів, які являють собою інструментарій статистичної науки.

Загальна теорія статистики містить принципи статистичної науки стосовно до різних сторін суспільного життя, тобто загальні правила і методи статистичного дослідження. Вона розробляє понятійний апарат статистичної науки, систему категорій, розглядає у загальному вигляді методи збирання, зведення, узагальнення і аналізу статистичних даних. Курс загальної теорії статистики побудований відповідно до стадій статистичного дослідження.

Предметом пізнання загальної теорії статистики є найбільш загальні властивості кількісних відносин соціально-економічних явиш,. В складі її вивчаються такі найважливіші розділи: статистичне спостереження, статистичне групування, середні величини. Вибіркове спостереження, ряди динаміки, індекси статистичні графіки.

Загальна теорія статистики розробляє загальні показники і методи вивчення структури явищ і змін їх у часі, закономірностей і тенденцій їх розвитку і причинно-наслідкових зв'язків між ними, а також принципи і методи статистичного моделювання і статистичного прогнозування. Показники і методи загальної теорії статистики використовуються всіма іншими галузями статистики.

Соціальна статистика – галузь статистики, яка вивчає кількісну і якісну сторони масових суспільних явищ і процесів, що відбуваються у соціальному житті, і розробляє Інтегровану систему показників здійснення соціальних процесів і явищ. Такі показники всебічно характеризують стан і розвиток соціальних умов життя, розкривають існуючі тенденції і закономірності розвитку соціальних процесів, дають повну картину устрою і способу життя людини у конкретних історичних умовах розвитку суспільства. Використовуючи статистичні методи, соціальна статистика вивчає політичну, планову й ідеологічну сторони життя, такі соціальні його аспекти, як формування особистості, сім'ї, добробут населення. До найважливіших показників соціальної статистики належать показники складу і устрою суспільства, структура і склад населення країни, рівень його освіти і культури, стан здоров'я і медичного обслуговування, зайнятість трудових ресурсів, рівень реальних доходів, споживання матеріальних благ і послуг, житлово-комунальні й побутові умови, умови праці та відпочинку тощо.

У більш вузькому розумінні соціальною статистикою називають кримінальну, клінічну, моральну, санітарну, статистику навколишнього середовища і т. ін. Умови соціальних перебудов у суспільстві вимагають відображення в показниках соціальної статистики демократизації суспільного життя, самоуправління, ринкових відносин і т. ін.

Економічна статистика як галузь єдиної статистичної науки, упираючись на положення загальної теорії статистики, вивчає кількісну сторону масових суспільних явищ і процесів у сфері матеріального виробництва з метою виявлення пропорцій тенденцій і закономірностей їх розвитку. Тобто вона кількісно характеризує дію економічних законів, досліджує обсяг, структуру і динаміку явищ, показує взаємозалежності економічних процесів з урахуванням конкретних природних та історичних умов розвитку суспільства. Економічна статистика досліджує всю економіку країни, даючи їй числову характеристику.

Об'єкт вивчення економічної статистики – процеси розширеного відтворення, його здійснення в умовах переходу до ринкових відносин і кінцеві результати для народного господарства в цілому, її предметом, як галузі практичної діяльності держави, е кількісна сторона масових економічних явищ, які в сукупності характеризують народне господарство.

Галузеві статистики вивчають показники процесу виробництва в галузях матеріального виробництва (сільському господарстві, промисловості), в галузях, де продовжується процес виробництва у сфері обігу (торгівля, зв'язок, транспорт тощо); показники роботи галузей невиробничої сфери (житлово-комунального господарства, науки, фізичної культури і спорту тощо). До галузевих статистик належать деякі розділи статистики, пов’язані із функціональним аспектом диференціації економічних наук: статистика праці, статистика фінансів. Ці статистики розвивають і доповнюють методи і систему показників, розроблених загальною теорією статистики і економічною статистикою стосовно до особливостей конкретних галузей.

Як наслідок інтеграції статистики з інформатикою і кібернетикою виникає розділ статистики – теорія автоматизованої статистичної інформаційної системи. У результаті інтеграції статистики з математичною статистикою і теорією ймовірностей намітився розділ статистики – статистичне моделювання і прогнозування.

§ 1.2. Статистичні сукупності

Вивчення статистичною наукою масових суспільних явищ означає, що статистичні показники завжди є наслідком узагальнення деякої сукупності фактів. Поняття сукупності у статистиці має дуже важливе значення. Під **статистичною сукупністю** розуміють масу однорідних у певному відношенні елементів (явищ, фактів і т. ін.), які мають єдину якісну основу, але різняться між собою за певними ознаками.

Під одноякісністю, або однорідністю, розуміють підпорядкованість елементів, що складають сукупність, загальному закону розвитку або їх закону розвитку або їх однотиповість (наприклад, Інформація про сукупність одиниць господарств, малих підприємств; про сукупність одиниць виробленої ними продукції і т. ін.).

Статистична сукупність складається з окремих одиниць (наприклад, у конкретному малому підприємстві є зведення про вид, вантажопідйомність, кількість днів роботи, витрати на ремонт по кожному автомобілю). Такі окремі первинні елементи, або індивідуальні явища, які складають статистичну сукупність, називають **одиницями сукупності**.

Залежно від мети досліджень однорідність сукупності можна вивчати у різноманітних аспектах розвитку. Так, по молочному стаду корів у господарстві – це породний склад, продуктивність, класність, захворюваність тощо.

Повне уявлення про статистичні сукупності можна мати лише при досконалому вивченні їх ознак. У навчальній літературі найбільш вдало ці питання розкриті в навчальному посібнику І. П. Суслова[1]. Розглянемо це питання в запропонованій автором послідовності.

Статистичні сукупності у сфері суспільного життя можна поділити на дві групи:

сукупності, створені самим життям, які утворюють єдність незалежно від того, чи підлягають вони вивченню статистикою (наприклад, вивчення у господарстві сукупності робітників за освітою, віком, участю у громадській роботі тощо);

сукупності, утворені спеціально з метою статистичного аналізу (наприклад, сукупності підприємств за видами їх комерційної діяльності, за чисельністю в них кваліфікованих робітників, кількістю унікальних видів вироблюваної продукції і т. ін.).

Як бачимо, сукупностям першої і другої груп притаманна однорідність, але в першій з них одиниці тісно і реально взаємопов'язані, а у другій групі досліджувані сукупності можуть бути безпосередньо не пов'язаними між собою.

---

[1] Суслов Й. П. Общая теория статистики. – М.: Статистика, 1978. – 250 с.

Таким чином, одиниці статистичної сукупності, утворюючи разом дещо ціле, за рядом своїх властивостей і особливостей різняться між собою. Тобто важливою рисою статистичної сукупності є наявність варіації між її ознаками. Вивчення статистичної сукупності на основі цієї варіації становить важливе завдання статистичної науки.

У формуванні однорідної статистичної сукупності вабливе місце відводиться застосуванню математико-статистичних критеріїв. Створення однорідної статистичної сукупності вважається вихідним (і важливим) моментом в організації обґрунтованого застосування апарату математичної статистики у соціально-економічних дослідженнях. Тут вирішальне значення має з'ясування специфіки формування однорідної статистичної сукупності у просторовому і часовому розрізах. В умовах, коли рівень забезпеченості технічними засобами недостатній (останнім часом спостерігається занепад роботи машинолічильних станцій і центрів на районному рівні), статистичні сукупності формують малі за обсягом з наміром подальшого переходу до великих за обсягом. Наприклад, по підсумовуючих показниках у малій сукупності одержують нові підсумки за цими ж показниками в більшій за обсягом сукупності. Але такий підхід знижує аналітичні можливості у використанні статистичної інформації.

Вирішення цього питання можливе лише через створення автоматизованих банків даних на всіх рівнях управління народним господарством країни (район, область, країна). У цьому зв'язку особливого значення набувають правила і принципи визначення статистичних характеристик сукупностей, утворених па підставі об'єднання підсукупностей. Серед таких характеристик слід назвати загальну середню, одержану на підставі групових середніх. У таких умовах об'єднання статистичних сукупностей знадобиться знання правила складання і розкладання дисперсій (цим моментам присвячені спеціальні розділи підручника). На підставі цього правила виникає нове правило поєднання коефіцієнтів варіації, рівнянь і коефіцієнтів кореляції, рядів динаміки, індексів помилок вибірки і ряд інших методичних нюансів.

Наукові дослідження у цьому напрямі дали змогу знайти загальний механізм співвідношення загальних статистичних характеристик (по всій сукупності) з локальними, одержаними по частинах всієї статистичної сукупності. З’ясування механізму цих співвідношень принципово важливе для розвитку ряду понять і категорій статистичної теорії. Зокрема, у даному випадку діє така схема утворення статистичної сукупності: від більшої за обсягом сукупності до меншої і далі – до окремого її елемента.

Таким чином, формування статистичної сукупності передбачає реалізацію одночасно діючих, протилежних один одному прийомів: об'єднання і роз'єднання елементів і частин статистичної сукупності.

Виникає запитання: навіщо потрібні такі операції у формуваннях сукупностей? Відповідь міститься у постановці і формулюванні таких завдань:

1) за встановленими правилами па підставі локальних статистичних характеристик визначити загальні характеристики;

2) виходячи із загальних статистичних характеристик сукупності, на підставі заданих критеріїв знайти локальні статистичні характеристики.

Це все свідчить про важливість категорії «статистична сукупність», адже особливості законів розвитку суспільних явищ вимагають статистичних методів пізнання досліджуваних сукупностей, а шлях цей досить складний і пролягає через методи, які розробляє статистична наука.

§ 1.3. Предмет статистки

1.3.1. Предмет статистики як суспільної науки

Визначити предмет будь-якої науки – означає вирішити питання про її зміст і місце серед інших наук, а також характер взаємовідносин з ними. Питання визначення предмета (також змісту і структури) статистичної теорії суспільних явищ і процесів являє собою об'єкт пожвавлених дискусій протягом усієї історії розвитку статистики. Актуальним воно залишається і на сучасному етапі розвитку статистичної науки. Управління народним господарством в умовах переходу до ринкових відносин висуває нові вимоги до статистико-економічного аналізу суспільних явищ, розкриття специфічних тенденцій і закономірностей у змінах цих явищ, зумовлених новими формами і методами господарювання. Слід враховувати також тенденції і закономірності в розвитку статистичної науки. Перш за все це загальний процес «математизації» всіх наук і статистичної зокрема. Істотних змін набувають організація і форми проведення статистичного спостереження у зв'язку з широкою автоматизацією статистичних розрахунків.

Питання про зміст статистичної науки (а звідси і про її предмет) вважається одним із складних. Для вірного його розуміння треба виходити насамперед із висловленої вище думки про співвідношення якісного і кількісного аналізу суспільних явищ.

Предметом статистики є розміри і кількісні співвідношення масових суспільних явищ, закономірності їх формування, розвитку та взаємозв’язку.

Отже, предметом дослідження статистики є масові явища соціально-економічного життя; вона вивчає кількісну сторону цих явищ у нерозривному зв’язку з їх якісним змістом у конкретних умовах простору та часу.

Важливо відзначити, що особливістю статистики є те, що вона вивчає масові суспільні явища, а це означає, що в статистичних показниках узагальнюються ті чи інші факти досліджуваних сукупностей. У своїх показниках статистика, характеризуючи конкретну міру явищ, встановлює загальні властивості, виявляє відмінності окремих рис, об’єднує окремі елементи в однорідні групи, виявляє певні типи досліджуваних явищ.

Масовому явищу притаманна особливість наявності в ньому множини елементів схожих за істотними властивостями. Наявність певної властивості у будь-якого поодинокого елемента є випадковістю. Об’єднання таких елементів у єдине ціле нівелює дію випадковості, тобто дає результат практично незалежний від випадку.

Спираючись на данні масових суспільних явищ, статистика за допомогою кількісних характеристик показує ступінь розвитку явищ (чи процесів), напрямки та темпи їх змін, характер (тісноту) взаємозв’язку та взаємозалежності.

Широке впровадження математико-статистичних методів в економічний аналіз, автоматизація процесів збирання, зберігання, передачі і обробки статистичної інформації зумовлюють нові підходи щодо визначення предмета, змісту і структури статистичної теорії суспільних явищ і процесів.

Найбільш вдало (на нашу думку) зазначене питання висвітлене вченим-статистиком В. І. Сіськовим. Його принциповий підхід до розробки змісту статистичної теорії і визначення її предмета ґрунтується на співвідношенні якісного і кількісного аналізу. Виходячи з цієї принципової позиції, розглянемо зміст теорії статистики, зокрема соціально-економічної, за такою логічно-структурною схемою: 1 – принцип побудови системи статистичних показників; 2 – класифікація і зміст типів статистичних задач; 3 – поняття і принципи формування однорідної статистичної сукупності; 4 – система статистичних методів аналізу; 5 – зміст і організація статистичної інформації.

Наявність у підручниках і навчальних посібниках переважно питань статистичної методології слід вважати об'єктивно необхідним, але коли мова йде про предмет статистики, то мається на увазі статистика не стільки як навчальна дисципліна, скільки як наука, завданням якої є пізнання кількісних сторін масових суспільних явищ у конкретних умовах простору і часу. Навчальний предмет «Статистика» включає як загальні закони для всіх масових явищ, так і статистичний метод. Але коли вивчають певний масовий процес, загальні закони масових явищ дають основу для методу його дослідження. Звідси можна зробити висновок, що статистика як навчальний предмет зводиться до вивчення в основному статистичного методу.

У науковій статистичній літературі е різні погляди відносно визначення предмета статистики як науки, Окремі автори навіть не визнають наявність предмета у статистиці, вважаючи її наукою методологічною. Безпідставність такої точки зору очевидна, адже немає безпредметних наук і самі методи як вже зазначалось, зумовлені її предметом. Прихильники цієї точки зору забувають про існування економічної статистики, галузевих статистик та статистик сфер діяльності, тобто статистик, пов'язаних з практичною діяльністю.

Переважна більшість вчених-статистиків є прихильниками визначення статистики як суспільної науки, а звідси і визначення її предмета згідно із концепціями, викладеними вище. Слід визнати аксеоматичним, що оскільки об'єктом вивчення статистики є суспільство, її відносять до суспільних наук. Але на відміну від інших суспільних наук (останні мають один об'єкт і предмет) статистика має специфічний предмет. Наприклад, якщо політекономія пізнає суспільство за допомогою таких категорій, як товар, гроші, вартість, кредит тощо, у статистиці в ролі основних категорій виступають узагальнюючі статистичні показники.

1.3.2. Предмет математичної статистики, її місце в системі статистичних дисциплін

В умовах широкого застосування методів сучасної математики в усіх галузях наукових досліджень, фундаментальних і прикладних, а також у вирішенні ряду практичних проблем суспільного життя увага надається математичній статистиці. Як галузь математичних знань вона, базується на теорії ймовірності і є наукою про методи умовиводу щодо властивостей дослідженої статистичної сукупності. Математична статистика гармонічно поєднана з загальною науковою методологією, з інтерпретацією явищ з позицій їх діалектичного розвитку та з особливими методами спеціальних галузей статистичної науки. Пропонуючи свою математичну техніку стосовно до імовірнісного характеру досліджуваних явищ і процесів, вона стає методом по відношенню до спеціальних наук, в яких вона застосовується. Її математичний апарат плідно використовується при вивченні явищ і процесів, що відбуваються в житті суспільства.

У даний час математична статистика знаходить широке застосування в економіці різних галузей народного господарства, біології, фізиці, хімії, медицині та ін. На основі її методів можна вирішувати і багато аналітичних задач в галузі економіки. Зокрема, кількісні характеристики, одержані в результаті математико-статистичного аналізу, дозволяють мати більш глибоке уявлення про характер причинно-наслідкових зв'язків явищ, а також одержати стійкі надійні параметри для здійснення економічних розрахунків і особливо з метою прогнозування.

Впровадження комп’ютерних технологій створює реальні можливості широкого впровадження методів математичної статистики для розв’язання різного роду економічних задач.

Як окрема наукова дисципліна математична статистика визначилася в другій половині XIX сторіччя. До цього часу значні успіхи вже були досягнуті в теорії ймовірностей, яка дала теоретичну основу для математичної статистики. Поштовхом до її розвитку були експериментальні дослідження в галузі наук про природу.

Як самостійна наукова дисципліна математична статистка є складовим елементом статистичної науки взагалі, її специфічним методом дослідження. Відповідаючи на питання різних наук, математична статистика сформувалася в найбагатший арсенал математико-статистичних прийомів обробки емпіричних даних. При цьому гармонічно поєднуються її наукова методологія, інтерпретація явищ на основі філософії і специфічні методи статистичної науки.

Сучасні наукові дослідження характеризуються розвитком і взаємопроникненням різних наук. Математична статистика виступає як наукова дисципліна, що передбачає свій метод по відношенню до спеціальних наук. Але не слід ототожнювати поняття «статистичні методи» і «математико-статистичні методи». Під останніми розуміють методи, які безпосередньо пов'язані з імовірнісною оцінкою результатів статистичного спостереження, що передбачають визначення величини математичної імовірності. Іншими словами, якщо мова йде не про описуючі функції статистики, то в усякому статистичному висновку визначальним початком є теорія ймовірностей. Математична статистика з її імовірнісною методологією створює (породжує) навіть деякі галузі теоретичних знань, відокремлені від певних наук. Наприклад, у фізиці та механіці (там, де мова йде про кількісне вивчення речовини) математична статистика інколи виявляється єдиним інструментом пізнання. У вивченні суспільних явищ вона створює математичний апарат соціально-економічної статистики, яка лише при його наявності здатна проникнути у специфічну природу досліджуваних явищ і процесів.

Відомий російський вчений статистик-математик професор Є.Є. Слуцький ще у 1912 році писав про значимість методів математичної статистики: «Ми приходимо, таким чином до кардинальної вимоги, яку життя ставить діячам статистики: статистик повинен бути математиком, бо його наука є наука математична»[2].

Математична статистика - основа для застосування власне математичних методів. Що ж складає її предмет?

Предмет математичної статистики – це формальна математична сторона статистичних методів дослідження, байдужа до специфічної природи об'єктів, які вивчаються. Згідно з цим визначенням предмет математичної статистики є чисто математичною теорією математико-статистичних методів незалежно від специфіки і сфери їх застосування.

Як відмічалося раніше, математична статистика – самостійна наукова дисципліна, заснована на теорії ймовірностей, остання є її теоретичною базою. Характеризуючи математичну статистику як

[2] Слуцький Є.Є.Теорія кореляції і елементи вчення про криві розподілу // Відомості Київського комерційного інституту. – К., 1912. Кн. XVI. с. 2

науку, яка займається розробкою методів одержання, опису і обробки даних статистичних спостережень з метою вивчення закономірностей випадкових масових явищ, відзначимо, що категорія «статистична закономірність» нерозривно пов'язана з основною категорією теорії статистики – «статистична сукупність». Оскільки в наступному викладі різних питань ця категорія буде зустрічатися часто, дамо їй визначення.

Статистична сукупність – це сукупність однорідних об'єктів чи явищ, об'єднаних за певними ознаками у єдине ціле. Окремі одиниці статистичних сукупностей відрізняються між собою. У цьому зв'язку виникає необхідність застосування деяких описуваних параметрів статистичної сукупності – статистичних характеристик. Назвемо найважливіші з них**:** середні величини (арифметична, гармонійна, геометрична, квадратична, мода, медіана та ін.) міри варіації (дисперсія, середнє квадратичне відхилення та ін.), міри асиметрії, моменти сукупності і ряд інших. Ці питання будуть предметом розгляду спеціальних тем.

Завдання математичної статистики в цілому пов'язані з вирішенням питань обробки даних статистичних спостережень. Залежно від характеру статистичного виміру явищ які вивчаються і мети аналізу їм надають тієї чи іншої форми.

Основні завдання математичної статистики, що є практично найважливішими можна об'єднати, виділивши три великі групи (категорії):

1. Встановлення законів розподілу різних випадкових змінних, одержаних у результаті статистичного спостереження. Оскільки аналізована сукупність являє собою не повний обсяг даних (вибірку і результати, одержані на їх основі, обумовлені елементами випадковості), потрібно знати, які саме риси досліджуваних явищ є стійкими, а які – випадковими, адже дані беруться не в повному обсязі. Вирішення цієї задачі можливе за умови правильного вибору методів обробки даних.

Методи, що використовуються, повинні встановити і зберегти характерні (типові) риси явища, і елімінуватися. Яке вивчається, несуттєві і другорядні властивості при цьому повинно. Отже, категорія завдань математичної статистики включає певну систему методів систематизації і перетворення даних статистичного спостереження. Математично ця задача може бути сформульована так: у результаті незалежних випробувань випадкової змінної величини - x одержано її значення: $x_1, x_2, x_3, \ldots, x_n$. Потрібно приблизно оцінити невідому функцію розподілу f(x) випадкової величини x .

2. Друга категорія завдань, які вирішуються математичною статистикою, – це **перевірка статистичних гіпотез,** яка є мовби логічним продовженням попередньої. Зокрема, маючи визначену сукупність даних (як правило, невелику за обсягом), дослідник зобов'язаний висунути ту чи іншу гіпотезу про характер закономірності, що притаманний явищу, яке вивчається Висунуту гіпотезу необхідно перевірити. Так, треба з'ясувати чи підтверджують дані спостереження гіпотезу про те, що середня їх величина дорівнюватиме відповідний середній для всієї сукупності, із якої проведена вибірка.

Одне з основних завдань цієї категорії – перевірка гіпотез відносно законів розподілу, тобто чи підтверджують дані вибірки гіпотезу про те, що досліджуване явище підпорядковане закону нормального розподілу (чи будь-якому іншому закону). У математичній постановці ця задачі може бути сформульована так: на підставі деяких передбачень можна вважати, що функція розподілу досліджуваної випадкової змінної величини x є f(x). Чи збігаються спостережувані значення з гіпотезою, якщо випадкова величина x дійсно має розподіл f(x)?

3. До третьої категорії завдань, які вирішує математична статистика, відноситься **оцінка невідомих параметрів різних розподілів.** Необхідність рішення цього роду завдань випливає з таких міркувань. Оскільки дослідник має справу не з усією сукупністю одиниць явища, яке вивчається, а тільки з їх частиною (вибіркою), рівень деяких статистичних характеристик для всієї сукупності залишається невідомим (наприклад, середня, дисперсія та ін.). У цьому випадку варто застосувати специфічний метод для оцінки тієї чи іншої характеристики, одержаної за даними вибіркового спостереження. Математична статистика має у своєму розпорядженні цілий арсенал методів для вирішення задач оцінок розподілу, а також їхньої надійності (точності).

У математичній постановці задача оцінки невідомих параметрів розподілу може бути сформульована так: випадкова змінна x має функцію розподілу певного виду, зумовлену деякими параметрами, з невідомими значеннями. За даними спостереження, необхідно знайти оцінки цих параметрів.

§ 1.4. Метод статистики

Статистична методологія являє собою сукупність прийомів, правил і методів дослідження. Під терміном «метод» розуміють спосіб теоретичного дослідження або практичного здійснення чого-небудь (наприклад, філософський метод, передовий метод). Коли мова йде про метод науки взагалі, то мають на увазі найбільш загальні способи підходу до вивчення будь-яких явищ. Під методом конкретної науки розуміють окремі специфічні прийоми і методи, пристосовані до дослідження її предмета. Загальнонауковим вважається діалектичний метод. Керуючись його принципами, статистика розробляє свої специфічні прийоми і методи, а також відпрацьовує таку систему показників, яка дозволяє вивчати кількісні сторони суспільних явищ і процесів.

Особливість (специфіка) статистичних методів полягає в їх комплексності, що зумовлено як різноманітністю форм статистичних закономірностей, так і складністю самого процесу статистичного дослідження. Специфіка методів пояснюється змістом виконуваної роботи у процесі дослідження тих чи інших соціально-економічних явищ. Природа останніх досить складна і непередбачена, тому вивчати їх треба у взаємозв'язку і взаємообумовленості. Для цього статистична теорія розробила досить широке коло методологічних І методичних засобів, які дозволяють кількісно вимірювати досліджувані зв'язки (особливо причинно-наслідкові).

Філософський підхід як методологічна основа статистичної науки вимагає розгляду явищ і процесів в їх русі, постійних змінах і розвитку. Із цією метою розроблена відповідна система показників, які дають змогу охарактеризувати варіації змін рівнів явищ, визначити тенденції і закономірності їх розвитку.

Важливим моментом філософського підходу до вивчення суспільних явищ є визначення границь переходу кількісних явищ у якісні форми їх прояву. Наприклад, в умовах переходу до ринкової економіки встановлення об'єктивно необхідних кількісних границь, які б дозволили відокремити якісний стан явищ (зокрема, це форми організації праці, види і форми підприємницької діяльності тощо), можна вирішити статистичними методами (групування, дисперсійний метод та ін.).

Філософська основа статистики знаходить своє втілення в її специфічних методах. Залежно від пристосування останніх до тих чи інших етапів статистичного дослідження прийнята така їх класифікація за стадіями дослідження.

1. Методи *статистичного спостереження*. Вони являють собою перший етап статистичного дослідження і виконують функції у збиранні і оцінці якості первинних статистичних даних. У зв'язку з масовим характером даних застосовується метод масового статистичного спостереження. Під останнім розуміють спостереження над множиною елементів, які складають статистичну сукупність. Вивчаючи такі сукупності за допомогою масових спостережень, статистика викриває притаманні їм загальні риси, процеси, закономірності. Робота ця досить копітка, складна і тому потребує наукової організації. До масових спостережень належать обстеження, збирання звітності, переписи і т. д.

2. Методи зведення і групування первинного статистичного матеріалу. Зведення включає методи перевірки, систематизації, обробки, підсумовування даних і представлення їх у формі статистичних таблиць.

Зведення забезпечує систематизацію первинної інформації, підрахунок чисельності одиниць сукупності і об'єму ознак, що їх характеризують. Важливим етапом цієї стадії дослідження є розподіл інформації (розчленування) за ознаками її відмінності, тобто групування статистичних даних. Специфіка методів групування зумовлюється завданнями дослідження і якістю первинної інформації.

Методи групування дозволяють розділити досліджувану сукупність на однорідні в певному відношенні частини. Наприклад, розчленування сукупності господарств певного регіону за формою їх виробничої діяльності (колективні, орендні, кооперативні тощо).

3. Методи (прийоми) визначення узагальнюючих зведених синтетичних показників. Вони становлять третю стадію статистичного дослідження і вирішують завдання визначення певних параметрів.

Стадія узагальнення і аналізу зведеного матеріалу передбачає виявлення характерних властивостей і закономірностей соціально-економічних явищ, взаємозалежностей факторних та результативних ознак і т. ін. На цьому етапі використовується весь арсенал статистичних прийомів і методів дослідження, розраховуються такі узагальнюючі статистичні показники: абсолютні, відносні й середні величнім. Окремі загальні риси формування названих вище показників визначають шляхом виміру їх варіації. Вимірювання варіації спеціальними прийомами дає змогу одержати характеристику умов, в яких здійснюються масові процеси, для цього досліджують закони розподілу.

Тенденції і закономірності у русі показників у часовому розрізі вивчають дослідженням рядів динаміки, спеціальних (математичних) прийомів їх обробки і моделювання. Щодо складних економічних явищ – для їх вивчення використовують індексний метод аналізу, який дозволяє синтезувати безпосередньо невимірювані величини.

Причинно-наслідкові зв'язки соціально-економічних явищ досліджуються методами аналітичних групувань, кореляційно-регресійного і дисперсійного аналізу. Значно розширюються аналітичні можливості статистичних методів при умілому використанні статистичних графіків (графо-аналітичиий метод), номограм, табличного, балансового і багатомірного методів аналізу.

Зазначені методи і прийоми досліджень та принципи їх використання у статистико-економічному аналізі являють собою предмет курсу загальної теорії статистики і математичної статистики.

Щодо теоретичних аспектів питання про статистичний метод, то слід відзначити, що він має ту ж саму основу, що й форми наукового методу, в яких застосовуються індукція і дедукція. Дослідник спочатку здійснює спостереження, а потім шляхом експерименту (чи аналізу) переходить за допомогою індукції до побудови теорії. Побудувавши теорію, він робить за допомогою дедукції передбачення результатів наступних експериментів, які в інших умовах передбачити було б неможливо. Потім збирають фактичні дані, які підтверджують або спростовують правильність передбачень, і процес продовжується шляхом розвитку основної теорії або побудови нової. У статистичних дослідженнях наведена спрощена схема реального процесу має свої особливості, адже статистика має справу переважно з невизначеними ситуаціями. Якщо ж ступінь невизначеності досить значний, результати матимуть обмежену сферу застосування. Наявність невизначеності ситуацій зумовлює в статистичному аналізі нескінченні труднощі. Тому вчені-статистики постійно працюють над винайденням нових або видозміною (удосконаленням) існуючих методів. Нові вимоги до економіки орієнтують на застосування досить складних і витончених методів статистики, зокрема методів математичної статистики.

Значення математики для розвитку статистики зросло останнім часом у зв’язку з широким використанням електронної техніки. Але це не означає, що використання складного математичного апарату може перетворити статистику в математику.остання досліджує просторові форми та кількісні співвідношення реального світу у «чистому вигляді», тоді як статистика вивчає матеріальний зміст явищ, використовуючи математику як інструмент дослідження. Це дозволяє упорядкувати інформаційні масиви, прискорити їх обробку, передачу та зберігання, а також перевести фактичні дані на формалізовану мову для здійснення масових розрахунків.

Зараз ведеться досить велика дослідницька робота щодо вивчення самого методу статистики. Тобто статистична теорія ще не є чимось закінченим і не може давати задовільних рішень усіх без винятку проблем. У статистиці, як і в інших науках, ще існує багато невідомого, але не буде перебільшенням сказати, що управління сучасною економікою було б неможливим без допомоги статистики. Статистика сприяє економічному прогресу в, коли її використовують як інструмент дослідницької роботи, вона є також важливим фактором наукових досліджень у цілому.

Статистичні дослідження завжди цікаві, навіть якщо вони і не доведені до кінця накресленого дослідником шляху (в науці це інколи трапляється), але на самому шляху досліджень пізнається дуже багато.

Майбутнє статистики як науки очевидне, хоча треба завжди пам'ятати, що будь-якій науці притаманні помилки, а тому вона завжди потребує удосконалення. Помилки, які зустрічаються в статистиці, найчастіше виникають не з причин недосконалості статистичної науки, а через невміле (неправильне) застосування статистичних методів.

ТЕМА 2. СТАТИСТИЧНЕ СПОСТЕРЕЖЕННЯ

§ 2.1. Поняття статистичного спостереження, основні вимоги щодо його здійснення

Щоб одержати інформацію про стан і розвиток економіки країни чи інші дані, що характеризують культурний і матеріальний рівень суспільства, здійснюють статистичне дослідження. Останнє складається з трьох послідовних етапів: статистичне спостереження зведення та групування зібраних матеріалів і аналіз результатів зведення.

Статистичне спостереження виступає як один із головних методів статистики і як одна з найважливіших стадій статистичного дослідження.

Статистичне спостереження – це планомірний, науково організований збір даних про явища і процеси суспільного життя шляхом реєстрації по заздалегідь розробленій програмі спостереження. У процесі статистичного спостереження одержують первинну статистичну інформацію, яка потрібна для здійснення функцій статистики.

Так, при статистичному спостереженні, наприклад, аграрних орендних підприємств в області реєструються дані про їх число, склад працівників, вироблювану продукцію, розподіл доходів, поголів'я худоби, розмір посівних площ і т. ін. Або інший приклад: при вивченні деяких явищ суспільного життя виявляється недостатньою наявність даних обліку і звітності, оскільки вони не завжди можуть дати повну і точну картину будь-якого явища чи процесу. Візьмемо, наприклад, визначення чисельності і складу населення в країні. Це питання має велике державне значення. А між тим, звітності, яка б давала точну відповідь щодо чисельності і складу населення, немає. Тому виникає потреба в організації спеціального спостереження. Або такий приклад: рівень цін на сільськогосподарському ринку. Спостереження за рівнем цін – питання державної політики і становить великий інтерес для економічного аналізу. В той же час необхідної звітності, по цьому питанню не існує.

У всіх подібних випадках здійснюється спеціально організоване статистичне спостереження. Останнє вважається фундаментом статистичного дослідження, адже у процесі його здійснення формується інформація, яка на наступних етапах дослідження підлягає обробці і аналізу. Інформація статистичного спостереження повинна бути об'єктивною і якісною, а отже, забезпечуватись правильною науковою організацією її одержання, належним виконанням самого спостереження.

Завдання статистичного спостереження зумовлюється завданнями, які ставляться перед дослідженням певних процесів і явищ і випливають з потреб управління ними. Суть їх полягає в одержанні у найкоротший строк повної і вірогідної інформації про досліджувані факти. Тобто найважливішим завданням статистичного спостереження є вірогідне, об'єктивне відображення спостережуваних (досліджуваних) явищ і процесів суспільного життя. Завдання статистичного спостереження (як і мету) слід чітко формулювати відповідно до результатів дослідження та з урахуванням об'єкта спостереження.

Наукова організація статистичного спостереження зумовлює дотримання певних вимог щодо його здійснення. Назвемо їх.

1. Явища, які підлягають спостереженню, повинні мати певне народногосподарське значення, а також наукову чи практичну цінність.

2. Оскільки суспільні явища знаходяться у постійній зміні й розвитку та мають різний якісний стан, статистичне спостереження повинне забезпечувати збір масових даних. в яких відбивається вся сукупність фактів. Неповнота зведень про досліджувані процеси призведе до помилкових висновків з результатів аналізу.

3. Складний взаємозв'язок і взаємопереплетіння економічних явищ зумовлює орієнтацію статистичного спостереження на збирання не тільки інформації, яка безпосередньо характеризує досліджуваний об'єкт, а й такої, що сприяє зміні його стану. Отже, дані спостереження повинні бути повними. Під повнотою даних розуміють повноту просторового охоплення одиниць досліджуваної сукупності, істотних сторін явищ, а також повноту охоплення у часі.

4. Інформація, одержувана за результатами статистичного спостереження, повинна бути вірогідною. Тобто спостережувані дані підлягають ретельній і всебічній перевірці з боку їх якості. Особливість зазначеної вимоги полягає у тому, що у разі одержання недостовірної інформації, не можна усунути її дефекти в процесі подальшої обробки, що ускладнює прийняття науково обґрунтованих рішень. Зрозуміло, що статистична інформація вважається якісною, якщо вона правдива, вірогідна і точна.

5. Статистичне спостереження здійснюється на науковій основі по заздалегідь розробленій програмі, яка забезпечує науковий підхід до вирішення методологічних і організаційних питань.

6. Дані статистичного спостереження повинні бути, порівнювані. Лише в такому разі забезпечується їх узагальнення і зіставлення у просторі і часі.

У випадках, коли статистична інформація необхідна для здійснення управлінських функцій, до неї ставиться така вимога, як своєчасність. Зрозуміло, що статистичні дані, навіть якщо вони достатньо точні (чи вірогідні), але надходять несвоєчасно, не можуть бути використані для прийняття управлінських рішень.

Крім стислості запитань, одна з важливих вимог полягає у дотриманні їх послідовності, оскільки певні відповіді повинні контролювати одна одну. Наприклад, при проведенні перепису населення у формулярі не випадково стоїть запитання про вік раніше, ніж про освіту, заняття, джерела засобів існування. Відомості про вік контролюють правильність відповідей.

Наукова організація статистичного спостереження передбачає визначення об'єкта і одиниці спостереження, розробку і відпрацювання програми. Здійснюється статистичне спостереження відповідно до організаційного плану його проведення.

Об'єкт статистичного спостереження – це сукупність суспільних явищ і процесів, які підлягають статистичному спостереженню. Наприклад, при вивченні сільського господарства об'єктом спостереження є сукупність сільськогосподарських підприємств.

Виділення об'єкта спостереження – завдання, як правило, складне і відповідальне. Масові суспільні явища і процеси наділені багатьма властивостями, вони тісно пов'язані між собою. Тому виділення об'єкта дослідження повинне ґрунтуватися на наукових принципах його визначення. Останнє має давати підстави для відокремлення даного об'єкта від суміжних з ним об'єктів, які являють собою предмет самостійного дослідження. Визначення об'єкта статистичного спостереження повинне мати точні вказівки щодо його рис і властивостей.

Наприклад, буде недостатнім вказати, що спостереженню підлягає сукупність сільськогосподарських підприємств. Виділення сільськогосподарських підприємств як об'єкта статистичного спостереження потребує точного встановлення системи ознак сільськогосподарського підприємства. Залежно від завдань спостереження такими ознаками можуть бути: форма власності, виробничий напрям, рівень технічного оснащення, організаційні форми господарювання і т. ін. Точне визначення об'єкта спостереження необхідне для одержання зіставних даних, уникнення можливих випадків подвійного обліку окремих фактів або пропуску певної категорії його елементів.

Для об'єкта статистичного спостереження характерне те, що його не можна вивчати безпосередньо у цілому, для цього потрібне виділення в його складі окремих одиниць.

Одиниця статистичного спостереження – це складовий елемент об'єкта дослідження, який є основою рахунку і носієм істотних ознак і властивостей, які підлягають реєстрації. Це первинний елемент об'єкта дослідження. Одиницю спостереження встановлюють виходячи із завдань спостереження і складності об'єкта дослідження.

Тому в кожному конкретному статистичному дослідженні масових фактів приймається одна або декілька одиниць спостереження. Так, при перепису населення одиницею спостереження, як правило, є людина. Але якщо дослідженню підлягає і сім'я, то в цьому випадку встановлюються вже дві одиниці спостереження: людина і сім'я.

При статистичному дослідженні галузі сільського господарства в різних випадках і залежно від завдань дослідження можуть бути прийняті різні одиниці спостереження. Наприклад, при вивченні продуктивності праці та її оплати одиницею спостереження буде окремий працівник: при вивченні структури сільського господарства за розмірами підприємств одиницею спостереження буде кожне окреме підприємство, тобто відокремлена адміністративно-господарська одиниця. До останньої належать колективні сільськогосподарські підприємства, держгоспи, орендні підприємства, фермерські господарства, підсобні господарства працівників аграрного сектора та ін.

Правильне визначення одиниці спостереження має істотне значення для організації і проведення статистичного дослідження. Цим значною мірою зумовлюється об'єктивність одержаних результатів.

Таким чином, визначення об'єкта і одиниці статистичного спостереження має ґрунтуватися на наукових принципах – це треба добре уявляти кожному, хто бере участь в його організації і проведенні.

§ 2.2. Програма статистичного спостереження

Програма статистичного спостереження являє собою перелік питань, на які треба одержати відповіді в процесі збирання статистичних зведень щодо кожної досліджуваної одиниці. Один і той самий об'єкт може бути обстежений з різних боків. Тому склад і зміст питань програми спостереження залежить від завдань дослідження і особливостей об'єкта. Вона повинна охоплювати широке і повне коло відомостей. Чим ширша програма, тим повніше висвітлюється досліджуване явище. Проте в неї не слід включати зайвих питань, які могли б ускладнити і розтягнути термін розробки даних. У той же час не слід складати програму надто вузько, адже в дослідження можуть не потрапити важливі питання.

При складанні програми велике значення має чітке формулювання питань, оскільки у більшості статистичних спостережень це складна і трудомістка робота, у виконанні якої беруть участь десятки і навіть сотні тисяч (при перепису населення) чоловік. Поставлені питання мають бути однаково зрозумілими для всіх.

Відповіді на питання програми спостереження записують у документ особливої форми – **статистичний формуляр**. Він являє собою первинний документ, в якому фіксують відповіді на питання програми по кожній з одиниць сукупності, це носій первинної інформації. Формуляри мають різні назви: форма первинного обліку або звітності, акт, бланк, табель, картка (фішка), анкета, опитувальний листок Для всіх перелічених видів формулярів характерні деякі обов'язкові елементи: змістовна частина, яка включає перелік питань програми, зведена графа або декілька граф для запису відповідей і шифрів (кодів) відповідей, титульна і адресна частини. На титульній сторінці записується назва статистичного спостереження (наприклад, «Перепис худоби в приватному секторі»), назва організації, яка проводить спостереження, а також зазначається, ким і коли затверджено формуляр або статистичне спостереження.

В адресну частину записують адресу обстежуваних (опитуваних) одиниць спостереження. Крім того, у формах статистичної звітності вказується, коли і куди треба надсилати одержану інформацію. Правильність даних обстеження стверджується підписами відповідальних осіб.

У практиці статистичного спостереження застосовують формуляри двох видів: картковий і списковий.

Картковим (або індивідуальним) називається статистичний бланк (фішка), який містить дані лише про одну одиницю спостереження. Загальна кількість карток повинна дорівнювати кількості одиниць досліджуваної сукупності.

Списковий формуляр – це статистичний бланк, в якому реєструються відомості по кількох одиницях спостереження. Наприклад, при перепису населення 1970 р. списковий формуляр було розраховано на реєстрацію у ньому зведень по шести особах, при перепису населення у 1989 р.– по двох.

Формуляри-картки зручні для ручної обробки занесених в них даних, але потребують значно більших трудових і матеріальних витрат, ніж формуляри-списки. Останні економічніші і зручніші для машинної обробки і контролю даних. Карткові формуляри використовуються при складанні звітів підприємств і закладів, які потребують широкої програми статистичного спостереження.

Спискові формуляри частіше використовуються при періодичних спостереженнях. Прикладом такого формуляра може бути переписний лист перепису населення 1989 р. Конструкція статистичних формулярів зумовлюється значною мірою впливом з боку технічних засобів обробки інформації. Це знаходить свій прояв у тому, що формуляри статистичного спостереження у деяких випадках виступають одночасно і безпосередніми носіями інформації при введенні їх в ЕОМ. Саме за таких умов вирішується питання безперервної технології автоматизації статистичних робіт, коли первинна інформація заноситься в автоматизований-банк даних (АБД) за допомогою перфострічок, магнітних стрічок, магнітних дисків і по каналах зв'язку, які з'єднують обчислювальні центри підприємств і державної статистики.

Вплив автоматизованої системи обробки статистичної інформації на конструкцію формулярів знаходить прояв у тому, що відповіді в них розташовані у зручному для шифрування (кодування) порядку.

Ефективність виконання розробленої програми спостережень значною мірою зумовлюється якістю інструктивного матеріалу. Для цього складають інструкцію (інколи її називають Статистичною інструкцією, або Доповідною інструкцією).

Інструкція – це документ, який пояснює питання програми статистичного спостереження, його мету, порядок заповнення статистичного формуляра і частково організаційні питання. (Інструкція є одним з найважливіших документів спостереження. Вона може містити вказівки відносно тих питань, які виникають у процесі проведення спостереження: об'єкт і одиниця спостереження, час і строки проведення, критичний момент спостереження і т. ін.

У багатьох випадках необхідні додаткові пояснення того, як правильно розуміти дане питання і як вірно записати відповідь на нього. Як правило, інструкції пишуть для осіб, які здійснюють перепис або заповнюють форми статистичної звітності. Вони дуже важливі для забезпечення однакового розуміння питання у всіх спірних і сумнівних випадках. Наприклад, при перепису населення 1989 р. ставилось запитання «національність». Для запису відповіді особам, які здійснювали перепис, треба було дати певні вказівки, щоб запобігти різнобою у відповідях. Для цього були дані інструкції, де вказувалося: «Записується національність, котру вказує сам опитуваний. Національність дітей визначається батьками. Лише в тих сім'ях, де батько і мати належать до різних національностей і батьки утрудняються самі визначити національність дітей, слід надати перевагу національності матері». Таке уточнення порядку записів на запитання перепису дає придатний для обробки первинний матеріал.

Відсутність в інструкції тлумачень певного питання призведе до того, що кожна особа (лічильник) буде розуміти запитання по-своєму, внаслідок чого зібрані матеріали зовсім знеціняться.

Вказівки інструкції повинні бути конкретні й чіткі, а текст стислим і лаконічним. Інструкції до форм статистичної звітності здебільшого друкуються на самій формі документа.

Отже, головне призначення інструкції полягає у поясненні змісту питань програми, як треба давати на них відповіді і заповнювати формуляр. Найбільш типові ситуації повинні розглядатися на прикладі.

Слід зазначити, що методологія розробки програми статистичного спостереження в останні 20 років зазнала досить значних змін у зв'язку з функціонуванням автоматизованої системи державної статистики (АСДС), яка зумовила створення автоматизованих банків даних (АБД). Наявність останніх дала змогу запровадити в галузі матеріального виробництва регістрову форму статистичного спостереження, яка полягає у створенні регістра, або автоматизованої картотеки сукупності одиниць статистичного спостереження певного типу. Перехід до автоматизованої статистичної інформаційної системи (АСІС) створює широкі можливості удосконалення програми спостереження, а саме: відбувається об'єднання інформаційних баз державної статистики, галузевих і регіональних органів управління, підприємств, об'єднань та інших ланок управління.

Наведемо приклад розробки програми і проведення статистичного спостереження.

Проводиться статистичне спостереження результатів роботи за господарський рік орендних сільськогосподарських підприємств району. Поставлено завдання вивчити загальне виробництво продукції цими підприємствами і його зв'язок з основними факторами виробництва. По кожному підприємству є дані річних звітів і якісної оцінки землі.

Розробити програму, провести статистичне спостереження і здійснити контроль даних.

Розглянемо послідовність виконання поставленого завдання за наступними етапами.

1. З'ясовують питання, що собою являє продукція підприємств і показники, якими визначається її обсяг. Відомо: а) орендні підприємства виробляють два види продукції – рослинницьку і тваринницьку; б) обсяг продукції по кожному продукту обліковують у натуральному виразі, а в цілому– по вартості у порівнянних фактичних цінах. Виходячи з поставленого завдання, вивчення загального обсягу виробництва продукції, в програму спостереження включають дані про виробництво не окремих видів продукції, а всієї продукції у грошовому виразі. Щоб мати можливість зіставлення обсягів виробництва продукції по спостережуваних підприємствах, вартість продукції беруть не в фактичних, а у порівнянних цінах.

2. З'ясовують основні фактори, які визначають обсяг виробництва продукції. Теоретично відомо, що основними факторами виробництва у сільському господарстві є земля, трудові ресурси, основні засоби виробництва, добрива і корми.

Отже, у програму спостереження включають показники обсягу зазначених факторів: 1) площа сільськогосподарських угідь, га; 2) площа орних земель, га; 3) середньорічна чисельність працівників підприємства, чол.; 4) грошові витрати на добрива; 4) грошові витрати на використані корми; 6) бал оцінки якості землі.

3. Вирішують питання про порядок запису в програмі спостереження обраних показників. Із метою зручності проведення обстеження розташовують показники відповідно до послідовності їх наведення у формах звітності: вказують номери сторінок, рядків, граф, де записані ці дані. Показник якості орних земель, який береться з матеріалів оцінки ґрунтів, записують у кінці програми.

Таким чином, одержуємо програму у вигляді: 1) вартість валової продукції сільського господарства у порівнянних цінах, всього, грн. (тут і далі в дужках указують номери сторінок, рядків і граф); 2) в тому числі продукція рослинництва, грн. (...); 3) середньорічна чисельність працівників на підприємстві, чол. (...); 4) середньорічна вартість основних виробничих фондів сільськогосподарського призначення, тис.грн. (...); 5) сума витрат на органічні і мінеральні добрива, грн. (...); 6) площа сільськогосподарських угідь, га (...); 7) у тому числі ріллі, га (...); 8) сума витрат на згодовані корми, грн. (...); 9) оцінка якості землі, балів.

4. Розробляють формуляр статистичного спостереження, в який записують значення ознак по кожному орендному підприємству. Таким формуляром може бути відомість. Але для нашого випадку краще для кожного підприємства відкрити картки у вигляді так званих фішок (рис. 1).

1	9600000	Підприємство «Світоч»	Район ___підгрупа__	Область ____група__	87	9
2	6316000					10
3	425					11
4	12810					12
5	46860					13
6	4 800					14
7	2 990					15
8	1278 000					16

Рис. 1. Формуляр статистичного спостереження (фішка)

Як бачимо, у формуляр занесено значення ознак у пронумеровані рядки (графи) у тій послідовності, в якій вони вказані у програмі спостереження: у рядок 1 записано вартість валової продукції всього, у рядок 9 – бал оцінки якості землі.

5. Здійснюють статистичне спостереження, а саме – заносять у фішки значення перелічених у програмі ознак по кожному підприємству, як це показано на прикладі орендного підприємства «Світоч».

6. Проводять контроль одержаної інформації (способи контролю матеріалів спостереження будуть розглянуті далі).

§ 2.3. Організаційний план статистичного спостереження, забезпечення точності даних

Організаційний план статистичного спостереження – це складова частина загального плану спостереження, в якій викладено порядок його організації і проведення. У ньому даються роз'яснення програмно-методологічних і організаційних питань. До перших належать формулювання мети і завдань спостереження, визначення його об'єкта і одиниці, розробка програми. До організаційних питань належать: фіксація місця, часу і строків спостереження, вказівка на те, хто його проводить, як воно проводиться і як здійснюється постачання статистичними формулярами осіб, які виконують спостереження, способи доставки заповнених формулярів у відповідні статистичні органи. Сюди відносять також ряд специфічних підготовчих робіт, зокрема таких, як підбір та навчання кадрів, що залучаються до проведення спостереження, підготовка графічного матеріалу та ін.

В організаційному плані статистичного спостереження конкретизуються права і обов'язки окремих установ і організацій, які беруть участь у заходах спостереження.

При плануванні спостереження насамперед визначають органи спостереження – організаторів і виконавців робіт, а також права і обов'язки кожного співвиконавця.

Програму і план статистичного спостереження розробляють органи державної статистики на рівні Міністерства Державного комітету статистики України. Низові органи державної статистики переважно виконують роботу по збору статистичних даних, їх первинному контролю і зведенню по певній програмі.

У плані вказують **строк проведення спостереження**, тобто час початку і закінчення збирання зведень. Цей час не можна ототожнювати **з часом спостереження**, тобто часом, до якого належать зведення. Статистичні показники характеризують досліджуване явище або за певний період, або на певний момент часу. Наприклад, дані про кількість виробленої продукції можна взяти тільки за період (день, декаду, місяць, квартал, рік), а показники запасів матеріальних цінностей можуть бути представлені на певний момент часу (на початок місяця, на початок кварталу, на початок або кінець року і т. ін.).

У плані має бути точно визначена територія, на якій здійснюється спостереження, а також особи і організації, відповідальні за проведення підготовчих робіт, збір, перевірку і обробку інформації по окремих ділянках території.

Серед організаційних питань значне місце у плані відводять проведенню підготовчих робіт. Насамперед треба скласти список звітних одиниць, тобто тих, які будуть обстежені. Цей список (колективних підприємств, державних підприємств, орендних підприємств і т. ін.) необхідний для перевірки повноти зведень, що надходять, а також визначення обсягу робіт і розрахунку необхідної кількості робітників для проведення статистичного спостереження.

Важливим підготовчим заходом є розрахунок потреби в кадрах для проведення спостереження, їх добір та інструктаж. Необхідно заздалегідь віддрукувати і розіслати бланки документів та інструкції щодо їх заповнення. Інструктаж вважається однією з найважливіших підготовчих робіт статистичного спостереження. Успіх проведення останнього багато в чому залежить від рівня підготовленості кадрів.

При підготовці складних статистичних спостережень, як правило, в плані передбачається проведення пробних спостережень з метою перевірки на практиці проекту плану і програми основного спостереження. Матеріали таких пробних спостережень використовують для уточнення, доповнення і конкретизації програми і плану спостереження, а також інструкцій.

Серед підготовчих робіт чільне місце повинне належати (особливо це стосується перепису населення) пропаганді спостереження серед населення. Засоби масової інформації повинні вести роз'яснювальну роботу щодо завдань і мети спостереження, що значною мірою сприятиме підвищенню ефективності вирішення планово-організаційних питань спостереження. Роз'яснювальна робота про мету і завдання проведення спостереження здійснюється через пресу, радіо, телебачення та інші засоби масової інформації.

Таким чином, організаційний план статистичного спостереження передбачає: визначення часу спостереження, часу і місця його проведення, порядок передачі матеріалів спостереження, комплекс підготовчих робіт, заходи, що забезпечують точність (вірогідність) статистичних даних. Кожній з названих вище категорій і етапів статистичного спостереження можна дати такі визначення.

Час спостереження – це момент або період часу, якого стосується статистична інформація (дані). Наприклад, інформацію про виробництво сільськогосподарської продукції збирають за певний період часу – день, декаду, місяць, квартал, рік.

Перепис населення передбачає належність даних до певного (критичного) моменту часу. Так, критичним моментом перепису населення 1989 р. було 24 години у ніч з 11 на 12 січня 1989р.

Час проведення спостереження, як правило, не збігається з часом спостереження. Наприклад, перепис населення 1989 р. здійснювався впродовж восьми днів – з 12 до 19 січня.

Місце проведення спостереження вибирають, виходячи зі способу спостереження і особливостей об'єкта спостереження. Так, інформацію про виробництво продукції сільськогосподарським підприємством реєструють у бухгалтерії підприємства, а при перепису населення – вдома, у поїздах, на вокзалах тощо.

Органами спостереження можуть бути різні установи системи державної статистики, відомства, підприємства і громадські організації. Серед них визначають організатора (установа системи державної статистики) і виконавця. Права і обов'язки кожного учасника статистичного спостереження повинні бути чітко визначеними.

Органи державної статистики повинні не тільки організовувати проведення статистичного спостереження, а й ретельно перевіряти точність одержаних результатів при його здійсненні. Перевірка вірогідності даних – найважливіша умова успішної роботи у справі спостереження. Щоб забезпечити вірогідність даних, необхідно повсякденно, систематично контролювати, чи вірно зрозумілі і застосовуються статистичні програми та інструкції, чи забезпечується повнота одержаних зведень, правильність даних бухгалтерського і оперативного обліку.

У процесі здійснення спостереження можуть бути допущені помилки, їх називають помилками спостереження. Види і природа виникнення таких помилок будуть розглянуті далі.

§ 2.4. Організаційні форми, види і способи статистичного спостереження

У статистичній практиці застосовуються різні форми статистичних спостережень. Із погляду організації спостереження розрізняють дві його основні форми: звітність і спеціально-організоване статистичне спостереження

Звітність як форма статистичного спостереження характеризується тим, що статистичні органи систематично одержують від підприємств і організацій (закладів тощо) в установлені строки зведення про умови і результати роботи за минулий період. Обсяг і зміст такої інформації визначаються затвердженими формами звіту. Як джерела даних для звітності використовують документи оперативно-технічного і бухгалтерського обліку. Обліково-статистичний апарат підприємств обробляє первинні записи у документах і результати заносить у форми звітів.

У нашій країні звітність е основною формою статистичного спостереження. Основну масу зведень, необхідних для управління народним господарством, а також для наукових досліджень, статистичні органи одержують у формі звітності.

Таким чином, **звітність** – це форма статистичного спостереження, при якій статистичні дані надходять у статистичні органи від підприємств і установ у вигляді обов'язкових і таких, що мають юридичну силу звітів про їх роботу. Організацію статистичної звітності здійснює Державний комітет статистики України. Він затверджує форму, порядок і строки подання звітності.

Перелік усіх форм із зазначенням їх реквізитів називають **табелем звітності**. Кожна форма звітності повинна мати такі відомості: назву, номер і дату затвердження; назву підприємства, його адресу і підпорядкованість; адреси, куди подається звітність, періодичність, дату подання, спосіб передачі; змістовну частину у вигляді таблиці; посадовий склад осіб, відповідальних за розробку і вірогідність звітних даних, тобто зобов'язаних підписати звіт.

Статистична звітність характеризується суворою регламентацією і відносною стабільністю вирішення всіх програмно-методологічних і організаційних питань спостереження. Подання її за передбаченими адресами і строками є обов'язковим для підприємств і організацій. Категорично забороняється всім державним органам управління вимагати, а підприємствам і організаціям подавати будь-які звіти, не передбачені державною звітністю.

Статистична звітність є основним джерелом інформації. яка забезпечує керівництво економікою на загальнодержавному, галузевому і регіональному рівнях управління. Слід розрізняти звітності: загально – і внутрішньовідомчу; міжгалузеву і галузеву; типову і спеціалізовану; первинну і зведену; особливо виділяється звітність тимчасова оперативна.

За способом, подання розрізняють звітність **поштову** і

термінову. Остання передається по телеграфу, телетайпу, радіо та іншими швидкими засобами. За, періодичністю розрізняють звітність **періодичну і одночасову**. Періодична подається через однакові проміжки часу або в точно визначені дати (наприклад, 5-го числа кожного місяця; не пізніше 1-го жовтня кожного року і т. ін.). Одночасова звітність подається у міру необхідності без певної періодичності. Періодична звітність поділяється на **поточну** – період її подання менше року (тиждень, місяць, квартал тощо) і річну – період її подання календарний рік.

Існує також статистична звітність, яка подається раз на рік поза зв'язком з початком чи закінченням календарного року, і звітність з періодом більше року (два рази у 5 років, один раз у 2 роки тощо).

Так звана **типова звітність** містить показники, загальні для різних видів діяльності (чи виробництва). Якщо збирають дані, специфічні для окремих видів діяльності (чи виробництв), їх відображують показниками у відповідній **спеціалізованій** звітності.

Як зазначалося вище, другою за значенням організаційною формою спостереження є **спеціально організоване статистичне спостереження**. Інколи його називають (до речі, помилково) просто «перепис». Застосовують спеціально організоване статистичне спостереження у таких випадках: 1) коли не можна застосувати звітність (наприклад, облік діяльності фермерських або особистих господарств населення); 2) складати звітність нераціонально; 3) необхідно детально вивчити явище поряд з вивченням його у формі звітності; 4) потрібно перевірити вірогідність даних звітності; 5) для вирішення самостійних науково-практичних завдань.

Спеціально організоване статистичне спостереження поєднує, в собі такі організаційні форми: а) перепис, б) суцільне і несуцільне обстеження.

Перепис як вид спеціально організованого статистичного обстеження проводиться з метою одержати дані про явище на певний момент часу Тобто обчислити чисельність і склад об’єкта статистичного спостереження за рядом характерних для нього ознак, які не збираються в порядку статистичної звітності (наприклад, перепис населення, перепис виробничого обладнання і т. ін.). У ряді випадків переписи доповнюють (істотно уточнюють) дані поточного обліку. Вони вимагають ретельної попередньої підготовки. Характерними особливостями перепису є: одночасність проведення його на всій передбаченій території; єдність програми спостереження;

стислість строків статистичного спостереження станом на один і той же момент часу – **критичний момент перепису**.

Існує два типи переписів: одні переписи проводять на підставі даних обліку і звітності підприємств і організацій, інші – на підставі спеціально організованої реєстрації фактів.

Переписи першого типу, як правило, проводять робітники підприємств і установ під керівництвом органів державної статистики. Цей тип перепису називають одночасовим обліком. Прикладом першого типу можуть бути переписи промислового обладнання, залишків важливих видів матеріалів (чорних, кольорових металів, будівельних матеріалів тощо), облік тракторного парку у сільському господарстві, заключний облік посівних площ за видами підприємств і т. ін.

Прикладом другого типу переписів, при яких статистичні формуляри заповнюються шляхом спеціально організованої реєстрації фактів, є перепис населення. Це спеціально організоване статистичне спостереження, метою якого є одержання інформації про чисельність, розміщення і склад населення. Реєстрація потрібних фактів здійснюється шляхом опитування. Наукові принципи проведення переписів передбачають встановлення критичного моменту перепису, періоду перепису, способу збирання зведень.

Залежно від повноти охоплення статистичної сукупності розрізняють суцільне і несуцільне статистичне спостереження.

При **суцільному статистичному спостереженні** обстеженню підлягають усі одиниці, які входять у склад досліджуваної сукупності. Однак не слід розуміти так, що суцільне спостереження обов'язково у всіх випадках повинне охоплювати досліджуване явище по всій країні. Досліджувана сукупність може обмежитись територіальними, відомчими, галузевими й іншими рамками. .Важливим тут є те що в межах цієї сукупності обов'язково реєструються зведення про кожну одиницю досліджуваного об'єкта. Прикладом суцільного спостереження є перепис населення.

При **несуцільному спостереженні** обстежується лише частина одиниць статистичної сукупності. При його організації ставиться завдання (як правило) поширити результати спостереження на всю сукупність. Прикладом такого виду спостереження може бути обстеження: втрат урожаю сільськогосподарських культур, схожості насіння, ступеня засміченості посівів, бюджету сімей населення і т. ін.

Основні види несуцільного спостереження такі: вибіркове, обстеження способом основного масиву, монографічне і анкетне.

Вибірковим називають таке статистичне спостереження, при якому обстеженню підлягає частина статистичної сукупності, відібраної на основі науково розроблених принципів відбору. Це найпростіший і найбільш досконалий вид несуцільного обстеження.

Обстеження основного масиву (або способом основного масиву) являє собою таке несуцільне обстеження, при якому з усієї сукупності одиниць для спостереження відбирається така їх частина, в якій обсяг досліджуваної ознаки становить питому вагу, більшу за 50 % загального обсягу сукупності (табл. 1).

Таблиця 1

Групи підприємств за рівнем середньорічної зарплати одного працівника

Середньорічний рівень зарплати , грн..	**Кількість підприємств у групі**	**Середньорічний рівень зарплати в групі**	**Питома вага підприємств до загальної сукупності, %**
До 3500	5	3200	15,1
3500 - 3700	7	3660	21,2
3700 - 3900	12	3812	36,4
3900 - 4100	6	4000	18,2
Понад 4100	3	4408	9,1
Разом	33	-	100

Як свідчать дані таблиці 1, у 33 підприємствах середньорічний рівень зарплати одного робітника варіює від 3200 до 4408 гривень. Основним масивом у даній сукупності можна вважати 21 підприємство з рівнем середньорічної зарплати понад 3700 грн.; оскільки питома вага їх у загальній кількості перевищує 50 % (36,4 + 18,2+9,1=63,7).

Державна статистика обстеженням основного масиву вивчає ціни на ринках продажу різного виду продукції і виробів.

Метод основного масиву вважається недосконалим суцільним методом обстеження. Інколи його називають групувально – цензовим методом. Недоліком методу основного масиву вважається те, що з обстеження вилучається частина одиниць сукупності, якою нехтують як неістотною.

Для **монографічного спостереження** характерним є детальне вивчення окремих одиниць статистичної сукупності або невеликих їх груп, подібних у певному відношенні. Одиниці або групи явищ повинні бути типовими, щоб на їх підставі можна було судити про характер цих явищ. Прикладом монографічного обстеження може бути вивчення досвіду передового підприємства або їх групи. У статистиці монографічне спостереження застосовують при вивченні і популяризації передового досвіду, а також процесу розвитку окремого трудового колективу, недоліків у роботі тощо.

Об'єктом монографічного спостереження, крім підприємства, може бути виробнича бригада, школа, вуз, місто, регіон, сім'я й інші об'єкти.

Таким чином, монографічний спосіб – це обстеження поодинокого прикладу, який повинен ілюструвати всю статистичну сукупність, що в свою чергу дозволяє конкретизувати наші знання цієї сукупності. Предметом монографічного дослідження можуть бути елементи, типові для даної сукупності, або ж елементи, які характеризують її розвиток.

Монографічне спостереження здійснюють з метою виявлення тенденції розвитку прогресивних явищ і поширення передового досвіду. Цей вид спостережень допомагає викрити невикористані резерви, що досягається монографічним описуванням передових підприємств і досвіду окремих бригад, ланок чи осіб. Монографічне обстеження використовують для коригування даних суцільного обстеження.

Анкетне обстеження ґрунтується на принципі добровільного заповнення окремими особами (адресатами) надісланих їм спеціальних анкет. Цей спосіб спостереження широко застосовується у конкретних соціологічних дослідженнях. його використовують редакції газет, журналів, установи зв'язку, науковці певних галузей, зокрема економічної науки.

Анкетне обстеження певною мірою близьке до вибіркового спостереження. Але при вибірковому одиниці статистичної сукупності підлягають безпосередньому обстеженню. При анкетних обстеженнях звертаються із запитаннями до тих осіб чи організацій, які можуть дати необхідну інформацію. Таким чином, характерна риса анкети – непряме спостереження.

Даний спосіб обстеження завжди здійснюється для висвітлення специфічної, чітко обмеженої проблеми. Воно завжди висвітлює будь-яке часткове, визначене питання. Запитання анкет не завжди мають суто статистичний, числовий характер. Інколи обстеження доповнюється запитаннями якісного описового характеру, наприклад, запитаннями про думку щодо причин спостережуваних фактів.

Наприклад, треба вивчити причини зниження обсягів виробництва тваринницької продукції фермерськими господарствами, а також з'ясувати, якими засобами можна допомогти справі. Якщо відповіді в одержаних анкетах не задовольняють, посилається анкетна комісія, яка проводить бесіди в підприємствах і на цій основі складає звіт. Бесіди мають бути спрямовані на одержання статистичних даних, інколи виходять за ці рамки. Комісія робить спробу встановити причини зниження обсягів виробництва і вислуховує численні думки. з цього питання. Як правило, головну частину анкети і складає з'ясування думок.

Анкетування проводиться як статистичними органами, так і науково-дослідними установами (останніми досить часто).

Анкетне опитування має давню історію. Наприклад, за часів Наполеона у Франції анкетним методом вивчались питання стосовно скоєних злочинів. У Великобританії анкети застосовувалися ще раніше – у XVII столітті.

Існує два типи анкетних обстежень: перший тип – анкети направляються певному, як правило, невеликому колу спеціалістів з даного питання; другий тип – ґрунтується на масовому збиранні відповідей і обробці їх статистичними методами з метою одержання «середньої» думки.

При анкетному методі обстежений результати його можуть бути викривлені, оскільки програма такого обстеження зачіпає інтереси опитуваних осіб. Останні в своїх відповідях можуть дати таку прикрасу повідомлюваним даним, яка їм вигідна, і навпаки – замовчувати невигідні факти. Крім того, відповіді надсилають лише ті, хто зацікавлений у збиранні даних, про які йдеться у питаннях анкети. Тому анкета не завжди дає репрезентативні результати. Окремими статистиками анкета вважається вкрай недосконалим засобом статистичного обстеження.

За часом проведення статистичні спостереження поділяють на поточні (безперервні), періодичні й одноразові.

Поточне спостереження полягає в безперервній реєстрації фактів і явищ у міру їх виникнення. Прикладом може бути реєстрація народжених дітей, шлюбів і розлучень у загсах, облік виробленої продукції на промислових підприємствах та ін.

Періодичним спостереженням називають таке, що повторюється через певні, заздалегідь установлені проміжки часу.

При даному спостереженні явища реєструють через певні періоди. Прикладом періодичного спостереження може бути щомісячний звіт підприємств про стан тваринництва, а також щорічні переписи худоби станом на 1 січня.

Одноразове спостереження проводять для вивчення якогось явища на певний момент часу (у разі потреби). Прикладом одноразових спостережень є переписи плодоягідних насаджень та ін.

Статистичне спостереження здійснюється такими способами: безпосереднім, документальним і опитуванням.

Спосіб **безпосереднього спостереження** характеризується тим, що представники органів державної статистики й інших організацій записують дані у статистичні формуляри після особистого огляду, підрахунку, вимірювання чи зважування одиниць спостереження.

Документальний – основний спосіб статистичного спостереження. Його здійснюють на підставі документів оперативно-технічного та бухгалтерського обліку. Цей спосіб використовують при складанні підприємствами і організаціями звітності на підставі документів первинного обліку. Оскільки джерелом зведень при складанні первинних документів є безпосереднє спостереження, то при належному контролі за веденням первинного обліку і правильністю заповнення статистичної звітності документальний спосіб спостереження дає найточніші результати.

Опитування – це спосіб спостереження, при якому відповіді на запитання статистичного формуляра записують зі слів чи письмових відповідей опитуваних осіб. Опитування може бути **організоване** по-різному: усне (експедиційний спосіб), самореєстрацією, кореспондентським і анкетним способами.

При *усному опитуванні* працівник, який проводить спостереження, сам заповнює статистичний формуляр, розмовляючи з опитуваною особою. Так було при перепису населення у 1989 році.

При способі *самореєстрації* опитувані особи особисто заповнюють бланк формуляра згідно із вказівками щодо його заповнення. Цей спосіб спостереження застосовують, наприклад, при обстеженні бюджетів сімей.

При *кореспондентському способі* відомості статистичним органам повідомляють добровільні кореспонденти.

Анкетний спосіб збору даних ґрунтується, як було зазначено, на принципі добровільного заповнення адресатами анкет (листів опитування). Як правило, заповнених анкет повертається менше, ніж розсилається. Крім цього, перевірити вірогідність даних практично неможливо. Тому анкетний спосіб застосовують у тих випадках, коли не вимагається висока точність зведень, а потрібні наближені характеристики. Цей спосіб використовують при вивченні попиту населення на окремі товари чи продукти харчування і ін. Державна статистика анкетний спосіб спостереження не застосовує.

Вибір форми, виду і способу статистичного спостереження визначається характером об'єкта, що вивчається, вимогами до ступеня точності показників, фінансовими можливостями й іншими факторами.

Створення автоматизованої системи державної статистики зумовило принципово новий підхід до техніки і технології збирання інформації, її передачі, обробки і зберігання. Єдина інформаційна база у вигляді автоматизованих банків даних створила необхідні умови для інтегрованої територіально-розподільчої системи, яка здійснює шляхом одноразової фіксації нагромадження, обробку, зберігання, пошук і видачу інформації для багатоцільового і багаторазового використання. При цьому раціональним стає не тільки сам процес збирання і обробки інформації, а й удосконалюються методологія, система показників і їх статистичний аналіз із використанням економіко-математичних методів.

§ 2.5. Помилки статистичного спостереження. Способи контролю інформації

Вірогідність статистичних даних – закон державної статистики. Забезпечується вона належним складанням програми і плану спостереження, науковою організацією збирання, обробки і аналізу інформації. Хоч як старанно не було б організоване статистичне спостереження, зібрані матеріали можуть мати різні за характером і виникненням неточності: неповне охоплення одиниць спостереження, що підлягають реєстрації; пропуски окремих записів; помилки поодиноких записів тощо. Якщо повноту охоплення одиниць спостереження і пропуски окремих показників встановити неважко, то знайти допущені похибки поодиноких записів, так звані помилки спостереження, справа не з легких.

Помилки в процесі спостереження призводять до зниження його точності.

Точністю статистичного спостереження називають ступінь відповідності величини будь-якого показника (ознаки), встановленої за допомогою спостереження, дійсній величині. Вона вимірюється різницею або співвідношенням цих величин.

Розбіжність між величиною будь-якого показника, встановленого шляхом спостереження, і дійсним його розміром називають **помилками статистичного спостереження**. Помилки спостереження поділяють на два види: помилки реєстрації і помилки репрезентативності.

Помилки реєстрації виникають внаслідок неправильного встановлення фактів або неправильного їх запису у формуляр.

Помилки репрезентативності мають місце лише при вибірковому обстеженні і виникають внаслідок того, що вибіркова сукупність недостатньо повно відтворює всю досліджувану сукупність. Докладніше помилки репрезентативності описані в §11.4.

Помилки репрезентативності можуть бути як при суцільному, так і при несуцільному спостереженні. Вони можуть бути навмисними і ненавмисними. **Навмисні** помилки є наслідком свідомого перекручення дійсності у бік збільшення або зменшення дійсних розмірів досліджуваної ознаки.

Ненавмисні помилки виникають незалежно від бажання осіб, які повідомляють або реєструють дані.

Ненавмисні помилки реєстрації можуть мати випадковий або систематичний характер.

Випадкові ненавмисні помилки реєстрації – це помилки, які виникають внаслідок різних випадкових причин: описка, обмовка і т. ін. Вони призводять до відхилень даних спостереження від фактичних розмірів ознаки з однаковою ймовірністю як у бік збільшення, так і в бік зменшення даних. При досить великій кількості одиниць спостережень випадкові помилки можуть взаємно погашатися і не справляти істотного впливу на результати спостереження.

Систематичні ненавмисні помилки реєстрації виникають з певних невипадкових причин і призводять до відхилень даних спостереження від фактичних розмірів ознаки в бік збільшення або зменшення. Причиною таких помилок може бути несправність вимірювальних приладів, нечітке формулювання питань, недосконалість статистичного інструментарію, схильність людей до округлення цифр і т. ін.

Навмисні помилки реєстрації завжди мають систематичний характер.

Логічно завершується статистичне спостереження прийманням матеріалів дослідження. Коли матеріал статистичного спостереження одержано повністю від усіх одиниць, що підлягають спостереженню, перевіряють повноту (якість) заповнення бланків. Якщо при прийманні матеріалу спостереження виявлено незаповнені (або частково заповнені) бланки, це означає що при статистичному спостереженні пропущена одиниця спостереження. Тому відповідальна особа, приймаючи статистичні формуляри (бланки) в першу чергу перевіряє повноту їх заповнення і у випадку необхідності вживає заходи для їх виправлення. .

Поряд з перевіркою повноти заповнення бланків здійснюється контроль за вірогідністю і правильністю відповідей. При прийманні матеріалів спостереження головна увага приділяється правильності заповнення відповідних бланків і перевірці вірогідності (точності) показників.

Контролю за вірогідністю статистичних даних статистичні органи приділяють особливу увагу. Такі функції (обов'язки) державна статистика виконує у тісному контакті з органами контролю, прокуратури і громадськими організаціями.

Із метою виявлення і усунення допущених при реєстрації помилок статистичні органи здійснюють арифметичний і логічний контроль зібраного матеріалу.

Арифметичний контроль полягає у перевірці точності арифметичних підрахунків і розрахунків: перевірка підсумкових показників у документах, перевірка правильності підрахунків процентів, середніх величин і т. ін.

Логічний контроль полягав у зіставленні відповідей на запитання і з'ясування їх логічної узгодженості. У процесі логічного контролю можуть бути встановлені нереальні або малоправдоподібні відповіді.

Розглянемо загальні прийоми логічного контролю.

1. Зіставлення відповідей на різні взаємопов'язані питання у формулярах. Наприклад, запис у формулярі про те, що дитина дошкільного віку має середню освіту, є помилковим.

2. Порівняння записів у документі, що перевіряється, з аналогічними записами в інших документах.

3. Зіставлення звітних показників за суміжні періоди.

4. Застосування методу балансової узгодженості показників.

Найчастіше використовують таку балансову рівність: наявність на початок періоду плюс надходження мінус вибуття дорівнює наявності на кінець звітного періоду.

5. Проведення безпосередньо за переписами контрольних перевірок – суцільних або вибіркових.

Зазначені прийоми перевірки статистичних даних шляхом арифметичного і логічного контролю використовують як при перевірці матеріалів спеціально організованих статистичних спостережень, так і звітності. Можна стверджувати, що арифметичний контроль чітко встановлює наявність помилки, а логічний – у більшості випадків лише виявляє можливість помилки. При цьому, якщо проведення арифметичного контролю вимагає від статистика елементарної грамотності, то логічний – може здійснюватися лише висококваліфікованими спеціалістами.

Значна вірогідність статистичних даних зумовлюється діючою системою заходів, спрямованих на зменшення і уникнення помилок. Серед них слід назвати такі: якісний первинний облік; розробка наукових рекомендацій з питань перевірки вірогідності даних; добір кваліфікованих кадрів-статистиків, автоматизація статистичних робіт і т. д.

ТЕМА 3. ЗВЕДЕННЯ І ГРУПУВАННЯ СТАТИСТИЧНИХ ДАНИХ

§ 3.1. Зміст і завдання статистичного зведення

Статистичне спостереження, даючи об'ємний, але різноманітний матеріал про окремі явища досліджуваної сукупності, ще не дозволяє зробити будь-які висновки про цю сукупність. Адже в результаті збирання фактів дійсність стає відомою, але ще не пізнаною. Щоб за науково зібраними фактами зробити об'єктивні висновки, глибоко пізнати дійсність, ці фактори необхідно узагальнити, теоретично обміркувати.

Як відомо з викладеного раніше, статистичне спостереження збігається з першим ступенем людського пізнання дійсності – емпіричним. Теоретичне узагальнення фактів є другим ступенем складнішого процесу пізнання світу, адже на цьому етапі здійснюється наукове узагальнення статистичних даних. Таке узагальнення дає можливість встановити внутрішні зв’язки між явищами, їх кількісно-якісні перетворення. Але було б помилковим вважати, що теоретичне обміркування і узагальнення фактів здійснюється тільки після того, як зібрані факти. Обміркування матеріалів починається уже в процесі їх збирання, але процес теоретичного узагальнення здійснюється лише після того, коли зібрані всі факти, які цікавлять дослідника.

Процес теоретичного узагальнення статистичних даних, зведеня фактів у єдине ціле в статистиці називають зведеням статистичних даних.

Статистичне зведення являє собою сукупність прийомів, які дозволяють одержати узагальнюючі статистичні показники як зведені ознаки масових явищ, що характеризують стан, взаємозв’язки і закономірності розвитку явищ в цілому.

Не треба змішувати поняття «статистичне зведення» і «зведення» у вузькому розумінні слова. Під останнім розуміють підсумовування даних про число одиниць сукупності і значень їх ознак. Тобто, це один з етапів статистичного зведення. В цілому статистичне зведення включає такі етапи: 1) статистичне групування; 2) підсумовування даних (зведеня у вузькому розумінні слова); 3) табличне і графічне оформлення одержаних даних.

Одержана в процесі зведення система статистичних показників підлягає подальшому аналізу в наукових і практичних цілях.

Таким чином, статистичне зведення – це первинна наукова обробка даних спостереження для характеристики суцільного явища узагальнюючими показниками.

Зведення являє собою другий ступінь статистичного дослідження і від його якості значною мірою залежить результат всієї статистичної роботи.

За допомогою статистичного зведення розв’язують такі завдання: 1) групування даних; 2) розробка системи показників для характеристики груп і всієї статистичної сукупності; 3) обчислення групових і загальних показників; 4) зведення результатів обчислення у статистичних таблицях.

Відповідно до цих завдань статистичне зведення передбачає виконання таких операцій: 1) вторинний контроль (логічний і арифметичний) якості матеріалів статистичного спостереження; 2) шифрування, тобто присвоєння певних номерів (або літер) значенням окремих ознак, за якими здійснюється групування матеріалів; 3) розкладання матеріалів на якісно однорідні групи; 4) підрахунок підсумовуючих даних, занесення підсумків у спеціальні зведенні формуляри.

Статистичне зведення здійснюють за спеціальною заздалегідь розробленою програмою. Залежно від мети і завдань дослідження програма встановлює групувальні ознаки для утворення однорідних у певному відношенні груп, їх число, макети розроблених таблиць, які містять об'єкти дослідження і показники, що їх характеризують.

Оскільки суспільні явища досить різноманітні, програма повинна бути складена так, щоб одержаний в результаті зведеня матеріал характеризував досліджуване явище з різних боків.

Крім програми зведеня, ще складають план його проведення; у ньому дають вказівки про послідовність і строки виконання окремих частин зведення і викладення його матеріалів, передбачають виконавців, вид зведення і способи контролю узагальнюючих даних тощо.

Таким чином, поняття статистичного зведення у широкому розумінні слова охоплює цілий комплекс статистичних операцій, спрямованих на об'єднання зареєстрованих при спостереженні поодиноких випадків у групи, подібні у тому чи іншому відношенні; підрахунок підсумків по виділених групах і по всій сукупності в цілому; оформлення результатів групувань і зведення у вигляді статистичних таблиць.

З погляду організації розрізняють два види статистичного зведення: централізоване і децентралізоване.

Централізоване зведення проводять в одному центральному органі, наприклад, Державному комітеті статистики України, куди заздалегідь надсилають усі матеріали статистичного спостереження.

Децентралізоване зведення здійснюють поступово в різних ланках системи статистичних органів; на рівні району, області, країни Централізований вид зведення має свої переваги перед децентралізованим у тому, що є можливість проведення його за єдиною методологією при значно менших затратах праці і високій точності розрахунків. До його недоліків слід віднести низьку оперативність використання результатів зведень на місцях і складність у виправленні виявлених помилок спостереження. Централізоване зведення має місце, як правило, у великих спеціально організованих статистичних спостереженнях (наприклад, переписи).

У статистичній практиці найчастіше застосовують децентралізоване зведення; інколи ці два види поєднують.

Розрізняють зведення первинне і вторинне.

Первинне зведення – це таке, при якому обробка і підрахунок даних здійснюються безпосередньо в процесі статистичного спостереження. При **вторинному зведенні** обробка і підрахунок матеріалів відбуваються за результатами первинного зведення. Як правило, це стосується децентралізованого зведення.

Якщо процес зведення зводиться до звичайного підсумовування даних, тобто до одержання підсумовуючих показників, його називають **простим зведенням**. Але ним обмежуються лише у випадках, коли немає істотної різниці між досліджуваними явищами, що на практиці зустрічається дуже рідко. Як правило, досліджувані статистикою соціально-економічні явища за своїм складом неоднорідні й істотно різняться між собою. Тому завдання статистичного зведення не обмежуються простим підрахунком показників, адже в такому випадку ігноруються існуючі істотні різниці між явищами. Статистика повинна виявляти ці різниці і досконало вивчати їх. У такому разі зведення статистичних даних здійснюють шляхом їх групувань за певними ознаками з подальшим розрахунком системи показників по кожній групі з підрахунком групових і загальних підсумовуючих показників. Зведення, що включає групування даних, називається **груповим.** Тільки методом групувань можна обробити зібрані матеріали, щоб одержати зведені узагальнюючі показники, які б об'єктивно відображували існуючі реалії у досліджуваній сукупності явищ.

Отже, статистичне групування даних спостереження є методологічною основою наукового зведення.

§ 3.2. Статистичне групування, його суть, завдання і види

Як відомо, масові суспільні явища або сукупності складаються з одиниць, які різняться між собою як якісно, так і кількісно. Ці різниці можуть бути істотними, тобто викликаними певними істотними факторами, і неістотними, що являють собою випадкові відхилення. Але і в сукупностях якісно однорідних, які мають незначну кількісну варіацію, у процесі розвитку спочатку збільшуються кількісні відмінності, а потім кількісні відмінності переходять у якісні з властивою їм уже другою мірою. Самі якісні відмінності в даних випадках виступають явно, видимо, в інших – приховано за кількісними змінами.

Тому одне з головних завдань статистики полягає в тому, щоб розділити складну, неоднорідну сукупність одиниць спостереження на однорідні всередині, але істотно різні між собою групи; своєчасно виявити за зростанням кількісних різниць якісні переходи. Таке завдання в статистиці вирішується за допомогою методу статистичних групувань.

Статистичне групування являє собою процес утворення подібних в тому чи іншому відношенні груп, який здійснюється за наявними статистичними даними. Причому процес утворення груп слід розглядати з різних боків. Якщо дослідник зустрічається з єдиною множиною різних суспільних явищ, то групування треба розглядати як уявне розчленування такої множини явищ на подібні групи (логічна операція аналізу). Якщо ж виходити з того, що, безпосередньо приступаючи до групувань, дослідник уже має дані про окремі одиниці спостереження, то групування можна розглядати як уявне об'єднання окремих одиниць у схожі групи (логічна операція синтезу). Отже, групування – це єдиний аналітико-синтетичний процес і розчленування, і об'єднання суспільних явищ, у результаті якого утворюються відповідні групи. У ньому практично знаходить свій прояв закон великих чисел, адже у показниках, обчислених на достатньо великій кількості одиниць сукупності в групах, взаємо погашається випадковість і лишається істотне, необхідне.

Таким чином, метод статистичних групувань, крім характеристики механізму взаємодії всієї різноманітності типів і форм тих чи інших явищ, доводить істотність показників, які повинні дати характеристику стану і розвитку сукупності. А з цього випливає, що цілий ряд статистичних методів (метод середніх величин, вибірковий, кореляційно-регресійний та інші) певною мірою спирається на метод статистичних групувань як на свою основу. Як не можна проводити статистичного групування без попереднього науково обґрунтованого статистичного спостереження, так само не можна ефективно використовувати методи середніх і відносних величин, аналізу зв’язків і т. ін., не проводячи попередньо статистичних групувань (безумовно, якщо воно не закладено в самій схемі статистичного спостереження).

Що ж становить основу статистичного групування, в чому його суть?

Групування – невід’ємний елемент зведення, його найважливіший етап. Це процес утворення груп одиниць сукупності, однорідних у певному істотному відношенні, а також тих, що мають однакові або близькі значення групувальної ознаки. Групуванням називають розподіл статистичної сукупності на групи (частини) за рядом характерних для них ознак. Це статистичний метод розчленування складного масового явища на істотно різні групи з метою всебічної характеристики його стану, розвитку і взаємозв'язків.

Із наведеного визначення випливає, що суть методу статистичних групувань полягає у тому, що складне масове явище розглядається не як єдине нероздільне ціле, а у ньому виділяються окремі групи одиниць із статистичними показниками, які дають кількісну характеристику якісно своєрідній частині одиниць усієї сукупності. Тобто кожна з одержаних груп об'єднує однорідні одиниці сукупності. Групи встановлюють на підставі аналізу суті досліджуваного явища. У результаті внутрішньогрупових підрахунків одержують характеристику у вигляді системи статистичних показників. Необхідність такого підходу до вивчення масових явищ зумовлюється тим, що в складних масових явищах існують якісні відмінності між окремими одиницями, адже кожна з них має свої індивідуальні риси. У цілому ж всі такі одиниці мають загальні риси, які формують дане явище. Наприклад, розглянемо урожайність зернових культур у підприємствах району. Щоб мати уявлення про урожайність, недостатньо знань лише про загальний її рівень, адже всередині даної статистичної сукупності (підприємства, району) містяться одиниці спостереження зі своїми індивідуальними властивостями. Це підприємства з конкретними природними умовами, специфікою технологій вирощування, різним рівнем забезпеченості матеріальними і трудовими ресурсами і т. ін.

Використання методу статистичних групувань знаходить місце і при вивченні якісно однорідних сукупностей, де ще не спостерігається якісних перетворень, але є кількісні відмінності. У таких випадках важливо відокремити групи з різними значеннями ознак і вивчити взаємозв'язки досліджуваних ознак з іншими ознаками даного явища. Наприклад, явище продуктивності молочного стада корів у підприємствах району розглядається як середній надій від однієї корови (за добу, декаду, місяць, рік) без врахування породного складу, року лактації, віку тварин тощо. Виділення груп підприємств за рівнем продуктивності корів дає змогу вивчати взаємозв'язки даної ознаки, тобто продуктивності корів, з іншими ознаками, такими як рівень годівлі, навантаження тварин на одного працюючого, продуктивність праці та ін.

Особливого значення метод статистичних групувань набуває у вивченні соціально-економічних явищ. Останні мають досить складний характер розвитку і якісних перетворень. До речі, сам метод групувань було розроблено стосовно до вивчення суспільних явищ і тому він знаходить широке застосування в аналізі саме соціально-економічних процесів, що відбуваються у суспільстві.

У цілому мета групувань зводиться до встановлення чисельності кожної окремо взятої групи (категорії) та вивчення впливу причин і залежностей явищ.

Метод статистичних групувань робить статистику одним з наймогутніших знарядь соціального пізнання і використовується для вирішення різних завдань, які виникають у процесі наукового статистичного дослідження. Їх можна звести до трьох основних: 1) виділення соціально-економічних типів явищ; 2) вивчення (характеристика) структури досліджуваного явища; 3) вивчення зв'язків і взаємозалежностей між явищами (характеристика взаємозв'язків між варіюючими ознаками).

Відповідно до вирішуваних завдань розрізняють такі види групувань: типологічні, структурні й аналітичні.

Із теорії групувань випливає, що зведені статистичні характеристики можна використовувати лише до якісно однорідних сукупностей. Отже, перше завдання методу групувань полягає у тому, щоб на підставі економічної теорії розподілити масу досліджуваних явищ на однорідні в соціально-економічному відношенні групи, виділити соціально-економічні типи явищ і процесів.

Таким чином, **типологічне групування** – це групування, за допомогою якого у досліджуваній сукупності явищ відокремлюються одноякісні в істотному відношенні групи, перш за все класи і соціально-економічні типи. Наприклад, групування населення за соціальною ознакою – класовим складом, підприємств – за формою власності, продукції – за економічним призначенням і т. д.

Таблиця 2

Групування господарств району за формою власності

Тип господарства	Роки		
	2002	2003	2004
Колективні сільськогосподарські підприємства	30	31	31
Державні сільськогосподарські підприємства	6	5	5
Фермерські господарства	36	42	46

Необхідність проведення типологічних групувань випливає з характеру суспільних відносин і законів, що діють у об'єктивній реальності. Якщо не проведене типологічне групування, тобто чітке розмежування соціально-економічних типів явищ, то ніякі найтонкіше прийоми наступної статистичної обробки даних досліджуваної сукупності не в змозі дати об'єктивні результати дослідження.

Прикладом типологічних групувань можуть бути дані таблиці 2.

Другим важливим завданням статистичних групувань є дослідження структури типових однорідних груп. Виділення якісно однотипної сукупності ще не значить, що в ній всі одиниці за всіма ознаками однакові. Наприклад, всі фермерські господарства як тип господарств однорідні, але вони істотно різняться за розмірами угідь, наявністю техніки та знарядь, чисельністю худоби і рядом інших ознак. Виявити і дати характеристику цим відмінам – завдання статистики.

Отже, **структурне групування** (інколи його називають варіаційним) – це групування, яке характеризує розподіл одиниць однотипних (однорідних) сукупностей за будь-якими ознаками. Тобто групування, яке дозволяє виявити склад (внутрішню будову) однорідної у якісному відношенні сукупності за певними ознаками. Цей вид групувань знаходить винятково широке застосування при дослідженні соціально-економічних явищ.

Таке групування дає інформацію про те, з яких частин складається досліджувана множина явищ, яка будова типів явищ і якими показниками характеризуються окремі частини.

На відміну від пізнавального значення типологічних групувань, структурне групування для одержання висновків щодо поточного стану господарства, для оперативного керівництва роботою підприємства та його галузей служить базою для розрахунку наявних у підприємстві резервів. Так, на рівні адміністративного району можна здійснити групування підприємств (однотипних) за ознаками зростання продуктивності праці та її оплати, зниження собівартості виробництва, витрат енергетичних ресурсів, сировини тощо.

Таблиця 3

Склад населення області за місцем проживання, %

Місце Проживання	Роки		
	2002	2003	2004
Місто	30,7	32,6	35,9
Село	69,3	67,4	64,1
Всього	100,0	100,0	100,0

За допомогою структурних групувань вивчається перш за все внутрішня структура типів статистичних сукупностей. Таке групування здійснюється після і на основі типологічних групувань. Але інколи вивчається структура загальних сукупностей, які включають неоднорідні в певному відношенні явища. Так, досліджується структура всіх сільськогосподарських підприємств області, структура працівників тваринницької галузі і т. ін.

Якщо структурне групування здійснюється в територіальному розрізі, його називають **територіальним.**

Типологічні і структурні групування, виконані на базі показників за кілька періодів, дають уявлення про відповідні зміни досліджуваних явищ у часі.

За допомогою структурних групувань вивчають склад населення за віком, статтю, місцем проживання, національністю і т. ін.; склад робітників за стажем роботи, віком тощо; склад аграрних підприємств за видами земельних угідь тощо.

Прикладом структурних групувань можуть бути дані таблиці 3.

Наступне завдання, яке вирішується за допомогою статистичних групувань – це дослідження взаємозв'язків варіюючих ознак у межах одноякісної сукупності. Такі групування називають аналітичними. **Аналітичне групування** дає змогу встановлювати та вивчати причинно-наслідкові зв'язки між досліджуваними явищами та їх ознаками (при цьому слід пам'ятати, що аналітичні функції притаманні всім видам групувань, і перш за все типологічним). Взаємопов'язані ознаки поділяють на **факторні і результативні**. При цьому групи утворюють, як правило, за факторною ознакою, а для кожної виділеної групи розраховується або середнє значення результативної ознаки, якщо вона кількісна, або відносні величини, якщо вона якісна. Взаємозв'язок проявляється у систематичній зміні результативної ознаки у зв'язку зі зміною факторної ознаки.

Прикладом аналітичного групування може бути групування підприємств за рівнем продуктивності праці, за рівнем оплати праці тощо, коли вивчається вплив цих факторних ознак на собівартість продукції – результативну ознаку.

Таблиця 4 являє собою аналітичне групування підприємств за рівнем затрат праці на виробництво 1 центнера продукції. Тут показник затрат праці є факторною ознакою, а собівартість центнера – результативною. Як бачимо, між продуктивністю праці (затратами праці) і собівартістю існує пряма залежність, адже збільшення затрат праці зумовлює зростання собівартості, і навпаки.

Таблиця 4

Вплив продуктивності праці на собівартість виробництва продукції

Групи підприємств за рівнем затрат на виробництво 1 ц, люд.-г.	Кількість підприємств у групі	Собівартість 1ц, грн.
I-7-10	10	12,60
II-10-13	20	16,18
III-13-15	6	19,36
Всього	36	17,26

Притаманним для аналітичних групувань є те, що кожна група факторної ознаки характеризується середніми величинами результативної ознаки.

Таким чином, аналітичне групування відіграє важливу роль у вивченні зв'язків між ознаками, виявляючи цей зв’язок і створюючи можливість його кількісної характеристики.

Основні завдання, які вирішують статистичні групування, тісно пов'язані між собою, взаємно переплітаються і доповнюють одне одного. Тому розмежування трьох видів групувань (типологічні, структурні і аналітичні), коротка характеристика яких наведена вище, є певною мірою умовним. В окремих випадках одне й те саме групування дає можливість виявити типи явищ, охарактеризувати їх структуру і констатувати наявність певних взаємозв’язків між ознаками.

Складні явища, які мають багато істотних сторін (наприклад, сільськогосподарське виробництво), не можна всебічно охарактеризувати одним групуванням; необхідна, як правило, система їх (наприклад, групування за виробничим напрямом, за продуктивністю праці, за рівнем її фондоозброєності і т. д.).

У зв'язку із зазначеним вище, назва «аналітичні групування» не зовсім вдала, адже таке визначення рівною мірою стосується типологічних і структурних групувань, оскільки вони несуть в собі також пізнавальні, тобто аналітичні функції. Більш вдало, на наш погляд, їх суть відображує назва «факторно-результативні групування», адже їх призначення – виявляти зв’язки між факторними і результативними ознаками. Основою такого виду групувань є вивчення дії зміни однієї або кількох факторних ознак на зміну результативної ознаки.

Розрізняють групування за однією ознакою – просте групування і групування за двома і більше ознаками – комбінаційне групування. При комбінаційному групуванні групи, виділені за однією ознакою (наприклад, за виробничим напрямом), розбивають на підгрупи за другою ознакою (наприклад, за рівнем продуктивності праці). Комбінаційне групування має більш широкі аналітичні можливості, ніж просте, його використовують переважно для вивчення взаємозв’язків між ознаками. Порядок комбінації ознак обґрунтовується економічно і може бути легко змінений (при необхідності). Якщо по кожній з ознак є підсумкова група, комбінаційне групування можна «згорнути» у будь-якому напрямі в просте.

Наведемо приклад (схему) комбінаційного групування сільськогосподарських підприємств за трьома ознаками: за розміром площ сільськогосподарських угідь, якістю та рівнем удобреності ґрунтів (табл. 5).

В останньому підсумковому розділі, який об'єднує підприємства незалежно від розмірів площ сільськогосподарських угідь, знаходяться дані по підприємствах, згрупованих лише за двома ознаками: за якістю ґрунтів і рівнем їх удобреності. Вибираючи три підсумкових рядки в кожному розділі, знайдемо групи за розміром площ сільськогосподарських угідь і рівнем удобреності у колонках, об'єднаних у вертикальну рубрику «всього», – комбінацію розміру угідь і удобреності площ. Просте групування за розміром площ сільськогосподарських угідь знаходимо по горизонтальних рядках «всього» у тій же вертикальній рубриці; просте групування за якістю ґрунтів – у тій самій вертикальній рубриці у її нижньому розділі (останніх чотирьох рядках); просте групування за рівнем внесених добрив на одиницю площі – у останньому рядку таблиці.

Таблиця 5

Макет таблиці комбінаційного групування за трьома ознаками

Групи підприємств за площею с.-г. угідь, га	Підгрупи підприємств за якістю ґрунту , балів	Підгрупи підприємств за кількістю внесених мінеральних добрив на 1 га, ц діючої речовини			
		C_1 – до 2,0	C_2 – 2,0-2,5	C_3 – понад 2,5	Усього
A_1 – до 4600	B_1 – до 60 B_2 – 60-70 B_3 – понад 70				
	Усього				
A_2 – 4600-5000	B_1 – до 60 B_2 – 60-70 B_3 – понад 70				
	Усього				
A_3 – понад-5000	B_1 – до 60 B_2 – 60-70 B_3 – понад 70				
	Усього				
Усього	B_1 – до 60 B_2 – 60-70 B_3 – понад 70				
	Усього				

У кінці таблиці праворуч внизу дається загальний підсумок по всій сукупності сільськогосподарських підприємств (Σ).

За групувальні ознаки можна брати синтетичні результативні показники господарської діяльності (наприклад, собівартість 1 ц продукції, рівень рентабельності і т. ін.) і ознаки, які є факторами виробництва. У першому випадку групування називають **результативним,** у другому **– факторним**.

Групування за результативними ознаками дозволяють досить надійно виділити виробничі типи і дати в середньому характеристику їх особливостям. Але вони не дають можливості виділити всю різноманітність форм і показати ступінь впливу того чи іншого фактора на результат виробництва. Групування за факторами дає змогу показати різноманітність виникаючих форм і ступінь впливу того чи іншого фактора на результативні показники.

За допомогою факторних групувань встановлюються і вивчаються причинно-наслідкові зв'язки між ознаками однорідних явищ, виявляються фактори розвитку сукупності. Зокрема, це стосується аналітичних групувань, хоча аналітичні функції (як вже було сказано раніше) притаманні типологічним і структурним групуванням.

Факторні групування основані на вивченні того, як у масових явищах зі зміною одного або кількох факторних ознак змінюється результативна ознака.

При вивченні соціально-економічних зв'язків за допомогою факторних групувань треба дотримуватись певних правил. На думку І. П. Суслова, їх можна звести до таких.

1.Факторні ознаки служать основою групувань; з метою зручності і можливості зіставлення їх зображують, як правило, у вигляді рівних закритих інтервалів; число інтервалів беруть від 3 до 6-8, встановлюють його шляхом проб різних варіантів (методичні підходи щодо встановлення числа груп розглянуто в наступному параграфі).

2. Результативну ознаку дають у вигляді інтенсивних статистичних показників. Для більшої визначеності і наочності ці ознаки часто підлягають деяким перетворенням.

3. Оскільки найважливішою підставою вивчення зв'язків за допомогою факторних групувань є положення про можливість погашення впливу інших причин, то правильні висновки можуть бути зроблені лише на підставі груп і підгруп досить великого обсягу.

Далі автор відмічає, що факторні групування дають можливість робити висновки про наявність зв'язку, про його форму (прямий, обернений) і наближено характеризувати тісноту зв'язку.

Результати аналітичних групувань треба по можливості оцінювати з погляду їх істотності. Це завдання вирішується за допомогою розрахунків тих чи інших математичних критеріїв. Питання про їх застосування розглянуто далі.

В аналітичній роботі часто виникає питання – яким ознакам віддавати перевагу при групуванні: факторним чи результативним?

Факторні ознаки мають важливіше аналітичне значення, даючи можливість кількісно оцінити вплив окремих факторів на досліджувані явища. Результативні ознаки, оскільки в основу групувань покладено результат впливу багатьох факторів, дозволяють у певних випадках «бачити» загальні типові різниці незалежно від форм, в яких вони виступають. Такі групування досить ефективні в аналізі соціально-економічних явищ, коли ставиться на меті принципова розвідувальна оцінка зв'язку різних факторів, які наводяться у присудку таблиці, з групувальною результативною ознакою.

На жаль, цей вид статистичних групувань ще не знайшов широкого використання у аналітичній роботі науковців і практичних працівників у галузі економіки.

При статистичному групуванні велике пізнавальне значення має поєднання факторних і результативних ознак. У такому разі будуються комбінаційні групування за формою факторно-результативних або результативно-факторних. Тобто, одна з групувальних ознак є факторною, друга – результативною. Вибір схеми «факторно-результативна» чи «результативно-факторна» залежить від мети дослідження, знання природи економічних явищ і досконалого володіння методикою статистичних групувань.

§ 3.3. Методологія статистичних групувань

Науковому статистичному групуванню передує теоретико-економічний аналіз досліджуваного явища. і разом з тим використання сучасних статистичних методів дозволяє кількісно оцінити ступінь однорідності виділених груп, здійснити відбір істотних групувальних ознак, удосконалювати методику визначення величини інтервалів групувань.

Групування статистичної сукупності починають з вибору групувальних ознак. Але процедурі відбору ознак передує досить важливий етап дослідницької роботи, пов'язаний із з'ясуванням тенденцій розвитку явища, специфіки розвитку досліджуваних об'єктів та ін.

Від вибору групувальної ознаки залежить розв'язання питання про утворення груп. Групування за атрибутивною ознакою обмежується кількістю значень ознаки. Наприклад, поголів'я спортивних коней можна поділити лише на таку кількість груп за породним складом, скільки фактично є таких порід.

Після відбору групувальної ознаки постає питання про кількість груп, на які буде розподілена досліджувана сукупність, і про межі груп. Розв'язання даного питання залежить від конкретних умов і завдань.

На цьому етапі встановлюють величину і границі кожного інтервалу. Оскільки характер реально існуючих сукупностей та їх розподіл досить різноманітні, то існують різні методичні підходи у вирішенні питання про кількість груп. Загальним принципом, з якого треба виходити, є характер матеріалу та чисельність досліджуваної сукупності. Характерні особливості розподілу не виявляються, якщо при невеликій сукупності одиниць спостереження взяти велике або дуже мале число груп. До цього питання існують різні підходи. Розглянемо їх.

Групувальна ознака може змінюватися дискретно, тобто перервно і безперервно. Якщо мінливість ознаки має дискретний характер, число груп варіаційного ряду, як правило, визначається числом цих дискретних значень (якщо їх небагато). Наприклад, групування підприємств за наявністю виробничих бригад – 1,2,3 і т. д.

При мінливості ознаки безперервного характеру звертають увагу на ранжирований ряд. Якщо зростання рівнів групувальної ознаки відбувається з плавними переходами, перевага віддається **рівним інтервалам**. У разі стрибкоподібних змін групувальної ознаки будують групи з **нерівними інтервалами**. Границі у таких випадках встановлюють, як правило, в точках різких переходів.

Таким чином, у процесі групування за кількісною ознакою для обмеження окремих груп утворюють рівні або нерівні інтервали.

Питання визначення кількості груп в умовах порівняно поступових змін групувальної ознаки (у ранжированому ряду) може вирішуватися з різних методичних підходів.

Орієнтовно число інтервалів (груп) можна визначити шляхом добування квадратного кореня з обсягу досліджуваної сукупності. При цьому число інтервалів не повинно бути меншим 5 і більшим 20. Так. при чисельності вибірки 50 одиниць спостереження число інтервалів дорівнює 7 ($\sqrt{50}$).

Якщо сукупність невелика за обсягом, інтервальний ряд будують таким чином, щоб у крайні групи (першу і третю) потрапило по 25 % одиниць сукупності, а в середню – 50 %. У цьому випадку групування складається з трьох нерівних інтервалів. Наприклад, сукупність з 28 підприємств матиме розподіл: І група – 7 одиниць, ІІ – 14, ІІІ – 7 одиниць.

Визначення числа груп, запропоноване Стерджессом, полягає у розрахунку формули: $\eta_{інт} = 1 + 3{,}322\lg n$, де $\eta_{інт}$ – число груп (інтервалів); n – чисельність сукупності. Застосовуючи цю формулу, будемо мати сукупності розміром 10-100 одиниць 4-7 груп: 100-1000 одиниць – 7-10; 1000-10000 – 12-14 груп. Як бачимо, відносне зростання числа груп із - збільшенням сукупності відбувається досить інтенсивно в інтервалі 10–100 одиниць і уповільнюється в інтервалі 100-1000 одиниць. Майже зовсім відсутнє таке зростання у інтервалі 1000-10000 одиниць сукупності.

Потрібно відмітити, що підхід досить формальний і небезпечний, який звільнює від можливості економічного мислення. Адже підводити умовно кожний своєрідний емпіричний розподіл під єдиний тип без врахування особливостей конкретних сукупностей не можна.

Слід визнати найвдалішими рекомендації В. П. Левинського, який пропонує своєрідні нормативи числа інтервалів, зумовлені обсягами досліджуваної сукупності (табл. 6).

Таблиця 6

Рекомендоване число груп для різної кількості спостережень

Кількість одиниць спостережень	Рекомендоване число інтервалів (груп)
до 40	3-5
40-60	6-8
60-100	8-10
100-200	10-12
200-500	12-17

Якщо число одиниць спостереження налічується до 40, число інтервалів становитиме 3 або 5. Розподіл сукупності на 4 групи небажаний, адже в такому випадку втрачається середня група (інтервал).

Перевага рекомендації В. П. Левинського у порівнянні з рекомендацією Стерджесса у тому, що вона не так жорстко пов’язує число груп з чисельністю одиниць спостереження. А в такому разі дослідникові надається можливість певного вибору числа груп залежно від характеру сукупності. В економічних дослідженнях найбільш поширений обсяг сукупності 100-500 одиниць. За формулою Стерджесса число груп дорівнюватиме 7-10, за рекомендацією В. П. Левинського – від 10 до 17 груп.

Слід пам’ятати, що кількість обраних інтервалів (груп) залежить від коливності групувальної ознаки: чим воно більше, тим більше треба утворювати груп. Треба також намагатися, щоб виділені групи були достатньо заповнені одиницями спостереження. Наявність незаповнених Інтервалів або потрапляння в них лише окремих одиниць сукупності – результат того, що невдало обрано інтервали, кількість їх взята, ймовірно, зайва. Наявність малонаповнених інтервалів (груп) має право на існування лише по краях групування, де концентруються характеристики як передових, так і відстаючих показників за розміром відносно середнього рівня. Особливо це стосується структурних групувань. Кількість груп тут не повинна бути досить великою чи досить малою. У першому випадку є ризик загубитися у дрібницях, у другому – не виявити досить важливі властивості досліджуваної сукупності. Оптимальна кількість інтервалів дозволяє викрити всі істотні особливості досліджуваної сукупності.

Отже, якщо вирішено питання про визначення числа груп, на яке буде поділена сукупність, вихідні варіанти розташовують у ранжирований ряд за групувальною ознакою. В умовах відсутності ускладнюючих обставин, тобто наявності порівно поступових змін факторної ознаки, найпростішим способом визначення величини інтервалу при побудові рівновеликих інтервалів буде відношення:

$$i = \frac{x_{\max} - x_{\min}}{n},$$

де i – величина інтервалу; $x_{\max}$ і $x_{\min}$ – відповідно максимальна і мінімальна варіанти; n – задане число груп (інтервалів).

У випадках, коли невелика частина сукупності значно віддалена за розміром групувальної ознаки від сукупності основного масиву, за $x_{\max}$ приймається максимальна варіанта основного масиву.

Існують рекомендації щодо встановлення величини інтервалу групувань з деякими поправками до попередньої формули. У цьому випадку формула набуває вигляду: $i = \frac{x_{\max} - x_{\min} + 1}{n}$.

У випадках, коли максимальне і мінімальне значення у ранжированому ряду групувальних ознак значно відрізняється від решти показників, за $x_{\max}$ приймається суміжне наступне значення ознаки $x_{\max+1}$, а за $x_{\min}$ суміжне попереднє її значення $x_{\min-1}$.

У ряді випадків вихідна величина інтервалу групувань задається дослідником, а число груп у такому разі є похідним, тобто: $\frac{x_{\max} - x_{\min}}{i} = n$

Якщо розрахована величина рівного інтервалу становить дробове число, його заокруглюють до цілого, цим самим розширюючи границі, якими охоплює інтервал розмаху коливності значень групувальної ознаки.

Маючи встановлене число інтервалів і величину інтервалу, визначають границі інтервалів (груп). Так, нижня границя першого інтервалу (x_1) встановлюється за мінімальною варіантою ($x_{\min}$). Верхня границя цього інтервалу (x_2) дорівнюватиме ($x_{\min+i}$). Нижня границя другого інтервалу відповідає (умовно) верхній границі першого інтервалу (x_2), а верхня границя другого інтервалу (x_3) дорівнюватиме ($x_2 + i$) і т. д. При встановленні границь інтервалів (груп) необхідно пам'ятати, що верхня границя завжди менша від нижньої границі наступного інтервалу на ціну поділки, тобто одиницю виміру.

Приклад. Дані про середньоденну зарплату 57 підприємств згрупувати, утворивши групи з рівними інтервалами.

На першому етапі визначають кількість інтервалів. Згідно з вищезазначеним, для сукупності одиниць спостереження 40-60 рекомендована кількість інтервалів дорівнює 6-8. Вибираємо число інтервалів 7, тобто поділимо сукупність на 7 груп за розмірами показників зарплати.

Вихідні дані: 29,3; 31,0; 21,5; 21,4; 28,3; 35,7; 37,6; 19,8; 23,8; 21,6; 32,8; 27,6; 42,7;27,2; 32,3; 30,1; 30,2; 25,8; 24,6; 25,4; 29,8; 28,4; 21,7; 27,5; 23,8;37,4; 26,7; 16,5; 29,0; 21,1; 36,2; 29,6; 21,1; 26,3; 21,5; 27,5;29.5; 24,3; 21,3; 30,4; 30,4; 39,5; 25,8; 26,6; 24,4; 32,3; 26,6;25,9; 32,8; 29,3; 32,3; 25,3; 32,6; 21,5; 23,3; 27,1; 29,6.

Розмістивши варіанти в ранжирований ряд, маємо: 16,5; 19,3; 19,8; ... 37,6; 39,5; 42,7.

Крок інтервалу дорівнюватиме:

$$i = \frac{x_{\max} - x_{\min}}{n_{iipm}} = \frac{42{,}7 - 16{,}5}{7} \approx 4$$

Заокругливши до цілих варіанти, розрахуємо нижні і верхні границі інтервалів: 1 – 16+4 – 20 (тобто 16–20); II – 20+4 (тобто 20–24) тощо.

Будуємо макет таблиці групованого розподілу частот результатів спостереження (табл. 7).

Таблиця 7

Групування підприємств за рівнем денної зарплати

Групи підприємств за рівнем зарплати, грн.	Кількість підприємств
I - до 20	3
II -20-24	11
III - 24-28	17
IV -28-32	14
V -32-36	7
VI - 36-40	4
VII –понад 40	1
Всього	57

Наведений у таблиці ряд пар чисел складає емпіричний розподіл частот n_i- за значеннями x_i.

Сума частот дорівнює обсягу вибіркової сукупності

($\sum n_i = n = 57$).

Як уже було зазначено раніше, застосування методу статистичних групувань у дослідженні соціально-економічних явищ (так само як і інших видів явищ) повинно ґрунтуватися на знанні теоретичних положень і їх вимог. Чисто емпіричний підхід до узагальнення матеріалів спостереження може призвести до того, що дані, зібрані за науковими принципами і ретельно перевірені, можуть виявитися непридатними для поглибленого вивчення того чи іншого явища.

Теорія групувань вимагає одержання всебічної характеристики досліджуваного явища або його типів. Виділити і охарактеризувати типи можна лише за умов попереднього теоретичного висвітлення факторів, при поєднанні статистичних методів узагальнення з теоретичними положеннями наук, що вивчають дане явище.

Одним з основних положень теорії групувань вважається виділення із всієї різноманітності зв’язків основного процесу, який визначає всі інші зміни явища і веде до якісних перетворень.

На наступному етапі теоретичного обґрунтування з’ясовують, які нові якісні зміни відбуватимуться в ході розвитку даного процесу, тобто, які нові типи даного явища знаходять свій прояв і які виявляються їх найбільш істотні риси.

Викладене вище дає підстави стверджувати, що практичному застосуванню методу статистичних групувань передує ретельний теоретичний аналіз факторів, виявлення головного напряму розвитку досліджуваного явища і виділення із складної сукупності окремих груп одиниць, які належать до різних типів.

Але тут слід відзначити, що попереднє теоретичне вивчення даних при групуваннях не є догмою і не означає, що метод групувань відіграє певну технічну, тобто пасивну, роль в аналізі. Це зовсім не так. Використання статистичних групувань дає змогу одержати кількісну характеристику стану досліджуваних явищ, виявити якісні перетворення, перевірити наукові гіпотези відносно напряму розвитку явища і цим самим збагатити теорію питання, поставленого на дослідження.

Якщо вивчено зміст основного процесу і встановлено типи явищ, приступають до з'ясування основних форм, в яких здійснюється розвиток типів явищ. Відповідно до форм розвитку явищ відбирають найбільш істотні ознаки, які дозволяють виділити групи із якісно однорідних одиниць спостереження. Врахування форм розвитку явища має велике значення при застосуванні статистичних групувань. Нехтування цим методичним положенням може призвести до суб'єктивних висновків за результатами групувань, адже в такому разі є ймовірність змішування явищ і викривлення дійсних кількісних характеристик. Наприклад, якщо згрупувати сільськогосподарські підприємства області за чисельністю поголів'я великої рогатої худоби, можна дійти висновків, що зі зменшенням поголів'я з розрахунку на одне підприємство підвищується ефективність виробництва, що слід вважати необ'єктивним. Такий суб'єктивний висновок пояснюється тим, що при групуванні не враховано форми розвитку типів підприємств, оскільки в групу нечисленних за кількістю худоби підприємств потрапили такі, що мають високорозвинений рівень виробництва взагалі або спеціалізуються на виробництві окремих видів рослинницької продукції, маючи для цього відповідні оптимальні умови виробництва. Відповідно до форм розвитку типів при такому групуванні повинні враховуватися ознаки, які характеризують безпосередньо і розмір тваринництва, і характер виробництва (його інтенсивність, концентрацію, спеціалізацію і т.ін.).

Таким чином, метод статистичних групувань дає об'єктивні результати в аналізі лише за умов, коли за виділеними групами буде розраховано комплекс найбільш істотних статистичних показників, що характеризують основні сторони і взаємозв'язки досліджуваних явищ. Відбір показників здійснюють з урахуванням теоретичних положень окремих наук, які розкривають якісні особливості суті досліджуваних процесів, а також з урахуванням вимог статистичної науки, яка вимагає наявності достатньо великої чисельності одиниць спостереження у групах і застосування найбільш істотної форми показників. Важливим моментом у практичному використанні результатів групувань слід вважати процес перевірки їх на вірогідність. Це питання потребує детального розгляду окремо.

На початковому етапі здійснення статистичних групувань перевіряють «сумнівні» варіанти на належність їх до ряду розподілу. Із цією метою використовують τ-критерій. Так, перш ніж розрахувати величину рівновеликого інтервалу, попередньо оцінюються крайні варіанти ранжированого ряду розподілу на належність їх до останнього. Критерієм належності сумнівних варіант до досліджуваної сукупності виступає стандартизоване відхилення значень сумнівних варіант (це, як правило, мінімальна і максимальна й близькі до них варіанти) від середньої. Розмір стандартизованого відхилення не повинен перевищувати число 3, тобто:

$$\tau = \frac{x_i - \bar{x}}{\sigma_x} \le 3,$$

де τ – критерій належності; x_i – максимальне і мінімальне значення групувальної ознаки; σ_x – середнє квадратичне відхилення.

Розглянемо розрахунок названого критерію на прикладі вибіркової сукупності показників середньоденної зарплати, наведеної вище, s визначимо належність максимальної (42,7) і мінімальної (16,5) варіанти до цієї сукупності (табл. 8). За даними робочої таблиці 8, обчислюємо $\bar{x}$ і σ_x.

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{1590}{57} = 28; \ \sigma = \sqrt{\frac{\sum (x_1 - x)^2 n_1}{\sum n_1}} = 5{,}39.$$

Таблиця 8

Розрахунок τ - критерію по вибірковій сукупності показників денної зарплати

Інтервал	Варіанти (центр) x_i	Частота	Розрахункові дані			
			$x_i n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
До 20	18 (умовно)	3	54	-10	100	300
20-24	22	11	242	-6	36	396
24-28	26	17	442	-2	4	68
28-32	30	14	420	2	4	56
32-36	34	7	238	6	36	252
36-40	38	4	152	10	100	400
Понад 40	42 (умовно)	1	42	14	196	196
Всього	-	57	1590	-	-	1668

$$\tau_{\max} = \frac{42{,}7 - 28}{5{,}39} = 2{,}73; \ \tau_{\min} = \frac{16{,}5 - 28}{5{,}39} = -2{,}13.$$

По одержаних результатах розрахунків τ - критерію робимо висновок, що максимальна і мінімальна ознаки у досліджуваній сукупності є типовими для неї, адже їх розміри не перевищують числа 3 (–2,13, 2,73<3).

(Оскільки обчислення спеціальних параметрів $\bar{x}$ і σ_x буде предметом розгляду спеціальних тем, розрахунок даних статистичних характеристик тут не коментується).

МОДУЛЬ 2

ТЕМА 4. УЗАГАЛЬНЮЮЧІ СТАТИСТИЧНІ ПОКАЗНИКИ

§ 4.1. Абсолютні показники, їх значення

У системі узагальнюючих статистичних показників мають широке застосування абсолютні показники, адже за їх допомогою можна одержати характеристики різних сторін соціально – економічних явищ: чисельність працівників певної галузі, кількість підприємств, фонд заробітної плати тощо.

Абсолютними показниками в статистиці називають показники, які виражають розмір (обсяг, рівень) кількісних ознак досліджуваних явищ. Це не абстрактні, а іменовані числа, які виражають розміри суспільних явищ у певних одиницях виміру.

Статистика розглядає індивідуальні і загальні (сумарні) абсолютні показники. **Індивідуальні** абсолютні показники відображають розміри кількісних окремих одиниць досліджуваної сукупності, їх одержують у процесі статистичного спостереження (наприклад, посівна площа сільсьгосппідприємства, обсяг виробництва валової продукції конкретним приватним підприємством тощо).

Загальні абсолютні показники характеризують розмір кількісної ознаки деякої (певної) сукупності одиниць (наприклад, посівні площі державних аграрних підприємств району поголів'я великої рогатої худоби у приватному секторі району і т.ін.). Сумарні абсолютні показники одержують, як правило, шляхом додавання індивідуальних абсолютних показників; інколи їх обчислюють шляхом множення. В окремих випадках абсолютні показники одержують шляхом відповідних розрахунків. Так, у проміжках між переписом населення чисельність його на певну дату визначають шляхом розрахунків.

Загальні абсолютні величини мають велике пізнавальне значення і знаходять широке використання в управлінні народним господарством країни. Зокрема, це абсолютні показники про чисельність трудових ресурсів, виробничі потужності, розміри посівних площ тощо.

Як було зазначено вище, абсолютні показники – це завжди іменовані числа. У ролі їх вимірників у статистиці застосовуються різноманітні одиниці виміру: натуральні, умовно – натуральні, вартісні і трудові.

Одиниці виміру, які застосовуються для визначення кількості окремих видів матеріальних благ у їх натуральному виразі, тобто як певних споживних вартостей, називають **натуральними.** Останні широко застосовуються при визначенні розмірів виробництва і споживання різних видів продуктів. Отже, натуральні показники відображують розміри тих чи інших явищ у фізичних мірах, тобто мірах маси (тоннах, центнерах, кілограмах), довжини і площі (метрах, гектарах), об'єму (кубометрах).

Для вимірювання об'ємів однорідних, але неоднакових явищ використовують **умовно-натуральні** одиниці виміру (наприклад, умовне поголів'я тварин, умовні еталонні трактори, умовні гектари тощо).

За одиницю виміру для умовно-натуральних показників беруть, як правило, будь-яку фізичну одиницю, наділену в певних розмірах властивістю, характерною для продукції конкретного виду. Так, при розрахунках потреби у кормах застосовують коефіцієнти перерахунку худоби в умовні голови: для корів і для бугаїв-плідників – 1; молодняка великої рогатої худоби – 0,6; свиней – 0,3; овець і кіз – 0,1; коней – 1,0; птиці – 0,02.

Умовно-натуральні одиниці виміру дають змогу привести певні різновидності явищ до порівнянного виду і тим визначити їх загальний обсяг у натуральному виразі.

Для відображення розмірів деяких окремих складних явищ застосовують **комбіновані** одиниці виміру. Так, для визначення обсягів транспортних робіт автомобільного парку виконану роботу вимірюють у тонно-кілометрах (1 т.-км дорівнює перевезенню 1 т вантажу на відстань 1 км).

При визначенні затрат робочого часу на виробництві певного виду продукції застосовують комбіновану одиницю виміру (людино-годину). Для забезпечення масовості даних дослідження шляхом підсумовування кількості підприємств за кілька років одержують одиниці спостереження в комбінованих одиницях виміру – підприємство-роки.

Поряд із зазначеними вище видами одиниць виміру в статистиці досить широко використовують трудові і вартісні вимірники.

До трудових вимірників належать комбіновані одиниці виміру: людино-година, людино-день, людино-рік. Абсолютні показники в зазначених одиницях виміру застосовують при визначенні обсягів трудових ресурсів, розрахунку показників продуктивності праці, затрат праці на різну продукцію чи послуги тощо.

За вартісний вимірник приймають грошові одиниці виміру – гривні, долари, франки та ін. У грошовому виразі визначають вартість виробленої і реалізованої продукції (за всіма її вилами) в підприємстві, вартість фондів, розміри доходів тощо. Абсолютні показники у вартісному вимірі дозволяють визначити загальні розміри багатьох народногосподарських економічних показників (наприклад, національний доход, обсяг виробленої і реалізованої продукції в країні і т. ін.).

Слід зазначити, що абсолютні показники інколи наводяться в укрупнених вимірниках. Так, дані про обсяг продукції можуть бути наведені у тисячах, мільйонах або мільярдах гривень. Пояснюється це тим, що укрупнені вимірники дають точніше уявлення про розміри досліджуваних явищ.

§ 4.2. Відносні показники, їх види і форми

Досліджуючи економічні явища чи процеси, статистика не обмежується розрахунком тільки абсолютних показників, яку б велику роль вони не відігравали в аналізі. Адже жодне явище не може бути зрозумілим, якщо його розглядати поза зв'язком з іншими явищами. Із цією метою абсолютним показникам дають порівняльну оцінку за допомогою **відносних показників**. Тобто останні є результатом зіставлення абсолютних показників. Значення відносних показників для аналізу досить велике, адже за їх допомогою порівнюють характеристики окремих одиниць груп і сукупностей у цілому, вивчають структуру явищ та закономірності їх розвитку, аналізують виконання плану, вимірюють темпи розвитку і інтенсивність поширення суспільних явищ.

За формою відносний показник являє собою дріб, чисельником якого є величина, котру порівнюють (в окремих випадках її називають поточною, або звітною), а знаменником – величина, з якою здійснюють порівняння. Знаменник відносної величини вважається **базою порівняння**. Так, питому вагу висококваліфікованих працівників підприємства розраховують діленням кількості осіб з високим рівнем кваліфікації на загальну чисельність працюючих. Базою порівняння у наведеному прикладі є загальна чисельність працюючих.

Якщо базову величину показника приймають за одиницю, формою її зображення буде коефіцієнт (кратне відношення), якщо за 100 – формою зображення відносних показників будуть проценти.

Коефіцієнт як форма виразу відносної величини показує, у скільки разів порівнювальна величина більша базової (чи яку частину від неї становить, якщо величина коефіцієнта менша за одиницю).

У статистичній практиці коефіцієнти, як правило, використовують для вираження відносних величин у випадках, коли порівнювальна величина перевищує базову більш як у 2–3 рази. Якщо таке співвідношення має менші розміри – застосовують процентні числа. У випадках, коли базову величину приймають за 1000, відносні показники виражають у проміле (‰). Наприклад, якщо питома вага осіб сільського населення району з вищою освітою становить 16 ‰, це означає, що на кожну 1000 сільського населення у середньому припадає 16 чоловік з вищою освітою.

В окремих випадках відносні показники розраховують на 10000 (продециміле), 100000, 1000000 одиниць (наприклад, у статистиці охорони здоров'я розраховують кількість ліжко-місць на 10000 населення).

Відносні величини, виражені на 1000, 10000, 100000 і т. д. одиниць, вживають з метою надання їм більш придатного для сприйняття вигляду, оскільки, підібравши вдало базу порівняння, можна запобігти дробовим числам.

Форму виразу відносного показника вибирають у кожному конкретному випадку залежно від характеру одиниць спостереження і результатів, які одержують при зіставленні однієї величини з іншою.

Залежно від пізнавального значення відносні показники, які використовує статистика, класифікують за такими ознаками:

1) відношення між однойменними показниками;

2) відношення між різнойменними показниками.

Перша група являє собою відносні величини, які не мають розмірності, їх виражають, як правило, у процентах або у коефіцієнтах. Показники цієї групи досить різноманітні, за призначенням їх поділяють на такі види: 1) відносні величини структури; 2) відносні величини виконання плану; 3) відносні величини виконання планового завдання; 4) відносні величини динаміки; 5) відносні величини порівняння.

Друга група відносних показників включає: 1) відносні величини інтенсивності; 2) відносні величини координації.

Відносні показники структури характеризують склад того чи іншого суспільного явища, тобто показують, яку питому вагу займають окремі частини в усьому явищі. Розраховують їх відношенням частини до цілого. Виражаються вони в процентах або частках одиниці. Наприклад, загальні витрати на виробництво продукції становлять 600 грн, а витрати на оплату 240 грн. Отже, питома вага витрат на оплату праці у загальній сумі витрат становить 240:600 = 0,4, або 40 %.

Відносні показники виконання плану – це відношення фактичного рівня показника до рівня, запланованого на той же період. Наприклад, якщо було заплановано одержати урожайність зернових культур 46 ц/га, а фактично одержано 49,8 ц з одиниці площі, то відносна величина виконання плану становить (49,8 : 46) • 100 = 107,8 %, тобто план виконано на 107,8 %, або перевиконання становить 7,8 %.

Відносні показники виконання планового завдання являють собою відношення величини показника, встановленого на плановий період, до його величини, яка досягнута фактично на цей період, або будь-якої іншої, прийнятої за базу порівняння. Тобто, це відношення планового рівня у наступному періоді до фактичного рівня звітного періоду, прийнятого за базу порівняння. Так, встановлюється завдання: підвищити продуктивність праці щодо попереднього періоду на 16 % або знизити собівартість на 10 %.

До речі, якщо підійти критично до цього питання, то відносні величини даного виду не є показниками статистики. А розглядають їх у статистиці щодо дійсного їх зв'язку із статистичними відносними величинами, зокрема з показниками виконання плану.

Відносні показники динаміки характеризують зміни суспільних явищ і процесів у часі. Розраховують їх відношенням рівня відповідного наступного періоду до рівня попереднього періоду, або будь-якого іншого, прийнятого за базу порівняння. Відповідно до обраної бази порівняння можуть бути ланцюгові і базисні. **Ланцюгові відносні величини динаміки** визначають відношенням рівнів наступного і попереднього періодів**. Базисні відносні величини динаміки** розраховують відношенням рівня відповідного наступного періоду до певного рівня, прийнятого за базу порівняння.

Відносні показники порівняння – це результат зіставлення одних і тих же характеристик двох різних сукупностей, груп чи одиниць. Порівнюють при цьому будь-які кількісні характеристики: обсяги сукупностей (або груп), середні або сумарні величини тієї чи іншої ознаки. Наприклад, порівнюючи кількість автомобілів станом на початок року по двох підприємствах, одержимо відносну величину порівняння, яка дорівнює $\frac{66}{88} = 0{,}75$ (або 75 %), тобто в порівняльному підприємстві чисельність автомобілів на 25 % менша, ніж у базовому підприємстві.

Відносні показники координації характеризують співвідношення між складовими частинами цілого. Одну з частин цілого приймають за базу порівняння і знаходять відношення до неї всіх інших частин. Наприклад, за результатами перепису населення встановлюють співвідношення народжень хлопчиків і дівчаток (у розрахунку на 100 народжених певної статі). Або інший приклад. За результатами спостереження встановлено, що в підприємствах району з розрахунку на 100 чоловіків працює 116 жінок. Окремі автори схильні вважати відносні величини координації відносними показниками структури, що не зовсім вірно. Адже вони не дають уявлення про структуру явищ, а лише визначають, скільки одиниць певної частини цілого припадає на іншу її частину, прийняту за базу порівняння.

Відношення між різнойменними (різноякісними) показниками називають **відносними показниками інтенсивності**, або статистичними коефіцієнтами. Вони відображують ступінь поширення одного явища порівняно з іншими взаємопов'язаними явищами. До них належать показники щільності поголів'я тварин у розрахунку на 100 га сільськогосподарських угідь (ріллі, площі зернових). Відносні показники цієї групи виражають завжди іменованими числами. При цьому у їх назву входять найменування одиниць виміру обох порівнюваних ознак.

Слід відзначити, що відносні показники можуть бути середніми величинами (наприклад, середній темп зростання, середній темп приросту, середній процент виконання плану тощо), їх розраховують двома способами: 1) як середні з окремих відносних показників; 2) як відношення двох сумарних абсолютних показників, що включають абсолютні величини, покладені в основу розрахунку окремих відносних величин. Розглянемо приклад обчислення середнього темпу зростання робітників підприємства за вихідними даними таблиці 9.

Таблиця 9

Чисельність робітників підприємства , чол.

Показники	Разом	У тому числі		
		Цех № 1	Цех № 2	Цех № 3
Середньорічна чисельність:				
базисний рік	3600	1609	816	1175
поточний рік	3900	2108	929	863
Темп зростання, %	108,3	131,0	113,8	73,4

Як бачимо, за розрахунками співвідношень абсолютних показників відносний показник динаміки загальної чисельності робітників дорівнює 108,3 % (3900 : 3600)•100. Відносні показники динаміки чисельності за номерами цехів становлять відповідно: по цеху № 1 – 131,0 %, цеху № 2 – 113,8, цеху № 3 – 73,4 %. Відносний показник динаміки загальної чисельності робітників можна розрахувати і як середній з відносних показників динаміки за відповідними номерами цехів, а саме: (1,310х1609 + 1,138х816 + 0,734х1175) : 3600 = 3898,8:3600 = 1,083, або 108,3 %.

Наведені розрахунки дають підстави зробити висновок, що середні відносні показники є загальними відносними показниками, адже вони характеризують загальну зміну по всій сукупності, загальне відношення однієї сукупності до іншої. Звідси можна зробити висновок, що як абсолютні, так і відносні показники можуть бути середніми. І ті, і другі наділені рядом важливих загальних властивостей, у зв'язку з чим у статистиці з абсолютними і відносними показниками відокремлюються як особливий вид **середн**і **показники**. Останні розглянуті у розділі.

Таким чином, з наведеного випливає, що зіставлення статистичних показників здійснюється в різних формах і напрямах. Відповідно до різних завдань і напрямів зіставлення статистичних показників застосовують різні види відносних величин. Розглянуті види відносних величин зводяться до класифікації, схематично представленої на рисунку 2.

Наведена класифікація наочно ілюструє можливості зіставлення статистичних даних, які належать до різних періодів, різних об'єктів і різних територій.

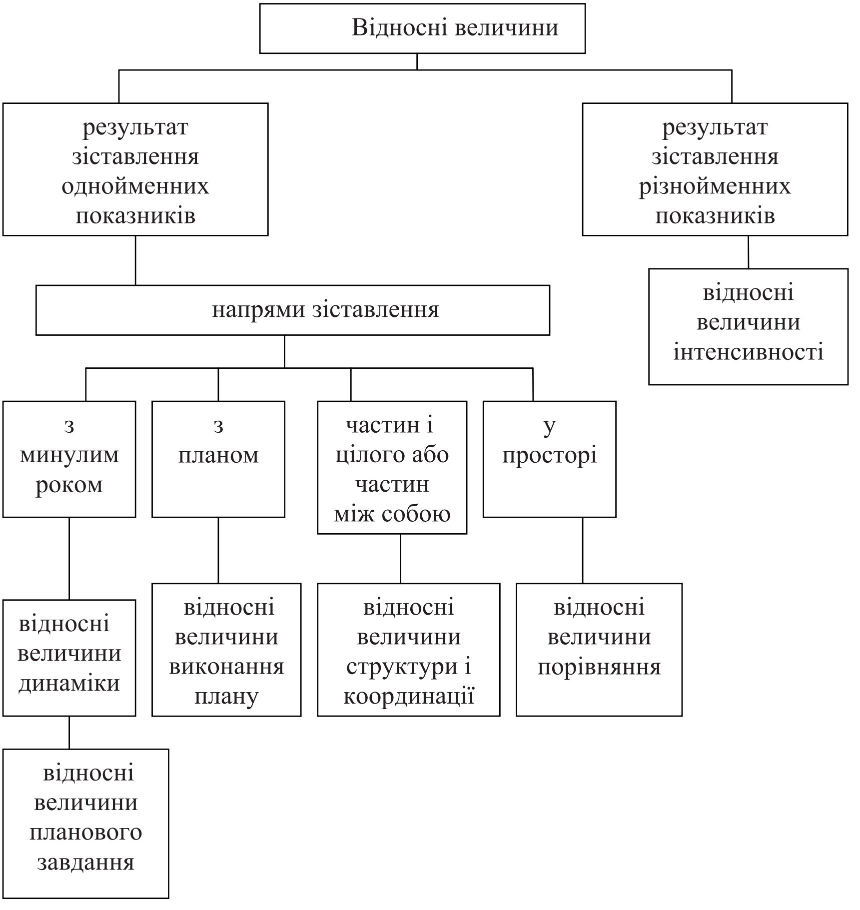


Рис. 2 Класифікаційна схема відносних величин

§ 4.3. Середні величини як характеристики ряду

При зоровому сприйнятті показників рядів розподілу і їх графіків переконуємося, що розмір варіант має деякі загальні закономірності, які проявляються в тому, що їх величини групуються навколо центра розподілу. За даними статистичного ряду при віддаленні від центра розподілу вгору і донизу, а при графічному зображенні при віддаленні вправо і вліво частоти постійно спадають.

Тенденція значень ознаки групуватися навколо центра розподілу частот, статистичною характеристикою якого є середня арифметична $(\bar{x})$, називається **центральною тенденцією.**

Таким чином, виникає необхідність розрахунку характеристик статистичних рядів розподілу.

Найважливішою характеристикою варіаційного ряду розподілу є **середня величина. Статистичні середні** відображують об'єктивну наявність певних умов, які проявляються в кожній одиниці досліджуваної сукупності; вони дають узагальнюючу кількісну характеристику статистичним сукупностям однотипних явищ по варіаційній ознаці. Середня узагальнює або являє собою весь діапазон даних і є результатом абстрагування відмінностей, що притаманні одиницям сукупності. В ній нівелюються випадкові відхилення, властиві індивідуальним значенням ознаки, яка вивчається, а також відображаються загальні умови, що формують досліджувану сукупність.

Середні величини можуть бути одержані в результаті багаторазових вимірювань однієї й тієї ж ознаки (величини). Середні одержують і при вимірюванні багатьох однорідних величин.

Розрахунок середніх передбачає обов'язковість обліку умов виникнення кожної індивідуальної величини, інакше обчислення можуть призвести до фіктивних середніх. Щоб середня величина відбражувала типове і загальне для всієї сукупності, остання повинна бути якісно однорідною.

Статистика розрізняє два типи середніх величин: **об'ємні** і **структурні.** Математична статистика поділяє об'ємні середні величини на види: **1) середня арифметична; 2) середня геометрична; 3) середня гармонійна; 4) середня квадратична (кубічна) і т.д.**

Названі вище види середніх величин можна одержати з формули **степеневої середньої.** Для незгрупованих даних формула степеневої середньої має вигляд: $\bar{x} = \sqrt[k]{\frac{\Sigma x^{к}}{n}}$.

Якщо дані згруповані і мають відповідні частоти (n_i), середня степенева визначається за формулою середньої зваженої:

$$\bar{x} = \sqrt[k]{\frac{\Sigma x_i^k n_i}{\Sigma n_i}},$$

де $\bar{x}$ – степенева середня;

k – показник степеня, що визначає вид середньої;

x – варіанта;

n – частота ($n = \Sigma n_i$).

Середня арифметична. Якщо в формулу степеневої середньої поставити значення к = 1, отримаємо середню арифметичну:

а) просту (незважену) $\bar{x} = \frac{\Sigma x}{n}$;

б) зважену $\bar{x} = \frac{\Sigma x_i n_i}{\Sigma n_i}$.

Так, якщо є $x_1, x_2, x_3, \ldots, x_n$,то середня арифметична проста вираховується за формулою:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}.$$

У розгорнутому вигляді формула середньої арифметичної зваженої має вигляд: $\bar{x} = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \ldots + x_n n_n}{n_1 + n_2 + n_3 + \ldots + n_n}.$

Приклад. За даними про виробництво валової продукції обчислити середню арифметичну. Дані берем з таблиці 10.

Таблиця 10

Вихідні і розрахункові дані

Вироблено продукції, млн.грн., (x_i)	Кількість підприємств (n_i)	$x_i n_i$
17,6	7	123,2
21,6	11	237,6
25,6	18	460,8
29,6	9	266,4
33,6	5	168,0
37,6	5	188,0
41,6	2	83,2
Всього	57	1527,2

Одержуємо: $\bar{x} = \frac{\Sigma x_i n_i}{\Sigma n_i} = \frac{1527,2}{57} = 26,8$.

Середня арифметична як математична функція має ряд математичних властивостей. Розглянемо їх.

1. Величина середньої арифметичної не змінюється, якщо частоти ряду розподілу замінити частостями.

2. Якщо до варіант ряду додати (або відняти) одну й ту ж величину, то середня арифметична, обчислена з нових (змінених) варіант, збільшиться (або зменшиться) на цю ж величину.

3. Якщо варіанти ряду помножити або поділити на одну ту ж величину, то середня арифметична зі змінених варіант буде відповідно більшою або меншою в стільки ж разів.

4. Алгебраїчна сума відхилень окремих варіант від середньої арифметичної ряду дорівнює нулю.

5. Сума квадратів відхилень від середньої арифметичної завжди менша, ніж сума квадратів відхилень від будь-якої іншої величини.

Перелічені вище властивості середньої арифметичної дозволяють застосовувати спрощені прийоми (способи) її розрахунку: ці питання докладно вивчаються в курсі загальної теорії статистики.

Середня гармонійна. Одержують її при підстановці у формулу степеневої середньої значення k = - 1.

$$\bar{x}_{гм} = \sqrt[-1]{\frac{\Sigma x^{-1}}{n}} = \frac{1}{\Sigma \frac{1}{x}} = \frac{n}{\Sigma \frac{1}{x}}.$$

Середня гармонійна зважена має вигляд:

$$\overline{x_{гм}} = \frac{\Sigma w}{\Sigma \frac{w}{x}}$$

Як видно, середня гармонійна являє собою обернену величину середньої арифметичної з обернених величин даних чисел.

Приклад. За даними про собівартість одиниці продукції і загальні витрати визначити середню гармонійну зважену (табл. 11).

Таблиця 11

Вихідні і розрахункові дані

Собівартість одиниці продукції, грн. (x_i)	Загальні витрати, грн. (w)	$\frac{w}{x}$
20	4000	200
25	5000	200
30	9000	300
x	Σw =18000	$\Sigma \frac{w}{x}$ =700

$$\overline{x_{гм}} = \frac{\Sigma w}{\Sigma \frac{w}{n}} = \frac{4000+5000+9000}{\frac{4000}{20}+\frac{5000}{25}+\frac{9000}{30}} = \frac{18000}{700} = 26$$ грн.

При розрахунку середньої гармонійної можна значно спростити обчислювальну роботу, якщо використовувати в розрахунках таблицю обернених чисел.

Середня геометрична. Цей вид середньої отримуємо, якщо у формулу степеневої середньої підставити значення k = 0.

$$\bar{x} = \sqrt[\kappa]{\frac{\Sigma x^{\kappa}}{n}} = \sqrt[0]{\frac{\Sigma x^{0}}{n}} = \left(\frac{\Sigma 1}{n}\right)^{\frac{1}{0}} = \left(\frac{n}{n}\right)^{\infty} = 1^{\infty}.$$

Розкриття невизначеності цього виду досягається логарифмуванням обох частин рівностей $\bar{x} = \sqrt[\kappa]{\frac{\bar{x}^{\kappa}}{n}}$.Таким чином, маємо:

$$\ln\bar{x} = \frac{1}{\kappa}(\ln\Sigma x^{\kappa} - \ln n) = \frac{\ln\Sigma x^{\kappa} - \ln n}{n}.$$

Диференціюючи по змінній **к,** одержуємо:

$$\lim \quad (\ln\bar{x}) = \underline{lim}\frac{\Sigma x^{\kappa} \times \ln x}{1 \times \Sigma x^{\kappa}} = \frac{\Sigma \ln x}{n}; \quad \ln\bar{x} = \frac{\Sigma \ln x}{n}.$$

Потенціюючи останній вираз, знаходимо середню:

$$\overline{x_{гр}} = \sqrt[n]{x_1 \times x_2 \times x_3 \times \ldots \times x_n}.$$

Отже, середня геометрична являє собою корінь степеня числа спостережень (n) з добутку даних чисел : $x_1, x_2, x_3, \ldots, x_n$. Після логарифмування маємо:

$$\lg\bar{x} = \frac{\lg x_1 + \lg x_2 + \lg x_3 + \ldots + \lg x_n}{n}.$$

Значення *x* дорівнює антилогарифму $\lg\bar{x}$.

Таким чином, середню геометричну можна розглядати як антилогарифм середньої арифметичної з логарифмів даних чисел. Цей вид середньої застосовують при розрахунку середнього коефіцієнта зростання за певний період в рядах динаміки.

Приклад. За даними про чисельність працюючих за 6 років знайти середній щорічний коефіцієнт зростання за період 1997-2002 рр. Проміжні розрахунки наведені в таблиці 12.

Таблиця 12

Вихідні і розрахункові дані для обчислення середньої геометричної

Рік	Чисельність працюючих	Коефіцієнт зростання, к	Логарифм числового значення коефіцієнта зростання
1997		-	-
1998	670	1,07545	0.0315
1999	728	I.08657	0,0363
2000	800	1,09891	0,0411
2001	883	1,10375	0,0429
2002	906	1.02605	0,0111
Сума	х	х	0,1629

За даними таблиці 12 знаходимо середнє значення з логарифмів числових значень коефіцієнтів зростання:

$\lg к = 0{,}1629 : 5 = 0{,}0326;$

$ant \lg 0{,}0326 = 1{,}077$.

Таким чином, розрахунок середнього коефіцієнта зростання чисельності працюючих можна подати в такій послідовності:

$$\lg \overline{K} \sqrt[5]{1{,}075 \times 1{,}087 \times 1{,}099 \times 1{,}104 \times 1{,}026} = \frac{0{,}0315 + 0{,}0363 + 0{,}0411 + 0{,}0429 + 0{,}0111}{5} = 0{,}0326;$$

$$\overline{K} = ant \lg 0{,}0326 = 1{,}077.$$

Середня квадратична. При підстановці у формулу степеневої середньої k = +2 одержуємо середню квадратичну. Формула її має наступний вигляд:

при незваженій формі – $\overline{x_{кв}} = \sqrt{\frac{\Sigma x^2}{n}} = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + ... + x_n^2}{n}}$;

при зваженій формі –

$$\overline{x_{кв}} = \sqrt{\frac{\Sigma x_i^2 n_i}{\Sigma n_i}} = \sqrt{\frac{x_1^2 n_i + x_2^2 n_2 + x_3^2 n_3 + ... + x_n^2 n_n}{n_1 + n_2 + n_3 + ... + n_n}}.$$

Практично середня квадратична величина використовується в тих випадках, коли варіанти ряду представлені у вигляді відхилень фактичних їх значень від середньої арифметичної (або від якоїсь нормативної величини).

Приклад. При змінній нормі виробітку трактора на оранці 7 га обчислити середню величину відхилень фактичних показників виробітку протягом робочого тижня (табл. 13).

Таблиця 13

Вихідні та розрахункові дані для обчислення середньої квадратичної

Фактичні показники виробітку трактора, га (x_i)	Відхилення від норми, га (x_i-$x_н$)=M	Кількість тракторів, n_i	Розрахункові величини	
			M^2	$M^2 n_i$
6	-I	1	1	1
7	0	5	0	0
8	1	6	1	6
10	3	3	9	27
12	5	6	25	150
Всього	X	21	x	184

За даними таблиці 13 знаходимо величину середньої квадратичної зваженої $\overline{x_{кв}} = \sqrt{\frac{\Sigma x_i^2 n_i}{\Sigma n_i}} = \sqrt{\frac{184}{21}} = \sqrt{8{,}76} \approx 2{,}96$.

Отже середня величина відхилень фактичних показників виробітку трактора на оранці від норми виробітку становить приблизно 3 га (2,96).

Формули для обчислення різних типів середніх величин наведено в таблиці 14.

Ознаки, які вивчаються, можуть бути **первинними** і **вторинними.** До **первинних** відносяться ті, які безпосередньо характеризують досліджувані об'єкти, наприклад, урожайність, собівартість тощо. Відношення двох чи декількох первинних ознак обумовлює характер вторинності – їх називають **вторинними.** За формою ознаки можуть бути **прямі** і **обернені.** Названі особливості досліджуваних ознак визначають статистичну розмірність (табл. 14).

Таблиця 14

Математичні особливості різних типів середніх величин

Характер ознаки	Форма величини (ознаки)	Статистична розмірність (показник степеня в формулі степеневої середньої)	Вид середньої величини	Форма середньої величини	Формула
Первинний	Обернена	-1	Гармонійна	Проста	$\overline{x_{гм}} = \frac{n}{\Sigma \frac{1}{x}}$
Вторинний	Обернена	-1	Гармонійна	Зважена	$\overline{x_{гм}} = \frac{\Sigma w}{\Sigma \frac{w}{x}}$
Вторинний	Пряма	0	Геометрична	Проста	$\overline{x_{гр}} = \sqrt[n]{x_1 \times .x_2 \times x_3 \ldots \times x_n}$
Первинний	Пряма	1	Арифметична	Проста	$\overline{x_a} = \frac{\Sigma x}{n}$
Вторинний	Пряма	1	Арифметична	Зважена	$\overline{x_a} = \frac{\Sigma x_i n_i}{\Sigma n_i}$
Первинний	Пряма	2	Квадратична	Проста	$\overline{x_{кв}} = \sqrt{\frac{\Sigma x^2}{n}}$
Вторинний	Пряма	2	Квадратична	Зважена	$\overline{x_{кв}} = \sqrt{\frac{\Sigma x_i^2 n_i}{\Sigma n_i}}$

Різні види середніх, розраховані для одного і того ж варіаційного ряду, різняться між собою за величиною. При цьому найменшою буде середня гармонійна, найбільшою – середня квадратична. Систематичність послідовного зростання у розрізі видів середніх випливає з «правила мажорантності» і зумовлюється показником у формулі степеневої середньої (табл. 15).

Таблиця 15

Мажорантність середніх величин

Статистична розмірність (показник степеня в формулі степеневої середньої)	Вид середньої величини	Примітка
-1	Гармонійна	Мінімальна
0	Геометрична	
1	Арифметична	
2	Квадратична	Максимальна
$\overline{x_{гм}} \langle \overline{x_{гр}} \langle \overline{x_{а}} \langle \overline{x_{кв}}$		

У дискретному і інтервальному рядах розподілу обчислюються так названі порядкові (структурні) середні – мода і медіана.

Мода – це варіанта, яка найчастіше зустрічається в даному варіаційному ряді. Для дискретного ряду розподілу мода визначається за частотами варіант і відповідає варіанті з найбільшою частотою. Для інтервального ряду розподілу з рівними інтервалами інтервал, що містить моду (модальний), визначається по найбільшій частоті. При нерівних інтервалах мода знаходиться за показником найбільшої щільності розподілу.

Для випадку розподілу з рівними інтервалами мода (Мо) внутрімодального інтервалу обчислюється за формулою:

$$\mathrm{M}_0 = x_{мо\min} + i\frac{n_{мо} - n_{м0-1}}{(n_{м0} - n_{m-1}) + (n_{м0} - n_{м0+1})},$$

де $x_{м0\min}$ – нижня границя модального інтервалу; і- інтервальна різниця (величина інтервалу); $n_{м0}$– частота модального інтервалу; $n_{м0-1}$– частота інтервалу, що передує модальному; $n_{м0+1}$ частота інтервалу, наступного за модальним.

Медіана – це значення варіаційної ознаки, яка приходиться на середину варіаційного ряду. Якщо кількість членів ряду парна, то медіана дорівнює середній арифметичній із двох серединних значень варіант.

Для обчислення медіани інтервального варіаційного ряду знаходять інтервал, який містить медіану, шляхом використання нагромаджених частот або частостей. Медіанному інтервалу відповідає перша з нагромаджених частот або частостей, що перевищує половину всього обсягу сукупності. Внутрі знайденого інтервалу медіана (M_e) розраховується за формулою:

$$M_e = x_{me\min} + i\frac{\frac{1}{2}\sum n_i - S_{Me-1}}{n_{Me}}$$

де $x_{me\min}$ – нижня границя медіанного інтервалу; і – інтервальна різниця (величина інтервалу); $\frac{1}{2}\Sigma ni$ – половина суми всіх частот або частостей; $S_{ме-1}$– нагромаджена частота або частость інтервалу, який передує медіанному; $n_{ме}$ – частота медіанного інтервалу.

Карл Пірсон встановив взаємозв'язок між модою, медіаною і середньою арифметичною, який виражається рівністю:

$$\mathrm{M}_e = \frac{1}{3}\mathrm{M}_0 + \frac{2}{3}\bar{x}_\alpha.$$

Приклад. За даними інтервального ряду розподілу підприємств по врожайності зернових культур знайти середній показник урожайності, використовуючи у розрахунку формули структурних середніх – моди і медіани.

Таблиця 16

Вихідні і розрахункові дані

Урожайність, ц з І га (x_i)	Кількість підприємств, частота (n_i.)	Нагромаджена сума частот (n'_i)
16-20	3	3
20-24	6	9
24-28	17	26
28-32	16	42
32-36	7	49
36-40	5	54
40-44	3	57
Всього	57	х

У складеному інтервальному варіаційному ряді (табл.16) модальним інтервалом, із найбільшим числом варіант (17), є третій інтервал з рівнем урожайності від 24 до 28 ц. Отже, мода повинна знаходитися всередині цього інтервалу. Підставивши в наведену вище формулу необхідні значення, одержимо:

$$\mathrm{M}_0 = x_{м0\min} + i\frac{n_{м0} - n_{м0-1}}{(n_{м0} - n_{м0-1}) + (n_{м0} - n_{м0+1})} = 24 + 4\frac{17-6}{(17-6)+(17-16)} = 27{,}7$$ (ц з 1 га).

Обчислимо медіану за даними цього ж інтервального ряду розподілу. Як уже було сказано, медіана (M_e) в статистичному ряді розподілу посідає центральне місце. Для її знаходження необхідно мати нагромаджений підсумок частот. Медіана знаходиться в тому інтервалі, нагромаджена сума частот якого дорівнює або більша за напівсуму частот ряду (57 : 2). У розглянутому прикладі таким інтервалом є четвертий, з рівнем урожайності від 28 до 32 ц, оскільки тут нагромаджена сума частот становить 42. Підставивши відповідні значення у формулу медіани (Ме) одержимо:

$$M_e = x_{me\min} + i\frac{\frac{1}{2}\sum n_i - S_{Me-1}}{n_{Me}} = 28 + 4\frac{28,5-26}{16} = 28,6$$ (ц з 1 га).

§ 4.4. Умови наукового застосування статистичних показників

Природа соціально-економічних явищ досить складна і специфічна. Пояснюється це тим, що розміри і кількісні їх взаємозв'язки зумовлюються значною різноманітністю факторів, що діють у часі І просторі, зумовлюючи неоднакову швидкість і напрями змін явищ. Отже, статистичне вивчення суспільних явищ повинне ґрунтуватися на наукових принципах, які виходять із знання суті досліджуваних явищ, економічних понять і категорій. Лише за таких умов можна переходити до вивчення системи економічних показників

Як зрозуміло з викладеного раніше, абсолютні і відносні показники відображують різні сторони суспільного життя. Так, індивідуальні абсолютні показники характеризують результати діяльності окремих працівників, окремих підприємств тощо, загальні – характеризують сукупні результати роботи окремих галузей, регіонів і країни в цілому. Відносні показники відображають міру взаємного зв'язку явищ в їх розвитку (відносні показники динаміки, виконання плану), внутрішнього зв'язку сторін одного явища (відносні показники структури), співвідношення між явищами (відносні показники порівняння). Міра ця відображується у вигляді чисел, що зумовлюють можливість зіставлення таких процесів і структур, які стосуються різних абсолютних мас (наприклад, порівняння темпів зростання продуктивності прані і оплати праці, чисельності поголів'я і його продуктивності тощо).

Щоб статистичні показники вірно виконували свої функції, їх слід розраховувати за науковими принципами. Існує два головних критерії науковості статистичних показників. Перший з них належить до теоретичної обґрунтованості показників, другий – до фактичної бази, на якій вони розраховані. Теоретична обґрунтованість показника полягає в утворенні його на підставі глибокого теоретичного аналізу соціальної дійсності, тобто філософського підходу до аналізу. Головною теоретичною основою статистичних показників тут виступають принципи, закони і категорії філософії, адже вони озброюють статистика знанням загальних закономірностей суспільного розвитку.

Теоретичні основи статистичних показників зумовлені також і спеціальними суспільними науками, за допомогою яких досліджуються ті чи інші суцільні явища. До таких наук відносять політекономію, конкретні економіки, правові науки, демографію тощо.

Особливе місце в утворенні статистичних показників належить статистичній теорії, розробленій статистичною наукою.

Таким чином, процесу утворення наукового статистичного показника передує знання філософії, політекономії, конкретних економік й інших спеціальних наук. Перший критерій науковості здебільшого визначає суть статистичного показника.

Другий критерій науковості статистичних показників полягає в утворенні їх на базі наукової інформації. Він пов'язаний з конкретним кількісним і якісним змістом показників. Статистичне дослідження (як і будь-яке інше) починається зі збирання інформації, і висновки, які резюмуються у статистичних показниках, мають сенс у разі обґрунтованості їх фактами – це загальнонауковий принцип.

Серед основних вимог, які ставляться до статистичних показників, такі: повнота вихідних даних, їх порівнюваність і вірогідність (чи точність).

Основна вимога до вихідної інформації – **повнота** вихідних даних. Під повнотою фактів розуміють: а) повноту просторового охоплення явищ або елементів досліджуваного процесу; б) повноту охоплення сторін явищ, тобто повноту вихідних даних щодо всіх істотних ознак явищ; в) повноту охоплення у часі. Тут слід передбачити наявність явищ за максимально тривалий час.

Вимога повноти даних зумовлюється тим, що окремі випадкові факти у існуючій складності взаємозв'язків (наприклад, економічних процесів) формуються під впливом як істотних, так і випадкових причин і обставин. Тобто, якщо підходити суб'єктивно, то для доведення того чи іншого положення можуть бути використані досить суперечливі факти. Обмеженість окремих факторів повинна долатися вичерпними факторами, об'єднаними у статистичні сукупності, адже тільки за таких умов забезпечується всебічність його вивчення та відтворення в цілому. Якщо, наприклад, вивчається питання про собівартість виробництва продукції у фермерських господарствах конкретної області, то обов'язково треба мати інформацію про весь обсяг продукції цих господарств, який встановлюється шляхом підсумовування даних про обсяги виробництва продукції по всіх фермерських господарствах області. Якщо не встановити загального обсягу продукції, то не можна буде обчислити собівартість продукції, продуктивність праці, прибуток на одного працюючого і ряд інших важливих показників ефективності виробництва.

Наступною вимогою до придатних для обчислення статистичних показників є їх **порівнюваність**. Остання зумовлює узагальнення показників у часі і просторі і стосується не тільки абсолютних, а й відносних показників. Порівнюваність окремих факторів потрібна перш за все для того, щоб дані можна було підсумувати. Проблема зіставності даних надзвичайно складна. Суспільною наукою і практикою відпрацьовано ознаки порівнюваності, тобто правила наукового порівняння. Найважливіші з них такі: спільний предметний зміст фактів, відображення порівнюваних величин у однакових одиницях виміру, обов'язковість однакових прийомів розрахунку, наявність однакового кола об'єктів, які входять у порівнювані величини, однаковість території, якою охоплюються порівнювані величини.

Вимога **вірогідності** статистичних показників передбачає ступінь їх наближення до відображуваної реальності. Поняття вірогідності інколи ототожнюють з поняттям **точності**. Під останнім треба розуміти не ступінь наближення статистичного показника до реального розміру, а повну його відповідність реальності. У більш вузькому розумінні поняття точності показника вживають при дослідженні явищ, які формуються як під впливом закономірностей, так і під впливом випадковостей. У такому разі поняття точності пов'язують з імовірністю розрахунків і його визначення доповнюється поняттям надійності оцінки точності, тобто ступенем ймовірності, з якою можуть статися відхилення між одержаним показником і тим, який визнається істинним. Величину такої різниці називають помилкою, або похибкою. Якщо, наприклад, при кількох вимірах залишків силосу у силосній траншеї одержані різні показники вимірів, то приймається за істинний розмір показника середній об'єм запасу силосу, одержаний за кількома вимірами.

Відхилення середнього розміру від окремих результатів розрахунків у цьому випадку вважаються помилками. Такі помилки називають випадковими, або статистичними. Детальний розгляд цього питання є предметом вивчення спеціальних розділів.

Утворення статистичних показників і побудова їх системи здійснюються через досить складний шлях, пов'язаний (у всіх його відтінках) із суспільною практикою. Остання є вихідною базою статистичного пізнання і дослідження, а отже, і вихідним початком формування статистичних показників. Статистична практика (як і будь-яка суспільна) є не тільки вихідним початком, основною і кінцевою метою утворення статистичних показників, вона є їх критерієм істинності. Зразу ж виникає питання, як довести істинність сформованих показників. Таке доведення може здійснюватися двома шляхами: а) емпіричним, тобто безпосереднім порівнянням з життям, з практикою; б) логічним, тобто порівнянням сформованих статистичних показників з будь-якими іншими показниками (наприклад, обчислений показник собівартості виробництва молока в господарстві можна порівняти з аналогічним показником у молочному скотарстві підприємств адміністративного району).

До прямих (конкретних) прийомів перевірки відповідності статистичних показників дійсності належать: експертна оцінка (її здійснюють висококваліфіковані фахівці); розрахунковий експеримент (він проводиться з використанням математичних методів); порівняння показників, розрахованих різними способами і методами, та деякі інші.

До непрямих характеристик відповідності статистичного показника дійсності належить його стійкість. Стійкими вважаються такі показники, які при обмежених за розміром змінах вихідної інформації або удосконалення і уточнення методів їх виміру і розрахунків залишаються незмінними чи змінюються .незначно. Однією з ознак стійкості статистичних показників є те, що розміри їх повторюються у дослідженнях, здійснюваних у різних регіонах і за різні періоди часу.

Сталі показники використовують, як правило, при нормуванні, екстраполяції й інших практичних і аналітичних розрахунках.

Із сказаного вище випливає, що питання абсолютних і відносних показників треба розглядати з точки зору комплексного їх використання, тобто з урахуванням їх особливостей, взаємозв'язків і різниць. Насамперед слід враховувати існування тісного зв'язку між абсолютними і відносними величинами. Оскільки відносні показники відображують відношення абсолютних показників, їх зміни залежать від останніх. Але по своїй суті і характеру змін відносні показники істотно відрізняються від абсолютних, причому останні можуть мати зовсім протилежний напрям змін щодо абсолютних показників.

Слід враховувати таку особливість, що відносні величини по-різному відображують зміни суспільних явищ залежно від їх абсолютних розмірів. Так, малі за розміром явища у процентному відношенні змінюються значно швидше, ніж аналогічні їм великі за розміром явища. Тому розрахунок приросту щодо малого за розміром початкового рівня може дати більший процент зростання, ніж тисячі одиниць при значному за розміром початковому рівні.

Таким чином, при дослідженні конкретних соціально-економічних явищ не можна обмежуватися розрахунками лише процентних співвідношень, треба враховувати, що криється за такими співвідношеннями. Необхідно поєднувати застосування абсолютних і відносних показників, ізольоване застосування відносних показників від абсолютних може призвести до помилкових висновків, особливо при аналізі рядів динаміки.

Комплексне використання абсолютних і відносних показників дає змогу поглибити аналіз суспільних явищ, виявити закономірності і особливості їх розвитку.

Отже, розробці науково обґрунтованої системи статистичних показників передує велика і кропітка робота щодо вивчення форми і змісту показників, вимог до їх формування зокрема і в комплексі. Кінцевим завданням статистики у цій справі слід вважати побудову суперсистеми статистичних показників. Останні у досконалій формі повинні забезпечити можливість порівнянь виробничих стосунків у нових ринкових умовах виробництва не тільки в межах країни, а й на більш високому регіональному рівні.

ТЕМА 5. АНАЛІЗ РЯДІВ РОЗПОДІЛУ

§ 5.1. Поняття про статистичні ряди розподілу

Маючи в розпорядженні дані статистичного спостереження, що характеризують те чи інше явище, перш за все необхідно їх впорядкувати, тобто надати характер системності.

Англійський статистик У.Дж.Рейхман з приводу неупорядкованих сукупностей образно сказав, що зіткнутися з масою неузагальнених даних рівнозначно ситуації, коли людину кидають у лісових хащах без компаса. Що ж собою являє систематизація статистичних даних у вигляді рядів розподілу?

Статистичний ряд розподілу - це впорядковані статистичні сукупності (табл. 17). Найпростішим видом статистичного ряду розподілу є ранжированний ряд, тобто ряд чисел, що знаходиться в порядку зростання чи спадання варіюючої ознаки. Такий ряд не дозволяє судити про закономірності, закладені в розподілених даних: біля якої величини групується більшість показників; які є відхилення від цієї величини; яка загальна картина розподілу. З цією метою групують дані, показуючи, як часто зустрічаються окремі спостереження в загальному їх числі (Схема 1).

Таблиця 17

Загальний вигляд статистичних рядів розподілу

Варіанта	Частота
x_1	n_1
x_2	n_2
…	…
x_m	n_m
	Σn_i

Варіанта	Частота
x_1-x_2	n_1
x_2-x_3	n_2
…	…
x_m-x_{m+1}	n_m
	Σn_i

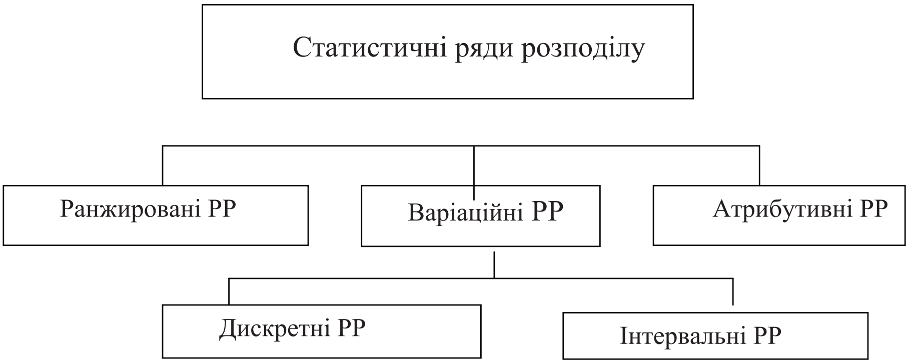


Схема 1. Схемастатистичних рядів розподілу

Розподіл одиниць сукупності за ознаками, що не мають кількісного виразу, називається **атрибутивним рядом** (наприклад, розподіл підприємств за їх виробничим напрямом).

Ряди розподілу одиниць сукупності за ознаками, що мають кількісний вираз, називаються **варіаційними рядами.** У таких рядах значення ознаки (варіанти) знаходяться в порядку зростання чи спадання.

У варіаційному ряді розподілу розрізняють два елементи: варіанта і частота. **Варіанта** – це окреме значення групувальної ознаки, **частота** – число, яке показує, скільки разів зустрічається кожна варіанта.

У математичній статистиці обчислюється ще один елемент варіаційного ряду – **частість.** Остання визначається, як відношення частоти випадків даного інтервалу до загальної суми частот. Частість визначається в частках одиниці, відсотках (%) в проміле (‰$_0$).

Таким чином, **варіаційний ряд розподілу** – це такий ряд, у якому варіанти розташовані в порядку зростання або спадання, вказані їх частоти або частості. Варіаційні ряди бувають дискретні (переривні) і інтервальні (неперервні).

Дискретні варіаційні ряди – це такі ряди розподілу, в яких варіанта як величина кількісної ознаки може приймати тільки певне значення. Варіанти різняться між собою на одну чи кілька одиниць.

Так, кількість вироблених деталей за зміну конкретним робітником може виражатися тільки одним певним числом (6, 10, 12 і т.д.). Прикладом дискретного варіаційного ряду може бути розподіл працівників за кількістю вироблених деталей (табл. 18).

Таблиця 18

Дискретний ряд розподілу

Вироблено деталей за зміну, шт. (x_i)	Кількість робітників, чол.,(n_i)
6	16
7	10
8	8
9	10
10	12
11	16
12	3

Інтервальні (неперервні) варіаційні ряди – такі ряди розподілу, в яких значення варіанти дано у вигляді інтервалів, тобто значення ознак можуть відрізнятися одне від одного на скільки завгодно малу величину. При побудові варіаційного ряду неперервної ознаки неможливо вказати кожне значення варіанти, тому сукупність розподіляється за інтервалами. Останні можуть бути рівні і нерівні. Для кожного з них вказуються частоти або частості (табл. 19).

В інтервальних рядах розподілу з нерівними інтервалами обчислюють такі математичні характеристики, **як щільність розподілу і відносна щільність розподілу** на даному інтервалі. Перша характеристика визначається відношенням частоти до величини того ж інтервалу, друга – відношенням частості до величини того ж інтервалу. Для наведеного вище прикладу щільність розподілу на першому інтервалі становитиме 3:5 = 0,6, а відносна щільність на цьому інтервалі – 7,5:5= 1,5%.

Таблиця 19

Інтервальний ряд розподілу

Чисельність працюючих, чол. (x_i)	Кількість цехів (n_i)	% до підсумку
20-25	3	7,5
25-30	9	22,5
30-35	16	40,0
35-40	8	20,0
40-45	4	10,0
Всього	40	100,0

§ 5.2. Графічне зображення рядів розподілу. Основні форми статистичних розподілів

Графічне зображення рядів розподілу (як і статистичних даних взагалі), крім досягнення наочності, переслідує й аналітичну мету. Графік дозволяє в найбільш простій і доступній формі піддати аналізу (візуально) статистичний ряд розподілу. Варіаційні ряди залежно від виду і поставленої задачі їх аналізу графічно можуть бути зображені у виді **полігону, гістограми, кумуляти, огіви.**

Полігон розподілу будується в прямокутній системі координат, при цьому на осі абсцис відкладається варіанта, а на осі ординат – частота або частість. За допомогою полігону розподілу, як правило, графічно зображуються дискретні варіаційні ряди (рис. 3, табл. 20). При побудові полігону для інтервальних рядів розподілу ордината, яка відповідає частоті (частості) встановлюється перпендикулярно осі абсцис у точці, що відповідає центру інтервалу.

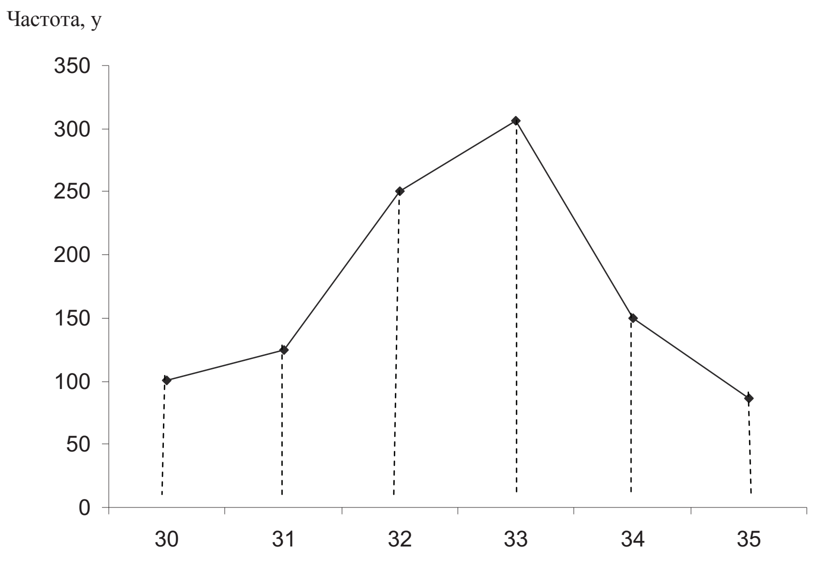


Рис.3 Полігон

Такий спосіб зображення інтервального ряду називають способом «навантажених ординат». Він заснований на допущенні умови рівномірного розподілу частот у межах інтервалів. Якщо це так, то частоти можна віднести до конкретного значення варіанти, яке знаходиться в центрі інтервалу. За таких умов середина (центр) інтервалу ніби «навантажується». Для графічного зображення інтервальних (неперервних) варіаційних рядів частіше використовуються **гістограми**. Остання являє собою ступінчасту фігуру у вигляді прямокутників, що примикають один до одного. Порядок побудови цього виду графіків такий. На осі абсцис відкладають інтервали значень варіанти. Вони ж є основами прямокутників, висота яких (ордината) пропорційна частоті (частості) інтервалів (рис. 4, табл. 21).

Таблиця 20

Розподіл робітників за змінною виробіткою продукції

Вироблено деталей за зміну, шт. (x_i)	Кількість робітників, чол. *(n_i)*
30	101
31	125
32	250
33	306
34	150
35	86
Разом	1018

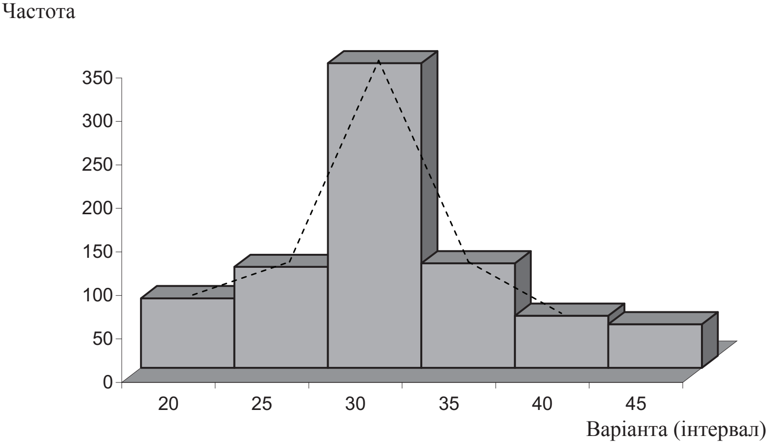


Рис. 4. Гістограма

Таблиця 21

Розподіл робітників за змінною виробіткою продукції

Вироблено деталей за зміну, шт.(x_i)	Кількість робітників, чол. (n_i)
20-25	80
25-30	116
30-35	350
35-40	120
40-45	60
45-50	50
Всього	776

Якщо необхідно зобразити на графіку інтервальний ряд розподілу з нерівними інтервалами, то гістограму будують не по частотах (частостях) інтервалів, а по показниках щільностей розподілу. При побудові гістограми по абсолютній щільності розподілу загальна площа її дорівнюватиме чисельності сукупності. При побудові графіка відносної щільності площа гістограми дорівнюватиме одиниці.

Інколи ідуть шляхом розчленування (подрібнення) інтервалів варіаційного ряду. У цьому випадку на графіку площа гістограми (дрібно ступінчастої) повинна відповідати попередній величині. Якщо процес подрібнення інтервалів продовжити, одержуючи все дрібніші їх параметри, то дрібноступінчаста гістограма в межі являтиме плавну криву.

При зображені варіаційного ряду з нагромадженими частотами (частостями) у прямокутні системі координат одержується так звана крива сум – кумулята. Якщо розподіл носить дискретний характер (переривний) на графіку на осі абсцис відкладають значення варіанти, на осі ординат – накопичені частоти (частості) (рис. 5, табл. 22).

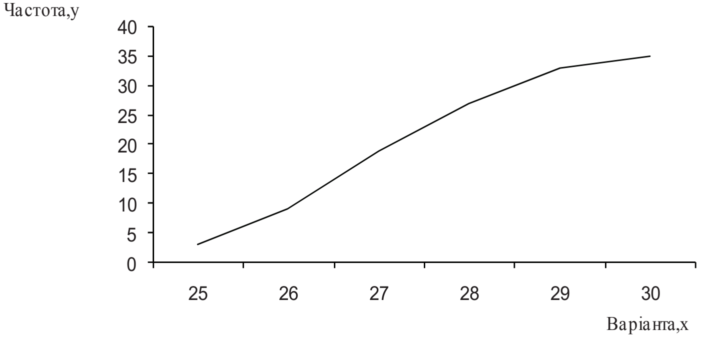


Рис. 5. Кумулята

Таблиця 22

Розподіл підприємств за рівнем рентабельності

Рівень рентабельності, % (x_i)	Кількість підприємств, (n_i)	Накопичена частота (n_i')
25	3	3
26	6	9
27	10	19
28	8	27
29	6	33
30	2	35
Всього	35	x

Примітка. При побудові кумуляти для інтервального ряду розподілу будують точки, абсциси яких – праві границі інтервалів, а ординати – відповідні їм нагромаджені частоти або частості.

По інтервальному (неперервному) ряду розподілу будують точки, абсциси яких – праві границі інтервалів, а ординати – частоти (частости), що їм відповідають.

Аналогічно кумуляті в прямокутній системі координат будують огіву. Різниця графіка лише в тому, що на осі абсцис наносять нагромаджені частоти, а на осі ординат – значення варіант. Щоб побачити форму огіви на графіку кумуляти, досить лист паперу з її зображенням повернути на 90° і подивитися на нього з протилежної сторони (на світло) (рис. 6, табл. 23).

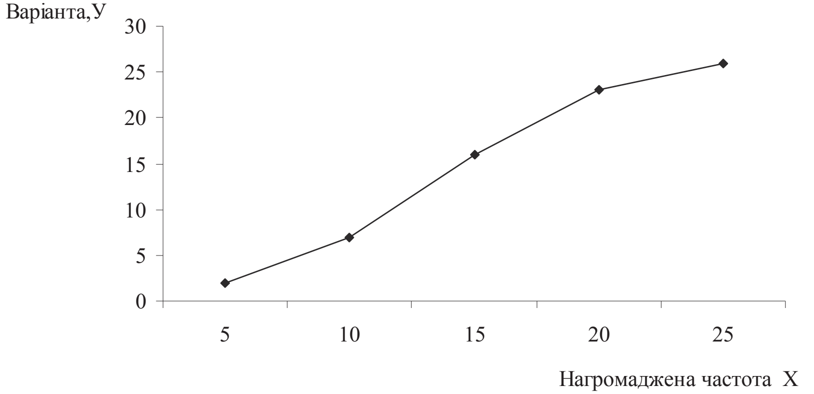


Рис.6.Огіва

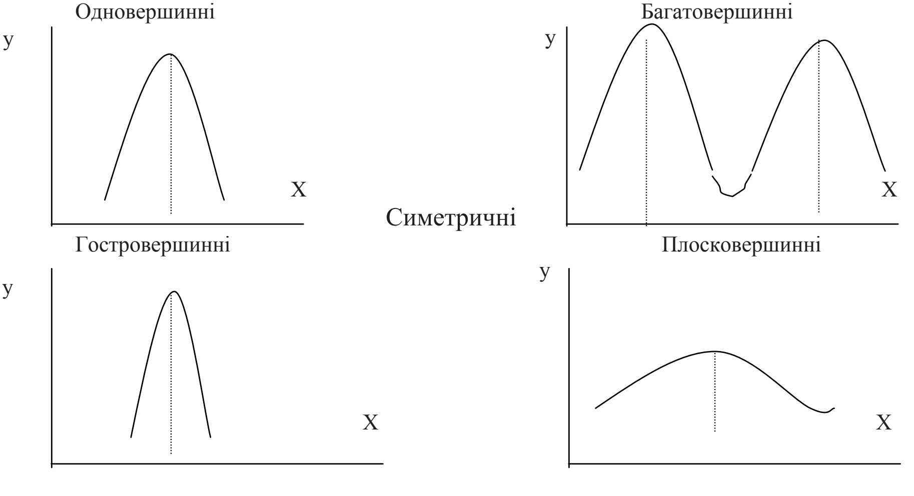
Таблиця 23

Розподіл підприємств за рівнем рентабельності

Рівень рентабельності, % (x_i)	Кількість підприємств (n_i)	Нагромаджена частота (n_i')
20-23	2	2
23-26	5	7
26-29	9	16
29-32	7	23
32-35	3	26
Разом	26	X

Якщо збільшити до нескінченності число одиниць спостережень і зменшити величину інтервалу, графік ряду розподілу буде мати форму кривої (рис. 7). На схемі 2 наведені форми статистичних розподілів.

2. Схема форм статистичних розподілів



Помірно асиметричні

Лівостороння скошеність (від'ємна асиметрія) — Правостороння скошеність (додатня асиметрія)

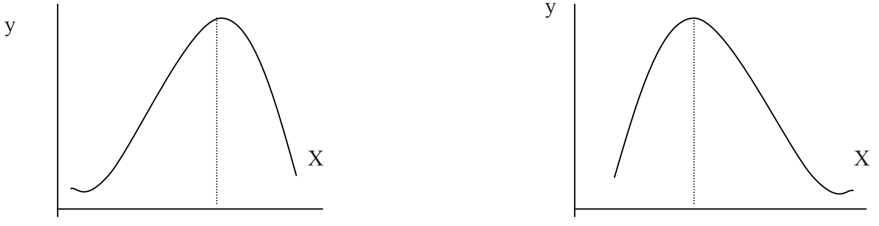
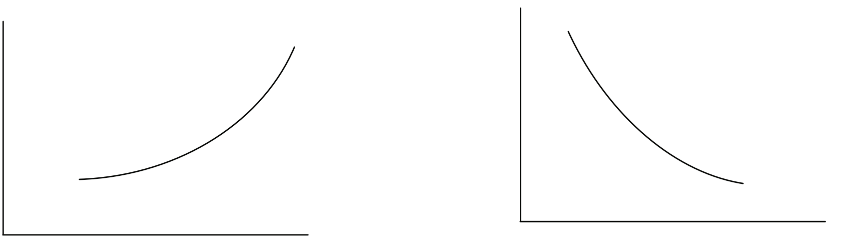


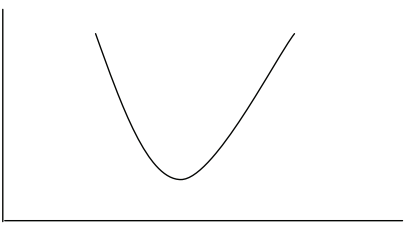
Рис. 7. Графіки форм статистичних розподілів

Крайньо асиметричні

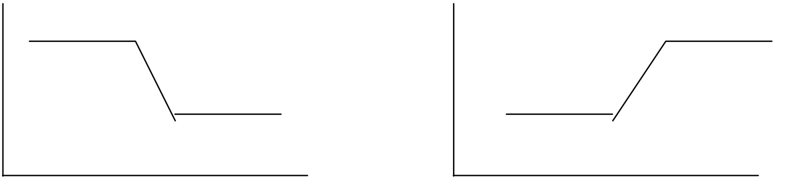
J –подібні



U-подібні



Z –подібні



Продовження рис. 7. Графіки форм статистичних розподілів

§ 5.3. Варіація ознак. Показники варіації

Розміри ознак, які характеризують кількісні зміни тих чи інших явищ, зазнають коливань.

Як відомо, у певних межах коливаються (варіюють) показники рівнів продуктивності праці та її оплати, собівартості та рентабельності виробництва продукції тощо. Ці коливання зумовлені певними факторами, які діють у різних напрямках. Для узагальнюючої характеристики статистичної сукупності за варіюючими ознаками розраховують середні величини. Але середня, характеризуючи варіаційний ряд у цілому, не враховує варіацію ознаки. Вона не показує, як розміщені навколо неї варіанти, тобто, чи зосереджені вони поблизу середньої, чи значно відхиляються від неї. Середня не показує характер варіації ознаки і степінь її коливань.

У деяких випадках та ж сама середня може характеризувати зовсім різні сукупності. Тобто в двох або декількох сукупностях середні величини однакові (за рівнем), а відхилення від цих середніх різні. У таблиці 24 наведено дані про виробничий стаж робітників двох цехів підприємств (А і Б).

Таблиця 24

Вихідні і розрахункові дані для обчислення середньої (x_i - стаж у роках, n_i - кількість робітників)

А			Б		
x_i	n_i	x_in_i	x_i	n_i	x_in_i
2	1	2	2	30	60
3	5	15	3	20	60
4	30	120	4	10	40
5	60	300	5	50	250
6	30	180	6	10	60
7	5	35	7	20	140
8	1	8	8	30	240
Всього	132	660	-	170	850

Середні, обчислені для обох сукупностей, будуть однакові
$\bar{x}_1 = \frac{\Sigma x_i n_i}{\Sigma n_i} = \frac{660}{132} = 5;\ \bar{x}_2 = \frac{850}{170} = 5.$

Відхилення від обчислених середніх мають різний характер. У першому цеху стаж 120 робітників (30+60+30) із 132 (тобто 91 %) відхиляється від середнього стажу (5 років) не більше як на 1 рік.

У другому цеху 70 випадків (10+50+10) із 170 мають таке ж відхилення – 41 %. Зрозуміло, що у першому випадку середня характеристика більш надійна (більш типова), ніж у другому. Якщо значення ознаки більше відхиляється від середньої (другий випадок), то досліджувана сукупність вважається менш однорідною, а середня менш надійною. Тому поряд з середніми величинами важливе теоретичне і практичне значення має вивчення відхилень від середніх. При цьому являють інтерес як крайні відхилення, так і сукупність всіх відхилень. Від розмаху і розподілу відхилень залежить надійність середніх характеристик. Останні, необхідно доповнювати показниками, які вимірюють відхилення від них, тобто показниками варіації.

Для кількісного виміру варіації ознаки математична статистика розробила ряд показників: розмах варіації, середнє лінійне відхилення, середній квадрат відхилень (дисперсія), середнє квадратичне відхилення, коефіцієнт варіації.

У таблиці 25 названі статистичні характеристики представлені структурними їх формулами. Серед коефіцієнтів варіації найбільш вживаний показник, який вираховується за середнім квадратичним відхиленням.

Таблиця 25

Формули розрахунку показників варіації

Статистична характеристика варіації	Форма, зумовлена відсутністю чи наявністю статистичних ваг	
	незважена	зважена
Розмах варіації	$R = x_{\max} - x_{\min}$	$R = x_{\max} - x_{\min}$
Середнє лінійне відхилення	$d_н = \frac{\Sigma \lvert x_i - \bar{x} \rvert}{n}$	$d_з = \frac{\lvert x_i - \bar{x} \rvert n_i}{\Sigma n_i}$
Дисперсія (девіата, середній квадрат відхилень)	$\sigma_н^2 = \frac{\Sigma (x_i - \bar{x})^2}{n}$	$\sigma_з^2 = \frac{\Sigma (x_i - \bar{x})^2 n_i}{\Sigma n_i}$
Середнє квадратичне відхилення	$\sigma_н = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$	$\sigma_з = \sqrt{\frac{\Sigma (x_i - \bar{x})^2 n_i}{\Sigma n_i}}$
Коефіцієнт варіації: по варіаційному розмаху (коефіцієнт осциляції)	$V_R = \frac{R}{\bar{x}} 100$	$V_R = \frac{R}{\bar{x}} 100$
по середньому лінійному відхиленню (нерівно та)	$V_d = \frac{d_н}{\bar{x}} 100$	$V_d = \frac{d_з}{\bar{x}} 100$
по середньому квадратичному відхиленню	$V_\sigma = \frac{\sigma_н}{\bar{x}} 100$	$V_\sigma = \frac{\sigma_з}{\bar{x}} 100$

Розмах варіації, являючи собою різницю між крайніми (екстремальними) значеннями ознаки варіаційного ряду, дає лише загальне уявлення про розміри варіації, тобто її наближену оцінку. Величина ця нестійка і значною мірою залежить від випадковостей. Вона не дає уявлення про розміри відхилень варіант одна від одної в проміжку між крайніми їх значеннями. Особливістю показника розмаху варіації (R) є те, що він не відображує відхилень усіх варіант, не враховує частоти, а величина його залежить від двох крайніх значень ознаки.

Тому для узагальненої характеристики розміру цих відхилень розраховують середню із відхилень.

Слід пам'ятати, що термін «відхилення від середньої» означає різницю між варіантою і середньої арифметичною в даній сукупності. У розрахунках завжди віднімають середню від варіант, а не навпаки.

Оскільки сума додатних і від'ємних відхилень завжди дорівнює нулю (властивість середньої арифметичної), умовно припускають, що всі відхилення мають однаковий знак. Сума таких відхилень, поділена на їх число, має назву **середнє лінійне відхилення** (d). Цей показник має значну перевагу перед розмахом варіації (R) у відношенні повноти коливання ознаки. Чим більша його величина, тим менш однорідною вважається сукупність. Показник середнього лінійного відхилення у статистиці застосовують рідко. Для виміру міри варіації частіше отримані відхилення підносять до квадрату, а з квадратів відхилень обчислюють середню величину. Одержана таким чином міра варіації називається **середнім квадратом відхилень** або **дисперсією** (σ^2).

Якщо добути корінь квадратний з дисперсії, одержимо **середнє квадратичне відхилення** (σ). Дана статистична величина характеризує абсолютну міру варіації, це іменоване число і виражається у тих же одиницях виміру, в яких виражені варіанти. Середнє квадратичне відхилення називають також <u>стандартним</u> **відхиленням, стандартом або просто «сигмою».**

Середнє квадратичне відхилення (σ) і дисперсія (σ^2) є загальноприйнятими показниками міри варіації ознаки, мають широке застосування у статистиці.

Здійснимо розрахунок названих статистичних характеристик за даними раніше розглянутого прикладу про середній стаж робітників (табл. 26).

Величина дисперсії відповідно для об'єктів А і Б становитиме:

$$\sigma_A^2 = \frac{\Sigma(x_i - \overline{x})^2 n_i}{\Sigma n_i} = \frac{118}{132} = 0.89; \; \sigma_Б^2 = \frac{720}{170} = 4,2.$$

Звідси знаходимо: $\sigma_A = \sqrt{0,89} = 0,94; \;\; \sigma_Б = \sqrt{4,2} = 2,05.$

Як бачимо, у другому випадку середнє квадратичне відхилення $\sigma_Б$ більш як у два рази перевищує величину σ_A. Отже, другий ряд розподілу характеризується більш високою варіацією ознаки, ніж перший.

Таблиця 26

Вихідні і розрахункові дані для обчислення показників варіації

А					Б				
x_i	n_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})n_i^2$	x_i	n_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})n_i^2$
2	1	-3	9	9	2	30	-3	9	270
3	5	-2	4	20	3	20	-2	4	80
4	30	-1	1	30	4	10	-1	1	10
5	60	0	0	0	5	50	0	0	0
6	30	1	1	30	6	10	1	1	10
7	5	2	4	20	7	20	2	4	80
8	1	30	9	9	8	30	3	9	270
Разом	132	х	х	118	х	170	х	Х	720

Середнє квадратичне відхилення використовується і як самостійна статистична характеристика, і як основа для побудови (обчислення) інших статистичних характеристик: коефіцієнтів варіації, помилок репрезентативності різноманітних характеристик розподілу, коефіцієнтів кореляції і регресії, елементів дисперсійного аналізу, формул регресії.

За своєю величиною σ залежить не тільки від ступеня варіації, а й від абсолютних рівнів варіант і середньої.

Тому порівнювати стандартні відхилення, розраховані за варіаційними рядами з різнойменними ознаками (як і з різними рівнями), безпосередньо не можна.

Можливість такого порівняння забезпечує показник процентного відношення середнього квадратичного відхилення і середньої арифметичної - **коефіцієнт варіації** (V). Цей показник характеризує відносну міру варіації і дозволяє порівнювати ступінь варіації ознак в рядах розподілу з різним рівнем середніх.

Наприклад, якщо для урожайності зернових культур в одній області σ_1 = 9ц і $\bar{x}_1$ =30ц, а в другій – $\sigma_2 = 8ц$, і $\bar{x}_2 = 20ц,$ то за абсолютною величиною варіація у першому випадку більша (9 >8) а відносна міра варіації менша:

$$V_1 = \frac{\sigma_1}{\bar{x}_1}100 = \frac{9}{30}100 = 30\%; \quad V_2 = \frac{\sigma_2}{\bar{x}_2}100 = \frac{8}{20}100 = 40\%.$$

Коефіцієнт варіації зручний для порівняння варіації різних явищ. Наприклад, якщо при порівнянні коефіцієнтів варіації віку робітників до рівня їх трудоучасті (сума відпрацьованого часу – люд.-г) виявиться, що коефіцієнт варіації віку V =5,3 %, а коефіцієнт варіації трудоучасті V= 14,7%, то робиться висновок про те, що рівень трудоучасті варіює більше, ніж вік.

Коефіцієнт варіації є оцінкою надійності середньої. При величині V = 5 % варіація вважається слабкою, V = 6-10 % – помірною, V = 16-20 % – значною;V = 21-50 % – великою; V > 50 % – дуже великою.

Для малих вибірок величина коефіцієнта варіації повинна бути не більше 33 %. Якщо $\bar{x}$ = 1; V = σ.

5.3.1. Найважливіші математичні властивості дисперсії

Знаючи математичні властивості дисперсії, можна спростити вирахування її величини. Розглянемо їх.

1. Якщо із усіх значень варіант відняти постійне число А, то величина дисперсії не зміниться

$$\sigma^2_{(x_i - A)} = \sigma^2.$$

Таким чином, середній квадрат відхилень можна обчислити не за величинами варіант, а за відхиленням їх від якогось постійного числа, тобто $\sigma^2 = \sigma^2_{(x_i - A)}$.

2. Якщо значення варіант поділити на постійне число А, то величина дисперсії зменшиться в A^2, а середнє квадратичне відхилення в А разів:

$$\sigma^2_{(\frac{x_i}{A})} = \sigma^2 : A^2.$$

Із цього випливає, що всі варіанти можна поділити на будь-яке постійне число, обчислити середнє квадратичне відхилення, а потім помножити його на це постійне число: $\sigma^2 = \sigma^2_{\left(\frac{x_i}{A}\right)} \cdot A^2$

3. Якщо вирахувати середній квадрат відхилень від будь-якої величини (А), що відрізняється в тій чи іншій мірі від середньої ($\bar{x}$), то величина його завжди буде більше середнього квадрата відхилень, обчисленого відносно середньої $(\sigma^2_A \rangle \sigma^2)$.

Отримане перевищення дорівнює квадрату різниці між середньою і умовно узятою величиною, тобто $/\bar{x} - A/^2$. Це все можна подати у такому запису:

$$\sigma^2_A = \sigma^2 + (\bar{x} - A)^2 \text{ або } \sigma^2_A = \sigma^2 - (\bar{x} - A)^2 .$$

Розглянута властивість середнього квадрата відхилень дозволяє зробити висновок про те, що дисперсія від середньої (σ^2) завжди менша за дисперсії, обчислені від будь-яких інших величин σ^2_A , тобто вона має властивість мінімальності.

4. Дисперсія постійної величини дорівнює нулю ($\sigma^2_{Aconst} = 0$). Ця властивість випливає з того, що дисперсія є показником розсіювання варіант навколо середньої арифметичної, а середня арифметична постійної величини дорівнює цій величині.

Ряд властивостей дисперсії ґрунтується на рівності $\sigma^2 = \overline{x^2} - (\bar{x})^2$, Тобто дисперсія дорівнює різниці між середньою арифметичною квадратів варіант і квадратом середньої арифметичної.

5.3.2. Загальна, міжгрупова і внутрішньогрупова дисперсія

Якщо всі значення ознаки статистичної сукупності (генеральної або вибіркової) розділити на декілька груп і розглядати кожну з них як самостійну (окрему) сукупність, то виникає необхідність обчислення трьох видів дисперсій: загальної, міжгрупової і внутрішньогрупової.

Загальна дисперсія – це середній квадрат відхилень значень ознак всієї сукупності відносно загальної середньої.

Міжгрупова дисперсія – це середній квадрат відхилень групових середніх відносно загальної середньої.

Внутрішньогрупова дисперсія – це середня арифметична часткових (групових) дисперсій, зважена обсягами груп.

У таблиці 27 наведена структурні формули обчислення названих видів дисперсій.

Таблиця 27

Формули для обчислення дисперсій

Вид дисперсії	Формула	Примітка
Загальна	$\sigma^2_{заг} = \frac{\Sigma(xi - \overline{x})^2 n_i}{\Sigma n_i}.$	n_i-частота ознаки
Міжгрупова	$\sigma^2_{мгр} = \frac{\Sigma(\overline{x}_j - \overline{x})^2 Nj}{\Sigma N_{j_i}}$	j-номер групи Nj-обсяг групи $\Sigma Nj = \Sigma n_i$ $\overline{x}_j$ - групова середня j - групи
Внутрішньогрупо ва	$\sigma^2_{вгр} = \frac{\Sigma \sigma^2_j N_j}{\Sigma N_j}$	$\sigma^2_j -$ часткові (групові) дисперсії

Приклад. За даними врожайності зернових культур 57 підприємств визначити загальну, міжгрупову і внутрішньогрупову дисперсії, утворивши сім груп підприємств за рівнем урожайності.

Для обчислення загальної дисперсії необхідно побудувати дискретний ряд розподілу (табл. 28).

За розрахунковими даними цього статистичного ряду визначаємо середню арифметичну ($\overline{x}$)і величину загальної дисперсії ($\sigma^2_{заг}$):

$$\overline{x} = \frac{\Sigma x_i ni}{\Sigma ni} = \frac{1535.5}{57} = 26.9; \ \sigma^2_{заг} = \frac{\Sigma(xi - \overline{x})^2 ni}{\Sigma ni} = \frac{1399.06}{57} = 24.5.$$

Таблиця 28

Вихідні і розрахункові дані для обчислення загальної дисперсії (дискретний ряд)

Варіанта , x_i	Частота n_i	Розрахункові дані			
		$x_i n_i$	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^2 n_i$
17,5	1	17,5	-9,4	88,36	88,36
17,6	2	35,2	-9,3	86,49	172,98
…	…	…	…	…	…
36,2	3	108,6	9,3	86,49	259,47
37,6	2	75,2	10,7	114,49	228,98
Разом	57	1535,3	х	х	1399,06

Для визначення міжгрупової дисперсії необхідно обчислити групові середні ($\overline{x}_j$) і знайти загальний об'єм їх варіювання відносно загальної середньої ($\Sigma(\overline{x}_j - \overline{x})^2 Nj$). За розрахунковими даними таблиці 29 визначаємо розмір міжгрупової дисперсії:

$$\sigma^2_{мгр} = \frac{\Sigma(\overline{x}_j - \overline{x})^2 Nj}{\Sigma n_i} = \frac{1347.5}{57} = 23.6.$$

Таблиця 29

Вихідні і розрахункові дані для обчислення міжгрупової дисперсії

Інтервал (група)	Середня по групі, $\overline{x}_j$	Обсяг груп, N_j	Розрахункові дані		
			$\overline{x}_j - \overline{x}$	$(\overline{x}_j - \overline{x})^2$	$(\overline{x}_j - \overline{x})^2 N_j$
17,5-20,5	18,9	9	-8,0	64,00	576,0
20,5-23,5	21,9	6	-6,0	25,0	150,0
23,5-26,5	25,4	9	-1,5	2,25	20,3
26,5-29,5	28,2	13	1,3	1,69	22,0
29,5-32,5	30,8	15	3,9	15,21	228,0
32,5-35,5	34,0	3	7,1	50,41	151,2
35,5-38,5	36,9	2	10,0	100,00	200,0
Разом	X	57	X	x	1347,5

Щоб визначити внутрішньогрупову дисперсію, необхідно розрахувати часткові дисперсії у розрізі семи груп. Маючи групові середні $|\overline{x}_j|$, знаходимо по кожної групі відповідну часткову дисперсію $|\sigma^2_j|$. За даними прикладу, який розглядається, необхідно обчислити сім таких дисперсій $(\sigma^2_I, \sigma^2_{II}, \sigma^2_{III}, ..., \sigma^2_{VII})$.

Необхідні проміжні дані для їх обчислення наведено в таблиці 30.

Таблиця 30

Вихідні і розрахункові дані для обчислення внутрішньогрупової дисперсії (розрахунок часткових дисперсій) (σ^2_j)

Інтервал (група)	Варіанта, x_i	Частота, n_s	Розрахункові дані				
			$x_i n_i$	$x_i - \overline{x}_j$	$(x_i - \overline{x}_j)^2$	$(x_i - \overline{x}_j)^2 n_s$	σ^2_j
17,5-20,5	17,5	1	17,5	-1.4	1,96	1,96	$\sigma^2_I =$
$(\overline{x}_I = 18{,}9)$	17,6	2	35,2	-1.3	1,69	3,38	$= \frac{\Sigma(x_i - \overline{x}_I)^2 n_i}{N_I}$
Всього	x	9	170,0	x	x	9,05	$= \frac{9.05}{9} 1.0$
…	…	…	…	…	…	…	…
35.5-38,5	36,2	I	36,2	-0.7	0,49	0,49	
$(\overline{x}_{VII} = 36.9)$	37,6	I	37,6	0.7	0.49	0,40	$= \frac{\Sigma(\overline{x}_i - \overline{x_{VII}})^2 n_i}{N_{VII}}$
Всього	x	2	73,8	x	x	0,98 =	$= \frac{0{,}98}{2} = 0{,}49$

Початку обчислень часткових дисперсій передує розрахунок групових середніх ($\bar{x}_j$). Так, для першого інтервалу $\bar{x}_I = \frac{\Sigma x_i n_i}{N_I} = \frac{170}{9}$ =18,9. Аналогічно розраховуємо середні для інших груп. Потім знаходимо окремі дисперсії σ_j^2, величини яких становлять: $\sigma_I^2 = 1{,}0;$ $\sigma_{II}^2 = 0{,}43;$ $\sigma_{III}^2 = 1{,}12;$ $\sigma_{IV}^2 = 0.68;$ $\sigma_V^2 = 1.09;$ $\sigma_{VI}^2 = 0{,}58;$ $\sigma_{VII}^2 = 0{,}49$ (послідовність розрахунку показано тільки для першого і сьомого інтервалів).

Маючи обчислені значення часткових дисперсій, знаходимо величину внутрішньогрупової дисперсії:

$$\sigma_{вгр}^2 = \frac{\sigma_I^2 N_I + \sigma_{II}^2 N_{II} + \sigma_{III}^2 N_{III} + ... + \sigma_{VII}^2 N_{VII}}{N_I + N_{II} + N_{III}\Sigma.... + N_{VII}} =$$

$$= \frac{1\times 9 + 0.43\times 6 + 1.12\times 9 + 0.68\times 13 + 1.09\times 15 + 0.58\times 3 + 0.49\times 2}{9+6+9+13+15+3+2} = \frac{49.57}{57} = 0.87 \approx 0.9$$

Відповідно до правила складання дисперсій, яке випливає з доказу, що якщо сукупність складається з кількох груп, то загальна дисперсія дорівнює сумі внутрішньогрупової і міжгрупової дисперсій, маємо:

$$\sigma_{общ}^2 = \sigma_{мгр}^2 + \sigma_{вгр}^2 = 23{,}6 + 0{,}9 = 24{,}5.$$

За раніше наведеними розрахунками, величина загальної дисперсії $\sigma_{заг}^2$ дорівнює 24,5, що підтверджує вірність виконаних обчислень.

Теоретичний і практичний інтерес правила додавання дисперсій полягає у тому, що, знаючи дві величини дисперсії, на основі наведеної рівності завжди можна знайти третю. Наприклад:

$$\sigma_{вгр}^2 = \sigma_{заг}^2 - \sigma_{мгр}^2.$$

Маючи величини міжгрупової і загальної дисперсій, можна мати уяву про силу впливу групувальної ознаки. Про це мова піде при вивченні питань кореляційного і дисперсійного методів аналізу.

5.3.3. Дисперсія альтернативних ознак

Перш ніж розглянути питання про дисперсію альтернативних ознак, слід нагадати, що під альтернативною ознакою розуміють таку ознаку, якою одні варіанти наділені, а другі – ні. Так, якщо у вибірці, яка складається **з** n одиниць і n'' одиниць, наділених даною ознакою, то їх частка w у вибірковій сукупності становитиме:

$$w = \frac{n''}{n}.$$

Розрахунок загальної, міжгрупової і внутрішньогрупової дисперсій для альтернативних ознак поданий за формулами в таблиці 31.

Таблиця 31

Формули обчислення дисперсій для альтернативних ознак

Вид дисперсії	Формула	Примітка
Загальна	$\sigma_w^2 = w(1-w)$	w – частка одиниць наділених даною ознакою
Міжгрупова	$\sigma_{мw}^2 = \frac{\Sigma(w_i - w)n_i}{\Sigma n_i}$	w_i – частка одиниць, наділених даною ознакою в і –й групі
		n_i - число одиниць в і - й групі
Внутрішньогрупова	$\sigma_{\sigma w}^2 = \frac{\Sigma w_i(1-w_i)n_i}{\Sigma n_i}$	Σn_i - об'єм вибірки ($\sum n_i = n$)

Розглянемо послідовність розрахунку названих видів дисперсій на конкретному прикладі. У таблиці 32 представлена вибірка 60 підприємств, розподілених за виробничим типом на дві групи з обсягом n_i кожної і виділенням альтернативної ознаки – кількості збиткових підприємств (n'').

Підставляючи розрахункові дані таблиці 32 у формули відповідних видів дисперсій, одержимо:

$$\sigma_w^2 = w(1-w) = 0{,}233/1-0{,}233/ = 0{,}179;$$

$$\sigma_{мw}^2 = \frac{\Sigma(w_i - w)^2 n_i}{} = \frac{0.134}{60} = 0{,}002;$$

$$\sigma_{\sigma w}^2 = \frac{\Sigma w_i(1-w_i)n_i}{\Sigma n_i} = \frac{10{,}6}{60} = 0{,}177.$$

Таблиця 32

Вихідні і розрахункові дані для обчислення дисперсій альтернативної ознаки

Група	Обсяг груп, n_i	Число одиниць у групі, наділених даною ознакою, n''	Розрахункові дані					
			$w_i = \frac{n''}{n_i}$	$w_i = (1-w_i)$	$w_i(1-w_i)n_i$	$w_i - w$	$(w_i - w)^2$	$w_i - w)^2 n_i$
I	40	8	0,200	0,16	6,4	-0,03	0,0009	0,036
П	20	6	0,300	0,21	4,2	0,07	0,0049	0,098
Всього	60	14	0,233 (14:60)	х	10,6	х	х	0,134

Ґрунтуючись на правилі додавання дисперсій, маємо:

$\sigma_w^2 = \sigma_{мw}^2 + \sigma_{вw}^2$ або 0,179 =0,002+0,177; 0,179= 0,179.

Середнє квадратичне відхилення альтернативної ознаки в даному випадку легко знайти шляхом добування кореня з σ_w^2, тобто : $\sigma = \sqrt{w(1-w)} = \sqrt{0.179} = 0.42$.

§ 5.4. Моменти статистичного розподілу

Варіаційний ряд розподілу може характеризуватися системою статистик, які мають загальний математичний вираз і носять назву **моментів розподілу.** В цій системі знаходять своє відображення (місце) такі узагальнюючі характеристики ряду, як середня і дисперсія.

Система моментів розподілу вперше була розроблена російським математиком П.Л.Чебишевим.

Загальний математичний вираз моменту розподілу (загального емпіричного моменту) має вигляд:

$$\text{М}_к = \frac{\Sigma(x_i - A)^к n_i}{\Sigma n_i},$$

де $\text{М}_к$ – момент к-го порядку;

x_i – варіанти ряду;

n_i – частоти ряду;

k, A – постійні числа [к-порядок (степінь), A – довільне постійне число (несправжній нуль)].

Слід пам'ятати, що при розрахунку моментів статистичних розподілів осредняєтся k – та степінь відхилень значень ознаки (варіанти) x_i від деякої постійної величини (A).

Залежно від того, яка величина прийнята за умовний початок (A), загальна система моментів може бути подана підсистемами **початкових, центральних і нормованих моментів.**

Якщо умовний початок A = o, отримують підсистему **початкових моментів**. Початковий момент k-го порядку (k-ї степені) виражається формулою:

$$\text{М}_к = \frac{\Sigma x^к n_i}{\Sigma n_i},$$

де $\text{М}_к$ – початковий момент k-го порядку;

$\Sigma x^к n_i$ – сума добутків варіант k -ї степеня на їх частоти;

Σn_i – сума частот.

При k= 0 момент називається початковим моментом нульового порядку, при k= 1 – початковим моментом 1-го порядку при k= 2 – початковим моментом 2-го порядку і т.д.

Розрахунок величин початкових моментів від нульового до четвертого порядків представлений схематично в таблиці 33.

При A=$\bar{x}$ одержуємо підсистему центральних моментів. Центральний момент k -го порядку виражається формулою:

$$\mu_к = \frac{\Sigma(x_i - \bar{x})^k n_i}{\Sigma n_i}$$

Як видно з наведеної формули, центральні моменти являють собою середні із різних степенів відхилень від середньої арифметичної.

Таблиця 33

Розрахунок підсистеми початкових моментів

Порядок (степінь) k,	Формула	Зміст
0	$M_0 = \frac{\Sigma x_i^0 n_i}{\Sigma n_i}$	1
1	$M_i \frac{\Sigma x_i^1 n_s}{\Sigma n_i}$	$\bar{x}$ (середня арифметична)
2	$M_2 = \frac{\Sigma x_i^2 n_i}{\Sigma n^i}$	$\overline{x^2}$ (середня квадратів варіант)
3	$M_3 = \frac{\Sigma x_i^3 n_i}{\Sigma n_i}$	$\overline{x^3}$ (середня кубів варіант)
4	$M_4 = \frac{\Sigma x_i^4 n_i}{\Sigma n_i}$	$\overline{x^4}$ (середня четвертих степенів варіант)

Схема розрахунку підсистеми центральних моментів від нульового до четвертого порядків наведена в таблиці 34.

Обчислення центральних моментів можуть бути значно спрощені, якщо знати властивості цієї підсистеми моментів. Розглянемо їх.

1. Якщо усі варіанти статистичного ряду зменшити або збільшити на якесь постійне число С, то величина центрального моменту k -го порядка не зміниться. Так, якщо центральний момент розрахувати за зменшеними ($x_i - C$) або збільшеними ($x_i + C$) варіантами, то $\mu' = \mu$.

Таблиця 34

Розрахунок підсистеми центральних моментів

Порядок (степінь),k	Формула	Зміст	Взаємозв'язок з початковими моментами
0	$\mu_0 = \dfrac{\Sigma(x_i - \overline{x})^0 n_i}{\Sigma n_i}$	1	-
1	$\mu_1 = \dfrac{\Sigma(x_i - \overline{x})^1 n_i}{\Sigma n_i}$	0	$\mathrm{M}_i - \mathrm{M}_i$
2	$\mu_2 = \dfrac{\Sigma(x_i - \overline{x})^2 n_i}{\Sigma n_i}$	σ^2	$\mathrm{M}_2 - \mathrm{M}_i^2$
3	$\mu_3 = \dfrac{\Sigma(x_i - \overline{x})^3 n_i}{\Sigma n_i}$	Використовується для характеристики асиметрії розподілу	$\mathrm{M}_3 - 3\mathrm{M}_2\mathrm{M}_1 + 2\mathrm{M}_1^3$
4	$\mu_4 = \dfrac{\Sigma(x_i - \overline{x})^4 n_i}{\Sigma n_i}$	Використовується для характеристики гостровершиності	$\mathrm{M}_4 - 4\mathrm{M}_3\mathrm{M}_1 + 6\mathrm{M}_2\mathrm{M}_1^2 - 3\mathrm{M}_1^4$

2. Якщо усі варіанти статистичного ряду зменшити або збільшити в одне і те ж число раз (z), то центральний момент k -го порядку зменшиться або збільшиться у Z^k разів. Тобто якщо μ_k' центральний момент k –го порядку ( ) розраховувати за збільшеними у z разів варіантами ($x_i Z$), то величина його становитиме $\mu_k^1 = \mu Z^k$. Аналогічно одержуємо за зменшеними в z разів варіантами: $\mu_k' = \dfrac{\mu}{\mathrm{Z}^k}$.

Отже, одержавши центральний момент (μ) За зміненим рядом [напр., $x_i = (x_i - c) : \mathrm{Z}$], можна обчислити центральний момент k -го порядку для початкового ряду. Для випадку, коли варіанта зменшена на С одиниць, і це значення ($x_i - c$) у свою чергу зменшити у Z разів $[(x_i - c) : \mathrm{Z}\ ]$, то центральний момент (μ) k -го порядку буде дорівнювати:

$$\mu = \mu^1 \cdot \mathrm{Z}^{\kappa}.$$

Знання розглянутих вище властивостей підсистеми центральних моментів дозволить значно скоротити обсяги обчислювальних робіт, особливо у тих випадках, коли вибіркова сукупність представлена громіздкими розмірами величин досліджуваних ознак.

Як було зазначено раніше, центральний момент другого порядку (μ_2) являє собою дисперсію (σ^2). Якщо корінь квадратний з дисперсії (тобто середнє квадратичне відхилення) прийняти за стандарт ($\sqrt{\mu_2}$), то відношення центрального моменту k-го порядку до стандарту в k-тій степеня буде називатися **нормованим моментом.**

Загальна його формула має вигляд:

$$m_k = \frac{\mu_k}{(\sqrt{\mu_2})^k} = \frac{\mu_k}{\sigma^k}.$$

Згідно з наведеною вище формулою нормовані моменти від першого до четвертого порядків можна записати у такому вигляді: $m_1 = \frac{\mu_1}{\sigma^1} = \frac{0}{\sigma} = 0;\ m_2 = \frac{\mu_2}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1;\ m_3 = \frac{\mu_3}{\sigma^3};\ \ m_4 = \frac{\mu_4}{\sigma^4}$.

Послідовність розрахунку моментів розподілу полягає у складанні на першому етапі робочих таблиць. В останні заносяться вихідні і розрахункові дані з тим, щоб у подальшому їх використати для розрахунку тієї чи іншої формули моменту розподілу.

Наведемо форми таких таблиць у вигляді макетів (табл. 35, 36).

Таблиця 35

Вихідні і розрахункові дані для обчислення початкових моментів ряду розподілу

Варіанта, x_i	Частота n_i	Розрахункові дані						
		$x_i^{\ 2}$	$x_i^{\ 3}$	x_i^4	$x_i n_i$	$x_i^2 n_i$	$x_i^3 n_i$	$x_i^4 n_i$
…	…	…	…	…	…	…	…	…
Всього	…	х	х	х	…	…	…	…

Таблиця 36

Вихідні і розрахункові дані для обчислення центральних моментів ряду розподілу

Варіанта, x_i	Частота, n_i		Розрахункові дані					
		$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$	$(x_i - \bar{x})n_i$	$(x_i - \bar{x})^2 n_i$	$(x_i - \bar{x})^3 n_i$	$(x_i - \bar{x})^4 n_i$
…	…	…	…	…	…	…	…	…	…
Всього	х	х	х	х	х	…	…	…	..

Розглянуті ваше підсистеми моментів використовуються як статистичні характеристики розподілу. У статистичних розрахунках іноді звертаються до так званих умовних моментів. Одержують цю форму моментів при $A = x_0$, де x_0 – деяка варіанта (умовний початок). За x_0 приймається величина досліджуваної ознаки, яка близька до середньої варіанти $(\bar{x})$, тобто до варіанти, розміщеної приблизно в середині варіаційного ряду. Така варіанта , як правило, має найбільшу частоту.

Умовний момент k-го порядку має вигляд:

$$m_k = \frac{\Sigma(x_i - x_0)^k n_i}{\Sigma n_i}.$$

Як видно з формули, умовні моменти являють собою середні різних степенів із відхилень варіант від умовного початку (несправжнього нуля). Умовні моменти першого, другого, третього і вищого порядків будуть виражатися відповідними формулами:

$$m_1 = \frac{\Sigma(x_i - x_0)^1 n_i}{\Sigma n_i};\quad m_2 = \frac{\Sigma(x_i - x_0)^2 n_i}{\Sigma n_i};\quad m_3 = \frac{\Sigma(x_i - x_0)^3 n_i}{\Sigma n_i}$$ і т.д.

Відзначимо, що умовні моменти першого і другого порядків використовуються для спрощення розрахунків відповідно середньої арифметичній і дисперсії.

Для інтервального варіаційного ряду з рівними інтервалами розрахунок умовних моментів може бути значно спрощений, якщо відхилення ($x_i - x_0$) розділити на величину інтервала (і). Знаючи, що в основі дискретного ряду розподілу лежить арифметична прогресія, за аналогією можна спростити розрахунки умовних моментів і для цього виду рядів розподілу.

Відношення $\frac{(x_i - x_0)}{i}$ називають <u>умовними варіантами.</u>

Останні, як бачимо, використовують для спрощених методів розрахунку зведених характеристик вибірки шляхом заміни початкових варіант умовними.

Приклад. Необхідно знайти умовні варіанти для дискретного ряду розподілу 50 працівників за середньоденним рівнем зарплати (табл. 37).

Таблиця 37

Дискретний ряд розподілу

Варіанта, x_i	Частота, n_i
22,50	3
26,50	8
30,50	25
34,50	10
38,50	4

За умовний початок (несправжній нуль - x_0) приймаємо варіанту 30,5 (ця варіанта розміщена в середині варіаційного ряду). Різниця між сусідніми (будь-якими) варіантами (і) дорівнює 4. Для інтервального варіаційного ряду - це величина інтервалу.

Умовна варіанта буде дорівнюватиме $x_i^1 = \frac{x_1 - x_0}{i} = \frac{22{,}5 - 30{,}5}{4} = -2.$

Аналогічно розраховуємо інші умовні варіанти:

$x_2' = -1;\ x_2 = 1;\ \ x_3 = 1;\ \ x_4' = 1;$

$x_5' = 2.$ Як бачимо, одержані значення умовних варіант - це цілі числа, невеликі за обсягом, з якими набагато спрощуються обчислювальні операції в порівнянні з варіантами початкового ряду (22,50; 26,50; 30,50; 34,50; 38,50).

Маючи обчислені значення умовних варіант, можна знайти умовний емпіричний момент, який являє собою початковий момент k -го порядку, обчислений для умовних варіант:

$$m_k^{'} = \frac{\Sigma x_i^k n_i}{\Sigma n_i} = \frac{\Sigma\left(\frac{x_i - x_0}{i}\right)^k n_i}{\Sigma n_i}.$$

Так, умовний момент першого порядку буде дорівнювати:

$$m_1' = \frac{\Sigma\left(\frac{x_i - x_0}{i}\right) n_i}{\Sigma n_i} = \left(\frac{\Sigma x_i n_i}{\Sigma n_i} - x_0 \frac{\Sigma n_i}{\Sigma n_i}\right)\frac{1}{i} = \frac{1}{i}(\bar{x} - x_0).$$

Звідси $\bar{x} = m_1' i + x_0.$

Отже, щоб знайти середню вибірки, необхідно умовний момент першого порядку помножити на величину інтервалу (і) і до одержаного добутку додати варіанту, прийняту за умовний початок (несправжній нуль).

Від умовних моментів можна перейти до розрахунку початкових моментів розподілу (M_K): $m_k^{'} = \frac{1}{i^k} \times \frac{\Sigma(x_i - x_0)^k n_i}{\Sigma n_i} = \frac{M_K}{i^k}$.

Звідси початковий момент k -го порядку дорівнює

$M_K = m_k^{'} \times i^k$.

Отже, щоб знайти початковий момент k -го порядку, достатньо умовний момент цього порядку помножити на величину інтервалу в k-ій степені.

§ 5.5. Характеристика асиметрії і ексцесу

Щоб оцінити (дати оцінку) відхилення емпіричного розподілу від нормального, розраховують такі статистичні характеристики, як **коефіцієнти асиметрії** і **гостровершинності – ексцесу**. Перший з названих коефіцієнтів характеризує ступінь скошеності варіаційного ряду розподілу щодо його симетрії вправо або вліво.

При зміщенні вправо від центра асиметрія буде характеризуватися додатнім числом, при зміщенні вліво – від'ємним.

Коефіцієнт асиметрії (A_s) розраховується як відношення центрального моменту третього порядку до куба середнього квадратичного відхилення : $A_s = \frac{\mu_3}{\sigma^3}$, тобто $A_s = m_3$.

Як бачимо, коефіцієнт асиметрії - це нормований момент третього порядку (m_3). Вважається, що криві з абсолютною величиною показника асиметрії $A_s > \pm 0,5$ характеризуються значним зміщенням. Якщо $A_s \langle \pm 0,25$ – асиметрія незначна.

Графічно (рис.8) асиметрія описується напрямом більш довгої вітки кривої («дзвін»).

Якщо $A_S = 0$ – розподіл симетричний, якщо $A_s > 0$ – розподіл має правосторонню асиметрію; якщо $A_s < 0$ – лівосторонню асиметрію.

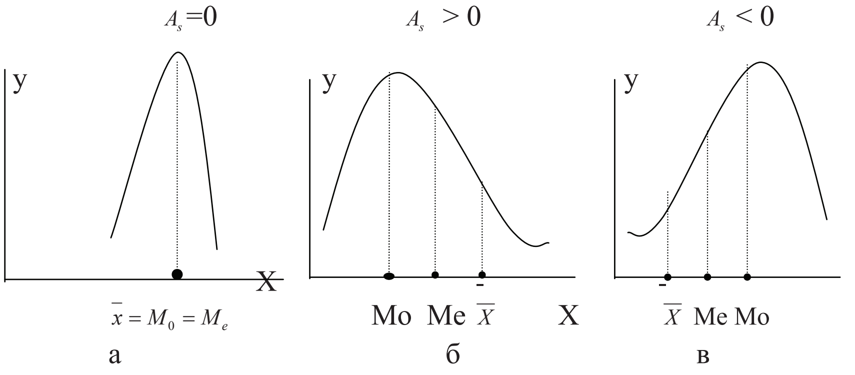


Рис. 8. Форми розподілу при різних значеннях коефіцієнта асиметрії (A_s)

Криві, зображені на рисунку 8, дозволяють ілюструвати симетрію і два найбільш поширених види асиметрії розподілу. При симетричному розподілі (а) середня арифметична, мода і медіана рівні між собою. Для асиметричних кривих ці статистичні величини неоднакові. Причому середня арифметична і медіана зміщені від центра вбік довшої вітки кривої. Оскільки середня арифметична ($\bar{x}$) «чутка» до «точного» положення більш віддалених від моди (M_0) точок кривої, а медіана (M_e) «нечутка», то середня ($\bar{x}$) зрушена більше, ніж медіана (M_e). У цьому випадку медіана знаходиться між модою і середньою арифметичною.

Як бачимо, напрямок асиметрії геометрично встановлюється дуже просто. Кількісна форма степеня асиметрії вимагає знаходження її алгебраїчної міри.

Приклад. За даними дискретного статистичного ряду розподілу господарств за врожайністю зернових культур потрібно кількісно виміряти асиметрію розподілу варіант у даній вибірці. Для знаходження величини коефіцієнта асиметрії за наведеною вище формулою необхідно виконати додаткові розрахунки. Останні наведені в таблиці 38.

Таблиця 38

Вихідні і розрахункові дані для обчислення коефіцієнтів асиметрії (A_s) і ексцесу (E_x)

Варіанта, x_i	Частота n_i	Розрахункові дані ($\bar{x}$ = 26,8)							
		$x_i n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$	$(x_i - \bar{x})^2 n_i$	$(x_i - \bar{x})^3 n_i$	$(x_i - \bar{x})^4 n_i$
17,6	7	123,2	-9,2	84,64	-778,7	7163,9	592,5	-5450,9	50147,3
21,6	11	237,6	-5,2	27,04	-140,6	731,1	297,4	-4546,7	8042,1
25,6	18	460,8	-1,2	1,44	-1,7	2,1	25,9	-30,0	37,8
29,6	9	266,4	2,8	7,84	21,9	61,3	70,6	197,1	551,7
33,6	5	168,0	6,8	46,24	314,4	2138,1	231,2	1572,0	10690,5
37,6	5	188,0	10,8	116,64	1259,7	13604,9	583,2	6298,5	68024,5
41,6	2	83,3	4,8	219,04	3241,8	47978,5	438,1	6483,6	95957,0
Всього	57	1527,2	x	x	x	x	2238,9	4523,6	233450,9

За попередніми розрахунками маємо: $\mu_2 = \dfrac{\Sigma(x_i - \bar{x})^2 n_i}{\Sigma n_i} = \dfrac{2238{,}9}{57} = 39{,}3$.

Оскільки $\mu_2 = \sigma^2$ величина середнього квадратичного відхилення становить:

$\sigma = \sqrt{\sigma^2} = \sqrt{39{,}3} = 6{,}3$.

Значення величини центрального моменту третього порядку $/\mu_3/$ одержуємо із виразу $\mu_3 = \dfrac{\Sigma(x_i - \bar{x})^3 n_i}{\Sigma n_i} = \dfrac{4523{,}6}{57} = 79{,}4$.

Підставивши значення μ_3 і σ^3 у формулу коефіцієнта асиметрії, маємо: $\mathrm{A}_s = \dfrac{\mu_3}{\sigma^3} = \dfrac{79{,}4}{6{,}3^3} = \dfrac{79{,}4}{250{,}0} = 0{,}318$.

Додатне значення показника A_s свідчить про правосторонню асиметрію розподілу господарств за урожайністю, а абсолютне його значення 0,25< |0,3181 < 0,5| означає наявність помірної асиметрії в досліджуваному ряді розподілу.

На завершення слід відмітити, що при переважаючій кількості варіант в ряді розподілу менших за величиною від вибіркової середньої коефіцієнт асиметрії буде від'ємним. Якщо в варіаційному ряді переважають варіанти за величиною більше середньої, буде мати місце додатна асиметрія.

Негативною стороною коефіцієнта асиметрії, як міри асиметрії, слід назвати ту, що цей показник не має ні верхньої, ні нижньої границі. Особливо великий розмір коефіцієнта асиметрії практично майже не має місця.

Крім розглянутого способу оцінки міри асиметрії, існують і інші методичні прийоми. Вони є предметом вивчення спеціального курсу.

Для встановлення міри відхилення від нормального розподілу вираховують показник ексцесу (E_X). Він характеризує відхилення від нормального розподілу варіант із виступанням або падінням вершини кривої розподілу. При виступанні вершини ексцес називають додатним, при її падінні – від'ємним.

Для кількісного виміру гостровершинності використовується центральний момент четвертого порядку (μ_4). Відношення останнього до середньоквадратичного відхилення в четвертому степені називають **коефіцієнтом гостровершинності** (ексцес). Тобто обчислюється нормований момент четвертого порядку ($\frac{\mu_4}{\sigma^4}$). Якщо за базу порівняння прийняти нормальний розподіл, то відношення $\frac{\mu_4}{\sigma^4} = 3$. Тому коефіцієнт гостровершинності або просто «ексцес» (Ex) буде виражатися формулою:

$$\mathrm{E}_x = \frac{\mu_4}{\sigma^4} - 3.$$

Якщо степінь гостровершинності нормальний, $\mathrm{E}_x = 0$. Для більш гостровершинних розподілів, ексцес буде додатнім ($\mathrm{E}_x \rangle 0$), для більш плосковершинних – від'ємним ($\mathrm{E}_x \langle 0$). Форми «вершинності» для вказаних випадків представлено на графіку (рис. 9).

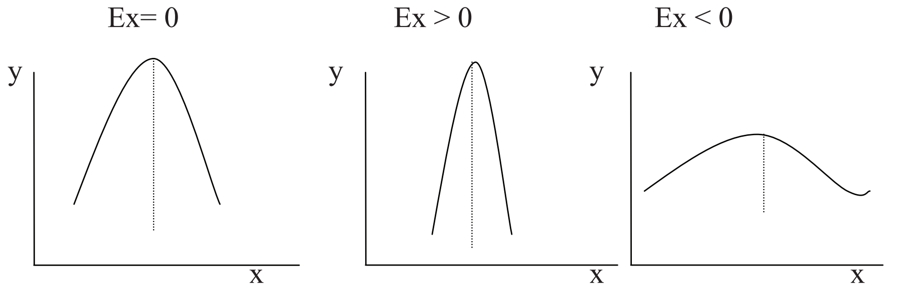


Рис. 9. Форми розподілу для різних значень ексцесу (E_x).

Якщо величина показника ексцесу $E_x = 0{,}4,$ крива вважається слабоексцесивною. Найбільша абсолютна величина від’ємного ексцесу становить мінус 2. При такому значенні вершина кривої опускається до осі абсцис і крива розподілу ділиться на дві самостійні одновершинні криві.

Слід відзначити, що термін «ексцес» грецького походження (kurtosis), і назви форми ексцесу в різних кривих розподілу мають корінь цього слова. Мається на увазі стрічкокуртичні, платокуртичні і мезокуртичні криві (рис. 10).

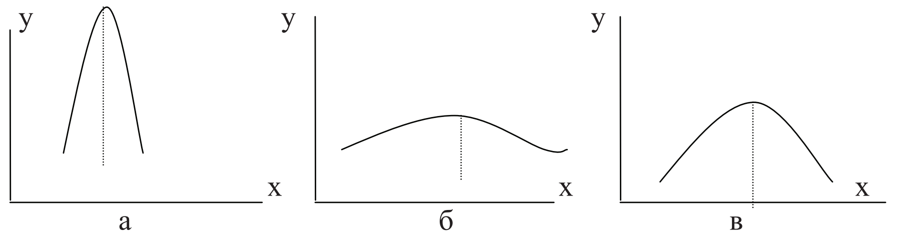


Рис. 10. Типи ексцесу: а - стрічкокуртичний, б - платокуртичний, в – мезокуртичний

Відповідно до графіка, стрічкокуртичної кривої (а) характерне розміщення більшості одиниць спостережень поблизу центра. У випадку платокуртичної кривої (форма силуету плато) варіанти значно віддалені від центра розподілу (б). Помірне розміщення навколо центра розподілу варіант визначає форма ексцесу у вигляді мезокуртичної кривої (в).

Для кривої нормального розподілу (при $\bar{x} = 0$ і $\sigma = 1$) значення коефіцієнта ексцесу (E_x) становить 0,263. При цьому його величина обчислюється за формулою:

$$E_x = \frac{\frac{1}{2}(Q_3 - Q_1)}{P_{90} - P_{10}},$$

де $Q_3 + Q_1$ - відповідно третя і перша квартилі;

P_{90}, P_{10} - дев'яностий і десятий перцентилі.

Таким чином, коефіцієнт ексцесу визначається відношенням половини розмаху між двома квартилями до різниці 90-го і 10-го перцентилей. (Величини, які характеризують розділення розподілу на чотири, десять і сто рівних частин, називаються відповідно квартилями, децилями і перцентилями).

Згідно з розподілом підприємств за врожайністю зернових культур (табл. 38), величина коефіцієнта ексцесу становить:

$$E_x = \frac{\mu_4}{\sigma^4} - 3 = \frac{233450 : 57}{6{,}3^4} - 3 = -0{,}52.$$

Відповідно до величини розрахованого показника ексцесу ($E_x = -0{,}52$), крива розподілу характеризується платокуртичною формою зі слабовираженою ексцесивністю, тобто форма кривої на графіку - плосковершинна.

ТЕМА 6. АНАЛІЗ ПОДІБНОСТІ РОЗПОДІЛІВ

§ 6.1. Статистична оцінка параметрів розподілу

Питання статистичної оцінки пов'язують в єдине ціле такі проблемні аспекти математичної статистики, як наукова методологія, випадкові величини, статистичні розподіли та ін. Для будь-якої вибірки притаманні помилки, зумовлені неповнотою охоплення одиниць, помилками вимірювання і тому подібними причинами. Такі помилки в реальному житті надають кожній гіпотезі (зокрема, сформульованій на базі економічних висновків) випадковий, стохастичний характер. Незалежно від кількості змінних, передбачених теоретичними гіпотезами, робиться припущення, що вплив різних видів помилок може бути достатньо точно описаний за допомогою лише однієї складової. Такий методологічний підхід дозволяє обмежитися одномірним розподілом імовірностей при одночасному оцінюванні декількох параметрів.

Статистична оцінка – це один із двох типів статистичного судження (другий тип – перевірка гіпотез). Вона являє собою особливого роду метод судження про числові значення характеристик (параметрів) розподілу генеральної сукупності за даними вибірки з цієї сукупності. Тобто, маючи результати вибіркового спостереження, ми намагаємося оцінити (з найбільшою точністю) значення визначених параметрів, від яких залежить розподіл ознаки (змінної), яка нас цікавить, у генеральній сукупності. Оскільки вибірка включає тільки частину одиниць генеральної сукупності (інколи дуже мале їх число), існує ризик допустити помилку. Незважаючи на зменшення такого ризику зі збільшенням числа одиниць спостереження, він все ж має місце при вибірковому спостереженні. Звідси, прийнятим за результатами вибірки рішенням надають імовірнісний характер. Але було б невірним розглядати статистичні судження тільки з позицій імовірностей. Такій підхід не завжди виявляється достатнім для побудови правильних теоретичних припущень відносно параметрів генеральної сукупності. Часто потрібен ще ряд додаткових суджень, які б забезпечили більш глибоке обґрунтування. Наприклад, потрібно оцінити з можливо більшим наближенням значення середньої чисельності кваліфікованих робітників у підприємствах регіону. При цьому оцінюється середня арифметична змінної x з генеральної сукупності, яка має нормальний розподіл. Одержавши вибірку по даній ознаці в кількості n одиниць, необхідно розв’язати питання: яку величину за даними вибірки необхідно прийняти як найбільш близьку до середньої в генеральній сукупності? Таких величин, математичне очікування яких дорівнює шуканому параметру (або близьке до нього), можна навести кілька: а) середня арифметична; б) мода; в) медіана; г) середня, обчислена за розмахом варіації, і т.д.

Із імовірнісної точки зору кожну з названих вище величин можна вважати такими, що дають найкраще наближення до шуканого параметра генеральної сукупності ($\bar{x}$), оскільки математичне очікування кожної з цих функцій (особливо для великих вибірок) дорівнює генеральній середній. Зумовлюється таке припущення тим, що при багаторазовому повторенні вибірки із тієї самої генеральної сукупності буде одержаний «в середньому» вірний результат.

Правильність «в середньому» пояснюється рівністю повторювань додатних і від’ємних відхилень виникаючих помилок оцінки генеральної середньої, тобто середня помилка оцінки буде дорівнювати нулю.

У практичних умовах, як правило, організовують одну вибірку, тому дослідника цікавить питання про більш точну оцінку шуканого параметра за результатами конкретної вибірки. Для вирішення такого завдання, крім висновків, які випливають безпосередньо з абстрагованого обчислення ймовірностей, потрібні додаткові правила мотивування найкращого наближення оцінки до шуканого параметра генеральної сукупності.

Існує достатня кількість способів оцінки констант за вибірковими спостереженнями. Які з них найкращі у рішенні конкретних завдань дослідження – займається теорія статистичного оцінювання. Вона досліджує умови, яким повинна підпорядковуватися та чи інша оцінка, орієнтує на оцінки, більш переважаючі при даних обставинах. Теорія оцінок вказує на перевагу однієї оцінки порівняно до іншої.

Як відомо, інформація, одержана на основі вибірки, не носить категоричного характеру у висновку. Якщо, наприклад, із досліджуваних 100 голів тварин щодо їх захворювання здоровими виявилися 99, то існує ймовірність, що одна тварина, яка залишилася необстеженою саме носить у собі вірус передбачуваного захворювання. Оскільки це мало ймовірно, робиться висновок про відсутність даного захворювання. У більшості випадків такий висновок повністю виправдовується.

Керуючись подібними висновками в практичній діяльності, експериментатор (дослідник) спирається не на вірогідність інформації, а лише на її ймовірність.

Другий бік вибіркового спостереження, як уже відзначалося, вирішує завдання можливо більш об'єктивного визначення ступеня надійності одержуваних вибіркових оцінок. Розв’язуванню цього завдання намагаються надати якомога точніший імовірнісний вираз, тобто мова йде про визначення ступеня точності оцінки. Тут дослідник визначає межі можливої розбіжності між оцінкою, одержаної при вибірці, і дійсним значенням її величини в генеральній сукупності.

Точність оцінки зумовлюється способом її розрахунку за даними вибірки і способом відбору одиниць у вибіркову сукупність.

Спосіб одержання оцінок передбачає будь-яку обчислювальну процедуру (метод, правило, алгебраїчну формулу). Це пріоритет теорії статистичного оцінювання. Способи відбору ведуть до питань техніки здійснення вибіркового дослідження.

Викладене вище дозволяє дати визначення поняттю «статистична оцінка».

Статистична оцінка – це наближене значення шуканого параметра генеральної сукупності, яке одержано за результатами вибірки й забезпечує можливість прийняття обґрунтованих рішень про невідомі параметри генеральної сукупності.

Припустимо, що $\widetilde{\theta}_n$ – статистична оцінка невідомого параметра θ теоретичного розподілу. За багаторазово здійснюваними однакового обсягу вибірками з генеральної сукупності знайдені оцінки $\widetilde{\theta}_1, \widetilde{\theta}_2, \widetilde{\theta}_3, ..., \widetilde{\theta}_n$, що мають різні значення. Тому оцінку $\widetilde{\theta}_n$, можна розглядати як випадкову величину, а $\widetilde{\theta}_1, \widetilde{\theta}_2, \widetilde{\theta}_3, ..., \widetilde{\theta}_n$ – як її можливі значення. Як випадкова величина, вона характеризується певною функцією щільності ймовірностей. Оскільки ця функція зумовлена результатом вибіркового спостереження (експерименту), то її називають **вибірковим розподілом.** Така функція описує щільність імовірності для кожної із оцінок, використовуючи певне число вибіркових спостережень. Якщо припустити, що, статистична оцінка $\widetilde{\theta}_n$, – це алгебраїчна функція від певного набору даних і такий набір буде одержаний при здійсненні вибіркового спостереження, то в загальному вигляді оцінка одержить вираз: $\widetilde{\theta}_n = f(x_1.x_2, x_3, ...x_m)$.

По закінченні вибіркового обстеження дана функція вже не є оцінкою загального вигляду, а приймає – конкретне значення, тобто стає кількісною оцінкою (числом). Інакше кажучи, з вищенаведеного виразу функції випливає, що будь який з показників, які характеризують результати вибіркового спостереження, можна вважати оцінкою. Вибіркова середня є оцінкою генеральної середньої. Розрахована за вибіркою дисперсія або обчислене з неї значення середнього квадратичного відхилення є оцінками відповідних характеристик генеральної сукупності і т.ін.

Як уже відмічалося, розрахунок статистичних оцінок не гарантує виключення помилок. Суть полягає в тому, що останні не повинні бути систематичними. Наявність їх має носити випадковий характер. Розглянемо методологічну сторону цього положення.

Припустимо, оцінка $\widetilde{\theta}_n$ дає неточне значення оцінки θ генеральної сукупності з нестачею. У цьому випадку кожне обчислене значення $\theta_i (i = 1,2,3,...,n)$ буде меншим за дійсне значення величини θ .

З цієї причини математичне очікування (середнє значення) випадкової величини $\widetilde{\theta}$ буде менше, ніж θ, тобто $(\mathrm{M}(\theta_n)\langle\theta)$. І, навпаки, якщо $\widetilde{\theta}_n$ дає оцінку з надлишком, то і математичне очікування випадкової $\widetilde{\theta}_n$ стане більшим, ніж θ .

Звідси випливає, що використання статистичної оцінки, математичне очікування якої не дорівнює оцінюваному параметру, призводить до систематичних похибок, тобто до невипадкових помилок, які викривляють результати вимірювань в один бік. Виникає природна вимога: математичне очікування оцінки $\widetilde{\theta}_n$ повинно дорівнювати оцінюваному параметру. Дотримання цієї вимоги не усуває помилок у цілому, оскільки вибіркові значення оцінки можуть бути більші або менші дійсного значення оцінки генеральної сукупності. Але помилки в один і другий бік від значень θ будуть зустрічатися (згідно з теорією імовірностей) з однаковою частотою. Отже, дотримання цієї вимоги, що математичне очікування вибіркової оцінки повинно дорівнювати оцінюваному параметру, виключає одержання систематичних (невипадкових) помилок, тобто

$$\mathrm{M}(\hat{\theta}) = \theta.$$

Вибір статистичної оцінки, яка дає найкраще наближення оцінюваного параметра, являє собою важливу задачу в теорії оцінювання. Якщо відомо, що розподіл досліджуваної випадкової величини в генеральній сукупності відповідає закону нормального розподілу, то за вибірковими даними необхідно оцінити математичне очікування і середнє квадратичне відхилення. Пояснюється це тим, що названі дві характеристики повністю визначають основи, на яких побудовано нормальний розподіл. Якщо досліджувана випадкова величина розподілена за законом Пуассона, оцінюють параметр λ, оскільки він визначає цей розподіл.

Математична статистика розрізняє такі методи одержання статистичних оцінок за вибірковими даними : метод моментів, метод максимуму правдоподібності.

При одержанні оцінок методом моментів моменти генеральної сукупності замінюються моментами вибіркової сукупності (замість ймовірностей за ваги використовують частоти).

Щоб статистична оцінка давала «найкраще наближення» до генеральної характеристики, вона повинна мати ряд властивостей. Про них мова піде нижче.

Можливість вибору найкращої оцінки зумовлюється знанням їх основних властивостей і вмінням класифікувати оцінки за цими властивостями. У математичній літературі «властивості оцінок» інколи називають «вимоги до оцінок» або «критерії оцінок».До основних властивостей статистичних оцінок належать: незміщеність, ефективність, спроможність, достатність.

Якщо прийняти, що вибіркова середня ($\tilde{x}$) і вибіркова дисперсія ($\sigma_в^2$) є оцінками відповідних генеральних характеристик ($\bar{x}, \sigma_г^2$), тобто їх математичним очікуванням, враховуємо, що при великій кількості одиниць вибірки названі характеристики ($\tilde{x}$,$\sigma_в^2$) будуть наближені до їх математичних очікувань. Якщо ж число одиниць вибірки невелике, ці характеристики можуть значно відрізнятися від відповідних математичних очікувань.

Якщо середнє значення вибіркових характеристик, вибраних як оцінки, відповідає значенню генеральної характеристики, оцінка називається **незміщеною.** Доказом того, що математичне очікування вибіркової середньої дорівнює генеральній середній ($M(\tilde{x}) = \bar{x}$), свідчить про те, що величина $\tilde{x}$ є незміщеною генеральною середньою. Інакше виглядає справа з вибірковою дисперсією ($\sigma_в^2$). Її математичне очікування $M(\sigma_в^2) = \frac{n-1}{n}\sigma_г^2$, не дорівнює генеральній дисперсії. Отже, $\sigma_В^2$ є зміщеною оцінкою $\sigma_г^2$. Щоб усунути систематичну помилку і отримати незміщену оцінку, вибіркову дисперсію множать на поправку $\frac{n}{n-1}$ (це випливає з утворення наведеного вище рівняння: $\frac{\frac{\sigma_в^2}{n-1}}{n} = \sigma_в^2 \frac{n}{n-1}$).

Таким чином, при нечисленній вибірці дисперсія дорівнюватиме:

$$\sigma_в^2 = \frac{\Sigma(x_i - \tilde{x})^2}{n} \times \frac{n}{n-1} = \frac{\Sigma(x_i - \tilde{x})^2}{n-1}.$$

Дріб ($\frac{n}{n-1}$) називають поправкою Бессела. Математик Бессел перший встановив, що вибіркова дисперсія є зміщеною оцінкою генеральної дисперсії і застосував вказану поправку для коригування оцінок. Для малих вибірок поправка ($\frac{n}{n-1}$) значно відрізняється від 1. Зі збільшенням числа одиниць спостереження вона швидко наближається до 1. При $n \rangle 50$ різниця між оцінками зникає, тобто $\sigma_в^2 = \sigma_г^2$.Із всього сказаного вище випливають такі визначення вимог незміщеності.

Незміщеною називають статистичну оцінку, математичне очікування якої при будь-якому обсязі вибірки дорівнює значенню параметра генеральної сукупності, тобто $\mathrm{M}(\tilde{\theta}) = \theta; \ \mathrm{M}(\tilde{x}) = \bar{x}$.

Категорію «математичне очікування» вивчають у курсі теорії ймовірностей. Це числова характеристика випадкової величини. Математичне очікування наближено дорівнює середньому значенню випадкової величини. **Математичним очікування дискретної випадкової величини** називають суму добутків усіх її можливих значень на їх імовірності. Припустимо, виконано n досліджень, в яких випадкова величина x прийняла m_1 разів значення x_1, m_2 разів значення $x_2,...,m_k$ разів значення x_k . При цьому $m_1 + m_2 + m_3 + ... + m_k = n$. Тоді сума всіх значень, прийнятих x ,дорівнює $x_1m_1 + x_2m_2 + x_3m_3 + ... + x_km_k$. .

Середня арифметична цих значень становитиме:

$$\bar{x} = \frac{x_1m_1 + x_2m_2 + x_3m_3 + ... + x_km_k}{n}$$ або $$\bar{x} = x_1\frac{m_1}{n} + x_2\frac{m_2}{n} + x_3\frac{m_3}{n} + ... + x_k\frac{m_k}{n}.$$

Оскільки $\frac{m_1}{n}$ - відносна частота w_1 значення x_1, $\frac{m_2}{n}$ - відносна частота значення x_2 і т.д., наведене вище рівняння набуде вигляду:

$$\bar{x} = x_1w_1 + x_2w_2 + x_3w_3 + ... + x_kw_k.$$

При великій кількості вибіркових спостережень відносна частота приблизно дорівнює ймовірності появи події, тобто

$w_1 = p_1; \ w_2 = p_2; \ w_3 = p_3,..., \ w_k = p_k$ або $\bar{x} \cong x_1p_1 + x_2p_2 + x_3p_3 + ... + x_kp_k$. Тоді $x \approx M(x)$ Імовірнісний зміст одержаного результату розрахунків полягає в тому, що математичне очікування наближено дорівнює (тим точніше, чим більша вибірка) середньому арифметичному спостережуваних значень випадкової величини $[\mathrm{M}(x_i) = \tilde{x} \ ]$.

Критерій незміщеності гарантує відсутність систематичних помилок в оцінці параметрів генеральної сукупності .

Зауважимо, що вибіркова оцінка ($\hat{\theta}$) – випадкова величина, значення якої може змінюватися від однієї вибірки до іншої. Міру її варіації (розсіювання) навколо математичного очікування параметра генеральної сукупності θ характеризує дисперсія $\sigma^2(\theta)$.

Нехай $\hat{\theta}_n$, і $\hat{D}_n$ – дві незміщені оцінки параметра θ , тобто $\mathrm{M}(\tilde{\theta}_n) = \theta$ і $\mathrm{M}(D_n) = \theta.$ Дисперсії їх $\sigma^2(\hat{\theta}_n)$ і $\sigma^2(\hat{D}_n)$. З двох оцінок варто віддати перевагу тій, яка має менше розсіювання навколо оцінюваного параметра. Якщо дисперсія оцінки $\tilde{\theta}_n$ менша дисперсії оцінки $\hat{D}_n$, то за оцінку θ приймається перша, тобто $\tilde{\theta}_n$.

Незміщена оцінка $\tilde{\theta}$, що має найменшу дисперсію серед усіх можливих незміщених оцінок параметра θ, обчислених за вибірками однакового обсягу, називається **ефективною оцінкою**. Це – друга властивість (вимога) статистичних оцінок параметрів генеральної сукупності. Треба, пам'ятати, що ефективна оцінка параметра генеральної сукупності, підпорядкованої певному закону розподілу, не збігається з ефективною оцінкою параметра другого розподілу.

При розгляді вибірок великого обсягу статистичні оцінки повинні мати властивість спроможності. Оцінка **спроможна** (застосовується також термін «придатна» чи «узгоджена») означає, що чим більше обсяг вибірки, тим більша ймовірність того, що помилка оцінки не перевищить скільки завгодно малого додатного числа Е. Оцінка $\hat{\theta}$ параметра θ називається **спроможною,** якщо вона підпорядковується закону великих чисел, тобто виконується така рівність:

$$\lim_{n\to\infty}\left\{P\left\|\hat{\theta}-\theta\right| < \mathrm{E}\ \right\} = 1.$$

Як бачимо, спроможною називають таку статистичну оцінку, яка при $n \to \infty$ наближається за ймовірністю до оцінюваного параметра. Іншими словами, це значення показника, одержане за вибіркою і яке наближається (збігається за ймовірністю) внаслідок закону великих чисел при збільшенні обсягу вибірки до свого математичного очікування. Наприклад, якщо дисперсія незміщеної оцінки при $n \to \infty$ наближається до нуля, то така оцінка виявляється і спроможною, оскільки має найменшу можливу дисперсію (при заданому обсязі вибірки).

Спроможними оцінками є:

1) частка ознаки у вибірковій сукупності, тобто частість як оцінка частки ознаки в генеральній сукупності;

2) вибіркова середня як оцінка генеральної середньої;

3) вибіркова дисперсія як оцінка генеральної дисперсії;

4) вибіркові коефіцієнти асиметрії і ексцесу як оцінка генеральних коефіцієнтів.

У літературі з математичної статистики чомусь не завжди можна зустріти опис четвертої властивості статистичних оцінок – достатність. Оцінка **достатня** (або вичерпна) – це оцінка, яка зумовлює (забезпечує) повноту обхвату всієї вибіркової інформації про невідомий параметр генеральної сукупності. Таким чином, достатня оцінка включає всю інформацію, яка міститься у вибірці стосовно досліджуваної статистичної характеристики генеральної сукупності. Жодна з розглядуваних раніше трьох оцінок не може дати необхідних додаткових відомостей про досліджуваний параметр, як достатня статистична оцінка.

Отже, середня арифметична вибіркова $\tilde{x}$ є незміщеною оцінкою середньої арифметичної генеральної $\bar{x}$. Фактор незміщеності цієї оцінки показує: якщо із генеральної сукупності взяти велику кількість випадкових вибірок, то їх середні $\tilde{x}_i$ відрізнялись би від генеральної середньої у більший і менший бік однаково, тобто, властивість незміщеності хорошої оцінки також показує, що середнє значення нескінченно великого числа вибіркових середніх дорівнює значенню генеральної середньої.

У симетричних рядах розподілу медіана є незміщеною оцінкою генеральної середньої. А за умови, що чисельність вибіркової сукупності наближається до генеральної ($n \to N$), медіана може бути в таких рядах і спроможною оцінкою генеральної середньої.Що ж стосується критерію ефективності відносно медіани як оцінки середньої арифметичної генеральної сукупності, можна довести, що у вибірках великого обсягу середньоквадратична помилка медіани (σ_{Me}) дорівнює 1,2533 середньоквадратичної помилки вибіркової середньої (σ_{x}). Тобто $\sigma^2_{Me} \rangle \sigma^2_{\bar{x}}$. Тому медіана не може бути ефективною оцінкою середньої арифметичної генеральної сукупності, оскільки її середня квадратична помилка більше середньої квадратичної помилки середньої арифметичної вибірки. До того ж середня арифметична задовольняє умовам незміщеності і спроможності, а, отже, є кращою оцінкою.

Можлива і така постановка. Чи може середня арифметична вибірки бути незміщеною оцінкою медіани в симетричних розподілах сукупності, для якої збігаються значення середньої і медіани? І чи буде вибіркова середня спроможною оцінкою медіани генеральної сукупності? В обох випадках відповідь буде позитивною. Для медіани генеральної сукупності (з симетричним розподілом) середня арифметична вибірки є незміщеною і узгодженою оцінкою.

Пам'ятаючи, що $\sigma_{Me} = 1{,}2533\sigma_{\bar{x}}$, приходимо до висновку: середня арифметична вибірки, а не медіана, є більш ефективною оцінкою медіани досліджуваної генеральної сукупності.

Кожна характеристика вибірки не обов'язково є найкращою оцінкою відповідної характеристики генеральної сукупності. Знання властивостей оцінок дозволяє вирішувати питання не тільки вибору оцінок, але й їх поліпшення. Як приклад можна розглянути випадок, коли розрахунки показують, що значення середніх квадратичних відхилень декількох вибірок із однієї генеральної сукупності у всіх випадках виявляються менше середнього квадратичного відхилення генеральної сукупності, причому величина різниці зумовлена обсягом вибірки. Помноживши значення середнього квадратичного відхилення вибірки на поправочний коефіцієнт, одержимо поліпшену оцінку середнього квадратичного відхилення генеральної сукупності. За такий поправочний коефіцієнт використовують поправку Бессела ($\frac{n}{n-1}$), тобто для усунення зміщення оцінки одержують $\sigma_{\bar{x}}\sqrt{\frac{n}{n-1}}$.Такий числовий вираз показує, що середнє квадратичне відхилення вибірки, використане як оцінка, дає занижене значення параметра генеральної сукупності.

Як відомо, статистичні характеристики вибіркової сукупності є наближеними оцінками невідомих параметрів генеральної сукупності. Сама оцінка може мати форму одного числа або якої-небудь певної точки. Оцінка, яка визначається одним числом, називається **точковою.** Так, вибіркова середня ($\tilde{x}$) є незміщеною і найбільш ефективною точковою оцінкою генеральної середньої ($\bar{x}$), а вибіркова дисперсія ($\sigma_в^2$) - зміщеною точковою оцінкою генеральної дисперсії (σ^2) .Якщо позначити середню помилку вибіркової середньої m_θ , то точкову оцінку генеральної середньої можна записати у вигляді $\bar{x} \pm m_\theta$. Це означає, що $\tilde{x}$ – оцінка генеральної середньої $\bar{x}$ з помилкою, яка дорівнює m_θ . Зрозуміло, що точкові статистичні оцінки $\tilde{x}$ і $\sigma_в^2$ не повинні мати систематичної помилки в бік завищення або заниження оцінюваних параметрів $\bar{x}$ і σ^2 . Як було сказано раніше, оцінки, які задовольняють таку умову, називаються незміщеними. Що ж являє собою помилка параметра m_θ ? Це середня з множини конкретних помилок:

$$m_{\tilde{x}} = \sqrt{\frac{\sum E_i^2 n_i}{n}} = \sqrt{\frac{\sum (\tilde{x} - \bar{x})^2 n_i}{\sum n_i}}; \quad m = \sqrt{\frac{\sigma^2}{n}}.$$

Точкова оцінка параметра генеральної сукупності полягає у тому, що з різних можливих вибіркових оцінок спочатку обирається та, яка має оптимальні властивості, а потім обчислюється значення цієї оцінки. Отримане розрахункове значення останньої розглядається як найкраще наближення до невідомого дійсного значення параметра генеральної сукупності. Додаткові розрахунки, пов'язані з визначенням можливої помилки оцінки, не завжди обов'язкові (залежно від вирішування задач оцінки), але, як правило, здійснюються практично завжди.

Розглянемо приклади визначення точкової оцінки для середньої досліджуваних ознак і для їх частки в генеральній сукупності.

Приклад. Посіви зернових культур району складають 20 000 га. При 10 %-му вибірковому обстеженні полів одержали такі вибіркові характеристики: середня врожайність – 30 ц з I га, дисперсія врожайності – 4, площа посівів високоврожайних культур – I200 гектарів.

Що можна знати про величину показника середньої врожайності зернових культур у районі і яке числове значення показника частки (питомої ваги) високоврожайних культур у загальній площі зернових досліджуваного регіону? Тобто необхідно дати оцінку названим параметрам ($\bar{x}$,p) у генеральній сукупності. Для розрахунку оцінок маємо:

$$N = 20000\,;\; n = 20000 \times 0{,}1 = 2000;\; \tilde{x} = 30\,;\; \sigma = \sqrt{4};\;\; w = \frac{1200}{2000} = 0{,}60.$$

Як відомо, вибіркова середня арифметична є ефективною оцінкою генеральною середньої арифметичної. Таким чином, можна прийняти, що найкраща оцінка генерального параметра (θ) є 30. Щоб визначити ступінь точності оцінки необхідно знайти середню (стандартну) її помилку:

$$m = \sqrt{\frac{\sigma^2}{n}(1 - \frac{n}{N})} = \sqrt{\frac{4}{2000}(1 - \frac{2000}{20000})} = 0{,}04.$$

Одержана величина помилки свідчить про велику точність оцінки. Значення m тут означає, що при багаторазовому повторенні таких вибірок помилка оцінки параметра становила б у середньому 0,04. Тобто за точковою оцінкою середня врожайність у господарствах району буде $\bar{x} = 30 \pm 0{,}04$ ц з I га.

Для одержання точкової оцінки показника частки посівів високоврожайних культур зернових у загальній площі зернових за кращу оцінку може бути прийнято показник частки у вибірці w=0,6. Таким чином, можна сказати, що за результатами спостережень найкращою оцінкою шуканого показника структури буде число 0,6. Уточнюючи обчислення, слід розрахувати середню помилку цієї оцінки:

$$m = \sqrt{\frac{w(1-w)}{n}(1-\frac{n}{N})} = \sqrt{\frac{0.6(1-0.6)}{2000}(1-\frac{2000}{20000})} = 0{,}01$$

Як бачимо, середня помилка оцінки генеральної характеристики дорівнюватиме 0,01.

Одержаний результат означає: якщо б багаторазово повторити вибірку з обсягом у 2000 га зернових, середня помилка прийнятої оцінки частки (питомої ваги) високоврожайних культур у площі зернових культур підприємств району була б ± 0,01. У такому разі Р = 0,6 ± 0,01. У процентному виразі частка високоврожайних культур у загальній площі зернових району складе в середньому 60 ± І.

Розрахунки показують, що для конкретного випадку найкращою оцінкою шуканого показника структури буде число 0,6, а середня помилка оцінки у той чи іншій бік буде приблизно дорівнювати 0,01. Як бачимо, оцінка досить точна.

Відомо кілька способів точкової оцінки середнього квадратичного відхилення у випадках, коли вибірка здійснена з генеральної сукупності одиниць з нормальним розподілом і параметр σ невідомий. Найпростішою (найбільш легкою в обчисленнях) оцінкою є розмах варіації ($R_в$)вибірки, помножений на поправочний коефіцієнт, взятий за стандартними таблицями і який залежить від обсягу вибірки (для малих вибірок). Параметр середнього квадратичного відхилення в генеральній сукупності можна також оцінити за допомогою обчисленої вибіркової дисперсії з врахуванням числа ступенів вільності. Корінь квадратний із цієї дисперсії дає величину, яка буде використана як оцінка генерального середньоквадратичного відхилення ($\sigma_г$).

Використовуючи значення параметра $\sigma_в$, обчислюють середню помилку оцінки генеральної середньої ($\bar{x}_г$) способом, розглянутим вище.

Як вказувалося раніше, відповідно до вимоги спроможності впевненість у точності тієї чи іншої точкової оцінки підвищується при збільшені чисельності вибірки. Продемонструвати це теоретичне положення на прикладі точкової оцінки дещо утруднено. Вплив обсягу вибірки на точність оцінки очевидний при обчисленні інтервальних оцінок. Про них мова піде нижче.

У таблиці 39 наведено найбільше часто використовувані точкові оцінки параметрів генеральної сукупності.

Таблиця 39

Основні точкові оцінки

Характеристика генеральної сукупності	Оцінка
Середня арифметична, $\overline{x}$	$\widetilde{x}$
Різниця середніх двох генеральних сукупностей, $\overline{x_1} - \overline{x_2}$	$\widetilde{x}_1 - \widetilde{x}_2$
Середнє квадратичне відхилення, $\sigma_г$	$\sigma_в$
Частка ознаки, $p_к$	w
Різниця частот двох ознак генеральних сукупностей, $P_1 - P_2$	$w_1 - w_2$
Сумарні параметри генеральної сукупності	N_x
Кількість елементів у групі генеральної сукупності	Nw

Обчислені різними способами значення оцінок можуть бути неоднакові за величиною. У цьому зв'язку в практичних розрахунках слід займатися не послідовним обчисленням можливих варіантів, а, спираючись на властивості різних оцінок, обрати одну з них.

При малій кількості одиниць спостережень точкова оцінка значною мірою випадкова, отже, мало надійна. Тому в малих вибірках вона може сильно відрізнятися від оцінюваної характеристики генеральної сукупності. Таке положення призводить до грубих помилок у висновках, які поширюються на генеральну сукупність за результатами вибірки. З цієї причини при вибірках малого обсягу користуються інтервальними оцінками.

На відміну від точкової інтервальна оцінка дає діапазон точок, всередині якого повинен знаходитись параметр генеральної сукупності. Крім того, в інтервальній оцінці вказується ймовірність, а, отже, вона має важливіше значення в статистичному аналізі.

Інтервальною називають оцінку, яка характеризується двома числами - границями інтервалу, який охоплює (покриває) оцінюваний параметр. Така оцінка являє собою деякий інтервал, у якому з заданою ймовірністю знаходиться шуканий параметр. За центр інтервалу приймається вибіркова точкова оцінка.

Таким чином, інтервальне оцінювання є подальшим розвитком точкового оцінювання, коли така оцінка при малому обсязі вибірки неефективна.

Задачу інтервального оцінювання в загальному вигляді можна сформулювати так: за даними вибіркового спостереження необхідно побудувати числовий інтервал, відносно якого з раніше обраним рівнем імовірності можна стверджувати, що в межах даного інтервалу знаходиться оцінюваний параметр.

Якщо взяти достатньо велику кількість одиниць вибірки, то, користуючись теоремою Ляпунова, можна довести ймовірність того, що помилка вибірки не перевищить деяку задану величину Δ, тобто $\left|\tilde{x} - \bar{x}\right| \leq \Delta$ або $|w - p| \leq \Delta$.

Зокрема, ця теорема дає можливість оцінювати похибки наближених рівностей :

$$\frac{n_i}{n} \approx p(n_i - частота); \tilde{x} \approx \bar{x}.$$

Якщо $x_1, x_2 x_3, ..., x_n -$ незалежні випадкові величини і $n \to \infty$, то ймовірність їх середньої ($\bar{x}$) знаходиться в межах від a до $в$ і може бути визначена рівняннями:

$$p(a \langle \tilde{x} \langle в) \int_{t_1}^{t_2} \frac{1}{\sqrt{2\Pi}} e^{-\frac{t^2}{2}} dt,$$

де $t_1 = \dfrac{\alpha - E(x)}{\sigma}; \; t_2 = \dfrac{в - E(x)}{\sigma}.$

Імовірність Р при цьому називають **довірчою імовірністю.** Таким чином, довірчою імовірністю (надійністю) оцінки генерального параметра по вибірковій оцінці називають імовірность, з якою здійснюються нерівності:

$$\left|\tilde{x} - \bar{x}\right| \leq \Delta; \; |w - p| \leq \Delta,$$

де Δ – гранична помилка оцінки, відповідно до середньої і частки.

Границі, в яких із цією заданою ймовірністю може знаходитися генеральна характеристика, називають **довірчими інтервалами** (довірчими границями). А границі цього інтервалу одержали назву границь довіри.

Довірчі (або толерантні) границі – це границі, вихід за межі яких даною характеристикою внаслідок випадкових коливань має незначну ймовірність ($p_1 \langle 0{,}5; p_2 \langle 0{,}01; \quad p_3 \langle 0{,}001$). Поняття «довірчий інтервал» введене Дж.Нейманом і К.Пірсоном (1950 р.). Це встановлений за вибірковими даними інтервал, який із заданою ймовірністю (довірчою імовірністю) охоплює (покриває) справжнє, але невідоме для нас значення параметра. Якщо за рівень довірчої ймовірності прийняти значення 0,95, то ця ймовірність свідчить про те, що при частих застосуваннях даного способу (методу) обчислень довірчий інтервал приблизно в 95% випадків буде покривати параметр. Довірчий інтервал генеральної середньої і генеральної частки визначається на основі наведених вище нерівностей, з яких випливає, що $\tilde{x} - \Delta \le \bar{x} \le \tilde{x} + \Delta;\ \ w - \Delta \le p \le w + \Delta.$

У математичній статистиці надійність того чи іншого параметра оцінюють за значенням трьох наступних рівнів імовірності (інколи називають «пороги ймовірності»): p_1 = 0,95; p_2= 0,99; p_3 = 0,999. Імовірності, якими вирішено нехтувати, тобто $\alpha_1 = 0.05;$; $\alpha_2 = 0.01;$ $\alpha_3 = 0{,}001$ **називають рівнями значимості, або рівнями істотності.** З наведених рівнів надійніші висновки забезпечує імовірність p_3 = 0,999. Кожному рівню довірчої ймовірності відповідає певне значення нормованого відхилення (див. табл. 27). Якщо немає в розпорядженні стандартних таблиць значень інтервалу імовірностей, то цю ймовірність можна обчислити з певним ступенем наближення за формулою:

$$P(t) = \frac{1}{\sqrt{2\Pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt.$$

На рисунку 11 заштриховані ті частини загальної площі, обмеженої нормальною кривою і віссю абсцис, які відповідають значенням $t = \pm 1;\ t = \pm 2;\ t = \pm 3$ і для яких імовірності дорівнюють 0,6287, 0,9545; 0,9973. При точковому оцінюванні розраховується, як уже відомо, середня помилка вибірки, при інтервальному – гранична.

Залежно від принципів відбору одиниць (повторного чи без повторного) структурні формули розрахунку помилок вибірки різняться за величиною поправки ($1 - \frac{n}{\mathrm{N}}$).

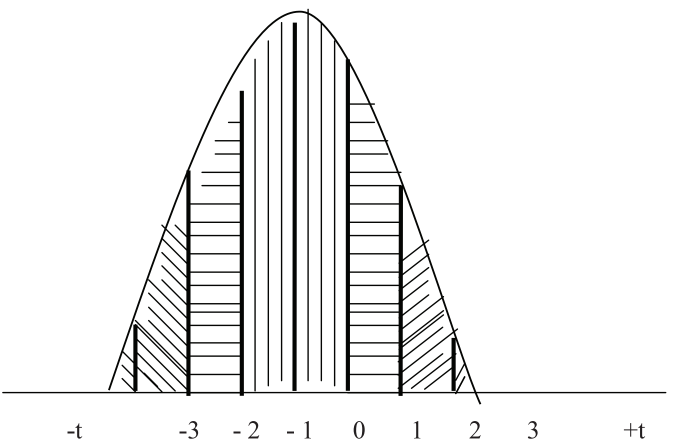


Рис. 11. Крива нормального розподілу ймовірностей

У таблиці 40 наведено формули розрахунків помилок оцінок генерального параметра.

Розглянемо конкретний випадок інтервальної оцінки параметрів генеральної сукупності за даними вибіркового спостереження.

*Приклад***.** При вибірковому обстеженні господарств району встановлено, що середньодобовий надій корів ($\widetilde{x}$) становить 10 кг. Частка чистопородної худоби у загальній чисельності поголів'я дорівнює 80 %. Помилка вибірки з довірчою ймовірністю Р = 0,954 виявилась рівною 0,2 кг; для частки чистопородної худоби 1 %.

Таким чином, межі, в яких може знаходитися генеральна середня продуктивність, будуть 9,8 < $\bar{x}$ < 10,2; для генеральної частки худоби – 79 < Р < 81.

Висновок: з імовірністю 0,954 можна стверджувати, що різниця між вибірковою середньою продуктивністю корів і генеральною продуктивністю становить 0,2 кг. Межа середньодобового надою – 9,8 і 10,2 кг. Частка (питома вага) чистопородної худоби в підприємствах району знаходиться в межах від 79 до 81 %, помилка оцінки не перевищує 1 %.

Таблиця 40

Розрахунок точкових і інтервальних помилок вибірки

Принцип відбору	Оцінюваний параметр			
	середня характеристика		частка ознаки	
	точкова оцінка	інтервальна оцінка	точкова оцінка	інтервальна оцінка
Повторний	$\sqrt{\frac{\sigma^2}{n}}$	$T\sqrt{\frac{\sigma^2}{n}}$	$\sqrt{\frac{w(1-w)}{n}}$	$t\sqrt{\frac{w(1-w)}{n}}$
Безповторний	$\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$	$T\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$	$\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$	$t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$

При організації вибірки важливе значення має визначення необхідної її чисельності (*n*). Остання залежить від варіації одиниць обстежуваної сукупності. Чим більша коливність, тим більшою повинна бути чисельність вибірки. Зворотний зв'язок існує між чисельністю вибірки та її граничною помилкою. Прагнення отримати меншу помилку вимагає збільшення чисельності вибіркової сукупності.

Необхідна чисельність вибірки визначається на основі формул граничної помилки вибірки (Δ) із заданим рівнем імовірності (Р). Шляхом математичних перетворень отримують формули розрахунку чисельності вибірки (табл. 41).

Таблиця 41

Розрахунок необхідної чисельності вибірки

Спосіб відбору	Чисельність вибірки (n) при визначенні і оцінці параметра	
	Середньої ($\bar{x}$)	частки (P)
Повторний	$\frac{t^2\sigma^2}{\Delta_{\tilde{x}}^{\;2}}$	$\frac{t^2 w(1-w)}{\Delta_w^2}$
Безповторний	$\frac{t^2\sigma^2 \mathrm{N}}{\Delta_{\tilde{x}}^2 \mathrm{N} + t^2\sigma^2}$	$\frac{t_w^2(1-w)\mathrm{N}}{\Delta_w^2 \mathrm{N} + t^2 w(1-w)}$

Слід відмітити, що все викладене відносно статистичних оцінок ґрунтується на припущені, що вибіркова сукупність, параметри якої використовуються при оцінці, одержана з використанням методу (способу) відбору, який забезпечує одержання ймовірностей вибірки.

При цьому, обираючи довірчу ймовірність оцінки, слід керуватися тим принципом, що вибір її рівня не є математичним завданням, а визначається конкретно вирішуваною проблемою. У підтвердження сказаному розглянемо приклад.

Приклад. Припустимо, на двох підприємствах імовірність випуску готової (якісної) продукції дорівнює Р= 0,999, тобто імовірність одержання браку продукції становитиме α= 0,001. Чи можна в рамках математичних міркувань, не цікавлячись характером виробленої продукції, вирішити питання про те, мала чи велика ймовірність браку α= 0,001. Припустимо, одне підприємство випускає сівалки, а друге – літаки для обробітку посівів. Якщо на 1000 сівалок трапиться одна бракована, то з цим можна миритися, бо переплавка 0,1 % сівалок дешевше, ніж перебудова технологічного процесу. Якщо ж на 1000 літаків зустрінеться один бракований, це, безумовно, приведе до серйозних наслідків при його експлуатації. Отже, у першому випадку ймовірність одержання браку α= 0,001 може прийматись, в другому випадку – ні. За цієї причини вибір довірчої ймовірності в розрахунках взагалі і при обчислюванні оцінок, зокрема, слід здійснювати виходячи з конкретних умов задачі.

Залежно від завдань дослідження може виникнути необхідність обчислення однієї або двох довірчих границь. Якщо особливості розв’язуваної задачі вимагають встановлення тільки однієї із границь, верхньої або нижньої, можна переконатись, що ймовірність, з якою встановлюється ця границя, буде вища, ніж при зазначенні обох границь для одного і того ж значення коефіцієнта довіри t.

Нехай довірчі границі встановлені з імовірністю Р= 0,95, тобто, в 95 % випадків генеральна середня ($\bar{x}$) буде не менше нижнього довірчого інтервалу $\bar{x}_{ниж.} = \tilde{x} - t\mu$ і не більше верхнього довірчого інтервалу $\bar{x}_{верх..} = \tilde{x} + t\mu$. У цьому випадку лише з імовірністю α= 0,05 (або 5 %) середня генеральна може вийти за вказані границі. Оскільки розподіл t симетричний, то половина з цього рівня ймовірності, тобто 2,5 %, припадатиме на випадок, коли $\bar{x} \langle \bar{x}_{ниж.}$ а друга половина – на випадок коли, $\bar{x} \rangle \bar{x}_{верх.}$. З цього випливає, що ймовірність того, що середня генеральна може бути менша, ніж значення верхньої довірчої границі $\bar{x}_{верх.}$, дорівнює 0,975 (тобто 0,95 +0,025). Отже, створюються умови, коли при двох довірчих границях ми нехтуємо значенням $\bar{x}$ як меншими $\bar{x}_{ниж.}$, так і більшими або $\bar{x}_{верх.}$ Називаючи тільки одну довірчу границю, наприклад, $\bar{x}_{верх.}$, ми нехтуємо лише тими $\tilde{x}$, які перевищують цю границю. Для одного і того ж значення коефіцієнта довіри t рівень значимості α тут виявляється в два рази меншим.

Якщо розраховуються тільки значення ознаки, які перевищують (або навпаки не перевищують) значення шуканого параметра $\bar{x}$, довірчий інтервал називається одностороннім. Якщо розглядувані значення обмежуються з обох сторін, довірчий інтервал носить назву двостороннього. Із сказаного вище випливає, що гіпотези і ряд критеріїв, зокрема критерій t-Ст'юдента, потрібно розглядати як односторонні і двосторонні. Тому при двосторонній гіпотезі рівень значимості для одного і того ж значення t буде в два рази більший, ніж односторонній. Якщо ми хочемо при односторонній гіпотезі залишити таким же рівень значимості (і рівень довірчої імовірності), як при двосторонній гіпотезі, то величину t слід взяти меншу. Ця особливість врахована при складанні стандартних таблиць критеріїв t-Ст'юдента (додаток 1).

Відомо, що з практичного боку частіше являють інтерес не стільки довірчі інтервали можливої величини генеральної середньої, скільки ті максимальні і мінімальні величини, більше або менше яких із заданою (довірчою) ймовірністю генеральна середня бути не може. У математичній статистиці їх називають гарантованим максимумом і гарантованим мінімумом середньої. Позначивши названі параметри відповідно через $\bar{x}_{\max}$ і $\bar{x}_{\min}$, можна записати : $\bar{x}_{\max} = \tilde{x} + t\mu$; $\bar{x}_{\min} = \tilde{x} - t\mu$.

При обчисленні гарантованих максимальних і мінімальних значень генеральної середньої, як границі одностороннього довірчого інтервалу в наведених вище формулах, величина t береться як критерій односторонній.

Приклад. По 20 ділянках вибірки встановлена середня врожайність цукрових буряків 300 ц/га. Дана вибіркова середня характеризує відповідний параметр генеральної сукупності ($\bar{x}$) з помилкою 10 ц/га. Відповідно до вибірковості оцінок генеральна середня урожайність може бути як більше, так і менше вибіркової середньої $\tilde{x}$= 300. Із імовірністю Р = 0,95 можна стверджувати, що шуканий параметр не буде більшим $\bar{x}_{\max} = 300 + 1{,}73 \times 10 = 317{,}3$ ц/га.

Величина t взята для числа ступенів вільності ν=20-1 при односторонній критичній області і рівні значимості α= 0,05 (додаток 1). Отже, із імовірністю Р= 0,95 гарантований максимально можливий рівень генеральної середньої врожайності оцінюється в 317 ц/га, тобто при найсприятливіших умовах середня урожайність цукрових буряків не перевищує вказаної величини.

У деяких галузях знань (наприклад, у природничих науках) теорія оцінки поступається перед теорією перевірки статистичних гіпотез. В економічній науці методи статистичної оцінки відіграють дуже важливу роль у справі перевірки надійності результатів досліджень , а також у різного роду практичних розрахунках. Передусім це стосується використання точкової оцінки досліджуваних статистичних сукупностей. Вибір якомога кращої оцінки – основна проблема точкової оцінки. Можливість такого вибору зумовлюється знанням основних властивостей (вимог) статистичних оцінок.

§ 6.2. Закони розподілу вибіркових характеристик

6.2.1. Загальне поняття законів розподілу

Закон розподілу характеризує випадкову величину з точки зору теорії ймовірностей. Розподіл імовірностей тісно зв'язаний з рядами розподілу частот. Якщо розглядати ряди розподілу (користуючись термінологією теорії ймовірностей) як перелік можливих результатів або груп вимірів і відповідних їм частот кожного результату, то аналогічне визначення можна дати і розподілу ймовірностей. Це перелік можливих результатів або груп вимірів, але, замість спостережуваної частоти, тут вказані ймовірності появи кожного результату.

У практичних і наукових розрахунках іноді доводиться аналізувати ознаку, яка є випадковою величиною з невідомим характером статистичного розподілу, тобто його законом. Щоб знайти цей закон розподілу, проводять статистичне спостереження за випадковою змінною у визначених умовах і одержують варіаційний ряд, який дає уявлення про її емпіричний розподіл. По цьому розподілу випадкової величини необхідно знайти невідомий її закон як загальний закон розподілу досліджуваної ознаки. Вирішення такого завдання у загальному вигляді вважається проблемним. Однак, виходячи з ряду загальних гіпотез, можна математично довести, якими повинні бути розподіли чисельностей ознаки досліджуваної сукупності. Такі розподіли називають **теоретичними.**

Слід пам'ятати, що законів, за якими розподіляється випадкова величина, існує багато. Але класичними прийнято вважати три теоретичних розподіли, які за своєю науковою важливістю займають чільне місце серед інших. Якщо розглядати в хронологічному порядку їх відкриття, то назви цих теоретичних розподілів розмістяться у такі послідовності: біноміальний (відкритий Я.Бернуллі, 1700), нормальний (Демуавр, 1773; Гаусс, 1809; Лаплас, 1812) та Пуассоновий (С. Пуассон, 1837). Серед названих важливим законом, на якому ґрунтується переважна більшість статистичних методів дослідження, є **закон нормального розподілу.**

Велика кількість теоретичних розподілів відкрита трохи пізніше. Але більшість цих відкрить значною мірою була зумовлена властивостями перших трьох розподілів (біноміальний, нормальний, Пуассонів) і особливо нормального. Названі три види розподілу являють логічно і теоретично відправний пункт теорії будь-яких спеціальних видів (типів) розподілів.

Окремі закони розподілу пов'язані з характером розподілу деяких випадкових величин, що застосовуються для вирішення конкретних задач. Закони названо іменами вчених, які визначили функції розподілу різних випадкових величин. Серед них широко використовуються закони розподілу (або просто «розподіли») Пірсона, Стьюдента, Фішера.

Під **законом розподілу** слід розуміти такий теоретичний розподіл, до якого прямує емпіричний розподіл при $n \to \infty$.Для чого потрібно знати закони розподілу? По-перше, для оцінки параметрів генеральної сукупності; по-друге, для перевірки статистичних гіпотез; по-третє, для здійснення прогнозних розрахунків.

Як говорилося вище, по емпіричному розподілу випадкової величини знаходять невідомий закон її розподілу. Але при рішенні ряду практичних задач відпадає необхідність розрахунку можливих значень випадкової величини і відповідних їм рівнів імовірностей. Зокрема, інколи зручніше використовувати деякі характеристики, що синтезують у собі інформацію про випадкову величину. Їх називають числовими (кількісними) характеристиками випадкової величини. Наведемо їх: середня (математичне очікування), дисперсія, мода, медіана, моменти різних порядків.

На основі всебічного аналізу цих параметрів, загальних теоретичних передумов і знання особливостей тих чи інших розподілів вибирають розподіл, який найкращим чином апроксимує емпирічний (фактичний) розподіл випадкової змінної. На наступному етапі дослідження знаходять параметри того закону розподілу, котрий характеризує досліджувану випадкову величину. Так, урожайність цукрових буряків («на корені») являє собою випадкову величину. Для визначення «видової» врожайності в регіоні відібрано 60 підприємств. Щоб визначити закон розподілу 60 показників урожайності підприємств, що знаходяться в однакових умовах, розраховано показники «видової» урожайності і впорядковані дані представлені у вигляді варіаційного ряду розподілу. Виходячи з теореми Ляпунова, слід припустити, що показник урожайності – випадкова величина, розподілена за нормальним законом. Потрібно обчислити параметри цього закону, які характеризують саме рівень урожайності. На подальшому етапі дослідження вирішується завдання перевірки правильності вибору виду розподілу. Інакше кажучи, встановлюється ступінь узгодженості припущеного (теоретичного) розподілу з емпіричним. Ці питання будуть розглянуті у наступних розділах.

6.2.2. Нормальний розподіл

Закон нормального розподілу, так званий **Закон Гаусса**, – один з найпоширеніших законів. Це фундаментальний закон у теорії ймовірностей і в її застосуванні. Нормальний розподіл найчастіше зустрічається у вивченні природних і соціально-економічних явищ. Інакше кажучи, більшість статистичних сукупностей у природі і суспільстві підпорядковується закону нормального розподілу. Відповідно можна сказати, що сукупності значної частини великих за обсягом вибірок підпорядковуються закону нормального розподілу. Ті із сукупностей, які відхиляються від нормального розподілу в результаті спеціальних перетворень, можуть бути наближені до нормального. У зв'язку з цим слід пам'ятати, що принципова особливість цього закону стосовно до інших законів розподілу полягає в тому, що він є законом границі, до якої наближаються інші закони розподілу в певних (типових) умовах.

Слід відмітити, що термін «нормальний розподіл» має умовний зміст, як загальноприйнятий у математичній і статистико-математичній літературі термін. Твердження, що та чи інша ознака будь-якого явища підпорядковується закону нормального розподілу, зовсім не означає непохитність норм, ніби притаманних досліджуваному явищу, а віднесення останнього до другого виду закону не означає якусь анормальність даного явища. У цьому розумінні термін «нормальний розподіл» не зовсім вдалий.

Нормальний розподіл (закон Гаусса-Лапласа) є типом безперервного розподілу. Де Муавр (1773, Франція) вивів нормальний закон розподілу ймовірностей. Основні ідеї цього відкриття були використані в теорії помилок вперше К. Гауссом (1809, Німеччина) і А.Лапласом (1812, Франція), які внесли вітчутний теоретичний вклад у розробку самого закону. Зокрема, К.Гаусс у своїх розробках виходив з визнання найбільш імовірним значенням випадкової величини-середню арифметичну. Загальні умови виникнення нормального розподілу встановив А.М.Ляпунов. Ним було доведено, що якщо досліджувана ознака являє собою результат сумарної дії багатьох факторів, кожен з яких мало пов'язаний з більшістю решти, і вплив кожного фактора на кінцевий результат набагато перекривається сумарним впливом всієї решти факторів, то розподіл стає близьким до нормального.

Нормальним називають розподіл імовірностей безперервної випадкової величини, яка має щільність:

$$f(x,\bar{x},\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-\bar{x})^2}{\sigma^2}};$$

де $\bar{x}$ – математичне очікування або середня величина.

Як видно, нормальний розподіл визначається двома параметрами: $\bar{x}$ і σ . Щоб задати нормальний розподіл, досить знати математичне очікування, або середню і середнє квадратичне відхилення. Ці дві величини визначають центр групування і форму кривої на графіку. Графік функції $f(x, \bar{x}, \sigma)$ називається нормальною кривою (крива Гаусса) з параметрами $\bar{x}$ і σ (рис. 12).

Крива нормального розподілу має точки перегину при t ± 1. Якщо уявити графічно, то між t=+*1 і* t=-1 знаходиться 0,683 частини всієї площі кривої (тобто 68,3%). У границях t=+*2* і t=- 2. знаходяться 0,954 площі (95,4 %), а між t=+3 і t= - 3 - 0,997 частини всієї площі розподілу (99,7%). На рис. 13 проілюстрований характер нормального розподілу з одно-, дво- і трисигмовою границями.

При нормальному розподілі середня арифметична, мода і медіана будуть рівними між собою. Форма нормальної кривої має вид одновершинної симетричної кривої, вітки якої асимптотично наближаються до осі абсцис. Найбільша ордината кривої відповідає $x = 0$. У цій точці на осі абсцис розміщується чисельне значення ознак, яке дорівнює середній арифметичній, моді і медіані. По обидві сторони від вершини кривої її вітки спадають, змінюючи в певних точках форму випуклості на увігнутість. Ці точки симетричні і відповідають значенням $x = \pm 1,$ тобто величинам ознаки, відхилення яких від середньої чисельно дорівнює середньому квадратичному відхиленню. Ордината, що відповідає середній арифметичній, ділить всю площу між кривою і віссю абсцис пополам. Отже, ймовірності появи значень досліджуваної ознаки більших і менших середньої арифметичної будуть рівні *0,50,* тобто $\left[P(x_i \langle \tilde{x} \langle x_i) = 0{,}50\right]$.

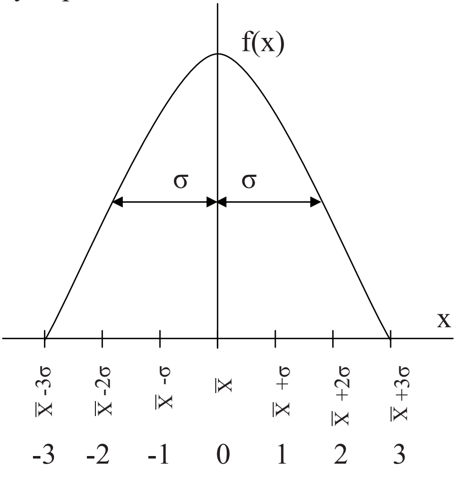


Рис.12. Крива нормального розподілу (крива Гауса)

Форму і положення нормальної кривої зумовлюють значення середньої і середнього квадратичного відхилення. Математично доведено, що зміна величини середньої (математичного очікування) не змінює форми нормальної кривої, а призводить лише до її зміщення вподовж осі абсцис. Крива зрушується вправо, якщо $\tilde{x}$ зростає, і вліво, якщо $\tilde{x}$ спадає.

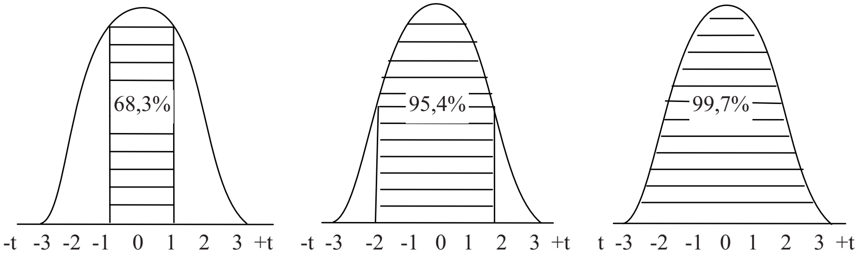


Рис.13. Нормальний розподіл з одно-, дво- та трьосигмовими границями

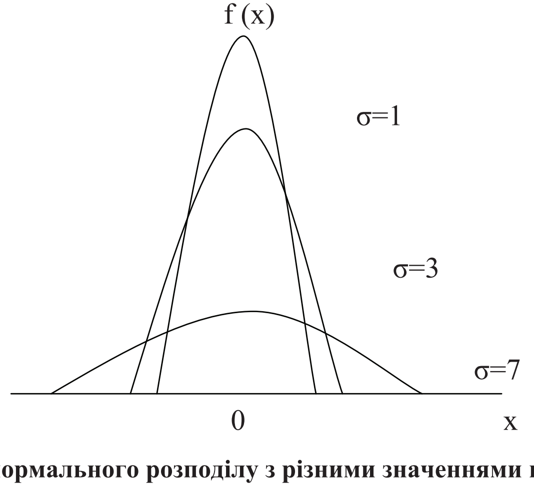


Рис.14. Криві нормального розподілу з різними значеннями параметра σ

Про зміну форми графіка нормальної кривої при зміні середнього квадратичного відхилення можна судити по максимуму диференціальної функції нормального розподілу, який дорівнює $\frac{1}{\sigma\sqrt{2\pi}}$. Як видно, при зростанні величини σ максимальна ордината кривої буде зменшуватися. Отже, крива нормального розподілу буде стискуватися до осі абсцис і приймати більш плосковершинну форму.

І, навпаки, при зменшенні параметра σ нормальна крива витягується в додатному напрямку осі ординат, а форма «дзвона» стає більш гостровершиною (рис. 14*)*. Відзначимо, що незалежно від величини параметрів $\tilde{x}$ і σ площа, обмежена віссю абсцис і кривою, завжди дорівнює одиниці (властивість щільності розподілу). Це наочно ілюструє графік (рис. 13).

Названі вище особливості прояву «нормальності» розподілу дозволяють виділити ряд загальних властивостей, які мають криві нормального розподілу:

1) будь-яка нормальна крива досягає точки максимуму ($x = \bar{x}$) *і* спадає безперервно вправо і вліво від нього, поступово наближаючись до осі абсцис;

2) будь-яка нормальна крива симетрична по відносно прямої, паралельної осі ординат і проходить через точку максимуму ($x = \bar{x}$)*;* максимальна ордината дорівнює $\frac{1}{\sigma\sqrt{2\pi}}$;

3) будь-яка нормальна крива має форму «дзвона», має випуклість, яка направлена вверх до точки максимуму. У точках $\bar{x} - \sigma$ і $\bar{x} + \sigma$ вона змінює випуклість, і , чим менше σ*,* тим гостріше «дзвін», а чим більше σ*,* тим більш похилішою стає вершина «дзвону» (рис.14). Зміна математичного очікування (при незмінній величині σ) не призводить до модифікації форми кривої.

При $\bar{x} = 0$ і $\sigma = 1$ нормальну криву називають нормованою кривою або нормальним розподілом у канонічному вигляді.

Нормована крива описується наступною формулою:

$$f(x) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{x^2}{2}}$$

Побудова нормальної кривої за емпіричними даними здійснюється за формулою:

$$n_m = \frac{ni}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}},$$

де n_m - теоретична частота кожного інтервалу (групи) розподілу;

n - сума частот, що дорівнює обсягу сукупності;

i - крок інтервалу;

π - відношення довжини кола до його діаметру, яке становить 3,1416;

е - основа натуральних логарифмів, дорівнює 2,71828;

t - нормоване відхилення, ($\frac{x-\widetilde{x}}{\sigma}$);

σ - середнє квадратичне відхилення.

Друга і третя частини формули ($\frac{1}{\sqrt{2\Pi}}e^{-\frac{t^2}{2}}$) є функцією нормованого відхилення f(t), яку можна розрахувати для будь-яких значень t . Таблиці значень f(t) звичайно називають «Таблицями ординат нормальної кривої» (додаток 3). При використанні цих функцій робоча формула нормального розподілу набуває простого вигляду:

$$n_T \frac{ni}{\sigma} f(t)$$
.

Приклад. Розглянемо випадок побудови нормальної кривої на прикладі даних про розподіл 57 працівників за рівнем денного заробітку (табл. 42).

За даними таблиці 42, знаходимо середню арифметичну:

$$\widetilde{x} = \frac{\Sigma x_i n_i}{\Sigma n_i} = \frac{1654}{57} = 29.$$

Розраховуємо середнє квадратичне відхилення:

$$\sigma = \sqrt{\frac{\Sigma(x_i - \widetilde{x})^2 n_i}{\Sigma n_i}} = \sqrt{\frac{2224}{57}} = \sqrt{39{,}02} = 6{,}25$$
.

Для кожної рядка таблиці знаходимо значення нормованого відхилення

$t = \frac{|x_i - \bar{x}\,|}{\sigma} = \frac{12}{6.25} = 1.92$ (для першого інтервалу і т.д.).

У графі 8 табл. 42 записуємо табличне значення функції f(t) з додатка , наприклад, для першого інтервалу t=1.92 знаходимо «1,9» проти «2» (0.0632).

Для обчислення теоретичних частот, тобто ординат кривої нормального розподілу, обчислюється множник:

$$\frac{ni}{\sigma} = \frac{57 \cdot 4}{6{,}25} = 36{,}5$$
.

Усі знайдені табличні значення функції *f(t)* множимо на 36,5. Так, для першого інтервалу одержуємо 0,0632x36,5 = 2,31 тощо. Прийнято нечисленні частоти *(*n_i *<5)* об'єднувати (у нашому прикладі – перших два і останніх два інтервали).

Якщо крайні теоретичні частоти значно відрізняються від нуля, розбіжність між сумами емпіричних і теоретичних частот може виявитися значною.

Графік розподілу емпіричних і теоретичних частот (нормальна крива) за даними розглянутого прикладу показано на рисунку 15.

Розглянемо приклад визначення частот нормального розподілу для випадку, коли в крайніх інтервалах відсутня частота (табл. 43). Тут емпірична частота першого інтервалу дорівнює нулю. Отримана сума неуточнених частот не дорівнює сумі їх емпіричних значень (56 ≠ 57). У цьому випадку розраховується теоретична частота для умовно отриманих значень центра інтервалу, нормованого відхилення і його функції.

У таблиці 43 ці величини обведено прямокутником. При побудові графіка нормальної кривої у таких випадках теоретичну криву продовжують. У розглянутому випадку нормальна крива буде продовжена в бік від'ємних відхилень від середньої, оскільки перша не уточнена частота дорівнює 5. Розрахована теоретична частота (уточнена) для першого інтервалу буде дорівнювати одиниці. По сумі уточнені частоти збігаються з емпіричними (57=57).

Таблиця 42

Розрахунок частот нормального розподілу
(вирівнювання емпіричних частот за нормальним законом)

Інтервал, *(i=4)*	Серединне значення (центр) інтервалу, x_i	Кількість одиниць, n_i	Розрахункові величини				Статистичні параметри			
			$x_i n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$	нормоване відділення, $t = \frac{\lvert x_s - \bar{x} \rvert}{\sigma}$	табличне значення функції, f(t)	теоретична частота нормального ряду розподілу, $f(t) \times \frac{ni}{\sigma}$	уточнене значення теоретичної частоти, n_T
A	1	2	*3*	4	5	6	7	8	9	10
15-19	17	4	68	-12	144	576	1,92	0,0632	$\left.\begin{array}{l}2{,}31\\6{,}42\end{array}\right\}$	9
19-23	21	6	126	*-8*	64	384	1,28	0.1758		
23-27	25	9	225	-4	16	144	0,64	0,3251	11,87	12
27-31	29	17	493	0	0	0	0	0,3989	14,56	15
31-35	33	13	429	4	16	208	0,64	0,3251	11,87	12
35-39 39-43	37 41	3 5	111 205	8 12	64 144	192 720	1,28 1,92	0,1758 0,0632	$\left.\begin{array}{l}6{,}42\\2{,}31\end{array}\right\}$	9
Всього	x	57	1654	0	x	2224	x	x	55,76	57

i=4	$\bar{x} = 29$	$\sigma = 6{,}25$	$\frac{ni}{\sigma} = 36{,}5$

Таблиця 43

Розрахунок частот нормального розподілу (вирівнювання емпіричних частот по нормальному закону)

Інтервал (і-2)	Серединне значення (центр) інтервалу, x_i	Кількість одиниць, n_i	Розрахункові величини				Статистичні параметри			
			$x_i n_s$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$	нормоване відхилення $t = \frac{\lvert x_s - \bar{x} \rvert}{\sigma}$	табличне значення функції, f(t)	теоретична частота нормального ряду розподілу $f(t) \times \frac{n_i}{\sigma}$	уточнене значення теоретичної частоти, n_m
А	1	2	3	4	5	6	7	8	9	10
19-21	$\boxed{20}$	-	-	-	-		$\boxed{2,49}$	$\boxed{'0,0180}$	-	$\boxed{1}$
21-23	22	5	110	-4	16	80	1,66	0,1006	5	5
23-25	24	15	360	-2	4	60	0,83	0,2827	13	13
25-27	26	20	520	0	0	0	0	0,3989	19	19
27-29	28	10	280	2	4	40	0,83	0,2827	13	13
29-31	30	5	150	4	16	80	1,66	0,1006	5	5
31-33	32	2	64	6	36	72	2,49	0,0180	I	I
Всього	x	57	1484	x	x	332	x	x	56	57

i=2	$\bar{x} = 26$	$\sigma = 2,41$	$\frac{ni}{\sigma} = 47,3$

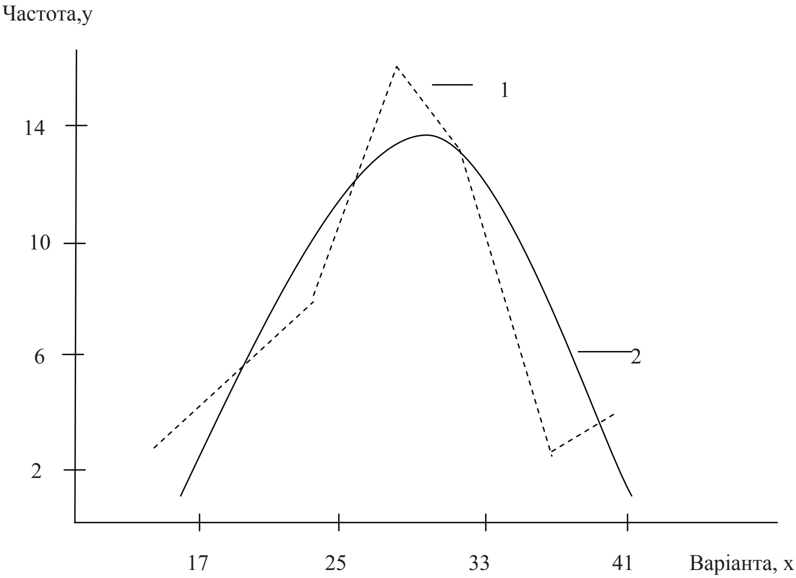


Рис. 15. Емпіричний розподіл (1) і нормальна крива (2)

Криву нормального розподілу по досліджуваній сукупності можна побудувати і іншим способом (на відміну, від розглянутого вище). Так, якщо необхідно мати наближену уяву про відповідності фактичного розподілу нормальному, обчислення здійснюють у такий послідовності. Визначають максимальну ординату, яка відповідає середньому розміру ознаки ($y_{\max}$), потім, обчисливши середнє квадратичне відхилення, розраховують координати точок кривої нормального розподілу за схемою, викладеною в таблицях 42 і 43. Так, за вихідними і розрахунковими даними таблиці 43 маємо середню $\tilde{x} = 26$ Ця величина середньої збігається з центром четвертого інтервалу (25-27). Отже, частота цього інтервалу «20» може бути прийнята (при побудові графіка) за максимальну ординату($y_{\max}$). Маючи обчислену дисперсію ($\sigma = 2{,}41$, див. табл. 43), розраховуємо значення координат всіх необхідних точок кривої нормального розподілу (табл. 44, 45). За отриманими координатами креслимо нормальну криву (рис. 16), прийнявши за максимальну ординату частоту четвертого інтервалу.

Узгодженість емпіричного розподілу з нормальним може бути встановлена також шляхом спрощених розрахунків. Так, якщо відношення показника міри асиметрії (A_S) до своєї середнєквадратичної помилки m_{A_s} або відношення показника ексцесу (Ex) до своєї середнєквадратичної помилки m_{Ex} перевищує за абсолютною величиною число «3», робиться висновок про невідповідність емпіричного розподілу характеру нормального розподілу (тобто, якщо $\frac{A_S}{m_{A_s}} > 3$ або $\frac{Ex}{m_{Ex}} > 3$).

Є й інші, нетрудомісткі прийоми встановлення «нормальності» розподілу: а) порівняння середньої арифметичної з модою і медіаною; б) використання чисел Вестергарда; в) застосування графічного способу за допомогою напівлогарифмічної сітки *Турбіна;* г) обчислення спеціальних критеріїв узгодження та ін.

Таблиця 44

Координати 7 точок кривої нормального розподілу

Точка	1	2і 3	4 і 5	6 і 7
Абсцис, x	$\bar{x}$	$\bar{x} \pm 0{,}5\sigma$	$\bar{x} \pm \sigma$	$\bar{x} \pm 1{,}5\sigma$
Ордината, у	y_{max}	$\frac{7}{8} y_{\max}$	$\frac{5}{8} y_{\max}$	$\frac{2.5}{8} y_{\max}$

Таблиця 45

Обчислення координат точок кривої нормального розподілу

x	$\bar{x} - 1{,}5\sigma =$ $= 22{,}4$	$\bar{x} - \sigma = 23{,}6$	$\bar{x} - 0{,}5\sigma =$ $= 24{,}8$	$\bar{x} = 26$	$\bar{x} + 0{,}5\sigma = 27{,}2$	$\bar{x} + \sigma = 28{,}4$	$\bar{x} + 1{,}5\sigma =$ $= 29{,}6$
y	6	12	17	20	17	12	6

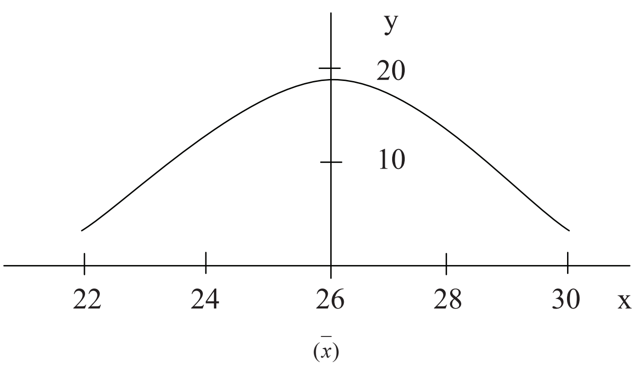


Рис .16. Крива нормального розподілу, побудована по семи точках

На практиці при дослідженні сукупності на предмет узгодження її розподілу з нормальним часто користуються «правилом 3σ».

Математично доведено ймовірність того, що відхилення від середньої за абсолютною величиною буде менше потрійного середнього квадратичного відхилення, дорівнюватиме 0,9973, тобто, ймовірність того, що абсолютна величина відхилення перевищує потрійне середнє квадратичне відхилення, дорівнює 0,0027 або дуже мала. Виходячи з принципу неможливості малоймовірних подій, можна вважати практично неможливим «випадок перевищення» 3σ. Якщо випадкова величина розподілена нормально, то абсолютна величина її відхилення від математичного очікування (від середньої) не перевищує потрійного середнього квадратичного відхилення.

У практичних розрахунках діють таким чином. Якщо при невідомому характері розподілу досліджуваної випадкової величини розраховане значення відхилення від середньої виявиться менше значення 3σ, то є підстави вважати, що досліджувана ознака розподілена нормально. Якщо ж вказаний параметр перевищить числове значення 3σ, можна вважати, що розподіл досліджуваної величини не узгоджується з нормальним розподілом.

Обчислення теоретичних частот для досліджуваного емпіричного ряду розподілу прийнято називати вирівнюванням емпіричних кривих по нормальному (або будь-якому іншому) закону розподілу. Цей процес має важливе як теоретичне, так практичне значення. Вирівнювання емпіричних даних розкриває закономірність в їх розподілі, яка може бути завуальована випадковою формою свого прояву. Встановлену таким чином закономірність можна використовувати для вирішення ряду практичних завдань.

З розподілом, близьким до нормального, дослідник зустрічається в різних сферах науки і областях практичної діяльності людини. В економіці такого роду розподіли зустрічаються рідше, ніж, скажімо, у техніці або біології. Зумовлено це самою природою соціально-економічних явищ, які характеризуються великою складністю взаємозалежних і взаємопов’язаних факторів, а також наявністю ряду умов, які обмежують вільну «гру» випадків. Але економіст повинен звертатися до нормального розподілу, аналізуючи будову емпіричних розподілів, як до деякого еталону. Таке порівняння дозволяє з'ясувати характер тих внутрішніх умов, які визначають дану фігуру розподілу.

Проникнення сфери статистичних досліджень в область соціально-економічних явищ дало змогу розкрити існування великої кількості різного типу кривих розподілу. Однак не треба вважати, що теоретична концепція кривої нормального розподілу взагалі мало придатна у статистико-математичному аналізі такого типу явищ. Вона може бути не завжди прийнятна в аналізі конкретного статистичного розподілу, але в області теорії і практики вибіркового методу дослідження має першочергове значення.

Назвемо основні аспекти застосування нормального розподілу у статистико-математичному аналізі.

1. Для визначення ймовірності конкретного значення ознаки. Це необхідно при перевірці гіпотез про відповідність того чи іншого емпіричного розподілу нормальному.

2. При оцінці ряду параметрів, приміром, середніх, методом максимальної правдоподібності. Суть його полягає у визначенні такого закону, якому підпорядковується сукупність. Визначається та оцінка, яка дає максимальні значення. Краще наближення до параметрів генеральної сукупності дає відношення:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \bar{x})^2}{2\sigma^2}}.$$

3. Для визначення ймовірності вибіркових середніх відносно генеральних середніх.

4. При визначенні довірчого інтервалу, в якому знаходиться наближене значення характеристик генеральної сукупності.

6.2.3. Розподіл Стьюдента

При розгляді питання середньої арифметичної у вибірках, які взяті з генеральної сукупності і підпорядковуються закону нормального розподілу, стає очевидним те, що цей розподіл залежить від середнього квадратичного відхилення (σ_2).

У практичних розрахунках значення генерального σ_2, як правило, невідоме, що призводить до певних розрахункових ускладнень. Ця обставина спонукала англійського статистика В.С.Госсета (він друкувався під псевдонімом Стьюдент) зайнятися пошуком такого розподілу середньої арифметичної, який не залежав би від параметра σ.

Поставлена задача Стьюдентом була вирішена у 1908 р. (у цей час він був службовцем на пивоварному заводі у м. Дубліні).

Відкритий закон розподілу підняв на нову сходинку теорію статистичного оцінювання і теорію перевірки статистичних гіпотез. У чому ж виражається розподіл, досліджуваний Стьюдентом? Він встановив, що імовірність нормованого відхилення $\frac{\bar{x}-\bar{x}}{\sigma}=t$ (пізніше Р.Фішер створив більш строгий теоретичний фундамент: $t=\frac{\tilde{x}-\bar{x}}{\sigma:\sqrt{n-1}}=\frac{\tilde{x}-\bar{x}}{\sigma}\sqrt{n-1}$) виражається рівнянням:

$$P(t)=C(1+\frac{t^2}{n-1})^{-\frac{1}{2}n},$$

де P(t) – імовірність того, що стандартизована різниця між $\tilde{x}$ і $\bar{x}$ має величину *t ;* C – деякий коефіцієнт, який залежить від обсягу вибірки.

Величина його становить:

$$C=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)\times\Gamma(\frac{n-1}{2})}},$$

де Г $\left(\frac{n}{2}\right)$ і Г$\left(\frac{n-1}{2}\right)$ - гама – функції.

Повна формула закону розподілу нормованого відхилення має вигляд:

$$P(t)=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\Pi(n-1)\Gamma(\frac{n-1}{2})}}\times(1+\frac{t^2}{n-1})^{-\frac{1}{2}n}.$$

Закону розподілу *t*-Стьюдента підпорядковуються малі вибірки, які одержані з нормального розподілу сукупностей. Характерною особливістю даного розподілу є те, що ймовірність значення t залежить від двох величин: обсягу вибірки *(n)* і нормованого відхилення *(t)* . Причому n береться числом ступенів вільності (υ=*n*-1).

При збільшенні чисельності вибіркової сукупності розподіл Стьюдента наближається до нормального:

$$P(t)=\frac{1}{\sqrt{2\Pi}}e^{-\frac{t^2}{2}}.$$

У спеціальній літературі є доведення, що при необмеженому зростанні обсягу вибірки розподіл Стьюдента прагне до нормального закону розподілу.

Якщо вибірка достотно мала (n < 15), розподіл імовірностей буде відрізнятися від нормального і тим більше, чим менший обсяг вибірки. Крива розподілу у таких випадках ніби розтягується (рис.15). Із збільшенням обсягу вибірки розподіл Стьюдента досить швидко наближається до нормального, зокрема, при n = 20 він практично не відрізняється від нього.

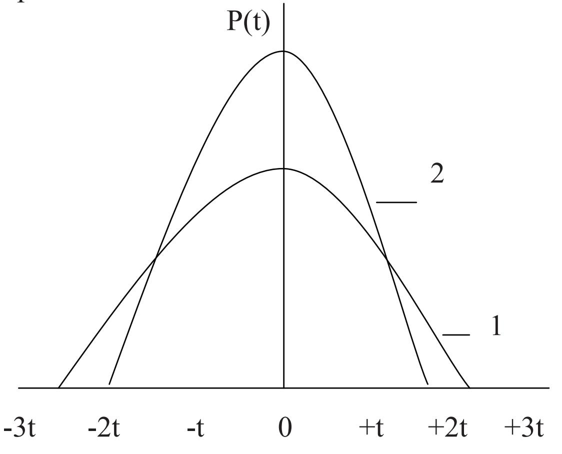


Рис. 17. Розподіл Стьюдента (I) на фоні нормальної кривої (2)

Із сказаного виходить, що розподіл Стьюдента являє собою частковий випадок нормального розподілу і відображає специфіку варіації для нечисленної вибірки, яка розподіляється за нормальним законом розподілу залежно від n .

Показники рівнів імовірності (P(t)), що розподіляється за законом Стьюдента, дано в стандартній математичній таблиці «Імовірності *t*-розподілу по Стьюденту для малих вибірок (у межах ± *t*)» (додаток2). У цій таблиці наведені рівні ймовірностей *P* для кожного значення нормованого відхилення *t* при визначеному обсязі вибірки, який береться «числом ступенів вільності». Тому положення про те, що кожному обсягу вибірки відповідає певне значення t, необхідно уточнити. Тут очевидна доцільність формулювання: кожному числу ступенів вільності відповідає t-розподіл. Розглянемо приклад.

Приклад. За результатами вибіркового обстеження 10 сімей підприємств харчової промисловості отримані середні рівні місячної заробітної плати, яка припадає на одного члена сім’ї, грн.: 100, 106, 85, 94, 88, 102, 120, 60, 95, 90.

Спираючись на дані вибірки, необхідно перевірити припущення, що середній розмір зарплати, який припадає на одного члена сім’ї працівників харчової промисловості (генеральна сукупність), дорівнюватиме 85 грн. ($\bar{x}$).

Ґрунтуючись на припущенні про нормальний характер розподілу досліджуваної ознаки в генеральній сукупності, виконуємо розрахунки: обчислюємо параметри $\bar{x} = 94;\ \sigma_в = 14{,}9$.

Прийнявши $\bar{x}$ = 85, визначаємо числове значення нормованого відхилення $t = \frac{\tilde{x} - \bar{x}}{\sigma}\sqrt{10} = \frac{94-85}{14{,}9}\sqrt{10} = 1{,}91$.Знаходимо за стандартною таблицею (додаток 2) імовірність для t = 1,9 і числа ступенів вільності υ= 10 - I. Для таких параметрів розрахункове значення ймовірності P = 0,955. Таким чином, для числового значення величини нормованого відхилення t = 1,91 імовірність P = 0,955. Тобто, імовірність появи значення t , більшим, ніж одержане при вибірці, буде α= I - 0,955 = 0,045, або приблизно один випадок з 20. Імовірність появи *t*, яке по абсолютною величиною буде більше спостережуваного значення, становитиме 2α =0,090, тобто приблизно один випадок з 10. Таке значення t слід визнати неістотним. Тому різниця між вибірковою і генеральною середньою не буде перевищувати 9 грн. (94-85).

Якщо визнати вибіркове значення *t* істотним, а таке припущення можна висунути, оскільки спостережуване значення t мало імовірне, то початкове припущення, зумовлене обчисленням значення t, буде невірним. Подібне міркування приводить до висновку про те, що середній розмір зарплати у розрахунку на одного члена сім’ї працівників досліджуваної галузі 85 грн. є сумнівним, а різниця між генеральною ($\bar{x}$ =85) и вибірковою ($\bar{x}$ = 94) середньою легко могла перевищити 9 гривень.

Крім розглянутої вище стандартної таблиці, широке практичне застосування знаходить інша математична таблиця значень критерію t для різних рівнів значимості $\alpha = 1\text{-}P$. Вона дозволяє (при певному рівні α) встановити можливі границі випадкових коливань вибіркової середньої ($\bar{x}$), а також знайти довірчий інтервал, який покриває середню арифметичну у генеральній сукупності (додаток 1).

Розглянемо випадок використання стандартної таблиці «Критичні точки розподілу Стьюдента (*t*-розподіл)» на наступному прикладі.

Внаслідок вибіркового обстеження 20 сільськогосподарських підприємств обласного регіону визначено середній показник вихододнів з розрахунку на одного працюючого, який виявився рівним 250, з середньої помилкою вибірки $m = \pm 5$. Потрібно визначити довірчий інтервал випадкових коливань шуканої середньої величини при рівні значимості α = 0,05.

У додатку I на перетині графи, яка відповідає α = 0,05, і рядку 20 - I =19 (число ступенів вільності) знаходимо значення t = 2,09. Отже, довірчий інтервал дорівнює $\Delta\tilde{x} = \tilde{x} \pm tm = 250 \pm 2{,}09 \times 5 = 250 \pm 10{,}45,$ або закруглено 250 ± 10.

Таким чином, межі генеральної середньої дорівнюватимуть 240-260, тобто в сільгосппідприємствах обстежуваного обласного регіону середній рівень показника кількості вихододнів з розрахунку на одного працюючого буде знаходитися між 240 і 260. Дане ствердження одержане при порозі ймовірності P =0,95. Рівень імовірності вибирається залежно від конкретних вимог вирішуваного завдання. В економічних дослідженнях використовують імовірність P = 0,95 (0,954).

Висловлене раніше положення про те, що при збільшенні обсягу вибірки розподіл Стьюдента наближається до нормального підтверджує порівняння даних двох стандартних таблиць: «Критичні точки розподілу Стьюдента (t-розподіл)» (додаток 1) і «Функція нормованого відхилення» (додаток 5). Достатньо розглянути фрагмент з обох таблиць для кількох вибірок, щоб переконатися у сказаному вище висновку (табл. 46).

Таблиця 46

Витяг з стандартної таблиці t-розподілів (додаток 1)

n	5 (υ=4)	20 (υ=19)	40 (υ=39)	60 (υ=59)	120 (υ=119)	∞
$t_{0,95}$	2,78	2,09	2,02	2,00	1,98	1,96

Так, у додатку 5 значенню рівня ймовірності 0,95 відповідає величина 1,96. Аналізуючи дані таблиці 46, бачимо, що при n= 60 розподіл Стьюдента майже не відрізняється від нормального:

2.00 - 1,96 = 0,04.

При невеликому обсязі вибірки ці два види розподілу мають значні чисельні відміни. Наприклад, для n =5 різниця у параметрах становитиме 2,78 - 1,96 = 0,82.

На завершення ще раз зробимо акцент на тому, що параметр $t = \frac{\tilde{x} - \bar{x}}{\sigma}\sqrt{n}$ має розподіл Стьюдента за умови, якщо досліджувана випадкова величина підпорядкована закону нормального розподілу, а середня обчислюється за вибірковими даними незалежних спостережень.

Назвемо також аспекти застосування розподілу Стьюдента.

1) При оцінці параметрів генеральної сукупності за даними малих вибірок (для визначення довірчих інтервалів);

2) При перевірці статистичних гіпотез відносно параметрів генеральної сукупності.

6.2.4. Розподіл X_i- квадрат

При перевірці статистичних гіпотез розглядаються питання про критерії узгодженості. Останні дозволяють вирішити задачу про відповідність або невідповідність певного закону розподілу, обраного для відображення досліджуваного емпіричного ряду розподілу.

Розраховані критерії згоди зумовлюють можливість (або неможливість) прийняття для досліджуваного ряду розподілу моделі, яка виражається деяким теоретичним законам розподілу. Та чи інша модель розподілу що відповідає визначеному закону може бути прийнята шляхом порівняння графічних зображень. Вченими-математиками розроблено ряд критеріїв узгодженості, обчислення яких дозволяє дати кількісну оцінку наближеності емпіричних і теоретичних розподілів. Окремі з них оцінюють імовірність розходження фактичного і теоретичного розподілу, а деякі дають пряму відповідь про можливість відображення досліджуваного емпіричного розподілу обраним теоретичним законом.

Для характеристики (оцінки) розходження емпіричних і теоретичних частот англійським статистиком Карлом Пірсоном (1900) розроблено критерій узгодженості, так званий «*хі* – квадрат». Даний критерій застосовується в тих випадках, коли необхідно визначити ступінь відмінності фактичного розподілу частот від теоретичного.

Теоретичний аспект визначення *хі* - квадрата як критерію може бути зведений до таких міркувань.

Якщо у вибірку з генеральної сукупності, розподіленої за нормальним законом ($x_1, x_2, x_3, ..., x_n$), ввести центровані і нормовані величини ($t_i = \frac{x_i - \bar{x}}{\sigma}$) і підсумовувати їх квадрати, одержимо значення величини «*хі* – квадрат» (χ^2) ;

$$\chi^2 = (\frac{x_1 - \bar{x}}{\sigma})^2 + (\frac{x_2 - \bar{x}}{\sigma})^2 + \frac{x_3 - \bar{x}}{\sigma})^2 + ... + (\frac{x_n - \bar{x}}{\sigma})^2 = \sum_{i=1}^{n}(\frac{x_i - \bar{x}}{\sigma})^2$$

У даному випадку величина χ^2, яка зумовлюється дисперсією σ^2

розподіляється за законом:

$$f(\chi^2) = \frac{1}{2\frac{\nu}{2}\Gamma(\frac{\nu}{2})}(\chi^2)^{\frac{\nu-2}{2}} e^{-\frac{\chi^2}{2}},$$

де υ - число ступенів вільності, яке дорівнює n-1 ; $\Gamma(\frac{\nu}{2})$ - гама-функція, зокрема $\Gamma(n+1)=n$.

Як видно з наведеного вище виразу, розподіл «*хі* – квадрат» визначається одним параметром – числом ступенів вільності.

Для різних обсягів вибірки (точніше - значень числа ступенів вільності) розподіл величини «*хі* – квадрат» буде асиметричним. При цьому, чим менша вибірка, тим сильніше проявляється асиметрія. Із збільшенням чисельності вибіркової сукупності асиметрія зменшується і розподіл «*хі* – квадрат» переходить у нормальний. Наочно характер такої зміни ілюструє графік (рис. 18).

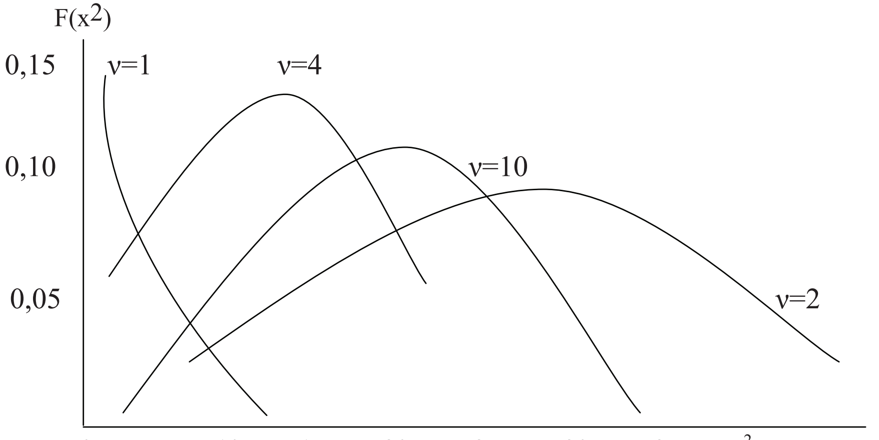


Рис. 18. Розподіл «*хі* – квадрат» при різних значеннях числа ступенів вільності

Якщо прийняти рід емпіричних і теоретичних частот відповідно за n_i і n_n, обчислення «*хі* – квадрат» – критерію виразиться формулою:

$$\chi^2 = \Sigma\frac{(n_i - n_m)^2}{n_m}.$$

Судячи по параметрах формули, величина критерію Пірсона являє собою суму відношень між квадратами різниць емпіричних і теоретичних частот до теоретичних частот.

Інтегрування диференціальної функції розподілу (за її складності) являє певні обчислювальні утруднення. У зв'язку з цим Р.Фішером розроблено стандартні математичні таблиці розподілу «*хі* – квадрат» (додатки 6, 7). Ці таблиці дають змогу обчислити ймовірність того, що випадкова величина, яка підпорядковується закону розподілу «*хі* – квадрат» з певним числом ступенів вільності, перевищить деяке фіксоване значення χ^2_v , або $P(\chi^2 \rangle \chi^2_v)$.

Другий аспект використання названих стандартних таблиць полягає у тому, що за їх допомогою можна встановити критичне значення «*хі* – квадрата», перевищення якого для відомого числа ступенів вільності буде свідчити про невідповідність досліджуваного розподілу нормальному закону. Є й інші аспекти практичного використання «*хі* – квадрат» - критерію. Розглянемо лише приклад для випадку встановлення ймовірності $P(\chi^2 \rangle \chi^2_v)$.

Для вибірки з числом ступенів вільності υ =21, що підпорядковується закону «*хі* – квадрат» розподілу (χ^2) необхідно визначити відхилення χ^2_v*,* імовірність перевищення якого дорівнює 0,05, тобто необхідно знайти «*хі* – квадрат» при υ = 21, для якого: $P(\chi^2 \rangle \chi^2_{21}) = 0{,}05.$ *Шукана величина буде знаходитись (додаток 7) на перетині рядка 21 і графи 0,95 і становитиме* χ^2_v = 32,7. *Звідси маємо:* $P(\chi^2 \rangle 32{,}7) = 0{,}05.$

Таким чином, величина χ^2_v, імовірність перевищення якої 0,05, буде 32,7

Слід відзначити деякі неточності, що існують у навчальній літературі при викладі питань «*хі* – квадрат – критерію». Відносно його відкриття, крім дати 1900 р. (Пірсон), слід пам'ятати і дату 1876 р. (Хельмерт).

Що ж стосується стандартної таблиці χ^2. - розподілу, то буде неточним її інформацію називати «Критерій Персона», бо розробка цієї таблиці належить Р.Фішеру. Останній вважав замість значень ймовірностей $P\chi^2$ *,* що відповідають деякому ряду χ^2, розраховувати значення χ^2, які відносяться до обраних рівнів імовірностей при різному числі ступенів вільності.

Ще одне зауваження щодо символіки написання «*хі* – квадрат». Звичайно прийнята форма χ^2 . Більш правильним буде $\chi^2_{\nu\alpha}$ Але, якщо виключена можливість невірного розуміння, її можна записувати з одним підрядковим числом (індексом). Якщо це не заважає правильному сприйманню змістовного навантаження параметра, запис його може бути і без підрядкових індексів.

6.2.5. Розподіл Фішера – Снедекора

У цілому ряді задач, вирішуваних математичною статистикою, зокрема у дисперсійному і кореляційно - регресійному аналізі , використовується «розподіл F», названий так по перший літері прізвища англійського статистика-математика Р.Фішера. Якщо $И_1$ *і* $И_2$ незалежні випадкові величини з розподілами χ^2, і з ν_1- і ν_2 ступенями вільності відповідно, то випадкова змінна. F буде дорівнювати: $F = \dfrac{И_1^2 : \nu_1}{И_2 : \nu_2} = \dfrac{И_1^2}{И_2^2} \times \dfrac{\nu_2}{\nu_1}.$

Одержана величина називається випадковою змінною з розподілом Фішера-Снедекора з ν_1 та ν_2 ступенями вільності. Приймаючи, що $И_1^2 \rangle И_2^2$ величина F буде мати лише значення, не менше як I.

Щільність імовірності випадкової змінної F , яка має розподіл Фішера -Снедекора з ν_1 і ν_2 ступенями вільності, має вигляд:

$$h(P) = \frac{F\dfrac{\nu_1+\nu_2}{2}(\dfrac{\nu_2}{\nu_1})}{\Gamma(\dfrac{\nu_1}{2})\Gamma(\dfrac{\nu_2}{2})} P^{\frac{\nu_2-2}{2}}(1+\frac{\nu_2}{\nu_1}P)^{-\frac{\nu_1+\nu_2}{2}}$$

.

Внаслідок великої складності розрахунку інтегралів доведення тут не наводиться. Але, як видно, розподіл F зумовлений і визначається двома параметрами, тобто числами ступенів вільності ν_1 і ν_2. Розподіл випадкової змінної F подано у вигляді спеціальних математичних таблиць. Останні побудовані так, щоб для різних рівнів довірчої ймовірності (в основному для Р = 0,95; Р = 0,99, Р = 0,999) і для різних сполучень числа ступенів вільності ν_1 і ν_2 даються значення F. Якщо прийняти позначення розрахункової і табличної величини F відповідно як F_p і F_T , то для них справедлива буде рівність $P\{F_p \rangle F_T\} = \alpha;$ Такі таблиці наведено в додатках 8, і 9. Практичне їх використання буде розглянуто у розділах «Дисперсійний аналіз»; «Кореляційно – регресійний аналіз»; «Методи багатомірного статистичного аналізу». Тут наведемо лише схематичний приклад.

Приклад. Вивчивши кількісний вплив фактора рівня продуктивність праці на її оплату по вибірці 60 підприємств, одержані наступні характеристики: факторна дисперсія $\sigma_x^2 =$ *3,06,* залишкова дисперсія σ_Z^2 = 0,15. Число ступенів вільності для факторної ознаки ν_x = 3 - I =2; для неврахованих факторів -ν_z 60 - 3 = 57.

Розрахункова величина F - критерію становитиме: $F_p = \frac{3.06}{0.15} 20.4$.

За стандартною таблицею F - розподілу знаходимо для рівня ймовірності P =0,95 і ступенів вільності $\nu_1 = 2.i$ $\nu_2 = 57$ табличне значення

F_T = 3,15 ; $F_p > F_r$ (20.4 >3,15).

Знайдені параметри свідчать про те, що в досліджуваних підприємствах вплив рівня продуктивність праці на її оплату виявився вірогідним із рівнем імовірності 0,95.

На закінчення відзначимо, чому розподіл F називають розподілом Фішера- Снедекора. Справа в тому, що Р.Фішер перший дослідив розподіл відношень двох вибіркових дисперсій, але предметом його вивчення був розподіл не відношень дисперсій, а логарифмічної величини $\frac{1}{2}\ln(\frac{\sigma_1^2}{\sigma_2^2}).$ Дещо пізніше американський статистик Дж.Снедекор розрахував таблиці розподілу змінної ($\frac{\sigma_1^2}{\sigma_2^2}$) , що виявилося значно зручніше для практичного використання в розрахунках. Цей розподіл він назвав на честь Фішера «Розподілом F». Пізніше даний вид розподілу почали називати «Розподілом Фішера-Снедекора».

МОДУЛЬ 3

ТЕМА 7. СТАТИСТИЧНІ МЕТОДИ ВИМІРЮВАННЯ ВЗАЄМОЗВ'ЯЗКІВ

§ 7.1. Дисперсійний аналіз

7.1.1. Загальнотеоретичні основи дисперсійного методу аналізу

В епоху бурхливого розвитку економіки використання методів математичної статистики в економічних дослідженнях стає нагальною необхідністю. Треба визнати, що останнім часом широкого застосування у багатофакторному аналізі набув кореляційно – регресійний метод, водночас майже зовсім не використовується досить ефективний спосіб статистико – математичної обробки даних дослідження – дисперсійний аналіз. Як і інші ймовірносно – статистичні методи, він набагато розширює можливості економістів аграріїв в аналізі виробництва й значно підвищує рівень наукових досліджень.

Головне призначення дисперсійного аналізу – статистично виявити вплив різних факторів на мінливість ознаки, що вивчається. Особливий інтерес становить використання методу в аналізі економічних процесів та явищ, коли мінливість результативної ознаки зумовлена одночасно дією кількох факторів із неоднаковою силою впливу. Зокрема, це спостерігається при аналізі результативних синтетичних показників економічної ефективності виробництва. Найбільш ефективний тут одночасний дисперсійний аналіз усіх відібраних факторів – багатофакторний аналіз. Можна, звичайно, зробити й попарне порівняння факторів, при якому всі інші ігноруються, однак такий підхід до розв'язання питання не дає змоги виявити існуючу в дійсності множинність ефектів взаємодії.

Прийняття на озброєння економістів дисперсійного методу дозволяє розв'язувати досить важливі завдання виходячи із сучасних вимог до рівня економічного аналізу. У сфері аграрно економічних досліджень цей ефективний статистико – математичний засіб повинен зайняти одне з провідних місць насамперед тому, що використання дисперсійного методу може мати самостійне значення. Зокрема, за його допомогою розв'язуються такі завдання: 1) кількісне вимірювання сили впливу факторних ознак та їх сполучень на результативну; 2) визначення вірогідності впливу та його довірчих меж; 2) аналіз окремих середніх та статистична оцінка їх різниці.

У поглибленому економічному аналізі дисперсійний метод може виконувати допоміжні функції. У цьому плані його використання відкриває широкі можливості щодо науково обґрунтованого підходу до застосування інших статистичних методів кількісного аналізу.

Як і інші статистико – математичні методи, дисперсійний аналіз являє собою чисто технічний засіб наукового пізнання. І використання його при вивченні аграрно – економічних процесів передбачає знання перш за все суті процесів, розуміння причинно – наслідкових зв’язків між явищами, що вивчаються, та вміння виділити найбільш важливі сторони взаємопов’язаних та взаємозумовлених економічних явищ.

Дисперсійний аналіз – це математико – статистичний метод вивчення результатів спостереження, що залежать від різноманітних одночасно діючих факторів. Він створений в двадцятих роках нашого століття зусиллями Р.Фішера. у подальшому суттєвого розвитку метод набув у працях Іейтса.

У нашій країні перший опис дисперсійного аналізу здійснено у 1933 р. М.Ф.Деревицьким у додаткових розділах до підручника В.Іогансена «Елементи точного вчення про змінюваність та спадковість».

Слід відзначити, що в економічних дослідженнях дисперсійний метод ще не набув такого широкого використання, як у біології, зоотехнії, та техніці. Натомість можливості його використання в сфері економіки досить широкі. Основне призначення дисперсійного аналізу –статистично виявити вплив факторів на варіацію ознаки, що вивчається. Особливий інтерес становить використання цього методу в тих випадках, коли зміна згаданої ознаки зумовлена одночасно дією – факторів, частка впливу яких різноманітна.

У дисперсійному аналізі використовується властивість суми квадратів центральних відхилень. Суть її полягає в тому, що коли кілька повністю незалежних факторів діють одночасно й зумовлюють загальну змінюваність ознаки, то сума окремих дисперсій, що вимірюють їх вплив, дорівнює загальній дисперсії: $\sum Д_1 + \sum Д_2 + \sum Д_3 + \ldots + \sum Д_п = \sum Д_{заг.}$

За однофакторною схемою вивчення здійснюють, маневруючи лише однією ознакою, вважаючи інші незмінюваними. Такий метод не дає змоги виявити взаємодію факторів при одночасній їх зміні. Цих недоліків позбавлений багатофакторний аналіз, при якому кожне спостереження служить для одночасної оцінки всіх факторів та їх взаємодій.

Якщо відобразити загальну мінливість рівня тієї чи іншої ознаки через C_y, то її можна показати як суму окремих дисперсій, що виникають під дією різних факторів. Загальна дисперсія (C_y) визначається як сума квадратів відхилень кожної варіанти від середньої арифметичної і може бути розкладена на складові: 1) C_x – факторна (міжгрупова) дисперсія, дисперсія, що виникає під впливом врахованих факторів; 2) C_z–залишкова дисперсія (внутрішньогрупова), зумовлена дією різних випадкових (неврахованих) факторів.

У загальному вигляді дисперсія (мінливість) ознаки виражається так: $C_y=C_x+C_z$.

Факторна дисперсія (C_x) являє собою суму квадратів окремих середніх значень ознаки (M_x), одержаних у групах статистичного комплексу діючих факторів, та загальної арифметичної ($M_{заг.}$), яка обчислюється для всього статистичного комплексу із показників варіюючої ознаки. Це можливо відобразити у вигляді: $C_x=(M_{част.}-M_{заг.})^2$ або $C_x=\sum n_x(M_{част.}-M_{заг.})^2$, де n_x – кількість спостережень за градаціями факторів.

Випадкова (C_z) дисперсія визначається як сума квадратів різниць варіюючої ознаки (V) відносно часткової середньої арифметичної: $C_z=\sum(V-M_{част..})^2$.

Відношення складових дисперсій до загальної характеризує ступінь факторних ознак у формуванні загальної мінливості результативної ознаки. Так, ступінь впливу врахованих факторів становить $\eta_z^2=\frac{C_x}{C_y}$.

Відзначене вище стосується вивчення мінливості ознак під впливом одного фактора. Якщо ж вивчається змінюваність результативної ознаки, зумовлена впливом кількох факторів, тоді факторна дисперсія C_x може бути представлена сумою дисперсій кожного фактора окремо (A, B,C і т.д.) та дисперсій спільної дії факторів, що аналізуються (AB, AC, BC, ABC, і т.д.) . Для випадку, коли досліджується вплив на результативну ознаку трьох факторів, ця дисперсія записується в такому вигляді

$$C_x=C_A+C_B+C_C+C_{AB}+C_{AC}+C_{BC}+C_{ABC}.$$

Наведене рівняння має силу, якщо статистичний комплекс за співвідношенням частот належить до рівномірного пропорційного комплексу. В аналізі економічних явищ найчастіше зустрічаються нерівномірні комплекси. Для них набирає сили нерівність $C_A+C_B+C_C+C_{AB}+C_{AC}+C_{BC}+C_{ABC} \neq C_x$.

Знаходження дисперсій спільного впливу факторів C_{AB}, C_{AC}, C_{BC}, C_{ABC} вимагає особливих прийомів, які будуть розглянуті пізніше на матеріалах конкретних даних.

Вивчаючи методичну сторону дисперсійного аналізу, можна виділити чотири етапи його здійснення : 1) обробка статистичного комплексу для одержання загальної факторної і залишкової дисперсій (C_y, C_x, C_z) ; 2) визначення частки кожної окремої дисперсії в загальній, для чого розраховують величини η_x^2 , η_z^2; 3) коригування одержаних дисперсій на число ступенів вільності, які знаходять для кожної дисперсії за певними формулами; 4) оцінка факторної дисперсії, тобто встановлення вірогідності впливу кожного вибраного фактора (x або A, B, C і т.д.) на варіюючу ознаку. Для цього використовують критерії Фішера (F – критерій).

Як зазначалося вище, в неортогональних комплексах система розрахунку часткових дисперсій C_A, C_B, C_{AB} і т.д. має деякі особливості, зумовлені непропорційними співвідношеннями частот одного фактора по групах (градаціях) другого фактора . Для усунення неортогональності перетворюють нерівномірний комплекс у рівномірний . існують різні способи перетворення нерівномірного комплексу в рівномірний – по Поморському Ю.П., Немчинову В.С. та ін. Точність цих методів деякою мірою різна, але можна користуватися будь – яким з них.

Серед діючих способів перетворення в практиці економічного аналізу найбільш зручним слід визнати метод Поморського Ю.П. В основі цього прийому лежить усереднення частот по групах факторів. По кожній групі досліджуваного фактора знаходять часткові середні з варіант підгруп і загальну середню по всьому дисперсійному комплексу. Спочатку часткові і загальну середню підносять до квадрату, а потім ці величини підставляють у формули дисперсій C_x, C_A, C_B, C_C, C_{AB} і т.д.

При обробці трифакторного дисперсійного нерівномірного комплексу формула дисперсії факторів має вигляд: $C_x = n\left(\frac{\sum M_x^2}{l_A l_B l_C} - M_{заг}^2\right)$,

де n – число одиниць спостереження, введених у комплекс;

$\sum M_x^2$ - сума квадратів часткових середніх арифметичних по всіх групах і по всіх факторах ; l_A, l_B, l_C - число груп по кожному з факторів; $M^2_{заг}$ - квадрат середньої арифметичної, що розраховується для всього комплексу.

Часткові дисперсії C_A, C_B, C_C обчислюють за такою формулою:

$$C_A = n\left(\frac{M_A^2}{l_A} - M^2_{заг.}\right),$$

де $\sum M_x^2$ -сума квадратів середніх арифметичних по групах факторів А, а відношення $\frac{\sum M_A^2}{l_A}$ дорівнює h_A.Аналогічно розраховують часткові дисперсії C_B, і C_C . Визначивши факторну (C_x,) і часткові (C_A, C_B, C_C) дисперсії, можна розрахувати дисперсії різних сполучень факторів ($C_{AB}, C_{AC}, C_{BC}, C_{ABC}$).

Величина загальної дисперсії дорівнюватиме $C_y = C_x + C_Z$. Величина загальної дисперсії із даного виразу буде дещо відрізнятися від тієї величини, яка одержана за вище наведеними розрахунками, але ця розбіжність діє несуттєво на наступні обчислення. Отже, наведеним способом розрахунків цілком можна користуватися.

Принцип розрахунку однофакторного і багатофакторного дисперсійних комплексів має деякі методичні особливості, зумовлені кількістю факторів, включених в аналіз. Ці особливості будуть викладені у логічній послідовності розглядуваних нижче розрахунків одно -, дво – і трифакторних дисперсійних комплексів.

7.1.2. Алгоритми рішення дисперсійних моделей

Приклад. Розглянемо послідовність розрахунку однофакторного дисперсійного комплексу на прикладі залежності середньорічного надою корів (V) від рівня годівлі (А) в 30 (*n*) підприємствах.

На першому етапі здійснюється групування підприємств за факторною ознакою. У даному прикладі сукупність підприємствах. поділена на три групи за рівнем використання кормів на корову в рік (А). Обробка вихідної інформації здійснюється за схемою таблиці 47.

На підставі даних таблиці 47 знаходимо загальну (C_y), факторну (C_x), залишкову (C_Z) дисперсії:

$$C_y = \sum V^2 - \frac{(\sum v)^2}{n} = 39740 - 39138 = 602;$$

$$C_x = \sum h - \frac{(\sum V)^2}{n} = 39433 - 39138 = 295;$$

$$C_Z = \sum V^2 - \sum h = 39740 - 39433 = 307.$$

Співвідношення складових дисперсій (C_x, C_Z) до загальної (C_y) показує ступінь участі факторних ознак у формуванні загальної їзмінюваності результативної ознаки. Так, ступінь впливу рівня годівлі корів на їх продуктивність становитиме: $\eta_x^2 = \frac{C_x}{C_y} = \frac{295}{602} = 0{,}49$ (49%).

Ступінь впливу суми інших неврахованих факторів на результативну ознаку обчислюється за таким співвідношенням: $\eta_Z^2 = \frac{C_Z}{C_y} = \frac{307}{602} = 0{,}51$ (51%).

Таким чином, у розглянутому прикладі факторна ознака (рівень годівлі) визначає 49% загальної варіації результативної ознаки (надою).

Таблиця 47

Вихідні і розрахункові дані однофакторного дисперсійного комплексу

Показник	Витрати кормів на 1 корову, ц к. од.			Суми ($\sum$)
	A_1 -до 47,4	A_2 -47,5-58,0	A_3 - понад 58,0	
V надій, ц	32,60	35,71	49,09	
	31,00	32,44	40,17	
	30,58	33,82	35,14	
	32,52	31,27	42,72	
	32,40	34,54	37,68	
	32,02	34,69	44,41	
	31,38	32,21	38,15	
		41,28		
		34,40		
		44,63		
		37,31		
		35,95		
		37,86		
		36,16		
		34,40		
		36,02		
$\sum V$	233,50	572,72	287,36	1083,58
n_x	7	16	7	30
$(\sum V)^2$	49952,25	32800,19	82575,76	
$h = (\sum V)^2 : n_x$	7136,04	20500,51	11796,54	39433,09
$\sum V^2$	7142	20667	11931	39740

Дисперсія як показник різноманітності залежить від кількості одиниць спостереження (підприємств) у групі. Для визначення впливу факторів ця обставина не має значення. В інших же випадках, зокрема, при встановленні вірогідності впливу факторів, необхідний показник, вільний від вказаної залежності, що допускає порівняння груп, різних за кількістю елементів, що входять до них. Таким показником є коригована дисперсія – девіата.

Девіатою називають дисперсію, яка припадає на один елемент вільного варіювання або на один ступінь вільності.

Корінь квадратний з девіати ($\sqrt{\sigma^2}$) являє собою звичайний показник математичної статистики – середнє квадратичне відхилення (σ).

У нашому прикладі число ступенів вільності варіації (ν) для факторної ознаки і для неврахованих факторів становитиме відповідно : $\nu_x = l - 1 = 3 - 1 = 2$; $\nu_Z = n - l = 30 - 3 = 27$, де l – кількість виділених груп; n – чисельність вибірки.

Розрахуємо девіати :

$$\sigma_x^2 = \frac{C_x}{\nu_x} = \frac{295}{2} = 147{,}50; \quad \sigma_Z^2 = \frac{C_Z}{\nu_Z} = \frac{307}{27} = 11{,}37$$

Критерієм вірогідності впливу факторної ознаки на результативну є співвідношення її девіати до девіати неврахованих факторів. Якщо розраховане співвідношення дорівнює чи більше визначеної стандартної величини, вплив вважається вірогідним з певним ступенем імовірності. Стандартні відношення девіат визначаються за спеціальними таблицями (додатки 8,9).

Знаходимо це співвідношення для факторної ознаки на такому прикладі:

$$F_p = \frac{\sigma_x^2}{\sigma_Z^2} = \frac{147{,}50}{11{,}37} = 12{,}97.$$

Одержаний критерій (F_p) порівнюємо з табличним його значенням при двох порогах ймовірності : 0,95; 0,99 (додатки 8,9).

Наведемо стандартні співвідношення девіат, що відповідають ступеням вільності варіації.

Імовірність Р	Критерій F
0,95	3,3
0,99	5,5

У нашому прикладі F_p(12,97) > F_T(3,3). Отже, в досліджуваних підприємствах вплив рівня годівлі корів на їх продуктивність виявився досить сильним і вірогідним. Про вірогідність результатів аналізу свідчить високий ступінь імовірності 0,99.

Приклад. Розрахунок двофакторного дисперсійного комплексу розглянемо на прикладі вивчення залежності собівартості 1ц молока (V) в 30 підприємств району від рівня концентрації поголів’я корів (А) і спеціалізації виробництва молока (В). з цією метою сукупність розділена на 2 групи з подальшим поділом на 2 підгрупи (табл. 48).

На підставі розрахункових даних таблиці 48 визначаємо загальну (C_y), факторну (C_x) і залишкову (C_Z) дисперсії:

$$C_y = \sum V^2 - \frac{(\sum V)^2}{n_x} = 23328{,}67 - \frac{(826{,}98)^2}{30} = 532{,}14;$$

$$C_x = \sum h - \frac{(\sum V)^2}{n_x} = 23023{,}33 - \frac{(826{,}98)^2}{n_x} = 226{,}80;$$

$$C_Z = \sum V^2 - \sum h = 23328{,}67 - 23023{,}33 = 305{,}34.$$

Таблиця 48

Вихідні і розрахункові дані двофакторного дисперсійного комплексу

Показник	Корів на 100 га с.-г. угідь, гол.				Сума ($\sum$)
	A_1 - до 50		A_2 - понад 50		
	Питома вага молока в обсязі товарної продукції, %				
	B_1 -до 30	B_2 -понад 30	B_1 -до 30	B_2 -понад 30	
V (Собівартість 1ц молока, грн.)	27,59	35,65	19,45	23,49	
	32,89	34,20	17,90	28,86	
	27,23	29,50	19,84	25,46	
	35,30	28,19	22,26	26,96	
	25,02	27,84	25,83	28,82	
	29,44	28,75	27,20	28,27	
	26,19	31,57	25,47	31,87	
	26,87		29,07		
$\sum V$	230,53	215,70	187,02	193,73	826,98
n_x	8	7	8	7	30
$(\sum V)^2$	53144,08	46526,49	34976,48	37531,31	-
$h = (\sum V)^2 : n_x$	6643,01	6646,64	4372,06	5361,62	23023,33
$M_x = \sum V : n_x$	28,81	30,81	23,38	27,68	-
M_x^2	830,01	949,25	546,62	766,18	3092,06
$\sum V^2$	…	…	…	…	23328,67

Ступінь впливу факторних ознак (концентрації її і спеціалізації виробництва) на результативну ознаку (собівартість виробництва молока) становитиме : $\eta_x^2 = \frac{C_x}{C_y} = \frac{226{,}80}{532{,}14} = 0{,}4262$ (42,6 %). Ступінь впливу неврахованих факторів на результативну ознаку буде: $\eta_Z = \frac{C_Z}{C_y} = \frac{305{,}34}{532{,}14} = 0{,}5738$ (57,4 %).

Для кількісної характеристики впливу кожного з факторів слід дисперсію їх сумарної дії розкласти на складові, тобто в факторній дисперсії (C_x) виділити дисперсії першого (C_A) і другого (C_B) факторів, а також їх сполучення (C_{AB}).Дисперсія C_{AB} характеризує ступінь зумовленості впливу першого фактора другим. Складемо допоміжну таблицю 49. Третя колонка цієї таблиці розраховується на основі даних таблиці 49.

59,62=28,81 +30,81; 52,19 = 28,81 + 23,38

Таблиця 49

Допоміжні розрахунки для обробки дисперсійного комплексу за факторами А, В

Градація, i	Число середніх , l	$\sum \text{M}_x$	$\text{M}_i = \frac{\sum \text{M}_x}{l}$	M_l^2
A_1	2	59,69	29,81	888,63
A_2	2	51,06	25,53	651,78
		110,68		$\sum \text{M}_A^2 = 1540{,}41$
B_1	2	52,19	26,10	681,2100
B_2	2	58,49	29,24	854,9776
		110,68		$\sum \text{M}_B^2 = 1536{,}1876$

Середня арифметична (загальна) по градаціях факторів становить: $\text{M}_{заг} = \frac{\text{M}_x}{\sum l} = \frac{110{,}68}{4} = 27{,}67.$

Ступінь різноманітності середніх арифметичних собівартості 1ц молока розраховуємо в такій послідовності :

для всіх градацій – $C'_x = n\left(\frac{\sum \text{M}_x^2}{l_A \cdot l_B} - \text{M}_{заг}^2\right) = 30\left(\frac{3092{,}06}{2 \cdot 2} - 27{,}67^2\right) = 221{,}70;$

для градації $A - C'_A = n\left(\frac{\sum \text{M}_A^2}{l_A} - \text{M}_{заг}^2\right) = 30\left(\frac{1540{,}41}{2} - 27{,}67^2\right) = 137{,}10$;

для градації $B - C'_B = n\left(\frac{\sum \text{M}_B^2}{l_B} - \text{M}_{заг}^2\right) = 30\left(\frac{1536{,}19}{2} - 27{,}67^2\right) = 74{,}10$;

для сполучення факторів

А і $B - C'_{AB} = C'_x - C'_A - C'_B =$ 221,70 – 137,10 – 74,10 =10,50

Для розкладу сумарної дисперсії досліджуваних факторів на складові розраховуємо поправочний коефіцієнт:

$$K = \frac{C_x}{C'_x} = \frac{226{,}80}{221{,}70} = 1{,}023.$$

Дисперсії, зумовлені дією досліджуваних факторів і їх сполучення, становлять: $C_A = C'_A K = 137{,}10 \cdot 1{,}023 = 140{,}25;$ $C_B =$ $C'_B K == 74{,}10 \cdot 1{,}023 = 75{,}81;$ $C_{AB} = C'_{AB} K = 10{,}50 \cdot 1{,}023 = 10{,}74.$

Кінцева дисперсійна структура двофакторного дисперсійного комплексу матиме вигляд: $C_y = (C_A + C_B + C_{AB}) + C_Z = 226{,}80 + 305{,}34 = 5320{,}14.$

Розраховуємо ступінь впливу факторів, що вивчаються, на формування змінюваності результативної ознаки. Зокрема, рівень концентрації поголів'я корів визначає варіацію собівартості 1 ц молока $\eta_A^2 = \frac{C_A}{C_y} = \frac{140{,}25}{532{,}14} = 0{,}2636,$ або 26,36 %; фактор спеціалізації – $\eta_B^2 = \frac{C_B}{C_y} = \frac{75{,}81}{532{,}14} = 0{,}1425,$ або 14,25; взаємодія факторів – $\eta^2_{AB} = \frac{C_{AB}}{C_y} = \frac{10,74}{532,14} = 0,0202,$ або 2,02 %.

Числа ступенів вільності для розрахунку девіати в двофакторному комплексі розраховуються в такій послідовності : $\nu_A = l_A - 1 = 2 - 1 = 1;$ $\nu_B = l_B - 1 = 2 - 1 = 1;$ $\nu_{AB} = l_A \cdot l_B = 1;$ $\nu_x = \nu_A + \nu_B + \nu_{AB} = 3;$ $\nu_Z = n - l_A \cdot l_B = 30 - 4 = 26.$

Визначаємо девіати :

$$\sigma^2_A = \frac{C_A}{\nu_A} = \frac{140,25}{1} = 140,25; \quad \sigma^2_B = \frac{C_B}{\nu_B} = \frac{75,81}{1} = 75,81;$$

$$\sigma^2_{AB} = \frac{C_{AB}}{\nu_{AB}} = \frac{10,74}{1}10,74; \quad \sigma^2_x = \frac{C_x}{\nu_x} = \frac{226,80}{3} = 75,60;$$

$$\sigma^2_Z = \frac{C_{AB}}{\nu_Z} = \frac{305,34}{26} = 11,74.$$

Розраховуємо F –критерій:

$$F_A = \frac{\sigma^2_A}{\sigma^2_Z} = \frac{140,25}{11,74} = 11,96; \quad F_B = \frac{\sigma^2_B}{\sigma^2_Z} = \frac{75,81}{11,74} = 6,45;$$

$$F_{AB} = \frac{\sigma^2_{AB}}{\sigma^2_Z} = \frac{10,74}{11,74} = 0,91; \quad F_x = \frac{\sigma^2_x}{\sigma^2_Z} = \frac{75,60}{11,74} = 6,44.$$

Одержані критерії порівнюємо з табличними їх значеннями при двох порогах імовірності 0,95 і 0,99 (додатки 8,9).

Стандартні відношення девіат, які відповідають ступеням вільності варіації неврахованих факторів ($\nu_Z = 26$) і розрахованим вище ступеням вільності варіації досліджуваних факторів становитимуть:

Імовірність Р	Критерій F	
	При $\nu_1 = 1$	При $\nu_3 = 3$
0,95	4,22	2,98
0,99	7,72	4,64

Результати аналізу кожного з факторів окремої чи сумарної їх дії слід вважати вірогідними при тих порогах імовірності, де

$F_p > F_T$. Недостовірними при $F_p < F_T$.

Приклад. Розглянемо послідовність розрахунку трифакторного дисперсійного комплексу на прикладі вивчення залежності собівартості виробництва 1ц яловичини (V) від рівня продуктивності праці (А) , рівня витрат кормів на 1 ц приросту (В) і собівартості 1ц кормових одиниць (С). З метою кількісної оцінки названих вище факторів на результативну ознаку будуємо трифакторний дисперсійний комплекс, основу якого становить комбінаційне групування 66 підприємств (табл. 50).

Досліджувану сукупність спочатку розподілено на дві групи: з рівнем затрат робочого часу на виробництво 1 ц яловичини до 90 людино –годин. (A_1) і понад 90 людино –годин (A_2). У кожній групі було виділено по дві підгрупи з середнім розміром витрат кормів на 1 ц продукції: менше 10 (B_1) і більше 10 ц кормових одиниць. (B_2). Потім кожна з них у свою чергу розподілена ще на дві підгрупи: з собівартістю 1 ц кормових одиниць, згодованих тваринам, до 14 грн.(C_1) і понад 14 грн. (C_2). У результаті досліджувана сукупність підприємств була розподілена на 8 підгруп, по кожній з яких наведено варіанти результативної ознаки (V) – рівень собівартості виробництва 0,1ц яловичини.

Оскільки у нашому прикладі розглядається трифакторний нерівномірний комплекс, обробку його здійснюємо в такій послідовності: спочатку будуємо звичайним чином кореляційну решітку, потім виконуємо допоміжні розрахунки, результати яких заносимо в цю ж таблицю. До них відносяться кількість (n=66) і сума ($\sum V$) варіант досліджуваного комплексу, сума часток від ділення квадратів сум варіант по кожній підгрупі на число варіант ($\sum h$=669887,82), сума квадратів середніх арифметичних по підгрупах $\left(\sum M_x^2 = 78826.00\right)$.

На підставі розрахункових даних таблиці 50 визначаємо загальну (C_y), факторну (C_x) і залишкову (C_Z) дисперсії:

$$C_y = \sum V^2 - \frac{(\sum V)^2}{n} = 672890,20 - \frac{6599,77^2}{66} = 12936,20;$$

$$C_x = \sum h^2 - \frac{(\sum V)^2}{n} = 669887,32 - \frac{6599,77^2}{66} = 9933,32;$$

$$C_z = \sum V^2 - \sum h^2 = 672890.20 - 669887.32 = 3002.88.$$

Встановлюємо частку всіх досліджуваних факторів у загальній варіації результативної ознаки. Так, ступінь впливу продуктивності праці, розміру витрат і вартості кормів на рівень собівартості становить : $\eta_x^2 = \frac{9933,32}{12936,20} = 0,768,$ або 76,8%; а суми неврахованих факторів – $\eta_Z^2 = \frac{3002,88}{12936,20} = 0,232,$ або 23,2%.

Як відзначалося вище, у багатофакторних комплексах дисперсія спільної дії врахованих факторів (C_x) підлягає розподілу на дисперсії кожного з факторів окремо (C_A, C_B, C_C), а також дисперсії різних варіантів їх сполучень $(C_{ABC}, C_{AB}, C_{AC}, C_{BC})$. У нерівномірних комплексах всі часткові дисперсії факторів відрізнятимуться від величин таких же дисперсій в рівномірному комплексі, тому позначимо їх через C'.

Допоміжні розрахунки для визначення окремих дисперсій показані у таблицях 51 і 52.

Таблиця 50

Обробка трифакторного дисперсійного комплексу

(А – групи підприємств за рівнем продуктивності праці, людино-годин на 1 ц ; В – розмір витрат кормів на виробництво 1ц яловичини, ц корм. од.; С- собівартість 1ц кормових одиниць, згодованих худобі, грн; V – рівень собівартості 0,1ц яловичини, грн.; n_i - кількість підприємств)

Групи та підгрупи за факторами			Вихідні і розрахункові дані							
$l_A = 2;\ l_B = 2;\ l_C = 2$			V	n_i	$\sum V$	$(\sum V)^2$	$\sum h = \frac{(\sum V)^2}{n_i}$	$\sum V^2$	$M_x = \frac{\sum V}{n_i}$	M_x^2
A_1 -до 90	B_1 – до 10	C_1 – до 14	76,76; 75,74…	4	278,01	77289,56	19322,39	19676,70	69,50	4830,25
		C_2 – понад 14	87,74; 92,09…	9	866,53	750874,24	83430,47	83792,93	96,28	9269,84
	B_2 – понад 10	C_1 – до 14	84,44; 92,84…	8	733,03	537332,98	67166,62	67376,62	91,63	8396,06
		C_2 понад 14	104,88; 102,08…	6	657,06	431727,84	71954,64	72144,87	109,51	11992,44
A_2 - понад 90	B_1 – до 10	C_1 – до 14	85,84; 98,66…	10	902,86	815156,18	81515,62	81919,98	90,29	8152,28
		C_2 – понад 14	107,17; 102,49…	9	966,59	934296,23	103810,69	104154,52	107,40	11513,29
	B_2 – понад 10	C_1 – до 14	10126; 103,22…	11	1117,69	1249230,94	113566,45	111971,19	101,61	10324,59
		C_2 – понад 14	123,48; 118,42…	9	1078,00	12920,44	129120,44	129673,92	119,78	14347,25
Сума			6599,77	66	-	-	669887,32	672890,20	-	78826,00

Ступінь відмінності по всіх факторах визначаємо за вище наведеною формулою:

$$C'_x = n\left(\frac{\sum M_x^2}{l_A \cdot l_B \cdot l_C} - M_{заг}\right);$$

$$M_{заг} = \frac{786,00}{8} = 98,25;$$

$$C'_x = 66\left(\frac{78826,00}{2 \cdot 2 \cdot 2} - 98,25^2\right) = 13212,54;$$

а окремо по факторах А, В, С за формулами:

$$C'_A = n\left(\frac{\sum M_A^2}{l_A} - M_{заг}^2\right) - 66\left(\frac{19391,14}{2} - 9653,06\right) = 2805,66;$$

$$C'_B = n\left(\frac{M_B^2}{l_B} - M_{заг}^2\right) = 66\left(\frac{19415,06}{2} - 9653,06\right) = 3595,02;$$

$$C'_C = n\left(\frac{M_C^2}{l_C} - M_{заг}^2\right) = 66\left(\frac{19505,78}{2} - 9653,06\right) = 6586,80$$

.

Таблиця 51

Допоміжні розрахунки для обробки дисперсійного комплексу за факторами А, В, С

Групи і підгрупи за факторами	Число середніх (l)	Число спостережень (n)	Розрахункові дані				
			$\sum M_x$	$M_i = \frac{\sum M_x}{l}$	M_l^2	$\sum V$	$M = \frac{\sum V}{n}$
A_1	4	27	366,92	8414,39	2534,63	2534,63	93,87
A_2	4	39	419,08	104,77	10976,75	4065,14	104,23
Показники по фактору		66	786,00	-	19391,14	6599,77	100,00
А							
B_1	4	32	363,47	90,87	8257,36	301,99	91,19
B_2	4	34	422,53	105,63	11157,70	3585,78	105,46
Показники по фактору В		66	786,00	-	19415,06	6599,77	100,00
C_1	4	33	353,03	88,26	7789,83	3031,59	91,87
C_2	4	38	432,97	108,24	11715,90	3568,18	108,18
Показники по фактору С		66	786,00	-	19505,73	6599,77	100,0

Таблиця 52

Допоміжні розрахунки для обробки сполучень факторів

Підгрупи за факторами	Число середніх (l)	Розрахункові дані		
		$\sum M_x$	$M_i = \frac{\sum M_x}{l}$	M_l^2
$A_1\ B_1$	2	165,78	82,89	6870,75
$A_1\ B_2$	2	201,14	100,57	10114,32
$A_2\ B_1$	2	197,69	98,84	9769,35
$A_2\ B_2$	2	221,39	110,69	12252,28
		786,00	-	$\sum M_{AB}^2$ =39006,70
$A_1\ C_1$	2	161,13	80,56	6489,91
$A_1\ C_2$	2	205,79	102,89	10586,35
$A_2\ C_1$	2	191,90	95,95	9206,40
$A_2\ C_2$	2	227,18	113,59	12902,69
		786,00	-	$\sum M_{AC}^2$ =39185,35
$B_1\ C_1$	2	159,79	79,89	6382,41
$B_1\ C_2$	2	203,68	101,84	10371,39
$B_2\ C_1$	2	193,24	96,62	9335,42
$B_2\ C_2$	2	229,29	114,64	13142,33
		789,00	-	$\sum M_{BC}^2$ =39231,55

За даними таблиць 51 і 52 визначаємо ступінь вільності середніх арифметичних для об'єднаних факторів : А і В – $C'_{AB} = n\left(h_{AB} - h_A - h_B + M_{заг}^2\right)$,

$$h_{AB} = \frac{\sum M_{AB}^2}{l_A \cdot l_B} = \frac{39006{,}70}{2 \cdot 2} = 9751{,}07;$$

де

$$h_A = \frac{\sum M_A^2}{l_A} = \frac{19391{,}14}{2} = 9695{,}57$$

$$h_B = \frac{\sum M_B^2}{l_B} = \frac{19415{,}06}{2} = 9707{,}53.$$

Отже,

$$C'_{AB} = 66(9751{,}67 - 9695{,}57 - 9707{,}53 + 9653{,}06) = 107{,}58.$$

Аналогічно розраховуємо часткові дисперсії для інших сполучень факторів:

А і С – $C'_{AC} - 66(9796{,}34 - 9695{,}57 - 9752{,}86 + 9653{,}06) = 64{,}02;$

В і С – $C'_{BC} = 66(9807{,}89 - 9707{,}53 - 9752{,}86 + 9653{,}06) = 36{,}96;$

А, В, С – $C'_{ABC} = 13212{,}54 - 2805{,}66 - 3595{,}02 - 6586{,}80 - 107{,}58 - 64{,}02 - 39{,}96 = 16{,}50.$

Знаходимо поправочний коефіцієнт $K = \frac{C_x}{C'_x} = \frac{9933{,}32}{13212{,}54} - 0{,}7518.$

Для виправлення часткових дисперсій $C'_x, C'_{\text{A}}, C'_{\text{B}}, C'_C, C'_{\text{AB}}, C'_{\text{A}C}, C'_{\text{B}C}, C'_{\text{AB}C}$ множимо на їх поправку 0,7512 і результати заносимо в другий рядок таблиці 54. Ступінь впливу досліджуваних факторів у формуванні загальної мінливості собівартості визначається відношенням часткових дисперсій по факторах $\left(C_{\text{A}};C_{\text{B};};C_C\right)$ і їх сполучень $(C_{\text{AB}};C_{\text{A}C};C_{\text{B}C};C_{\text{AB}C})$ до загальної дисперсії $\left(C_y\right)$. У нашому прикладі для фактора А (затрати живої праці на 1ц приросту) $\eta_A^2 = \frac{C_A}{C_y} = \frac{2109,29}{12936,20} = 0,163$ тобто в умовах досліджуваних підприємств варіація продуктивності праці становить 16,3 % варіації собівартості виробництва продукції.

Фактор В (розмір витрат кормів на виробництво 1ц яловичини) становить 20,9 % коливання показника рівня собівартості, а фактор С (вартість 1ц корм. од.) – 38,3 %. Частка впливу у зміні рівня собівартості взаємодії факторів характеризується такими даними: А і В – 0,6 %; А і С – 0,4 %; В і С – 0,2 %; А, В і С – 0,1 %.

Знаходимо число ступенів вільності варіації, які в трифакторному дисперсійному комплексі обчислюють у такому порядку : $\nu_{\text{A}} = l_{\text{A}} - 1 = 1;$ $\nu_{\text{B}} = l_{\text{B}} - 1 = 1;$ $\nu_C = l_C - 1 = 1;$ $\nu_{\text{AB}} = \nu_A \cdot \nu_{\text{B}} = 1;$ $\nu_{\text{A}C} = \nu_{\text{A}} \cdot \nu_C = 1;$ $\nu_{\text{B}C} = \nu_{\text{B}} \cdot \nu_C = 1;$ $\nu_{\text{AB}C} = \nu_A \cdot \nu_{\text{B}} \cdot \nu_C = 1;$ $\nu_x = \nu_{\text{A}} + \nu_{\text{B}} + \nu_C + \nu_{\text{AB}} + \nu_{\text{A}C} + \nu_{\text{B}C} + \nu_{\text{AB}C} = 7.$

Сума часткових ступенів вільності повинна давати їх число для загальної дисперсії $\nu_y = \nu_x + \nu_{\text{Z}} = 7 + 58 = 65.$

Девіати, розраховані за даними нашого прикладу, наведені у таблиці 53 по рядку 5.

Вірогідність дії факторів і їх сполучень визначаємо, як і раніше відношенням факторних девіат і їх сполучень до залишкової девіати. Для нашого прикладу наведені по рядку 5 таблиці величини девіат ділимо на залишкову дисперсію 51,77. Обчислені значення коефіцієнтів F записуємо по рядку 6.

Зіставляючи обчислені та табличні значення F критеріїв бачимо, що загальнофакторна дисперсія σ_x^2 і дисперсії, викликані кожним з досліджуваних факторів, достовірні при всіх порогах імовірності (Р=0,95 ; Р=0,99 ; Р=0,999), оскільки $F_p > F_T$. Дисперсії, зумовлені сполученнями (при всіх можливих варіантах) факторів, виявились невірогідними.

Таблиця 53

Зведена інформація результатів лічильної обробки трифакторного дисперсійного комплексу

Статистичні характеристики	Умовні позначення	A	B	C	AB	AC	BC	ABC	x	z	y
Дисперсія											
невиправлена	C	2805,66	3595,02	6586,80	107,58	64,02	36,96	16,90	13212,54		
виправлена	$C = C' \cdot K$	2109,29	2702,74	4951,96	80,88	48,13	27,79	12,40	9933,32	3002,88	12936,20
Коефіцієнт співвідношення	η^2	0,163	0,209	0,383	0,006	0,004	0,002	0,001	0,768	0,232	1,000
Число ступенів вільності	ν	1	1	1	1	1	1	1	7	58	65
Девіата	σ^2	2109,29	2702,74	4951,69	80,88	48,13	27,79	12,40	1419,04	51,77	-
Критерій Фішера	F_p	40,74	52,21	95,65	1,56	0,93	0,54	0,24	27,41	-	-
розрахунковий	0,999	12,1	12,1	12,1	12,1	12,1	12,1	12,1	4,3	-	-
табличний	F_T 0,99 0,95	7,1 4,0	7,1 4,0	7,1 4,0	7,1 4,0	7,1 4,0	7,1 4,0	7,1 4,0	3,0 2,2	- -	- -

7.1.3. Аналіз абсолютних змін досліджуваної ознаки

З аналітичної точки зору являє певний інтерес зіставлення груп у дисперсійному комплексі при вивченні впливу на результативну ознаку факторних ознак у різному їх сполученні (поєднанні). У трифакторному комплексі мають місце подвійні та потрійні взаємодії факторів. Наприклад, для розглядуваного прикладу розрахунку трифакторного комплексу середній рівень собівартості виробництва 1ц яловичини, сформованого під впливом факторів А і В при їх рівнях A_1 і B_1 становитиме $M_{A_1B_1} = 88{,}04$ грн. (278,00 + 866,53) : (9 + 4).

Аналогічно обчислюють названу результативну ознаку для всіх можливих сполучень вивчаючих факторних ознак. Нижче наведені середні рівні залежної змінної, одержані під впливом незалежних змінних у різних варіантах їх сполучень. Тобто маємо середній рівень собівартості одиниці продукції, зумовлений впливом різних варіантів взаємодії факторів продуктивності праці, рівня затрат та їх вартості.

$M_{A_1B_1}=88{,}04$; $M_{A_1B_2}=99{,}29$; $M_{A_2B_1}=98{,}39$; $M_{A_2B_2}=109{,}78$; $M_{A_1C_1}=84{,}25$; $M_{A_1C_2}=101{,}57$; $M_{A\ C_1}=96{,}22$; $M_{A_2C_2}=113{,}59$; $M_{A_1C_1}=84{,}33$; $M_{B_1C_1}=101{,}84$; $M_{B_2C_1}=97{,}41$; $M_{B_2C_2}=115{,}67$; $M_{A_1B_1C_1}=69{,}50$; $M_{A_1B_1C_2}=96{,}28$; $M_{A_1B_2C_1}=91{,}63$; $M_{A_1B_2C_2}=109{,}51$; $M_{A_2B_1C_1}=90{,}29$; $M_{A_2B_1C_2}=107{,}40$; $M_{A_2B_2C_1}=101{,}61$; $M_{A_2B_2C_2}=119{,}78$.

При аналізі загальної дії досліджуваних факторів спочатку вивчають вплив на результативну ознаку кожного фактора окремо, а потім їх сполучення. Судячи по даних розглядуваного прикладу, виявився досить сильний вплив фактора $C\left(\eta_C^2 = 38{,}3\%\right)$. Як виразився вплив цього фактора, показує основний ряд часткових середніх M_x, показаний графічно на рисунку 19. Із графіка і числового ряду добре видно, що фактор С при всіх градаціях факторів А і В діяв однаково : при C_1 до 14 грн. рівень собівартості приросту був порівняно низький, при C_2 - понад 14 грн. він підвищився. Найнижчий (69,50 грн.) рівень собівартості виробництва яловичини проявляється в групі $A_1B_1C_1$, оскільки сполучення факторів зумовлюючих такий рівень, містить найкращі показники продуктивності праці, витрат і вартості кормів у досліджуваній сукупності. Зіставлення груп $A_1B_1C_2$ і $A_1B_2C_2$ показує різницю середніх рівнів собівартості в підприємствах з однаково високою продуктивністю праці, низькою собівартістю витрачених кормів, але з різним рівнем їх витрат на виробництво 1ц яловичини.

Ця різниця в розглядуваному прикладі становить $M_{A_1B_2C_2} - M_{A_1B_1C_2}$ =109,51 – 96,28 = 13,23 грн.Зміна в абсолютних рівнях результативної ознаки, викликана підвищенням продуктивності праці (А) і зниженням вартості кормів (С) при однаково низькому рівні їх затрат (B_1), показує різницю $M_{A_2B_1C_2} - M_{A\ B_1C_1}$ = 107,40 – 69,50 = 37,90 грн.

Як бачимо, значний вплив на результативну ознаку виявив цей фактор і в сполученнях з факторами А і В. Частка впливу цих сполучень становила відповідно 22,2 і 17,6 %.

Аналізуючи дію факторів А і В, не важко переконатися, що нижчий рівень результативної ознаки (собівартості) у підгрупі A_1B_1, а найвищий – підгрупі A_2B_2, оскільки ці підгрупи містять відповідно кращі і гірші показники факторних ознак – продуктивності праці і витрат кормів на 1ц продукції. Різниця в рівнях собівартості тут становить 21,74 грн. (109,78 – 88,04).

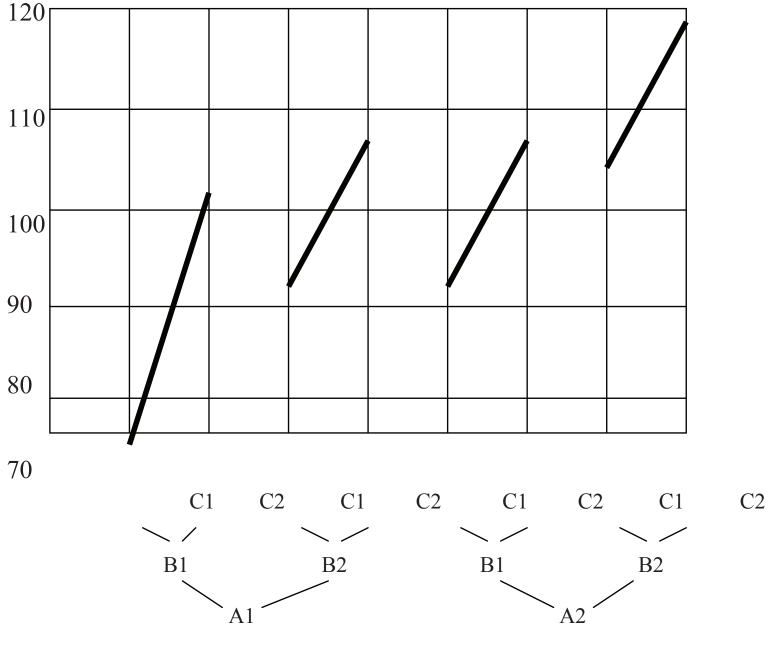


Рис. 19. Графічне зображення впливу факторів А,В,С та їх сполучень на результативну ознаку

Подвійні сполучення факторів А і В показують зміни результативної ознаки при різних рівнях факторів А і В. Так зіставлення груп підприємств, які мають однаковий рівень по фактору А (продуктивність праці), дозволяє визначити, як впливає на абсолютну величину результативної ознаки фактор В (витрати кормів на 1ц приросту). Різниця рівнів собівартості у даному випадку становить : $\mathrm{M}_{A_1B_1} - \mathrm{M}_{A_1B_2}$ =99,29 – 88,04= 11,25 грн.

Вплив сполучень факторів А і В показано графічно на рис.20, який свідчить, що достатньо фактору В змінитися, як при градації першого фактора $(A_1$ і $A_2)$ результативна ознака різко збільшується (рис. 20). Аналогічно проявляється і дія фактора А. При обох градаціях другого (B_1і B_2) зміни фактора А призводять до різкого збільшення значення результативної ознаки (рис. 20). Аналогічно можна порівнювати між собою рівні результативної ознаки, зумовлені впливом всіх можливих сполучень факторів в аналізованому дисперсійному комплексі.

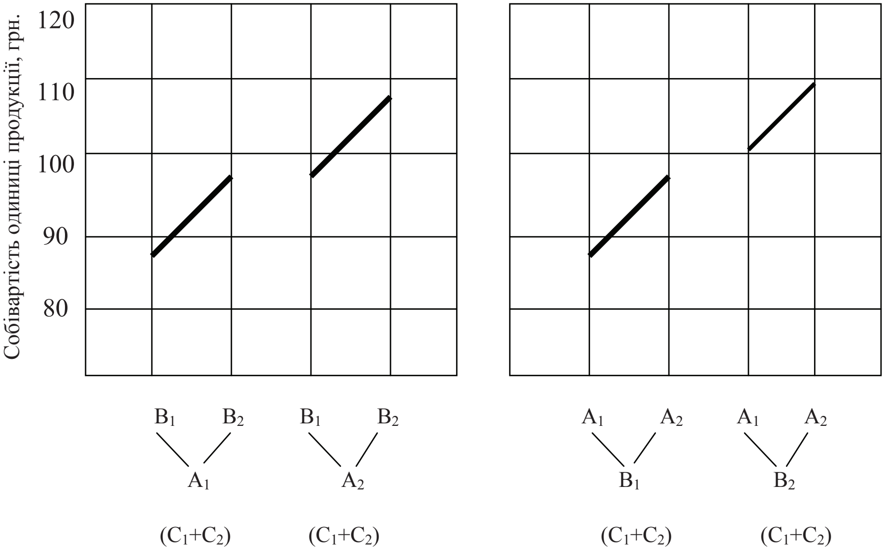


Рис.20. Графічне зображення впливу факторів А і В при усередненому рівні фактора С

Таким чином, за допомогою дисперсійного методу аналізу можна визначити не тільки частку дії досліджуваних факторів на результативну ознаку, але й абсолютну зміну останнього під впливом того чи іншого фактора і їх взаємодій.

Аналізуючи вплив факторів на результативну ознаку як окремо, так і різних їх поєднань, потрібно мати на увазі, що не виключені випадки, коли дія окремих факторів дуже мало впливає (або зовсім не впливає) на результативну ознаку, тоді як вплив їх взаємодії досить значний.

Це пояснюється це так. Вплив різних поєднань факторів, що вивчаються, помітним чином проявляється тільки у тих випадках, коли є різниця в дії одного фактора при різних градаціях іншого. Особливий вплив поєднань у дисперсійному комплексі проявляється тоді, коли при одній градації першого фактора другий діє дуже мало або навіть негативно, а при іншій градації – сильно і сприяє позитивному напряму у зміні результативної ознаки. Наприклад, вивчаючи прибутковість будь – якої галузі виробництва, можна виявити, що при одних, здавалось би, достатніх рівнях забезпеченості її технікою, робочою силою і т.д. – галузь збиткова, а при інших рівнях (які на перший погляд здаються недостатніми) спостерігається підвищення прибутковості. У зв’язку з цим виникає необхідність викривати і вимірювати ступінь впливу не тільки окремих факторів, але також і їх взаємодій, як частини загального сумарного впливу.

Оскільки завжди є деякі відмінності в діях одного фактора при різних рівнях (градаціях) іншого, сумарний вплив всіх врахованих факторів у кожній підгрупі дисперсійного комплексу складається з дій кожного фактора окремо і специфічного впливу їх поєднань.

Необхідно пам’ятати, що при аналізі відносних характеристик дисперсійної моделі виникає необхідність оцінки вірогідності одержаних різниць між частковими середніми. У даному випадку розраховується коефіцієнт вірогідності:

$$F_p = \frac{P_C}{\sigma_Z^2} \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} \text{ при } \nu_1 = 1;\ \nu_2 = \nu_Z,$$

де P_C – різниця між порівнюваними середніми комплексу;

σ_Z^2- залишкова дисперсія; n – число спостережень у порівнюваних групах комплексу ; $\nu_2 = \nu_Z$ – число ступенів вільності для залишкової дисперсії.

Як приклад визначимо вірогідність різниці у собівартості між групами $\mathrm{A}_2\mathrm{B}_2C_2$ і $\mathrm{A}_2\mathrm{B}_1C_1$. Вона становитиме : $\mathrm{M}_{\mathrm{A}_2\mathrm{B}_2C_2}$ – $\mathrm{M}_{A_2B_1C_1} =$ = 119,78 - 90,29 = 29,49. Підставляючи у наведену вище формулу відповідні дані, одержимо:

$$F_p = \frac{29{,}49^2}{51{,}77} \cdot \frac{10 \cdot 9}{10 + 9} = 79{,}63.$$

Знаходимо табличне значення F_T при $\nu_1 = 1;$ ν_2=58 (додатки 8, 9). Воно дорівнює 4,0 і 7,1 при рівнях імовірності відповідно 0,95 і 0,99 . Оскільки $F_p > F_T$, вирахована різниця визнається вірогідною. Звідси зниження собівартості виробництва яловичини залежно від рівня витрат кормів (В) і їх вартості (С) при постійному рівні продуктивності праці (А) потрібно визнати істотним. За таким же принципом встановлюється вірогідність різниці у будь – яких варіантах аналізованих факторів.

7.1.4. Можливості і обмеження застосування дисперсійного методу в статистико-економічному аналізі

Викладене вище не вичерпує можливостей дисперсійного аналізу. Знання його особливостей дозволяє безпосередньо оцінити вірогідність тих чи інших розрахунків при використанні методів статистичних групувань, кореляції, регресії. Особливо широкі можливості при оцінці множинних кореляційних залежностей (мова про них піде у наступній частині видання). Маючи порівняно невелике число одиниць спостереження, можна вводити у дисперсійний аналіз ряд ознак – факторів, обчислюючи випадкову помилку на достатньо великому числі ступенів вільності.

Зіставляючи кореляційні моделі з двома та більше змінними на невеликій сукупності об’єктів, за допомогою дисперсійного аналізу можна вирішити два дуже важливих питання: по-перше, в якому взаємозв’язку знаходяться включені в модель фактори, і, по – друге, чи будуть істотними висновки, зроблені на невеликій вибірці змінних. Неврахування цього положення відніме багато часу у пошуках істотних факторів – аргументів, а іноді навіть знецінює економічні дослідження.

Відмічаючи позитивні сторони дисперсійного аналізу, потрібно підкреслити, що він має ряд переваг, які вигідно відрізняють його від інших статистико – математичних методів. Назвемо головні з них. Використовуючи даний метод у багатофакторному аналізі економічних явищ, можна отримати картину, яка показує вплив кожного фактора у різних умовах, створюваних змінами різних факторів. При цьому застосування найрізноманітніших комбінацій факторів, що вивчаються, дає більш надійну основу для практичних рекомендацій, які залишаються придатними і при змінюваних умовах.

Аналізуючи економічні явища, де фактори інколи знаходяться у складному переплетінні, дисперсійний метод дозволяє об'єктивно пояснити складну картину, що виникає при такій взаємодії.

Разом з тим, потрібно пам'ятати про деякі обмеження дисперсійного аналізу. Так, суттєвим недоліком цього методу є те, що на результати досліджень впливає рівень показників підгруп (по досліджуваних факторах), що становить дисперсійний комплекс.

Отже, дисперсійні моделі, побудовані при одних рівнях факторних градацій, можуть мати вірогідний вплив, а при інших рівнях такий вплив відсутній. Одночасно потрібно наголосити, що результат оцінки по факторах залежить від того, як згруповані дані дослідження в статистичному комплексі.

Необхідно вказати і на обмеження у визначенні оцінки вірогідності впливу факторів. Якщо величина вирахуваного коефіцієнта F_p перебільшує його табличне значення F_T то вплив досліджуваного фактора вважається вірогідним, а якщо не перебільшує межу своїх випадкових коливань, то фактор не є суттєвим і не впливає на результат. Отже, не слід поспішати з висновком, оскільки причиною його невизначеності є недостатня кількість одиниць у вибірці для його переконливого підтвердження, а не різкий вплив факторів.

Інколи величина F_p може опинитись менше свого табличного значення не тільки через недостатньо різкий вплив фактора, що вивчається, або недостатню чисельність вибірки. Причиною може бути і те, що помилка кожного з показників, взятих окремо, дуже велика в результаті завищеної неоднорідності досліджуваних даних. Величину F_p (занижену) зумовлюють і властивості самих факторів, такі як функціональні і близькі до них зв'язки між факторами, використання в аналізі однорічних даних та ін. У результаті показники значно відрізняються від 0 або від 1, що збільшує їх можливі випадкові коливання. Це відбивається на величині їх помилки, а від останньої залежить значення розрахованого коефіцієнта F_p.

Поспішний висновок стосовно несуттєвості впливу фактора може тільки гальмувати подальші пошуки. Можливо цим і пояснюється переконання окремих дослідників відносно статистичної оцінки вірогідності дослідження взагалі. Недоказаність істотності впливу фактора повинна не стримувати, а навпаки, стимулювати подальші пошуки, покращення експеримента як у відношенні техніки обробки, так і підбору самого матеріалу. У такому випадку одержані позитивні результати стають ще більш неспростовними.

Щоб у дисперсійному аналізі мати об’єктивні результати, необхідно дотримуватись певних правил побудови (організації) дисперсійних комплексів. Якщо поділити групи на підгрупи (градації) таким чином, що в кожній з них рівні показників виявляться близькими по величині, а між групами різко різняться, то дисперсійний аналіз може призвести до негативної відповіді на питання про істотність досліджуваних факторів. Це є наслідком того, що у загальній кількості показників у середній групи буде багато таких із них, які мало відрізняються один від одного, що може погасити відмінності між іншими. Різкі ж відмінності між середніми груп ніби зникнуть у великій кількості подібних один до одного середніх.

Аргументуючи сказане, потрібно підкреслити, що дана обставина, обмежуючи можливості застосування дисперсійного аналізу у техніці, біології, тощо не так вже й небезпечна в галузі економіки. Тут оцінка в загальному і в цілому всіх відмінностей у характеристиках одиниць спостереження майже не має сенсу. В економічному аналізі дуже важлива оцінка відмінностей між кожною групою.

Із факту наявності у дисперсійному методі аналізу недоліків не випливає, що потрібно якось обмежити застосування цього методу в економіці. Мова йде не про обмеження, а про правильне використання його в аграрно – економічних дослідженнях, оскільки даний метод тільки у вказаному випадку є високоефективним. У цілому він повинен зайняти одне із перших місць серед інших статистико – математичних методів багатофакторного кількісного вивчення економічних процесів і явищ у будь – якій сфері людської діяльності.

§ 7.2. Кореляційно-регресійний аналіз

7.2.1. Загальнотеоретичні основи кореляційно-регресійного методу аналізу

Будь – яке явище природи і суспільства не може бути усвідомленим і зрозумілим без обґрунтування його зв’язків з іншими явищами. Щоб пізнати сутність явищ, необхідно розкрити їх взаємовідносини, кількісно визначити вплив тих або інших об’єктивних і суб’єктивних факторів.

Вплив цих факторів на рівень економічних показників в сільському господарстві до недавнього часу визначався в основному за допомогою методу статистичних групувань (цей метод буде розглянуто в темах загальної теорії статистики). Співвідношення ознак, виявлених в результаті статистичних групувань, відрізняються від співвідношень, які мають місце при функціональних зв’язках, коли кожному значенню аргумента відповідає визначене значення функції. Метод статистичних групувань дозволяє встановити тільки наявність зв’язку між явищами, не визначаючи при цьому його порівняльні кількісні параметри. Через це поряд з методом групувань, які відіграють винятково важливу роль у економічних дослідженнях, для рішення подібних питань необхідно застосовувати і інші методи, зокрема, метод кореляції. Термін «кореляція» вперше застосував Ж.Кюв'є в праці «Лекции по сравнительной анатомии» (1800-1805 рр.). Початкові математичні побудови методу кореляції були дані О.Браве у 1846 р. («кореляція» – від латинського «correlation» відношення, що означає співвідношення, відповідність предметів або понять).

Кореляцією називається неповний зв’язок між досліджуваними явищами. Це така залежність, коли будь - якому значенню однієї змінної величини може відповідати декілька різноманітних значень іншої змінної. Вона відображає закон множини причин і наслідків і є вільною неповною залежністю.

У дослідженнях важливо вивчати не стільки міру кореляції, скільки форму її і характер зміни однієї ознаки в залежності від зміни другої. Ці задачі розв’язуються методами регресійного аналізу. Перші спроби застосування цього методу в економіці були зроблені у кінці XIX і на початку XX століття) в Росії – роботи Е.Е.Слуцького, А.А.Чупрова, на Заході – роботи В.Парето, Гукера та ін.).

Кореляційний аналіз є свого роду логічним продовженням (розвитком) методу статистичних групувань, його поглибленням. Він допомогає вирішити цілий ряд нових завдань у економічному аналізі. Розрахунки на основі кореляційних моделей підвищують ступінь точності аналізу, часто виявляють недоліки попереднього аналізу. Перевага цього методу складається також і в тому, що він дає можливість розв’язувати задачі, які не можна вирішити за допомогою інших методів економічного аналізу – як, наприклад, розділ впливу багатьох факторів, які діють взаємопов’язано і взаємозумовлено.

Використання методу кореляції і регресії дозволяє вирішити такі основні завдання : 1) встановити характер і тісноту зв’язку між досліджуваними явищами; 2) визначити і кількісно виміряти ступінь впливу окремих факторів і їх комплексу на рівень досліджуваного явища; 3) на підставі фактичних даних моделі залежності економічних показників від різних факторів розраховувати кількісні зміни аналізованого явища при прогнозуванні показників і давати об'єктивну оцінку діяльності підприємств.

Відомо, що існує два типи залежності явищ: функціональний і кореляційний. При функціональному зв’язку зміна однієї ознаки чи показника на певну величину викликає за собою зміни другої ознаки чи показника на чітко визначену величину. Такого роду залежність в її чистому вигляді зустрічається в математиці, фізиці, хімії.

При кореляційній залежності будь - якому значенню однієї змінної величини може відповідати декілька чи навіть безліч різноманітних, тобто варіюючих значень іншої змінної величини.

Головна відмінність кореляційної залежності від функціональної полягає в тому, що функціональний зв’язок має місце в кожному окремому випадку спостереження, а кореляційний проявляється так само лише в середньому або в цілому для всієї даної сукупності спостережень і є неточним у відношенні окремих спостережень.

Кореляційний зв’язок величин полягає в тому, що при завданні однії з них встановлюється не одне точне значення, а ймовірності різноманітних значень іншої. Таким чином, залежність виявляється не між самими величинами, а між кожною з них і відповідним математичним очікуванням іншої.

Вивчення взаємозв’язків кореляційного типу має істотне значення особливо при аналізі явищ, які складаються під впливом великої кількості певних умов.

За своїми математичними особливостями кореляційні залежності можуть бути додатними і від'ємними, прямолінійними і криволінійними, простими і множинними.

Коли визначається зв'язок між двома ознаками, кореляція називається простою; якщо ж явище розглядається як результат впливу декількох факторів – множинною. За формою кореляційна залежність буває прямолінійною і криволінійною, за напрямком – прямою (додатною) і оберненою (від'ємною).

Необхідно підкреслити дві особливості, властиві кореляційному аналізу:

1) при використанні кореляційного методу вирішальне значення має всебічний, економічно усвідомлений попередній аналіз даних господарської діяльності. Слід пам'ятати, що зв'язок між ознаками і властивостями не є результатом математичних розрахунків, а лежить в природі самих економічних явищ і за допомогою методів математичної статистики можна лише виразити об'єктивно існуючі закономірності економічних процесів;

2) кореляцію можна виявити, лише досліджуючи достатньо велику сукупність спостережень, оскільки кореляційні зв'язки виявляються в формі спряженого варіювання двох або кількох зіставлених ознак.

Кореляційно - регресійний аналіз включає три етапи: 1) математико – економічне моделювання ; 2) рішення прийнятої моделі шляхом знаходження параметрів кореляційного рівняння (кореляційне рівняння, за первинною пропозицією англійського статистика – математика Ф. Гальтона, називають також рівнянням регресії); 3) оцінка і аналіз одержаних результатів.

Статистичне дослідження кореляційної залежності включає завдання визначення форми зв'язку і знаходження кількісної характеристики цієї форми. Процес встановлення форми зв'язку і вибору математичного рівняння, яке могло б найбільш повно відображати характер взаємозв'язку між ознаками досліджуваного явища, має вирішальне значення в кореляційному аналізі.

Питання вибору форми зв'язку та математичного рівняння можна вирішити на основі кількісного соціально – економічного аналізу явищ, що вивчаються, використовуючи при цьому такі методи статистичного аналізу, як графічний, статистичні групування, дисперсійний аналіз та ін .

При прямолінійному зв'язку збільшення факторної ознаки (x) викликає безперечне збільшення (чи зменшення) результативної ознаки (у) у середньому на певну величину.

Повну характеристику лінійного зв'язку можна одержати, користуючись критерієм лінійної кореляційної залежності акад. В.С.Немчинова[3]. Цей критерій являє таку схему:

1) $\overline{yx} = \overline{y} \cdot \overline{x} =$ повна відсутність лінійного кореляційного зв'язку;

2) $\overline{yx} \succ \overline{y} \cdot \overline{x}$ - прямий зв'язок між ознаками;

3) $\overline{yx} \prec \overline{y} \cdot \overline{x}$ - зворотний зв'язок між ознаками;

4) $\overline{yx} \cdot \overline{y} \cdot \overline{x} = \sigma y \cdot \sigma x$ - повна лінійна функціональна залежність.

У випадку, коли в кореляційному аналізі використовують групові середні, характер зв'язку між ознаками визначають за зміною останніх. Більш чи менш правильна систематична зміна їх від групи до групи свідчить про наявність прямолінійної залежності.

Показником тісноти зв'язку є лінійний коефіцієнт кореляції, величина якого визначається за такою формулою:

$$ч_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}.$$

Перетворення цієї формули призводить до вигляду: $r_{yx} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \sigma_y}$. Коефіцієнт кореляції коливається в межах від 0 ± 1.

7.2.2. Рівняння регресії, визначення його параметрів

Рівняння, що відображує зміну середньої величини однієї ознаки (у) в залежності від другої (х), називається рівнянням регресії або рівнянням кореляційного зв'язку.

При простій кореляції це рівняння має вигляд:

$$\overline{y}_x = \alpha_0 + \alpha_1 x,$$

де $\overline{y}_x$ − середнє теоретичне значення y при даному значенні x; $\alpha_0 \alpha_1$ - параметри рівняння.

Кореляційне рівняння пов'язує результативну ознаку з факторною у вигляді рівняння прямої лінії, де параметр α_1 визначає

[3] Немчинов В.С. Избранные произведения. Т.-2 – М.: Наука, 1967.- С.439

середню зміну результативної ознаки (y) при зміні факторної ознаки (x) на одиницю її натурального виміру.

Невідомі параметри α_0 і α_1 знаходять за способом найменших квадратів, який ставить умову, щоб сума квадратів відхилень y від аплікат $\bar{y}_x$, обчислених за рівнянням регресії, була найменшою, або, інакше кажучі, щоб при зображенні в прямокутній системі координат теоретична лінія регресії проходила б максимально близько до фактичних даних. Такій умові відповідає пряма, параметри якої є коренями системи нормальних рівнянь:

$$\begin{cases} a_0 n + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum xy. \end{cases}$$

Приклад. Розглянемо кореляційну залежність між затратами праці на виробництво одиниці продукції (у) і рівнем автоматизації процесів в 64 підприємствах.

За даними спостереження вирахуємо допоміжні величини (табл. 54). Підставивши в систему нормальних рівнянь замість буквених позначень обчислені сумарні значення, одержимо :

$$\begin{cases} 64\alpha_0 + 40{,}96\alpha_1 = 156{,}38; \\ 40{,}96\alpha_0 + 27{,}527\alpha_1 = 98{,}9720 \end{cases}.$$

Таблиця 54

Вихідні і розрахункові дані для обчислення параметрів кореляційного рівняння

Дані спостереження по 64 підприємствах		Розрахункові величини		
Затрати праці на одиницю, люд.-днів у	Показник рівня автоматизації х	ху	x^2	y^2
1,39	0,82	1,2398	0,6724	1,9321
1,54	1,00	1,5400	1,0000	2,37716
…	…	…	…	…
Всього 156,38	40,96	98,9720	27,5276	419,9248

Порівняємо коефіцієнт при невідомому a_0 і віднімемо з першого рівняння друге:

$$\begin{array}{r} \alpha_0 + 0{,}6400\alpha_1 = 2{,}4434 \\ -\ \alpha_0 + 0{,}6721\alpha_1 = 2{,}4163 \\ \hline -0{,}0321\alpha_1 = 0{,}0271 \end{array},$$

звідки

$$\alpha_1 = \frac{0{,}0271}{0{,}0321} = -0{,}8442.$$

Підставимо значення параметра

$\alpha_1 = -0{,}8442$ у перше рівняння, обчислимо значення параметра α_0:

$\alpha_0 + 0{,}6400(-0{,}8442) = 2{,}4434;\ \alpha_0 = 2{,}4434 + 0{,}5403;\ \alpha_0 = 2{,}9837$.

Таким чином, рівняння лінії регресії прийме такий аналітичний вигляд: $\overline{У_x} = 2{,}9837 - 0{,}8442x.$

Перевірка правильності рішення: $\overline{У} = 156{,}38 : 64 = 2{,}44;$

$\bar{x} = 40{,}96 : 64 = 0{,}64;\ \bar{y} = 2{,}98 - 0{,}84 \cdot \bar{x} = 2{,}98 - 0{,}84 \cdot 0{,}64 = 2{,}98 - 0{,}54 = 2{,}44.$

Для зручності інтерпретації виразимо x (факторну ознаку) в відсотках. У цьому випадку коефіцієнт при x зменшиться в 100 разів.

Таким чином, у загальному вигляді рівняння матиме вигляд:

$$\bar{y}_x = 2{,}98 - 0{,}008x.$$

Економічна інтерпретація даного кореляційного рівняння така: збільшення рівня автоматизації процесу на 1% призводить до зменшення витрат живої праці на одиницю продукції в середньому на 0,008 людино –дня.

7.2.3. Криволінійна регресія

При криволінійній формі зв’язку збільшення факторної ознаки призводить до нерівномірного збільшення (або зменшення) результативної ознаки, або ж зростання її величини змінюється спаданням, а зменшення – збільшенням.

Для визначення зв’язку між ознаками, взаємовідношення яких передбачає можливість існування оптимальних розмірів операцій, використовують рівняння параболи:

$$\bar{y}_x = \alpha_0 + \alpha_1 x + \alpha_2 x^2.$$

Одна з особливостей цього типу кривої та, що вона завжди має точку перетину (критичну точку), яка характеризує оптимальний варіант розміру величини результативної ознаки, і змінює напрямок свого руху лише один раз. Якщо в рівнянні величина α_1 виражена від’ємним числом, а α_2 - додатним, то крива змінюватиме напрямок спаду на зростання.

Для розрахунку параметрів рівняння параболи другого порядку використовується така система нормальних рівнянь:

$$\begin{cases} \alpha_0 n + \alpha_1 \sum x + \alpha_2 \sum x^2 = \Sigma y; \\ \alpha_0 \sum x + \alpha_1 \sum x^2 + \alpha_2 \sum x^3 = \Sigma xy; \\ \alpha_0 \sum x^2 + \alpha_1 \sum x^3 + \alpha_2 \sum x^4 = \Sigma x^2 y. \end{cases}$$

Приклад. Розглянемо залежність собівартості одиниці продукції від рівня спеціалізації виробництва (x, %) на прикладі 20 підприємств. Для рішення рівняння параболи підставляємо в систему нормальних рівнянь відповідні, попередньо розраховані, підсумкові дані. Отже, одержимо

$$\begin{cases} 20\alpha_0 + 1171{,}58\alpha_1 + 75489{,}78\alpha_2 = 2033{,}04; \\ 1171{,}58\alpha_0 + 75489{,}78\alpha_1 + 5180169{,}86\alpha_2 = 121585{,}39; \\ 75489{,}78\alpha_0 + 5180169{,}86\alpha_1 + 371409983{,}95\alpha_2 = 8001658{,}22. \end{cases}$$

Розв'язавши цю систему рівнянь, знаходимо значення параметрів:

$$\alpha_0 = 132{,}95; \ \alpha_1 = -1{,}79; \ \alpha_2 = 0{,}02.$$

Рівняння зв'язку у даному випадку має такий аналітичний вигляд : $\bar{y}_x = 132{,}95 - 1{,}79x + 0{,}02x^2$.

Застосування гіперболічної форми кореляційного зв'язку розглянемо на прикладі залежності рівня виробничих витрат на одиницю продукції (у) від об'єму її виробництва (х).

Відомо, що загальна сума затрат виробництва ділиться на дві групи. До першої відноситься сума витрат, яка залежить від кількості виробленої продукції. Ця група включає вартість використаної сировини і затрати на оплату праці. Друга група витрат містить в собі суму витрати, які не залежить або непрямо залежних від об'єму виробленої продукції . До них відносять амортизацію, поточний ремонт, інші основні і накладні витрати. Позначимо суму витрат першої групи, яка припадає на одиницю продукції, через Z_0, кількість виробленої продукції – через х. Загальна сума витрат цієї групи становить Z_0x. Суму витрат другої групи позначимо через Z_1. Тоді загальна сума витрат буде $xy = Z_0x + Z_1$. Щоб визначити рівень витрат, які припадають на одиницю продукції, потрібно отриманий вираз поділити на х. Ця залежність матиме вигляд: $У = Z_0 + Z_1\frac{1}{x}$.

Даний вираз нагадує рівняння гіперболи:

$$У_x = \alpha_0 + \alpha_1\frac{1}{x}.$$

Таким чином, залежність рівня витрат виробництва продукції від виробленої її кількості може бути виражена рівнянням двочленної гіперболи. Параметри цього кореляційного рівняння визначаємо за такою системою рівнянь:

$$\begin{cases} \alpha_0 n + \alpha_1\Sigma\frac{1}{x} = \Sigma У; \\ \alpha_1\Sigma\frac{1}{x} + \alpha_1\Sigma(\frac{1}{x})^2 = \Sigma\frac{1}{x}У. \end{cases}$$

Використовуючи попередньо розраховані підсумкові дані по 66 підприємствах, запишемо наведену систему рівнянь у вигляді:

$$\begin{cases} 66\alpha_0 + 111{,}8938\alpha_1 = 840{,}9200; \\ 111{,}8938\alpha_0 + 260{,}7022\alpha_1 = 1419{,}7429. \end{cases}$$

Розв’язавши ці рівняння, одержимо наступні числові значення параметрів: $\alpha_0 = 12{,}8826;\ \alpha_1 = 0{,}0834.$

Шукане кореляційне рівняння зв’язку собівартості одиниці продукції та об’єму її виробництва матиме вигляд:

$$\overline{y}_x = 12{,}88 + \frac{0{,}08}{x}.$$

Якщо інтервал зміни факторної ознаки значний, використовують рівняння тричленної гіперболи:

$$\overline{y}_x = \alpha_0 + \alpha_1 x + \alpha_2 \frac{1}{x}.$$

Система нормальних рівнянь у даному випадку буде наступною:

$$\begin{cases} \alpha_0 n + \alpha_1 \Sigma x + \alpha_2 \Sigma \frac{1}{x} = \Sigma y; \\ \alpha_0 \Sigma x + \alpha_1 \Sigma x^2 + \alpha_2 n = \Sigma xy; \\ \alpha_0 \Sigma \frac{1}{x} + \alpha_1 n + \alpha_2 \Sigma (\frac{1}{x})^2 = \Sigma \frac{1}{x} y \end{cases}.$$

У прикладі, який розглядається, ця система має вигляд:

$$\begin{cases} 66\alpha_0 + 51{,}2349\alpha_1 + 111{,}8938\alpha_2 = 840{,}9200; \\ 51{,}2349\alpha_0 + 52{,}5867\alpha_1 + 66\alpha_2 = 671{,}0019; \\ 111{,}8938\alpha_0 + 66\alpha_1 + 260{,}7022\alpha_2 = 1419{,}7429 \end{cases}.$$

Розв’язавши наведену систему рівнянь, одержимо числове значення параметрів: $\alpha_0 = 12{,}6407;$ $\alpha_1 = 0{,}8694;$ $\alpha_2 = 0{,}3388.$ Відповідно, кореляційне рівняння зв'язку матиме вигляд:

$$\overline{y}_x = 12{,}64 + 0{,}87x + \frac{0{,}34}{x}.$$

Для аналітичного виразу явищ, які відносяться до вивчення темпів росту, використовують рівняння експонентної кривої:

$\overline{y}_x = \alpha_0 \cdot \alpha_1^x,$ де α - фактор – аргумент (порядковий номер року, який в аналітичних рівняннях динаміки одержує значення 1,2, 3 і т. д.), α_0 - показник базисного року; α_1 – середньорічний темп зростання.

Невідомі параметри α_0 і α_1 у наведеній вище формулі визначають логарифмуванням, перетворивши показникову функцію в пряму: $Lgy = Lga_0 + xLga_1.$

Система нормальних рівнянь при цьому має вигляд:

$$\begin{cases} n \lg \alpha_0 + \lg \alpha_1 \Sigma x = \lg y; \\ \lg \alpha_0 \Sigma x + \lg \alpha_1 \Sigma x^2 = \lg y \Sigma x \end{cases}.$$

Приклад. Розглянемо зміну затрат праці на одиницю продукції (*у)* у підприємствах адміністративного району за десятирічний період

(1994-2003 рр.) Як відомо, показник затрат праці на одиницю продукції має завжди позитивне значення. При зниженні його рівня крива асимптотично (поступово) наближається до осі абсцис, але ніколи не може стати прямою, перетворитися в нуль або перетнути горизонтальну вісь (рис. 23), оскільки суспільство не може виробляти продукцію без затрат праці. У цьому випадку показники динаміки змінюються в геометричній прогресії.

Підставляючи попередньо розраховані сумарні показники для нашого прикладу в наведену вище систему нормальних рівнянь; одержимо:

$$\begin{cases} 10\lg\alpha_0 + 55\lg\alpha_1 = 3.77; \\ 55\lg\alpha_0 + 385\lg\alpha_1 = 20.44. \end{cases}$$

Розв'язавши цю систему рівнянь, одержимо логарифми числових значень невідомих параметрів : $\lg\alpha_0 = 0.39592;\ \lg\alpha_1 = 0.00347.$

Визначивши антилогарифми, знайдемо значення параметрів:
$\alpha_0 = 2{,}49;\ \alpha_1 = 0{,}99.$

Одержане рівняння зв'язку матиме аналітичний вигляд: $\overline{Y}_x = 2{,}49 \cdot 0{,}99^x$. Дане рівняння має таку економічну інтерпретацію: середні затрати праці на одиницю продукції в підприємствах району в нульовому періоді (1993 р.) досліджуваного відрізку часу складають 2,49 люд.-дня, а зниження їх в кожному наступному періоді становить 1 %.

В аналізі економічного явища часто використовують степеневу функцію виду $y = \alpha_0 x^{\alpha_1}$ Нелінійність відносно своїх констант зумовлюють її перетворення (шляхом логарифмування) в логарифмічно – лінійну функцію виду $\lg y = \alpha_0 + \alpha_1 \lg x.$

Таке перетворення дає можливість розв'язувати систему нормальних рівнянь методом найменших квадратів.

Застосовують логарифмічну лінійну функцію для явищ, характерних тим, що в міру приросту абсолютної величини факторної ознаки її вплив на результативну ознаку знижується. Для цього типу функції характерна пропорційність не абсолютних приростів (як для рівняння прямої лінії), а відносних приростів економічних показників, які вивчаються.

Якщо природа взаємовідношень економічних явищ така, що середня арифметична результативної ознака *(у)* пов'язана із середньою геометричною факторні ознаки *(х)*, то математичний вираз подібної залежності, тобто оцінку однієї змінної по другій, дає логарифмічна крива, гіпотетичне рівняння якої : $y = \alpha_0 + \alpha_1 \lg x.$

Відсутність числових значень логарифмів для нульових і від'ємних чисел обмежує можливості у використанні логарифміческих функцій в окремих випадках економічного аналізу.

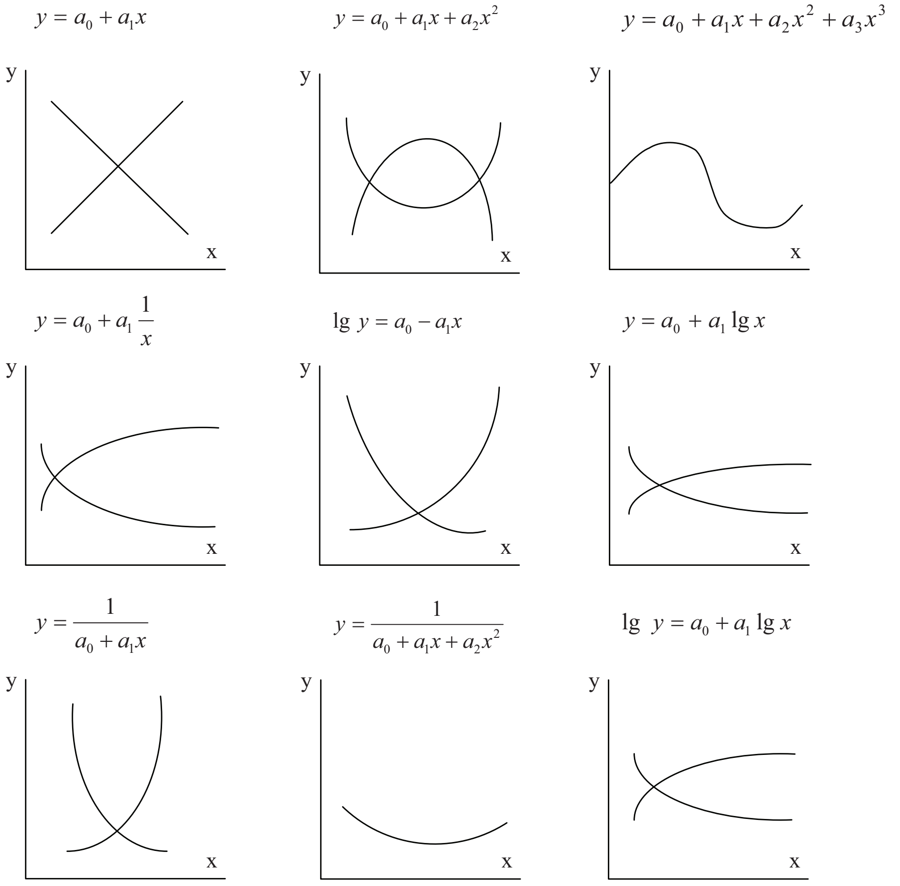


Рис 21. Лінії основних математичних функцій

На рисунку 21 представлені графічні зображення ліній основних математичних функцій.

Наявність випадкових факторів зумовлює ймовірнісний характер висновків про ступінь тісноти зв’язку. При цьому значення коефіцієнта кореляції, як і інших кореляційних характеристик, оцінюють на вірогідність. Суть такої оцінки міститься в зіставленні систематичної варіації результативної ознаки, зумовленої варіацією факторної ознаки, з випадкової варіацією (помилкою) результативної ознаки. З цією метою використовують критерії Ст’юдента і критерії Фішера.

Критерій Ст'юдента розраховується за формулою $t_p = \frac{|r|}{\mu_r}$,

де μ_r - середня помилка вибірки коефіцієнта кореляції, яка обчислюється за відношенням: $\mu_r = \frac{1 - r^2}{\sqrt{n-1}}$.

Одержане значення t_p зіставляють з його табличним значенням (додаток 1) і роблять висновки про вірогідність коефіцієнта кореляції. Його величина визнається вірогідною, якщо $t_p > t_т$.

У практичних розрахунках, для оцінки надійності коефіцієнта кореляції як правило, використовують таблиці вірогідних значень коефіцієнтів кореляції для відповідної кількості спостережень, тобто даного об'єму і рівня ймовірності (додаток 11).

Оскільки сам коефіцієнт кореляції є свого роду критерієм надійності досліджуваної залежності факторних і результативних ознак то висновок про його вірогідність може бути поширений на інші параметри кореляційно – регресійної моделі в цілому.

7.2.4. Множинна кореляція

До цих пір розглядалися моделі простої кореляції, тобто кореляційної залежності між двома ознаками. Проте в практиці економічного аналізу часто доводиться вивчати явища, які складаються під впливом не одного, а багатьох різних факторів, кожний з яких окремо може не справляти вирішального впливу. Сукупний же вплив факторів інколи виявляється достатньо сильним, щоб по їх змінах можна було робити висновки про величини показника досліджуваного явища. Методи вимірювання кореляційного зв'язку одночасно між двома, трьома і більше кореляційними ознаками створюють вчення про множинну кореляцію (питання множинної кореляції вперше досліджувались англійським вченим Ф.А.Еджвортом у кінці XIX ст.).

У моделях множинної кореляції залежна змінна «У» розглядається як функція кількох (*у* загальному випадку n) незалежних змінних «*x*».

Припущення про наявність лінійного зв'язку рівняння множинної регресії може бути показано в такому вигляді:

$$\overline{y}_{x_p x_n} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \ldots + \alpha_n x_n .$$

Із геометричної точки зору це рівняння визначає у просторі площини відповідних змінних $x_{1,} x_2, x_3, \ldots, x_n$ і *y*.

Множинне кореляційне рівняння встановлює зв'язок між досліджуваними ознаками і дозволяє вирахувати очікувані значення результативної ознаки під впливом включених в аналіз ознак – факторів, зв'язаних даним рівнянням .

Для оцінки ступеня тісноти зв'язку між результативною і факторними ознаками обчислюють коефіцієнт множинної кореляції. Величина його завжди додатне число, яке знаходиться в межах від 0 до 1.

У множинних кореляційно-регресійних моделях коефіцієнт простої кореляції між результативною ознакою і факторними, а також між самими факторними ознаками розраховують за формулами:

парні -

$$ryx_1 = \frac{\overline{yx_1} - \bar{y} \cdot \bar{x}_1}{\sigma_y \sigma_x}; \quad ryx_2 = \frac{\overline{yx_2} - \bar{y} \cdot \bar{x}_2}{\sigma_y \sigma_{x_2}} \quad rx_1x_2 = \frac{\overline{x_1x_2} - \bar{x}_1 \cdot \bar{x}_2}{\sigma_{x_1} \sigma_{x_2}};$$

часткові –

$$ryx_1 \cdot x_2 = \frac{ryx_1 - ryx_2 \cdot rx_1x_2}{\sqrt{1-r^2 yx_2}\sqrt{1-r^2 x_1x_2}}; \quad ryx_2 \cdot x_1 = \frac{ryx_2 - ryx_1 \cdot rx_1x_2}{\sqrt{1-r^2 yx_1}\sqrt{1-r^2 x_1x_2}};$$

Множинні (для двофакторної моделі):

$$Ryy_1x_2 = \sqrt{1-(1-r^2yx_1)(1-r^2yx_2 \cdot x_1)}.$$

Оцінку вірогідності множинного коефіцієнта кореляції (так як і кореляційного рівняння в цілому) одержують шляхом розрахунку F – критерію :

$$F_p = \frac{R}{P-1} : \frac{1-R^2}{n-P},$$

де P- кількість параметрів кореляційного рівняння.

Розрахункові значення F – критерію зіставляють з табличними (додатки 3,4). Якщо одержана величина F – критерію більше його табличного значення, коефіцієнт кореляції визнається вірогідним. Аналогічний висновок робиться по інших загальних характеристиках кореляційної моделі, таким як параметри рівняння, коефіцієнти детермінації та ін.

7.2.5. Загальнотеоретичні передумови застосування методів кореляційно-регресійного аналізу економічних явищ

Розгляд природи основних статистико – математичних методів і теоретичних передумов їх застосування в економічному аналізі почнемо зі стверджуючого припущення, що функціональні зв'язки в галузі економіки відсутні. Як уже було сказано раніше при функціональній залежності кожному конкретному значенню аргумента відповідає одне конкретне значенню функції. Такі залежності в абсолютно чистому вигляді демонструються абстрактними математичними формулами. В конкретних же економічних явищах, які зумовлюються множинністю причин, присутні неповні зв'язки або кореляційні. Їх ще прийнято називати статистичними. Багатозначність цих зв'язків породжується випадковими явищами.

Існують також підстави припустити, що економічні показники не обов'язково підпорядковуються закону нормального розподілу. Наприклад, відомі англійські вчені – статистики, які мають богатий досвід дослідження статистичних зв'язків, Е.Юл. і М. Дж. Кендел у частині закону нормального розподілу наводять слова К.Пірсона: «Цей закон не є загальний закон природи, ми повинні буквально полювати за подібними випадками»[4].У працях Е.Юла є твердження, що в економічній статистиці дуже важко підібрати навіть дві взаємопов'язані змінні, які характеризуються симетричним розподілом. Хоча потрібно сказати, що з проникненням статистики в галузь технологічних процесів (особливо промислового виробництва) відкрито немало розподілів, близьких до нормального.

Говорячи про природу кореляційно - регресійного методу, потрібно пам'ятати, що кореляційні розрахунки є чисто математичним прийомом, що зовсім не виявляють фізичну картину взаємозв'язків . Одержана на основі цього прийому числова оцінка зв'язків і залежностей інколи виявляється формальною, що показує лише поверхню явищ. Незнання цієї особливості методу веде за собою неправильне користування ним. Якщо до цього ж порушуються правила формування статистичних сукупностей, то дослідник потрапляє в полон логічних помилок, викликаних несправжньою кореляцією. На жаль, до цього часу завершеної теорії несправжньої кореляції не створено, незважаючи на те, що вона невід'ємна від природи кореляційного аналізу.

Кореляція в її формально – статистичному розумінні не розкриває причин зв'язку, а констатує лише його наявність, даючи оцінку сили і тісноти, встановлює ступінь вірогідності міркувань про наявність такого. Разом з регресійним кореляційний аналіз вирішує такі завдання: оцінка сили зв'язку і її кількісне вимірювання,

[4] Юл Е., Кендєл М.Дж. Теория статистики. М.: Росстатиздат, 1960.- С.225.

визначення форми зв'язку і реальності його існування. При вивченні економічних явищ дослідник, керуючись правилами кореляційно – регресійного аналізу, насамперед повинен виходити з економічного змісту досліджуваних залежностей. Лише після цього може бути встановлений їх причинно – наслідковий характер. Одержані результати розрахунків поширюються лише на ті об'єкти, кількісні характеристики яких включені в розрахунки. Звідси кореляційний аналіз повинен задовольняти вимогам об'єктивності на противагу формально – логічному підходу.

Приймаючи на озброєння методи кореляцій і регресій, необхідно обмежити дослідження від внесення в них викривлень, породжених суб'єктивною природою методів.Нерозуміння або недооцінка її, як правило, призводить до необґрунтованості висновків, суб'єктивізму в рішеннях, помилок в підборі одиниць спостережень і планування дослідження, а також до того, що досліджуваний зв'язок ставиться в залежність від обставин, що не мають до нього об'єктивного відношення.

Застосування теорії кореляції вимагає знання, насамперед, природи показників тісноти зв'язку.

Відомо, що в економічних дослідженнях твердо встановилась думка про можливість використання коефіцієнтів парної кореляції як свого роду критерію оцінки впливу відібраних факторів в парних і множинних моделях на результативну ознаку. Тобто, мова йде про те, що ряд економістів вважають високу абсолютну величину коефіцієнтів кореляції ознакою наявності сильного причинного зв'язку між явищами. Це методичне положення не завжди під собою має об'єктивну основу, природа двох змінних величин не виключає існування стохастичних зв'язків, які, полягають у тому, що можливі значення однієї змінної мають імовірності, які в змінюються залежно від значення, прийнятого іншої змінною. Проте останнє не вказує на наявність причинного зв'язку, хоча коефіцієнт кореляції може досягти при цьому значної величини.

У спеціальній літературі по теоретичній математиці про можливості трактування коефіцієнта кореляції як міри тісноти зв'язку говориться з обережністю. В економічній же літературі цього не спостерігається. Хоча майже всі автори – економісти повторюють слова математиків з теорії ймовірностей про необхідність обережного трактування величин коефіцієнтів кореляції, вони практично ігнорують це положення. Дійсно, в теорії імовірності коефіцієнт кореляції вводиться як параметр, дійсність величини якого вказує на наявність стохастичного зв'язку, але не визначає міри причинного зв'язку. Так, А. Хальд писав: «Визначивши коефіцієнт кореляції і перевіривши потім гіпотезу про нульову кореляцію, можна інколи довести існування стохастичного зв'язку між змінними. Проте необхідно підкреслити, що стохастична залежність не вказує з необхідністю на наявність функціонального зв'язку[5]. Коефіцієнт кореляції хоч і може вказувати на стохастичний зв'язок між x_1 і x_2, але при допомозі його не можна визначити, чи є велична x_1 причинно обумовленою величиною x_2, або x_2 – величиною x_1 , або ж їх зв'язок пояснюється тим, що обидві вони причинно зумовлені іншими факторами. Таким чином, і при значному коефіцієнті кореляції для визначення функціонального зв'язку потрібне додаткове дослідження. При подальшому дослідженні, яке передусім повинно ґрунтуватися на знанні особливості проблеми, регресійний аналіз часто грає важливу роль як засіб перевірки зроблених гіпотез»[6].

Іноді створюється помилкове враження присутності тісного стохастичного зв'язку і відсутності причинного між явищами стохастично і причинно незалежними. Про це також говориться у наведеній вище роботі А.Хальда: «В той час як стохастична незалежність може ховати причинний зв'язок, дві події можуть бути стохастично залежними, навіть якщо вони причинно (функціонально) незалежні»[7]. Тут мається на увазі той випадок, коли дві події стохастично і причинно незалежні, але кожна з них окремо залежить від третьої . У такому випадку дві події часто здаються стохастично залежними, якщо їх зв'язок з третьою не помічений.

Це ще раз підкреслює відмінності понять стохастичного і причинного зв'язку, а звідси і необхідність особливо старанного економічного усвідомлення зв'язків явищ, для визначення ступеня тісноти яких використовується коефіцієнт кореляції. Свою точку зору в цій частині А.Хальд визначив так: ... зміст, який може мати коефіцієнт кореляції, крім чисто описувального, залежність від знання особливостей походження зв'язку між величинами. Коефіцієнти кореляції можуть опинитися, таким чином, небезпечною зброєю при аналізі спостережень даних, оскільки вони можуть вести

[5] Під функціональним зв'язком А. Хальд в даному випадку розумів причинний зв'язок

[6] Хальд А. Математична статистика з технічними додатками : Пер. з англ. – М.: Вид-во іноземн. Літ., 1956 .- 584 с.

[7] Там само

до змішування стохастичного і функціонального взаємозв'язків і таким чином до помилкових висновків»[8].

Показники тісноти зв'язку (коефіцієнт кореляції, кореляційне відношення і та ін.), як уже було сказано, будуються для явищ, які прямо або непрямо піддаються дії складної комбінації взаємосплетених причин. Назвемо цей комплекс «умовами», в яких дане явище існує. Результати дії умов на явища висловлюються в формі хаотично змінюваних за напрямком і силою коливань величини явища біля певного рівня (постійного або змінюючого). Загальну характеристику комплексної дії умов дає пропонований теорією ймовірності показник середнього квадратичного відхилення (σ). Він таким чином вимірює не самі умови, а потужність їх впливу на явища. Якщо дослідженням охоплено два економічних явища, які мають сумісні (хоч би частково) умови, то ця спільність умов призводить до деякої подібності коливань обох явищ. З допомогою середньоквадратичного відхилення можна оцінити силу дії тієї частини умов, яка є загальною для даних явищ і порівняти її з загальною дією умов для аналізованих явищ. Ця логічна схема веде до природи показника тісноти зв'язку – кореляційного відношення.

Дослідник, який використовує в економічному аналізі показники тісноти зв'язку, повинен пам'ятати, що коефіцієнт кореляції являє собою тільки спрощений спосіб обчислення кореляційного відношення для випадку прямолінійного зв'язку. Оскільки природа показників зв'язку нероздільна із середньоквадратичним відхиленням, пізнавальна значимість показників зв'язку обмежена тими «умовами», які формують дане середньоквадратичне відхилення. Тому поширення висновків на другі недосліджені випадки (вище говорилось про їх можливу наявність) правомірно лише настільки, наскільки вивчені умови типові і повторюються в координатах простору і часу. Для економічних досліджень це обмеження може бути подолане знову таки шляхом проникнення у зміст самого показника зв'язку. А природа показника кореляції така, що він дає лише вихідну інформацію для висновків про причинні зв'язки. Якщо розподіл однієї з змінних кореляційної моделі не може бути охарактеризований за допомогою середньоквадратичного відхилення через слабку варіацію явищ, то в цьому випадку втрачає зміст вирахування коефіцієнта кореляції і кореляційного відношення.

[8] Хальд А. Математична статистика з технічними додатками : Пер. з англ. – М.: Вид-во іноз. Літ., 1956 .- С.585.

Такі випадки зустрічаються тоді, коли в комплексі причин, що формують варіацію, окремі з них проявляються у формі еволюторної або періодичної послідовності в часі або просторі. Якщо динаміка ряду не очищена від таких компонентів, вимірювання сили зв'язку втрачає зміст.

Отже, використання коефіцієнта простої кореляції як критерію оцінки вірності підбору факторів для моделі множинної регресії не завжди виправдане. Даний статистичний показник вимагає особливої обережності використання його в ролі критерію, оскільки взаємозв'язок одних і тих же факторів з урахуванням і без урахуванням впливу інших причин може проявитися по-різному. А парна залежність ігнорує дію інших факторів, приписуючи її повністю тільки одному. Тому найбільш методично обґрунтованим буде визначення не тільки парних, але й часткових коефіцієнтів кореляції. Часткова кореляція дозволяє виконувати більш глибоке дослідження зв'язків між явищами, даючи можливість виділити вплив в окремо конкретних причин на зміну величини результативної ознаки.

В економічних дослідженнях частковою кореляції майже не користуються, а обмежуються парною і множинної. Між тим природа часткового коефіцієнта кореляції розкриває дійсним зв'язок і взаємозалежність окремих факторів, який міг би виявитися затушованим при використанні лише коефіцієнтів парної кореляції. Так, при вивченні коливань врожайності картоплі в сільгосппідприємствах (101 господарство) було відібрано три фактори, які з економічної точки зору найбільш вагомо визначають варіацію рівня врожайності – це кількість внесених мінеральних добрив на одиницю площі, якість землі і рівень фондозабезпеченості (табл. 55).

Парні коефіцієнти, як видно з даних таблиці, вказали на наявність слабкого (0,227 – 0,276) кореляційного зв'язку врожайності з досліджуваними факторами. Проведена перевірка їх істотності при порозі ймовірності 0,99 підтвердила вірогідність факторів удобреності ґрунтів і фондозабезпеченості (додаток 11).

У цілому для моделі статистичні характеристики тісноти зв'язку, одержані на підставі парної кореляції, наводять на сумнів в частині категоричності їх трактування. Економічна природа даної залежності залишає бажати наявності більш тісного зв'язку досліджуваних змінних з результативним показником урожайності. Це побуджує продовжити дослідження вибраних факторів, виявити їх чистий вплив, розрахувавши коефіцієнти часткової кореляції в розрізі досліджуваних факторів. Напротивагу показникам парної кореляції, значення часткових коефіцієнтів кореляції (табл. 55) вказує на істотний тісний зв'язок факторів з результативною ознакою. Це відповідає висловленим теоретико – логічним припущенням.

Одержані коефіцієнти часткової кореляції дозволили виявити чистий вплив всіх розглядуваних факторів при постійності інших, що відповідає меті дослідження залежностей складних економічних явищ.

Таблиця 55

Характеристики парної, часткової і множинної кореляції факторів з показниками врожайності картоплі

Незалежні змінні (фактори)	Коефіцієнти кореляції		Коефіцієнти регресії	
	парної	часткової	парної	множинної
Внесено мінеральних добрив на 1 га ріллі, ц діючого речовини (x_1)	0,268	0,885	12,151	12,284
Якісна оцінка землі в балах (x_2)	0,227	0,778	0,819	0,958
Вартість основних виробничих фондів на 1 га ріллі, грн. (x_3)	0,276	0,882	0,084	0,098

7.2.6. Логіка побудови множинних кореляційно – регресійних моделей

Як було сказано, геометрична природа рівняння множинної регресії визначає положення в просторі площини відповідних змінних $x_1,x_2,x_3,\dots, x_n$ і y. Саме рівняння характеризує кількісний зв'язок між досліджуваними ознаками і дозволяє вирахувати очікувані значенням результативної ознаки під дією включених в аналіз ознак – факторів, пов'язаних даним рівнянням.

Найважливіше питання кореляційно – регресійного аналізу є підбір типу аналітичної функції при вивченні множинних зв'язків. Наскільки важливе це положення свідчить розгорнута ще в 60- х роках минулого століття нині триває дискусія. Наприклад, відомі американські економетрики Е.Хедді і Д.Диллон вважають, що для «… економічних явищ є неймовірним, щоб всім умовам найбільш відповідав один тип... функції. До того ж різні люди можуть привести в однаковому ступені обґрунтовані доведення на користь вибору того чи іншого типу функції». Деякі автори стверджують, якщо теоретично неможливо обґрунтувати тип функції, то це можна зробити емпірично, на основі графічного аналізу парних зв’язків . (Лукомський Я.І., 1961) Таке твердження треба вважати невірним. Економічні явища, як ніякі інші, взаємозв’язані. Отже, графічний аналіз парних зв’язків між функцією і аргументами мало, що дає для обґрунтування форми множинного зв’язку.

Окремі економісти і статистики пропонують використовувати для побудови кореляційних моделей, степеневу функцію виду: $y = \alpha_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \ldots \cdot x_n^{a}$.

Як аргумент виставляється те, що : зручною формою взаємозв’язку економічних показників є добуток показників. Це підтверджується всім комплексом існуючих формул. Показники норми амортизації, виробітку, рентабельності, ланцюгові аналітичні показники – отримують методом алгебраїчного множення. Поняття форми добутків полегшить також наступний аналіз змін норм виробітку впливу різних об’єктивних причин [3].

В.П.Хайкін та інші, наприклад, обґрунтовують застосування степеневих моделей тим, що при плануванні, головним чином, враховують найменші відносні відхилення фактичних значень від розрахункових, одержаних за рівнянням регресії.

Подібна аргументація непереконлива. Відносні відхилення можна одержати, не використовуючи логарифмічної лінійною форми зв’язку, наприклад, уявляючи вихідні дані у вигляді індексів, прийнявши при цьому гіпотезу про лінійну залежність між результативною ознакою і ознаками – факторами. Але справа в тому, що застосування кореляційного і регресійного аналізу ефективне лише тоді, коли модельована сукупність представлена широкою варіацією рівнів показників, що входять в модель. При використанні відносних чисел ця умова часто порушується.

Незважаючи на ряд переваг степеневої функції (простота лінеаризації, зручність інтерпретації і т.п.), вона має певні недоліки. Так , окремі ознаки – фактори, які входять вхідні в кореляційну модель (такі, як рівень механізації рівень рентабельності та ін.) можуть виявитися рівними нулю. Якби в даному випадку вибрати тип функції у вигляді добутку факторів, то вона в ряді випадків виявилася б рівної нулю, а рішення системи рівнянь було б пов'язано із значними обчислювальними труднощями. Так, при лінеаризації доводиться стикатися з тим, що в рівняннях ні один із членів не піддається логарифмуванню, оскільки lg 0= - ∞ .

Отже, математична природа кореляційних моделей свідчить про те, що функція повинна насамперед обґрунтовуватись економічно. Якщо цього зробити не можна, то тип функції визначається емпірично, тобто шляхом побудови ряду функцій і оцінки їх адекватності певними критеріями.

Е.Хедді і Д.Діллон у даному випадку вказують: «Для емпіричного дослідження недостатньо тільки логічно обґрунтовувати модель. Вона до того ж повинна піддаватися математичній обробці. Дослідник повинен піти на компроміс з теоретично ідеальною моделлю. По – перше, число розглядуваних окремих змінних повинне бути визначене з врахуванням як можливості одержання даних, так і наявності ресурсів для проведення розрахунків. По – друге, необхідно використати такий функціональний вираз, який статистично припустимий як при оцінці, так і при випробуванні істотності. Тому побудова економічної моделі даного економічного явища залежить від проблем, пов’язаних з одержанням даних їх статистичної оцінки» (Хедді Є. і Діллон Д., 1965).

Природа множинних кореляційних моделей і процес їх побудови зобов'язують враховувати об'єктивні особливості, наприклад, сільськогосподарських підприємств, де, як правило, розглядають дві групи факторів. Перша з них зв'язана із природними, друга - з матеріальними умовами, які є продуктом діяльності людей.

Фактор природніх умов суттєво впливає на процес виробництва, а разом з ним і на процес утворення вартості. Дія цього фактору підвищує або знижує рівень затрат на виробництво одного і того ж обсягу споживної вартості.

До другої групи відносять фактори, які випливають з виробничо господарської діяльності. Це, насамперед, напрям і рівень спеціалізації підприємств, розмір і структура виробничих фондів, об’єм виробництва і т.п. Дана група факторів істотно впливає на ефективність виробництва і тому повинна особливо враховуватися при кореляційному моделюванні.

Факторам – аргументам, відібраним для включення в регресійну модель, пред'являються також, крім основної вимоги про відображення об’єктивних особливостей сільськогосподарських підприємств, також і деякі інші вимоги. Насамперед фактори повинні бути кількісно вимірні, оскільки кореляційні формули за своєю природою відображають зв’язок тільки між кількісно визначеними ознаками. У випадку включення в кореляційну модель якісних ознак їм необхідно надати кількісну визначеність. Це може бути зроблено, наприклад, за допомогою бальної оцінки, шляхом присвоєння рангів і т.д.

Проблема аналізу ще більш ускладнюється, коли якісно варіює залежна змінна величина. Які б перетворення при цьому не застосовувалися, природа кореляції орієнтує тут лише на наближеність у відображенні досліджуваних залежностей.

Природа кореляції вимагає, дотримання умови, обов'язкової для підбору залежної і незалежної змінних. Жодна із змінних не повинна знаходитись у функціональній залежності від іншої, або їх групи. З одного боку, ця вимога випливає з того, що немає сенсу шукати кореляційну залежить там, де заздалегідь відомо існування функціональної залежності. З іншого боку, при існуванні функціональних зв'язків між включеними в кореляційну модель показниками, які утворюються в ході вирішення моделі, система нормальних рівнянь може вийти погано або в зовсім не обумовленою, а одержані результати – ненадійними.

Необхідно звернути увагу ще на один момент методологічного порядку. При побудові кореляційних моделей у них не можна включати групу факторів, лінійна комбінація яких дорівнює постійній величині або близька до неї. У цьому випадку система нормальних рівнянь для визначення коефіцієнтів регресії або не має рішення, або його одержують в результаті випадкових відхилень. У подібних випадках, якщо парний коефіцієнт кореляції між двома ознаками – факторами перевищує 0,8 (з певним довірчим рівнем), то включати в кореляційну модель можна один з факторів.

Неможливо не відмітити, що відбір вихідних даних для розрахунків кореляційного аналізу вимагає великої уваги і обережності. Справа в тому, що, з одного боку, надійність кореляційних формул безпосередньо залежить від обсягу статистичної сукупності. Адже в основу кореляційних розрахунків покладено усереднення - усереднюються як характер впливу кожного врахованого фактора на залежну змінну, так і загальний вплив решти, неврахованих причин. Загальновідомо, що середні тим надійніші, чим за більшим обсягом даних вони розраховувались.

З іншого боку, включення в кореляційну модель додаткових даних, якщо воно було зроблено без належного якісного відбору, може призвести до того, що формулою неможливо буде користуватися. Відомо, що середні лише тоді мають реальний економічний зміст, коли вони ґрунтуються на якісно однорідному матеріалі. Теорія середніх величин вчить нас застосовувати їх для кількісної характеристики тільки однорідної сукупності.

Як правило, економічні явища складуються під дією багатьох факторів. Але бажання враховувати їх у кореляційній моделі в можливо більшій кількості досить рідко себе виправдує. Така кореляційна модель занадто громіздка, причому вплив великої частини факторів виявляється статистично неістотним.

Таким чином, природа кореляції і регресії вводить певні обмеженні в частині практичного використання кореляційно – регресійного методу в аналізі соціально – економічних процесів. Одержання вірогідних висновків за результатами кореляційно – регресійного аналізу можливе тільки при дотриманні певних, вимог. Останні випливають із самої природи кореляції. Назвемо основні з них: визначеність характеру залежності (прямолінійної, криволінійної), статистична однорідність досліджуваної сукупності, кількісний вимір ознак, достатній обсяг вибірки.

Інколи дослідники з метою одержання корисної практичної інформації намагаються виявити залежності в ідеальному їх вигляді, коли дуже високі коефіцієнти кореляції. В результаті має місце така серйозна помилка: одночасно розглядається дуже велика кількість факторів, причому деякі з них тісно пов'язан між собою. Зміна одного фактора в такому випадку безумовно викличе зміну іншого, в результаті чого важко відокремити чистий вплив одного фактора від впливу іншого і задовільнити природу, на якій ґрунтується теорія множинної кореляції . Через це введення в аналіз великої кількості факторів з метою вивчення їх впливу на результативну ознаку іноді зовсім не так доцільно, як це здається з першого погляду. Методологічна буде більш правильним відбирати ті з них, які є основними.

Для успішного практичного використання кореляційних моделей як об'єктивного критерію найкращого рівняння зв'язку можуть бути використанні коефіцієнт множинної кореляції і стандартна помилка оцінки за рівнянням множинної регресії при задовільній економічній інтерпретації самої моделі множинної регресії. Зокрема, напрям і сила впливу окремих факторів на залежну змінну, яка характеризується параметрами рівняння, повинні відповідати емпіричним уявам про цей вплив, тобто крім підтвердження рівня значимості спостережуваної взаємозалежності статистичними методами необхідно ретельно вивчити її логічну обґрунтованість.

Враховуючи, що взаємодія одних і тих же факторів з врахуванням і без врахування впливу інших причин може виявлятися по - різному, всілякі висновки про можливу форму зв'язку у багатофакторній моделі, зроблений на підставі аналізу парних залежностей, не повинні трактуватися як абсолютно вірогідності, до них необхідно відносити дуже обережно. У цьому відношенні переваги віддаються методу часткової кореляції.

Застосування даного методу в економічному аналізі носить у відомій мірі умовний характер. Причини, які впливають на досліджувані явища в галузях народного господарства, дуже різноманітні. Тому необхідно завжди пам'ятати, якщо виключити вплив одного фактора ті що залишилися несуть на себі дію ряду інших умов, не врахованих в дослідженні. І все ж таки застосування методу часткової кореляції має важливе значення при поглибленому аналізі множинної кореляції.

Існують й інші принципові (і дискусійні) питання теорії, які орієнтують на правомірність використання методу кореляційно – регресійного аналізу соціально - економічних явищ. Розгляд їх виключено, у зв'язку з метою і завданнями даного посібника. Тезисність викладення окремих питань даного розділу обумовлена тими ж причинами.

МОДУЛЬ 4

ТЕМА 8. АНАЛІЗ ІНТЕНСИВНОСТІ ДИНАМІКИ

§ 8.1. Статистичні ряди динаміки, основні правила їх побудови

Явища суспільного життя знаходяться в постійних змінах і розвитку як у просторі, так і у часі. Одне з основних завдань статистики полягає в дослідженні процесів змін і розвитку явищ у часі, тобто вивчення процесу їх розвитку. Числові дані, які характеризують такі процеси і явища, утворюють ряди динамік (інколи їх називають динамічними, хронологічними або часовими рядами).

Рядом динаміки у статистиці називається ряд чисел, який характеризує зміну величини суспільного явища у часі. Це ряд послідовно розташованих у хронологічному порядку значень показника, який у своїх змінах відображує хід розвитку досліджуваного явища.

Кожний ряд динаміки складається з двох елементів : 1) **ряду рівнів**, які характеризують величину явища, його розмір; 2) **ряду періодів** або моментів часу, до яких належать рівні ряду. Обидва елементи називаються членами ряду динаміки.

Прикладом ряду динаміки можуть бути дані, наведені в таблиці 56.

Таблиця 56

Виробництво продукції підприємством

Рік	1999	2000	2001	2002	2003	2004
Виробництво, тис. грн.	1800	2100	2400	2200	2300	2350

Ряди динаміки дають матеріал для аналізу розвитку соціально – економічних явищ і процесів. Приклади їх використання можна знайти в різних сферах економічної роботи. Значення рядів динаміки зростає якщо вони ведуться неперервно на протязі довгого часу. Дослідження такого ряду дає можливість вивчати процес розвитку явищ, виявити основні тенденції і закономірності цього розвитку.

Статистичні дані, що входять до складу рядів динаміки повинні бути порівняними між собою. Використання їх в аналізі передбачає попередню ретельну перевірку і перерахунки. Слід підкреслити, що побудова ряду динаміки передбачає дотримання певних вимог.

Розглянемо їх.

1. Усі показники ряду динаміки повинні бути вірогідними, точними, науково обґрунтованими. Якщо в ряду показників є хоча б один не правильно обчислений, то порівняння з ним призведе до помилкових висновків.

2. Показники ряду динаміки повинні бути порівнянні за змістом, тобто вони повинні обчислюватися за єдиною методологією (наприклад, виробництво валової продукції подається в єдиних порівнянних цінах).

3. Порівнянність за територією, до якої належать показники ряду динаміки. Така вимога враховується у випадках, коли протягом періоду, яким охоплено ряд динаміки, відбувалися зміни меж району, області і т.д. Наприклад, у 2004 році до території адміністративного району була приєднана частина земель сусіднього району. В такому випадку при побудові ряду динаміки певного явища за попередні роки робиться поправка на приєднану територію.

4. Порівнянність у часі. За цією умовою (вимоги) показники ряду динаміки повинні бути обчислені в однакові періоди часу або на одну й ту ж саму дату. Якщо дана вимога не витримується, здійснюють відповідні перерахунки. Наприклад, не можна будувати ряд динаміки за 1994 – 2004 рр., якщо є інформація за 1994 – 2003 рр. про валове виробництво м'яса за рік, а в 2004 р. – за 10 місяців . У даному випадку беруть дані кожного року за 10 місяців, або будують ряд після завершення 2004 року.

5. Порівнянність рядів динаміки за колом охоплюваних ними об'єктів (наприклад, фермерських господарств). Так, якщо до 1995 року у складі району налічувалося 300 фермерських господарств, потім приєдналося ще 116 господарств, то при побудові ряду динаміки за 2002 - 2004 рр., необхідно всі показники обирати, виходячи із складу фермерських господарств до 1995 року, тобто по 300 господарствах.

Крім зазначених вище вимог, без урахування яких неможливо побудувати ряд динаміки, необхідно дотримуватися одних і тих самих одиниць виміру. Не можна також, в одному ряду динаміки поєднувати періоди і моменти часу. Наприклад, середньорічну чисельність працівників на підприємстві не можна порівнювати з їх чисельністю на початок місяця, року і т.д. За своєю сутністю всі вимоги зводяться до однієї: показники ряду динаміки повинні бути порівнянними між собою в усіх відношеннях.

§ 8.2. Види рядів динаміки, їх аналітичні показники

Залежно від реєстрації фактів ряди динаміки бувають дискретними і неперервними.

Дискретні ряди містять дані, одержані через певні проміжки часу (місяць, квартал, рік і т.д.). Слід розглядати три види дискретних рядів динаміки: моментні, інтервальні (періодичні) і ряди середніх.

Моментні ряди динаміки - це ряди статистичних величин, які характеризують розміри досліджуваного явища на певний момент часу (на початок місяця, кварталу, року і т.д.). До таких показників відносяться чисельність працюючих, парк автомобілів, поголів'я худоби та інші. Характер цих показників такий, що їх величини можна визначити лише на той чи інший момент часу.

Інтервальні ряди динаміки характеризують розміри досліджуваного явища за певні проміжки – інтервали (періоди) часу, тобто характеризують процеси за той чи інший період часу (добу, місяць, рік і т.д.) . До таких показників відносяться обсяг виробленої продукції, фонд заробітної плати та інші.

Ряди середніх характеризують зміну середніх рівнів досліджуваного явища у часі (наприклад, місячний рівень заробітної плати, денна продуктивність праці та ін.).

Неперервні ряди динаміки одержують у випадках, коли відбувається неперервний запис змін явища за допомогою відповідних приладів (механічних, електричних, електронних). Методика статистичного аналізу рядів динаміки базується на передбаченні неперервності досліджуваних процесів. Але перешкодою тут є обчислювальні труднощі, і тому неперервні ряди дискретизують, і результати аналізу виводять на підставі дискретних (перервних) послідовностей.

Залежно від виду узагальнених показників ряди динаміки можуть бути представлені абсолютними, відносними і середніми величинами. До перших відносяться показники земельних площ, валової продукції, виробництва автомобілів та ін., до других – питома вага площі окремих культур у загальній площі посіву , коефіцієнти зростання (наприклад, цін), показники виконання планових завдань тощо; до третіх – собівартість одиниці продукції, продуктивність тварин , заробітна плата одного працівника та ін.

Різний характер моментних і інтервальних показників зумовлює певні особливості відповідних рядів динаміки. Так, рівень інтервального ряду залежить від тривалості періоду часу, який він характеризує. Величина інтервального рівня тим більша , чим більша тривалість періоду. Рівні моментних рядів динаміки не залежать від проміжку часу між датами. Підсумок рівнів інтервального ряду дає результат за більш тривалий період часу. Так, від декадних рівнів можна перейти до щомісячних, від щомісячних – до квартальних, від останніх – до річних і т. д.. Іноді шляхом послідовного додавання рівнів інтервального ряду одержують нагромаджувальні підсумки за певний період (наприклад, нараховано заробітної плати, відпрацьовано людино – годин, оброблено площі тощо). Підсумовування рівнів моментального ряду динаміки само по собі не має змісту , адже одержані величини не мають економічного значення.

При аналізі рядів динаміки користуються рядом статистичних показників , які визначають характер, напрямок, і інтенсивність кількісних змін явищ. До таких показників відносяться : рівень ряду , середній рівень , абсолютний приріст, коефіцієнт (темп) зростання , темп приросту , абсолютне значення одного проценту приросту.

Рівнем ряду є кожен член ряду динаміки. Перший показник ряду називається початковим рівнем ,останній – кінцевим рівнем . Часто виникає потреба у визначенні середнього рівня динаміки. При цьому для інтервального ряду і рядів середніх величин середній рівень розраховується як середня арифметична проста з окремих рівнів : $\bar{y} = \frac{\sum y_i}{n}$.

Дещо інакше розраховується середній рівень моментного ряду динаміки. При його обчисленні виходять з розгляду найпростішого випадку, коли є дані на початок і кінець будь – якого періоду . Середній рівень у такому випадку визначають як середню арифметичну просту з цих двох показників. У моментних рядах кожен рівень можна розглядати як показник, що відноситься одночасно до початку одного і закінчення іншого періоду . Розглянемо приклад з моментним рядом динаміки про чисельність працюючих на підприємстві на 1 січня.

Роки	2000	2001	2002	2003	2004
Чисельність	789	811	837	860	901

За наведеними вихідними даними середньорічна чисельність працюючих становить, чол:

у 2000 р.: $\frac{789+811}{2}=800$;

у 2001р.: $\frac{811+837}{2}=824$;

у 2002 р.: $\frac{837+860}{2}=848$

Для кількох років середній рівень визначають за середньорічними як середню арифметична з них . Для нашого прикладу середня чисельність за 3 роки становила:

$$\frac{800+824+848}{3}=824 \text{ (чол.) },$$

або підставивши безпосередньо рівні моментного ряду, одержимо:

$$\frac{\frac{789+811}{2}+\frac{811+837}{2}+\frac{837+860}{2}}{3}=\frac{\frac{789}{2}+811+837+\frac{860}{2}}{3}=824 \text{ (чол.)};$$

Як бачимо, для розрахунку середньої чисельності за три роки використовується чотири рівні моментного ряду, з яких перший і останній беруться в напівсумі.

Отже, у загальному вигляді розрахунок середнього рівня для моментного ряду, який має *n* рівнів можна виразити формулою:

$$\bar{y}=\frac{\frac{y_1}{2}+y_2+y_3+...+\frac{y_n}{2}}{n-1} \text{ або } \bar{y}=\frac{\frac{y_1+y_n}{2}+\sum_{n-2}^{n-1}yi}{n-1}.$$

Одержана середня відома в статистиці як середня хронологічна.

Слід відзначити, що застосована формула буде вважатися правомірною у випадках, коли ряди мають рівні проміжки часу між датами (моментами) до яких відносяться рівні ряду. Якщо інтервали між датами неоднакові, середню хронологічну розраховують як середню арифметичну зважену, прийнявши за вагу відрізки часу між датами.

Середній рівень ряду динаміки дає узагальнюючу характеристику окремих рівнів ряду, які варіюють навколо нього. Визначають варіацію за допомогою вже відомих статистичних характеристик – середнього квадратичного відхилення (σ) і коефіцієнта варіації(ν):

$$\sigma=\sqrt{\frac{\sum(y_1-\bar{y})^2}{n}};\ \ v=\frac{\sigma}{y}\times100.$$

Коефіцієнт варіації використовується в аналізі рядів динаміки як відносний показник, що характеризує варіацію у кількох рядах.

Середній рівень ряду і зазначені вище показники варіації дають узагальнюючі характеристики рядам динаміки, але не дозволять визначити напрямок і розмір змін рівнів ряду в часі. Таке завдання вирішується за допомогою аналітичних показників ряду динаміки: абсолютного приросту, коефіцієнту (темпу) зростання, темпу приросту, абсолютного значення 1 % приросту.

Абсолютний приріст (А) – це абсолютна величина розміру змін досліджуваного явища, характеризується вона різницею між двома рівнями ряду динаміки. Абсолютні прирости можуть бути *базисними і ланцюговими*. *Базисні* визначають як різницю всіх рівнів ряду до одного, прийнятого за базу порівняння. *Ланцюгові* абсолютні прирости одержують як різницю наступного і попереднього рівнів. Величина абсолютного приросту показує на скільки одиниць рівень одного періоду більше або менше будь – якого іншого (як правило, попереднього) періоду, а отже він може мати знак «+» чи «-».При базисному способі обчислення абсолютних приростів (А) рівнів ряду динаміки (У) маємо : $A_1 = У_1 - У_0$; $A_2 = У_2 - У_0$; $A_n = У_n - У_0$.При ланцюговому способі обчислення: $A_1 = У_1 - У_0$; $A_2 = У_2 - У_1$; $A_n = У_n - У_{П-1}$.

Коефіцієнт зростання (К)- це відношення наступного рівня до базисного або попереднього. Він показує, у скільки разів рівень даного періоду більше або менше будь – якого рівня, прийнятого за базу порівняння. Коефіцієнт зростання, виражений у відсотках, називають **темпом зростання**. Залежно від мети дослідження за базу порівняння може прийматися постійний для всіх рівень ряду або кожний той, що передує йому. Розглянемо схематично розрахунок базисних (а) і ланцюгових (б) коефіцієнтів зростання:

а) $K_1 = \frac{У_1}{У_0}$; $K_2 = \frac{У_2}{У_0}$; $K_n = \frac{У_n}{У_0}$;

б) $K_1 = \frac{У_1}{У_0}$; $K_2 = \frac{У_2}{У_1}$; $K_n = \frac{У_n}{У_{n-1}}$.

Темп приросту (Т) – це відношення абсолютного приросту до попереднього або початкового (базисного) рівня, виражене у відсотках: $T_1 = \frac{A_1}{У_0}100$; $T_2 = \frac{A_2}{У_1}100$; $T_n = \frac{A_n}{У_{n-1}}100$.

Якщо база порівняння постійна, наведені формули мають вигляд:

$$T_1 = \frac{A_1}{У_0}100 \; ; \quad T_2 = \frac{A_2}{У_1}100 \; ; \quad T_n = \frac{A_n}{У_0}100 \; .$$

Коефіцієнти зростання і темпи приросту знаходяться у такому співвідношенні : $T_n = K_n \times 100 - 100$; $K_n = \frac{T_n + 100}{100}$.

Абсолютне значення 1% приросту являє собою частку від ділення абсолютного приросту на відповідний показник темпу приросту: $n_1 = \frac{A_1}{T_1}$; $n_2 = \frac{A_2}{T_2}$; $n_n = \frac{А_n}{T_n}$.

Дана величина являє собою соту частину попереднього рівня.

Розрахунок розглянутих вище аналітичних показників схематично наведено у таблиці 57.

Таблиця 57

Розрахунок аналітичних показників ряду динаміки

Показники ряду	Символи	Роки				
		2000	2001	2002	2003	2004
Рівень	у	789	811	837	800	860
Абсолютний приріст	$A_1 = y_1 - y_0$	-	22	26	-37	60
Коефіцієнт зростання	$K_1 = \frac{У_1}{У_0}$	-	1,028	1,032	0,056	1,075
Темп приросту	$T_1 = \frac{A_1}{У_0} \times 100$	-	2,8	3,2	-4,4	7,5
Абсолютне значення 1% приросту	$n_1 = \frac{А_1}{T_1}$	-	7,9	8,1	8,4	8,0

Для наведення аналітичних показників ряду динаміки в свою чергу можна розраховувати узагальнюючі показники у вигляді середніх величин. Так, за даними абсолютних приростів розраховують **середній річний абсолютний приріст**, як середню арифметичну просту:

$$\overline{А} = \frac{А_1 + А_2 + ... + A_n}{n} \; .$$

Із індивідуальних коефіцієнтів зростання, розрахованих за ланцюговим способом, середній коефіцієнт зростання обчислюють за, формулою середньої геометричної, тобто $\overline{K} = \sqrt[n]{K_1 \times K_2 \times ... \times K_n}$.

Середній коефіцієнт зростання можна розрахувати і за рівнями руду динаміки: $\overline{K} = \sqrt[n-1]{\frac{У_n}{У_0}}$, тобто за крайніми членами ряду динаміки. Із природи наведеної формули видно, що при однакових крайніх рівнях динаміки, але з різним характером змін у ньому можна одержати один і той же середній коефіцієнт зростання. Тому, якщо аналізують довгі і неоднакові за характером змін періоди, їх обов'язково дроблять на частини, для яких розрахунок середніх матиме зміст.

Крім розглянутих показників, поглиблений статистичний аналіз ряду динаміки передбачає визначення таких кількісних характеристик як автоковаріація, автокореляція і тренд.

Автоковаріація – це математичне очікування (середня арифметична) добутків відхилень рівнів ряду, зрушених між собою на період, *L* від середнього рівня. Кількісне вимірювання автоковаріації здійснюють за формулою:

$$C_x(L) = E\left[\left(x_l - \bar{x}\right)\left(x_{i+l} - \bar{x}\right)\right]$$

де $C_x(L)$ автокореляція; E - оператор математичного очікування;

L – лаг (часове зрушення); L = 1,…, T.

При L= 0 одержують дисперсію ряду динаміки: $C_x(0) - E\left[(x_i - \bar{x})(x_i - \bar{x})\right] = \sigma_x^2$.

Функція, що складається із значень $C_x(L),$ називається автоковаріацією і, як правило, задається таблицею.

Автокореляція це зв'язок між рівнями ряду динаміки. Тісноту такого зв'язку можна визначати через показники автокореляції, а саме $R_l \dfrac{C_x(L)}{C_x(0)} = \dfrac{C_{x[}(L)}{\sigma_x^2}$.

Тренд – це тенденція розвитку явища (див. вище).

В рядах динаміки, які несуть інформацію про економічні явища, спостерігається тенденція розвитку трьох видів : 1) тенденція середнього рівня; 2) тенденція дисперсії; 3) тенденція автокореляції.

Тенденція середнього рівня – аналітично виражається за допомогою математичної одиниці, навколо якої варіюють фактичні значення досліджуваного явища. У цьому разі значення тренда в окремі моменти часу є математичним очікуванням ряду динаміки. Тенденцію середнього рівня ще називають детермінованою компонентою досліджуваного явища і зображують у вигляді формули:

$$x_t = f(t) + Et$$

де x – рівень ряду динаміки в момент часу t ;

f(t) – детермінована компонента (аналітична функція);

Et – випадкова компонента

Тенденцію середнього рівня легко уявити у вигляді графіка.

Тенденція дисперсії являє собою тенденцію змін відхилень між емпіричними значеннями і детермінованою компонентою ряда. Цей вид тенденції також легко зображується графічно.

Тенденція автокореляції – це тенденція змін зв'язку між окремими рівнями ряду динаміки. Графічно такі зміни прослідкувати неможливо. Враховують їх у прогнозних розрахунках.

Щоб визначити тенденцію в русі показників динаміки використовують різні статистичні прийоми (методи), про них мова піде нижче.

ТЕМА 9. АНАЛІЗ ТЕНДЕНЦІЙ РОЗВИТКУ ТА КОЛИВАНЬ

§ 9.1. Прийоми аналітичного вирівнювання рядів динаміки

Ряди динаміки, рівні яких впродовж тривалого часу не змінюються зустрічаються досить рідко. Як правило, рівні ряду з часом змінюються, коливаються. Ці коливання для різних явищ неоднакові і можуть бути зумовлені різними причинами. Серед них: випадкові причини, вплив сезонності, дія будь – яких визначальних факторів.

Маючи справу з показниками ряду динаміки, дослідник завжди намагається виявити головну закономірність розвитку явища в окремі проміжки часу, тобто виявити головну тенденцію в зміні рівнів ряду, звільнену від дії різних випадкових факторів. З цією метою ряди динаміки піддають певній статистичній обробці. Остання може бути елементарно простою і більш складною, із застосуванням математичних методів. Завдання щодо їх використання зводиться до одного: елімінування дії випадкових, другорядних причин, а також встановлення характеру дії основних причин, які визначають динаміку досліджуваних явищ. На цій основі можна передбачити динаміку явищ і в майбутньому , що має аналітичне і практичне значення щодо управління процесами виробництва.

Мета інших прийомів (методів) аналізу рівнів динаміки зводиться до побудови математичних функцій динаміки на підставі ізольованих (дискретних) емпіричних спостережень. Їх називають **методами вирівнювання статистичних рядів.** Така функція динаміки і відповідні їм згладжені криві динаміки дають значно більш узагальнююче і наочне уявлення про досліджувану сукупність чи спостережуване явище, ніж звичний матеріал у вигляді статистичних рядів розподілу. Математичні функції дають змогу глибше вивчити фактори, які впливають на формування статистичної сукупності або на динаміку явища у часі.

Змикання рядів динаміки. Цей прийом обробки рядів динаміки застосовують у випадках, коли досліджувані рівні за одні роки неспівставні з рівнями за інші роки. Неспівставність може бути зумовлена різними причинами: територіальними змінами реорганізаційними факторами переходом до інших одиниць виміру і т.д. у таких випадках здійснюють перерахунок рівнів ряду. Наприклад, є дані про середньорічні надої молочного стада корів в господарствах району, територія якого була зменшена в 2004 році (табл. 58).

Оскільки границі адміністративного району у 2002 році змінилися дані за 2000-2002 рр. виявилися неспівставними з даними за 2002-2004 рр. Щоб побудувати ряд динаміки зівставних показників, приймають за базу порівняння рівень 2000 року, для якого є дані у старих і нових межах району. В результаті одержують ряди відносних величин з однаковою базою порівняння, які можна замінити одним зімкненим рядом динаміки. За даними такого ряду розраховують коефіцієнти (темпи) зростання по відношенню до будь – якого року. Так, середній надій у 2004 році по відношенню до 2000 року збільшився на 37,7 % (109,5 : 79,5 =1,377).

Таблиця 58

Зведення рядів динаміки до однієї бази порівняння

Роки	Середній надій на корову		2000 р. – 100 %		Зімкнений ряд динаміки, %
	до територіальних змін	після територіальних змін	до територіальних змін	після територіальних змін	
2000	3100		79,5		79,5
2001	3600		92,3		92,3
2002	3900	4200	100,0	100,0	100,0
2003		4400		104,8	104,8
2004		4600		109,5	109,5

Прийоми виявлення загальної тенденції. Найважливішим завданням аналізу ряду динаміки є виявлення закономірності розвитку явища виявлення загальної тенденції динаміки, а також її характеру, типу .

Під загальною тенденцію динаміки розуміють тенденцію до зростання, стабільності чи зниження рівня певного явища, а під *характером (типом) динаміки* розуміють ту чи іншу тенденцію зміни аналітичних показників динаміки: абсолютного приросту, коефіцієнта (темпу) зростання або темпу проросту.

Приклад. Розглянемо показники динаміки врожайності окремих культур у сільськогосподарському підприємстві за 6 років (табл. 59).

Таблиця 59

Динаміка врожайності сільськогосподарських культур у сільськогосподарському підприємстві

Роки	Картопля			Цукровий буряк			Ячмінь ярий		
	урожай-ність, ц/га	приріст		урожай-ність, ц/га	приріст		урожай-ність, ц/га	приріст	
		ц*	%**		ц	%		ц	%
1999	102	-	-	192	-	-	13,6	-	-
2000	120	18	12	228	36	12	15,0	1,4	10
2001	140	20	12	269	41	12	17,0	2,0	13
2002	160	20	11	312	43	12	19,6	2,6	15
2003	178	18	9	359	47	11	25,3	5,7	29
2004	197	19	8	411	52	11	31,0	5,7	22

Тут і далі : * абсолютний приріст; **темп приросту.

Наведені дані свідчать про загальну тенденцію зростання врожайності всіх трьох видів сільськогосподарських культур. Але характер такого зростання неоднаковий. Так, абсолютні прирости урожайності картоплі відносно стабільні, а щодо темпів приросту, то тут спостерігаєся тенденція до деякого зниження. Інший характер динаміки мають показники врожайності цукрових буряків: абсолютні прирости тут щорічно збільшуються, тоді як темпи приросту стабілізуються на рівні 11 – 12 %. Урожайність ярого ячменю зростає прискорено, як в абсолютному так і у відносному виразі.

Таким чином, за даними наведеного прикладу переконуємося, що зростання рівнів ряду динаміки може здійснюватися з неоднаковою інтенсивністю. Отже, завдання аналізу зводиться до того, щоб виявити загальні для всього досліджуваного періоду риси змін.

Серед різноманітних форм характеру динаміки виділяють такі її типи: I- абсолютні прирости спадають; II - абсолютні прирости стабільні; III – темпи зростання стабільні; IV – темпи зростання збільшуються.

Для I типу динаміки характерним є те, що при зростанні рівнів ряду спостерігається зниження абсолютного приросту; II тип динаміки показує, що зростання рівнів ряду супроводжується стабільністю абсолютних приростів і зниженням темпів приросту; III тип означає, що при стабільних темпах зростання абсолютні прирости збільшуються; IV - характеризується інтенсивністю зростання рівнів при систематичному підвищенні темпів зростання.

Таким чином, при постійному зростанні рівнів ряду характеристика динаміки у наведених її типів досить різноманітна , тобто інтенсивність зростання тут неоднакова. Наочно ці типи динаміки ілюструє графік (рис. 22).

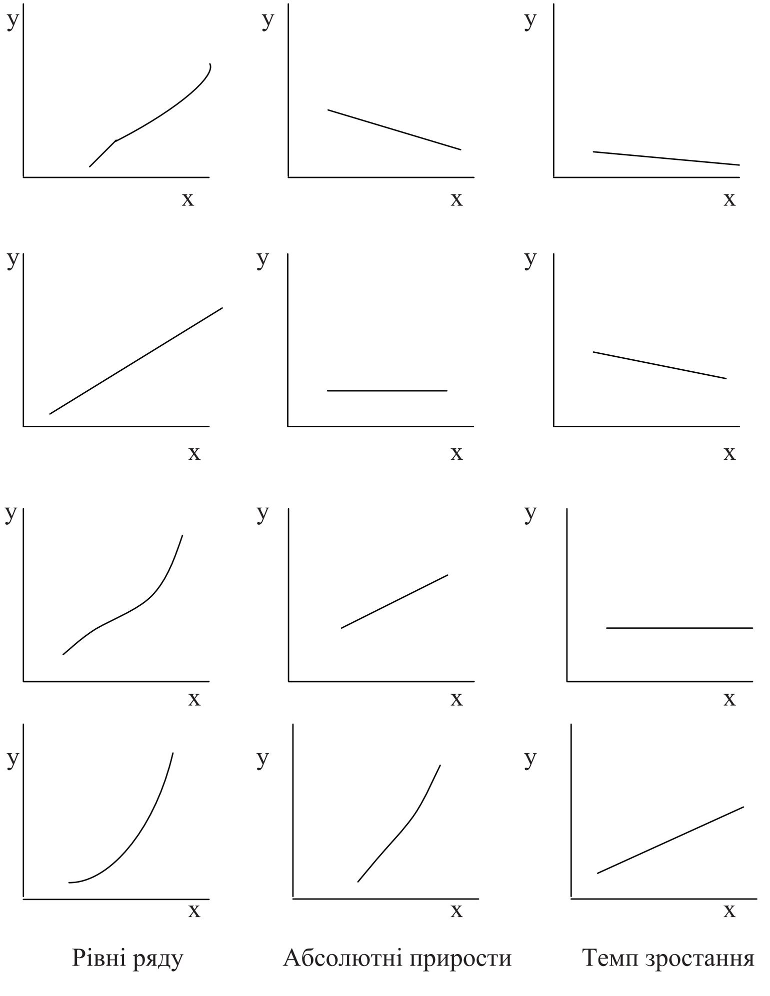


Рис.22. Типи динаміки

Інколи виявити загальну тенденцію розвитку і характер динаміки за ланцюговими показниками не вдається. Це трапляється в тих випадках, коли рівні або одержані ланцюгові показники динаміки значно варіюють, то підвищуючись, то знижуючись. У такому разі основна тенденція розвитку явища ніби затушовується. Щоб її виявити, статистика застосовує такі прийоми: згладжування шляхом укрупнення інтервалів; чи за ковзною (рухомою) середньою.

Згладжування (вирівнювання) шляхом укрупнення інтервалів. Цей найпростіший спосіб виявлення закономірності зміни рівнів динаміки полягає в одержанні середніх або підсумкових показників для укрупнення періодів (інтервалів) часу. Так, наприклад, в рівнях ряду динаміки показників урожайності цукрових буряків по роках спостерігається значна варіація, зумовлена природно – економічними факторами окремих років. Для встановлення тенденції в русі показників урожайності цієї культури розраховують середні значення за трихріччя, п'ятиріччя чи інші періоди.

Згладжування способом ковзної середньої є одним з ефективних методів виявлення загальної тенденції розвитку явища в часі. Суть його полягає в тому, що середній рівень обчислюється спочатку з певного числа перших рівнів ряду, потім - з такої ж кількості рівнів, але починаючи з другого, далі – починаючи з третього і т.д. Розраховані таким чином середні рівні ряду ніби ковзають по ряду динаміки від його початку до кінця, при цьому щоразу відкидається один рівень спочатку і додається наступний. Звідси назва – «ковзна» (рухома) середня. Згладжування таким способом можна здійснювати за будь – яким числом членів ряду. Наприклад, для згладжування ряду динаміки способом ковзної середньої з 5 членів, необхідно послідовно додати 5 членів ряду і результати поділити на 5.

Перш ніж розглянути процес розрахунків ковзних середніх, який проілюстровано таблицею 60, зупинимося на деяких технічних особливостях його здійснення.

Кожна ланка ковзної середньої умовно відноситься (записується чи наноситься на графік) до середини відповідного періоду. При цьому, якщо охоплено парне число рівнів ряду, то середина періоду не збігатиметься з жодним вихідним періодом або датою. У нашому прикладі має місце такий випадок. Одержані ланки ковзної середньої **центрують** шляхом розрахунку на їх основі двочленних ковзних середніх.

Як бачимо з даних, наведених у таблиці, після згладжування загальна тенденція зростання урожайності зернових культур проявляється досить виразливо. Водночас за такими розрахунками можна більш детально (ніж при звичайному укрупненні інтервалів) прослідкувати і **характер динаміки**. Так, показники графи 3 таблиці свідчать про те, що зростання здійснювалося нерівномірне і лише центрування згладжує цю нерівномірність.

Таблиця 60

Розрахунок ковзної середньої (6 – річної) динаміки врожайності зернових культур (рівень ряду - в ц з 1 га)

Рік	Рі-вень ряду	Шести-річчя (роки)	Рівень ряду		Приріст серед-нього рівня	Центрування	
			сума за 6 років	се-ред-ня за рік		сере-дина періо-ду	центровані ланки ковзної середньої
А	Б	В	1	2	3	4	5
1995	18,5						
	23,0						
1996	23,4						
1997		1995-2000	140,5	23,4	-		
1998	25,4	1996-2001	149,4	29,4	1,5	1998	(23,4+24,9):2 =24,1
1999	21,2	1997-2002	146,8	24,4	-0,5	1999	(24,9+24,4):2=24,6
2000	29,0					2000	(24,4+25,2):2=24,8
		1998-2003	151,3	25,2	0,8		
2001	27,4					2001	(25,2+25,6):2=25,4
		1999-2004	153,8	25,6	0,4		
2002	20,4						
2003	27,9						
2004	27,9						

В даному випадку розглянуто приклад згладжування за допомогою 6- річної (6 – членної) ковзної середньої. Питання про кількість років, охоплених ланкою ковзної середньої вирішується в залежності від конкретних особливостей досліджуваного ряду. При цьому, чим довший період, за який обчислюється кожна ланка рухомої середньої, тим сильніше буде згладжено ряд.

Недоліком такого способу вирівнювання є те, що згладжений ряд «укорочується» в порівнянні з фактичним на $\frac{n-1}{2}$ члена з початку і кінця ряду динаміки (n - число членів, з яких розраховується ковзна середня). У нашому прикладі цей показник становить $\frac{(6-1)}{2}=2,5$, тобто разом ряд динаміки укорочується на 5 рівнів (2,5 + 2,5), про що свідчить графа 2 таблиці 60.

Отже, одержане число ланок завжди менше вихідних рівнів, а це звужує можливості виявлення характеру (типу) динаміки, оскільки маємо малу кількість аналітичних показників ряду динаміки.

Аналітичне вирівнювання рядів динаміки вважається найбільш удосконаленим способом обробки ряду з метою встановлення тенденції розвитку явища. Завдання такого вирівнювання полягає у знаходженні простої математичної формули (апроксимуючої функції), яка б відображала загальну тенденцію ряда динаміки. Рівні ряду тут розглядаються як функція часу, а завдання (вирівнювання) зводиться до визначення виду функції, її параметрів за емпіричними даними та розрахунку теоретичних рівнів за знайденою формулою.

Суть такого вирівнювання полягає в наступному:

а) на підставі економічного аналізу виділяють певний етап розвитку явища і виявляють характер динаміки протягом цього етапу;

б) виходячи з характеру динаміки обираються той чи інший математичний вираз закономірності, тобто аналітичне рівняння, якому на графіку відповідає певна лінія - пряма, парабола, гіпербола, експонента, логарифмічна крива і т.д.

Підбір емпіричної формули, за допомогою якої здійснюється вирівнювання, відбувається у два етапи : 1) вибір виду функції, яка дає найкраще наближення; 2) визначення параметрів обраної функції.

Як підібрати потрібний вид функції, яка б давала найкраще наближення теоретичної лінії до емпіричних даних? Насамперед слід виходи з економічної суті досліджуваних залежностей і логіко – теоретичних передумов характеру зміни явищ. Зокрема :

1) якщо динаміка характеризується більш чи менш стабільним абсолютним приростом, тобто коли рівні ряду змінюються приблизно у арифметичній прогресії, використовується рівняння прямої лінії виду $\overline{y}_t = a_0 + a_1 t$, де $\overline{y}_t$ - теоретичний рівень (читається «ігрек», вирівняний по t); t - час, a_0 і a_1 - параметри прямої;

2) у випадках, коли абсолютні прирости рівномірно збільшується і при згладжуванні крива має один вигин, наближеним математичним виразом цієї тенденції можна обрати параболу другого порядку : $\overline{y}_t = a_0 + a_1 t + a_2 t^2$;

3) якщо криві при згладжуванні мають S – подібну форму (два вигини), використовують рівняння параболи третього порядку:

$$\overline{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3;$$

4) якщо коефіцієнти зростання, розраховані по відношенню до попереднього періоду, більш чи менш постійні, тобто ряд динаміки відображує розвиток у геометричній прогресії, розраховують показову функцію. Для вирівнювання ряду динаміки в даному випадку використовують рівняння $\bar{y}_t = a_0 a_1^t$.

Логарифм показової функції ($\lg \bar{y}_t = \lg a_0 + t \lg a_1$) являє собою рівняння прямої. Параметри a_0 і a_1 знаходять за системою нормальних рівнянь:

$$\begin{cases} n \lg a_0 + \lg a_1 \sum t = \sum \lg y \\ \lg a_0 \sum t + \lg a_1 \sum t^2 = \sum t \lg y \end{cases}$$

5) серед інших способів обробки рядів динаміки особливе місце займає вирівнювання за допомогою ряду Фур'є, який виражається рівнянням: $\bar{y}_t = a_0 + \sum_{\kappa=1}^{m} \left(a_\kappa \cos \kappa t + в\kappa \sin \kappa t \right)$

де к – гармоніка ряду (к=1; к=2; і т.д.) ; $a_0 = \frac{\sum y}{n}$; $a_\kappa = \frac{2\sum y \cos \kappa t}{n}$; $в_\kappa = \frac{2\sum y \sin \kappa t}{n}$.

Параметри аналітичного рівняння лінії зв'язку визначають за способом найменших квадратів. Тобто, сума квадратів відхилень фактичних рівнів динаміки від вирівняних повинна бути мінімальною: $\sum (y_t - \bar{y}_t) = \min$.

Вирівняні рівні ряду ($\bar{y}_t$) розраховують на підставі аналогічного рівняння шуканої прямої чи кривої. На графіку вони розташовані на прямій чи кривій лінії відповідного типу.

Таким чином, якщо розглядати технічний бік вирівнювання, то він зводиться до заміни фактичних рівнів теоретичними, які в середньому мінімально відхилися б від фактичних рівнів, але мали б певний аналітичний вираз.

Хоча спосіб вирівнювання ряду динаміки містить в собі певні умовності, в ряду випадків він є досить корисним технічним прийомом, який полегшує виявлення загальної тенденції і вивчення характеру ряду динаміки. Зокрема, це стосується вивчення сезонних коливань. Про це мова йтиме пізніше.

Приклад. Розглянемо найбільш поширеного і простого випадку аналітичного вирівнювання ряду динаміки врожайності проса за рівнянням прямої лінії (1987-2004 рр., *n*=18). Рівняння прямої має вигляд $\bar{y}_t = a_0 + a_1 t$, де t –

час, тобто порядковий номер періоду або моменту часу; a_0 і a_1- параметри шуканої прямої.

Параметри a_0 і a_1 знаходяться за системою нормальних рівнянь:

$$\begin{cases} a_0 n + a_1 \sum t = \sum y \\ a_0 \sum t + a_1 \sum t^2 = \sum ty, \end{cases}$$

де у – фактичні рівні ряду;

n= число рівнів.

Для зручності розрахунків будують робочу таблицю 61. На підставі даних цієї таблиці одержуємо таку систему нормальних рівнянь:

$$\begin{cases} 18a_0 + 171a_1 = 263{,}9 \\ 171a_0 + 2109a_1 = 2959{,}1 \end{cases}.$$

Рішення:

$$171a_0 + 2109a_1 = 2959{,}1$$
$$-171a_0 + 1624{,}5a_1 = 2507{,}1$$
$$484{,}5a_1 = 452;$$

a_1=0,93; $18a_0 + (171 \times 0{,}93) = 263{,}9$; $a_{0==}$=5,80.

Знайдені параметри дають змогу побудувати рівняння прямої:

$\overline{y}t$=5,8 + 0,93 t.

Таблиця 61

Вихідні та розрахункові дані вирівнювання ряду динаміки врожайності проса за рівнянням прямої

Роки	Порядковий номер року	Фактичний рівень урожайності, ц/га	Розрахункові величини		Теоретичний рівень урожайності, ц/га
n	t	y	ty	t^2	$\overline{y}t$
1987	1	7,5	7,5	1	6,7
1988	2	4,7	9,4	4	7,7
1989	3	10,2	30,6	9	8,6
…	…	…	…	…	…
2004	18	23,2	417,6	324	22,6
n=18	Σt=171	Σy=263,9	Σty=2959,1	$\sum t^2$ =2109	$\sum \overline{y}t$ =263,9

Дані одержаного рівняння свідчать про те, що середня врожайність проса в нульовому році (1986) досліджуваного періоду становила 5,8 ц , а підвищення її складає в середньому 0,93 ц у кожному наступному році, тобто теж саме означає середньорічний приріст урожайності впродовж досліджуваного періоду.

Підставивши в рівняння почергово значення t, одержуємо вирівняний ряд динаміки врожайності, який абстрагований від випадкових коливань і характеризується систематичним підвищенням рівнів.

В аналогічній послідовності здійснюється вирівнювання рядів динаміки за іншими типами аналогічних функцій (парабола, гіпербола, експонента та ін.).

§ 9.2. Статистичні прийоми виміру сезонних коливань

Досить значна кількість суспільних явищ має сезонний характер, тобто сезонні коливання. Рівень їх рік у рік у певні місяці підвищується, а в інші – знижується. Наприклад, витрати палива у весняно – літні місяці значно більші ніж в осінньо – зимові місяці; досить неоднаковими впродовж року виявляються ціни на сільськогосподарську продукцію на ринку і т.д. Такі внутрішньорічні коливання, які мають періодичний характер, називають **сезонними**. Вони завжди пов'язані з впливом природних факторів, особливо в сільському господарстві.

Сезонність – явище негативне, адже вона обумовлює нерівномірність здійснення виробничих процесів, призводить до зниження продуктивності праці та підвищення собівартості виробництва продукції. Тому подолання сезонності є важливим резервом підвищення економічної ефективності виробництва. Звідси випливає питання про необхідність вивчення сезонності та кількісного виміру сезонних коливань (сезонної хвилі), що є одним із важливих завдань аналізу рядів динаміки.

Розглянемо деякі методи, розроблені статистикою для виявлення та виміру сезонної хвилі.

Перший спосіб. А. Для рівнів ряду розраховується середня арифметична величина ($\bar{y}$), потім із нею порівнюють (у%) рівень кожного місяця (y_i). Одержане процентне відношення називається **індексом сезонності** : $I_C = \frac{y_i}{y} \times 100$.

Б. Вплив на місячні дані випадкових коливань зумовлює необхідність розрахунку для кожного місяця середніх показників за триріччя. Потім знаходять процентне відношення середніх для кожного місяця до загального середнього рівня, тобто: $I_C = \frac{\bar{y}_i}{\bar{y}} 100$, де $\bar{y}_i$ - середня для кожного місяця за 3 роки; $\bar{y}$ - загальний середній рівень за 3 роки. Схему цього способу розрахунку наведено в таблиці 62.

Таблиця 62

Розрахунок індексів сезонності за першим способом

Місяць	Середня денна виробітка на трактор, га умовної оранки				Індекси сезонності $(\overline{y_i} : \overline{y}) \times 100$ %
	2002р.	2003р.	2004р.	В середньому	
Січень	4,6	4,5	4,2	4,4	(4,4:5,7)=100=77
Лютий	5,0	4,8	4,5	4,7	(4,7:5,7)=100=84
Березень	4,9	5,1	6,0	5,3	(6,3:6,7)=100=93
...	...	...	...	...	...
Грудень	4,2	4,3	4,8	4,4	(4,4:5,7)=100=77
Середній рівень ряду	5,9	5,5	5,7	$\overline{y}=5,7$	100

Другий спосіб. При наявності даних за три роки розраховують індекси сезонності для кожного року за першим способом (А), а потім за одержаними індексами знаходять середню арифметичну.

Розглянемо даний спосіб на прикладі, здійснивши розрахунок для січня. Індекси сезонності для кожного року становлять : 2002 рік – (4,6:5,9)100= 78 %; 2003 рік – (4,5: 5,5)100= 82 %; 2004 рік - (4,2:5,7) 100= 74 %. Звідси середній індекс сезонності для січня становить : $\frac{78+82+74}{3}=78\%$.

Аналогічно здійснюють розрахунки для лютого, березня і т.д.

Наведені результати розрахунків свідчать про те, що індекси сезонності (для січня) майже не різняться між собою. Це пояснюється стабільністю місячного рівня в різні роки. У випадках, коли спостерігається тенденція до збільшення чи зменшення рік у рік місячних рівнів, то перевагу віддають другому способу.

Третій спосіб полягає в обчисленні відношень фактичних помісних рівнів до ковзної середньої, розрахункові для 12 місяців . На підставі таких відношень (індексів сезонності) за ряд років знаходять середню арифметичну для кожного місяця. Ці середні вважаються індексами сезонних коливань.

За аналогічною схемою розрахунків індекси сезонності можна побудувати на підставі відношень фактичних помісячних рівнів до рівнів, вирівняних за математичними формулами (прямої, параболи, гіперболи і т.д.). Існують і інші більш складні способи (методи) розрахунку індексів сезонності.

§ 9.3. Особливості кореляційного аналізу рядів динаміки та методичні основи статистичного прогнозування їх рівнів

Об’єктом кореляційного аналізу можуть бути не тільки статистичні (просторові) сукупності, а й сукупності, які характеризують зміну явищ у часі, тобто динамічні. Розроблена методологія кореляції для аналізу явищ у просторі не прийнятна для динамічних сукупностей. Тому при використанні кореляційного методу необхідно знати особливості та межі його використання. Насамперед це стосується перевірки передбачень та інтерпретації результатів аналізу рядів динаміки. Як кореляційна модель, так і її статистичні характеристики мають конкретний економічний зміст і висвітлюють економічне явище з певного боку. Оскільки зміст статистичних показників залежить від дотримання вимог щодо їх обчислення, то дослідник повинний вміти користуватися методологією аналізу і пояснювати одержані результати. Особливо це стосується досліджень кореляції між рядами динаміки.

Під **кореляцією рядів динаміки** розуміють метод вивчення зв'язку між показниками, представленими їх значеннями в послідовні моменти або періоди часу. Кореляція рядів динаміки має свої особливості, які зумовлені тим, що ряд динаміки, по-перше, має короткочасні коливання (місячні, квартальні, річні) і, по-друге, містить у собі такий компонент, як загальна тенденція в зміні показників ряду – «вісь кривої», або **тренд.** Під останнім розуміють зміну, яка визначає загальний напрям розвитку, основну тенденцію рядів динаміки. Лінію тренда можна порівняти з лінією регресії. Якщо остання являє собою плавну зміну результативної ознаки під впливом факторної, звільненої від дії всіх сторонніх (неврахованих) причин, то лінія тренда характеризує плавну у часі зміну явищ, викликаних різними обставинами короткочасних відхилень від загальної тенденції.

Наявність тренда ускладнює кореляційний аналіз рядів динаміки. Так, якщо вивчається кореляція рядів без виключення загальної тенденції в них, то показник тісноти залежності характеризуватиме зв'язок не лише між короткочасними коливаннями, а й між трендами. В іншому випадку, коли тренди будуть виключеними із корельованих рядів динаміки, одержаний коефіцієнт кореляції характеризуватиме тісноту залежності лише між короткочасними коливаннями.

Тренд, відображуючи загальний напрям змін явища, що відбуваються у часі, водночас визначає й залежність між членами ряду динаміки. Ця залежність, яка визначається формою лінії тренда, має таку ж саму статистичну природу, як і лінія регресії. Зазначену кореляційну залежність між сусідніми (попередніми і наступними) членами ряду називають **автокореляцією.**

Наявність автокореляції зумовлюється різними причинами, а саме:

1) у кореляційній динамічній моделі не врахований істотний фактор;

2) у моделі не враховано кілька неістотних факторів, взаємний вплив яких є істотним внаслідок збігу фаз та напрямів їх змін;

3) обрано неправильний тип моделі;

4) специфічна структура випадкової компоненти.

Щоб виявити лінії трендів з метою наступного їх виключення з аналізу, здійснюють вирівнювання ряду за допомогою ковзної середньої або аналітичне вирівнювання ряду динаміки за певною математичною функцією (прямої, параболи, експоненти та ін.).

Зазначені особливості кореляційного аналізу рядів динаміки розглянемо на прикладі показників продуктивності молочного стада корів *(у* – середньорічний надій, т) і рівня їх годівлі (х – річні витрати кормів на голову, ц к. од.): (табл. 63).

Таблиця 63

Вихідні і розрахункові дані для обчислення коефіцієнта кореляції в рядах динаміки

№ п.п. року	у	х	ух	$у^2$	$х^2$	Темпи зростання, %	
						х	у
1	2,7	39	105,3	7,29	1521	–	–
2	2,5	30	75,0	6,25	900	93	77
3	2,3	31	71,3	5,29	901	92	103
4	3,8	49	186,2	14,44	2401	165	158
5	3,5	40	140,0	12,25	1600	92	82
6	3,9	41	159,9	15,21	1681	111	102
7	3,2	39	124,8	10,24	1521	82	95
8	4,5	47	211,5	20,25	2209	141	121
9	5,1	60	306,0	26,01	3600	113	128
10	4,8	52	249,6	23,04	2704	94	87
$\sum$	36,3	428	1629,6	140,27	19098	–	–

Якщо розглядати кореляційну залежність в рядах динаміки, прийнявши рівень годівлі тварин за ознаку-фактор а рівень їх продуктивності – за результативну ознаку, то величина коефіцієнта кореляції становитиме:

$$r_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]\left[\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right]}} = \frac{1629{,}6 - \frac{36{,}3 \cdot 428}{10}}{\sqrt{\left[140{,}27 - \frac{36{,}3^2}{10}\right]\left[19098 - \frac{428^2}{10}\right]}} = 0{,}934$$

Одержана величина коефіцієнта кореляції свідчить про наявність досить тісної кореляційної залежності між продуктивністю тварин та рівнем їх годівлі. Така сильна кореляція зумовлена тим, що в обох порівнюваних рядах динаміки короткострокові коливання мають однакову тенденцію, отже, і тренди відображують однаковий напрям змін у часі як продуктивності тварин, так і рівня їх годівлі.

Форма і напрям лінії трендів для обох досліджуваних рядів: наведені на рис. 23. Лінії трендів надоїв і рівня годівлі визначені шляхом аналітичного їх вирівнювання за формулою прямої лінії: $\bar{y}_t = a_0 + a_1 t$.

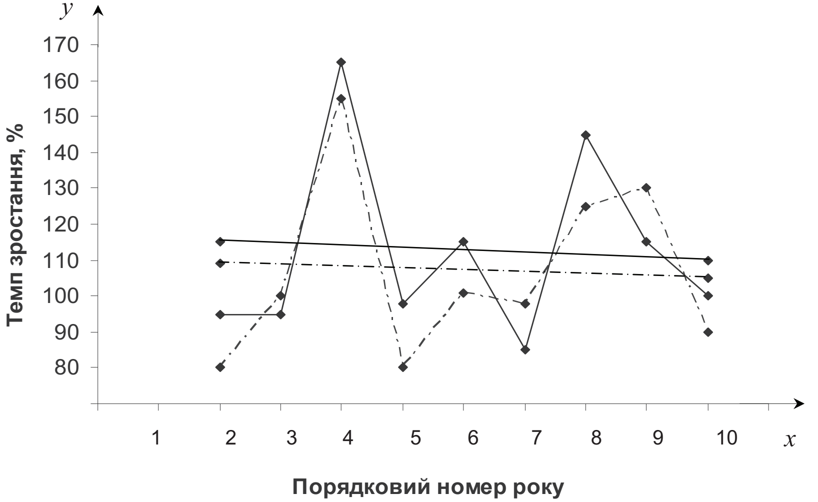


Рис. 23. Тренди продуктивності тварин і рівня їх годівлі:
1 – емпіричні дані надоїв; 2 – тренд надоїв;
3 – емпіричні дані рівня годівлі; 4 – тренд рівня годівлі.

Якщо ставиться завівання визначити, якою мірою щорічні коливання показників рівня продуктивності молочного стада залежали від рівня годівлі, то його вирішення потребує виключення з обох рядів динаміки трендів. Із цією метою визначають характеристики автокореляції. Для рядів надоїв і рівнів годівлі автокореляція визначається шляхом порівняння даних, які стосуються двох сусідніх років (табл. 64, 65).

Таблиця 64

Вихідні і розрахункові дані для обчислення коефіцієнта кореляції за показниками надоїв (автокореляція)

x	y	xy	x^2	y^2
2,7	2,5	6,75	7,29	6,25
2,5	2,3	5,75	6,25	5,29
2,3	3,8	8,74	5,29	14,44
3,8	3,5	13,3	4,44	12,25
3,5	3,9	13,65	12,25	15,21
3,9	3,2	12,48	15,21	10,24
3,2	4,5	14,4	10,24	20,25
4,5	5,1	22,95	20,25	26,01
5,1	4,8	24,48	26,01	23,04
31,5	33,6	122,5	117,23	132,98

Таблиця 65

Вихідні і розрахункові дані для обчислення коефіцієнта кореляції за показниками рівня годівлі (автокореляція)

x	y	xy	x^2	y^2
39	30	1170	1521	900
30	31	930	900	961
31	49	961	1519	2401
49	40	1960	2401	1600
40	41	1640	1600	1681
41	39	1599	1681	1521
39	47	1833	1521	2209
47	60	2820	2209	3600
60	52	3120	3600	2704
376	389	16033	16952	17577

Розрахуємо коефіцієнти кореляції по рядах динаміки надоїв і рівнів годівлі:

$$r_{yx} = \frac{122{,}5\frac{31{,}5\cdot 33{,}6}{9}}{\sqrt{\left[117{,}23-\frac{31{,}5^2}{9}\right]-\left[132{,}98-\frac{33{,}6^2}{9}\right]}} = 0{,}676;$$

$$r_{yx} = \frac{16033-\frac{376\cdot 389}{9}}{\sqrt{\left[16952-\frac{376^2}{9}\right]\left[17577-\frac{389^2}{9}\right]}} = -0{,}224$$

Як свідчать числові значення знайдених коефіцієнтів кореляції, в ряді динаміки надоїв існує досить значна додатна автокореляція; у ряді рівнів годівля тварин вона слабка і від'ємна.

Визначимо лінії трендів з тим, щоб потім виключити їх з аналізу. Зокрема, здійснимо аналітичне вирівнювання ряду динаміки за прямою $\overline{y}_t = a_0 + a_1 t$, де *t* – одиниця часу (у даному випадку рік).

Параметри a_0 і a_1 знаходять шляхом вирішення системи нормальних рівнянь:

$$\begin{cases} a_0 n + a_1 \sum t = \sum y; \\ a_0 \sum t + a_1 \sum t^2 = \sum yt. \end{cases}$$

Для показників динаміки надоїв (табл. 66) і рівнів годівлі (табл. 67) системи рівнянь мають вигляд:

$$\begin{cases} 10a_0 + 55a_1 = 36{,}3; \\ 55a_0 + 385a_1 = 223. \end{cases} \qquad \begin{cases} 10a_0 + 55a_1 = 428; \\ 55a_0 + 385a_1 = 2543. \end{cases}$$

Таблиця 66

Вихідні і розрахункові дані для обчислення лінії тренда показників надоїв

Рік (t)	Надій (y)	yt	t^2	$\bar{y}_t$
1	2,7	2,7	1	2,37
2	2,5	5,0	4	2,65
3	2,3	6,9	9	2,93
4	3,8	15,2	16	3,21
5	3,5	17,5	25	3,49
6	3,9	23,4	36	3,77
7	3,2	22,4	49	4,05
8	4,5	36,0	64	4,33
9	5,1	45,9	81	4,61
10	4,8	48,0	100	4,89
55	36,3	223	385	36,3

Таблиця 67

Вихідні і розрахункові дані для обчислення лінії тренда показників рівня годівлі

Рік (t)	Рівень годівлі (y)	yt	t^2	$\bar{y}_t$
1	39	39	1	32,49
2	30	60	4	34,78
3	31	93	9	37,07
4	49	196	16	39,36
5	40	200	25	41,65
6	41	246	36	43,94
7	39	273	49	46,23
8	47	376	64	48,52
9	60	540	81	50,81
10	52	520	100	53,10
55	428	2543	385	428

Після рішення цих систем одержимо аналітичні рівняння зв'язку: для рядів динаміки надоїв $\bar{y}_t = 2{,}09 + 0{,}28t$; для рядів динаміки рівнів годівлі $\bar{y}_t = 30{,}2 + 2{,}29t$.

Щоб виключити знайдений тренд із аналізу рядів динаміки, розраховують відхилення від нього емпіричних даних. Такі відхилення, які являють собою короткочасні коливання у «чистому» вигляді, можна корелювати. У прикладі з показниками продуктивності молочного стада та рівня годівлі визначено в обох рядах відхилення від трендів, розраховані за формулою прямої лінії (табл. 68).

За одержаними відхиленнями від тренда розраховують коефіцієнт кореляції (табл. 69):

$$r_{xy} = \frac{\sum \Delta x \Delta y}{\sqrt{\sum \Delta x^2 \sum \Delta y^2}}; \qquad r_{xy} = \frac{22{,}4675}{\sqrt{1{,}8930 \cdot 346{,}6185}} = 0{,}877.$$

Величина розрахованого коефіцієнта кореляції найбільш точно відображує тісноту зв'язку в рядах динаміки показників надоїв і годівлі. Адже наявність високої автокореляції в рядах показників надоїв ($r_{xy} = 0{,}676$) певною мірою посилювала зв'язок між щорічними надоями та рівнем годівлі, зумовивши високий рівень коефіцієнта кореляції (0,934) в досліджуваних рядах динаміки.

Таблиця 68

Відхилення від трендів

№ п.п. року	Середньорічний надій (Δy)	Рівень годівлі (Δx)
1	0,33	6,51
2	-0,15	-4,78
3	-0,63	-6,07
4	0,59	9,94
5	0,01	-1,65
6	0,13	-2,94
7	-0,85	-7,23
8	0,17	-1,52
9	0,49	9,19
10	-0,09	-1,10

Таблиця 69

Вихідні і розрахункові дані для обчислення коефіцієнта кореляції за відхиленнями від тренда

$\Delta y \Delta x$	Δ_y^2	Δ_x^2
2,1483	0,1089	42,3801
0,7170	0,0225	22,8484
3,8241	0,3969	36,8449
5,6876	0,3481	92,9296
-0,0165	0,0001	2,7225
-0,3822	0,0169	8,6436
6,1455	0,7225	52,2729
-0,2584	0,0289	2,3104
4,5031	0,2401	84,4561
0,0990	0,0081	1,2100
22,4675	1,8930	346,6185

Питання кореляції рядів динаміки тісно пов'язані з питаннями прогнозування. Економічний прогноз – це науково обґрунтоване передбачення розвитку будь-якого економічного показника тієї чи іншої галузі або народного господарства в цілому на перспективу, ступінь вірогідності якого залежить, від методів прогнозування. При економічному прогнозуванні використовують цілий комплекс методів (подібність і відмінність, супутні зміни, аналогія, моделювання та ін.), серед яких найпростішим і доступним широкому колу економістів-практиків є метод прогнозування шляхом екстраполяції рядів динаміки, який ґрунтується на гіпотезі стійкості закономірності розвитку явища (тренд). В його основу покладено зафіксовану в минулому і нинішню тенденцію, яка передбачається і в майбутньому, якщо не очікується змін зовнішніх та внутрішніх факторів, які зумовлюють її. Однак ця тенденція не завжди точно відображує дійсність, оскільки її збереження залежить насамперед від урахування взаємодії з Іншими тенденціями, що не завжди можливо. Незважаючи на це, метод екстраполяції тенденції широко використовують у прогнозуванні економічних показників, бо він дає змогу визначити заходи, спрямовані на запобігання шкідливим та посилення корисних тенденцій. Статистичний прогноз – це засіб запобігання раптовостям у дійсній реальності, що особливо важливо для економічних явищ.

Передумовою успішного застосування екстраполяції рядів динаміки (продовження рівнів ряду на майбутнє на підставі виявленої закономірності змін рівнів за досліджуваний проміжок часу) є наявність необхідної статистичної інформації, яка дає змогу перевірити гіпотезу стійкості розвитку. Таким чином, питання полягає в об'єктивному підході до вибору економічно обґрунтованого типу лінії вирівнювання ряду динаміки. Правильний вибір типу тренда дозволяє одержати точнішу характеристику коливання показників динаміки і тенденції їх змін. Від типу лінії значно залежить і величина очікуваного показника, а отже, і якість прогнозу.

У практиці статистичного прогнозування економічних показників найчастіше використовують аналітичні рівняння прямої лінії, показової кривої (експоненти) та параболи другого і вищих порядків. Вирівнювання здійснюють виходячи з логіко-теоретичного аналізу ряду динаміки, завдяки якому встановлюють характер динаміки й тип необхідної лінії аналітичного рівняння. При цьому беруть до уваги характер динаміки факторів, що зумовлюють основну тенденцію змін показників, які досліджуються. Але оскільки факторним ознакам також властиве певне (іноді істотне) коливання, характер якого не завжди зрозумілий і потребує в свою чергу поглибленого аналізу, логічніше буде при виборі типу лінії виходити з характеру динаміки результативної ознаки.

У зв'язку з тим, що завдання вибору типу аналітичної лінії потребує абстрагування від індивідуальних особливостей коливань багатьох факторів, тобто, передбачає узагальнене відображення їх дій, досить корисним слід вважати попередній аналіз первинних емпіричних показників у системі координат. При цьому метою вирівнювання рядів динаміки, є не вибір лінії, а встановлення тенденції розвитку» Наприклад, при вивченні показників динаміки врожайності в розрізі окремих культур тенденції змін її не збігаються, тому будуть різними й аналітичні рівняння зв'язку. Для встановлення стабільності тієї чи іншої характеристики динаміки (абсолютний приріст, темп росту та його прискорення) розраховують істотність їх різниць. Так, якщо у рядах величин приростів,, темпів росту чи прискорення не спостерігається певних закономірностей у розташуванні, тобто показники зазначених характеристик «розкидані» по ряду динаміки, можна стверджувати про істотність їх різниць. У випадку, коли розраховані характеристики динаміки за розміром їх величин концентруються в певних частинах ряду, різниця між ними буде істотною й потребуватиме статистичної оцінки. Вважається, що при ймовірності понад 0,9 різниця буде істотною, а тому екстраполяцією за відповідним типом лінії (абсолютний приріст – пряма, темп росту – показова крива, прискорення – парабола) виконувати не можна.

Проілюструємо це на прикладі рядів динаміки врожайності зернових культур за попередні 23 роки (табл. 70).

Як свідчать наведені дані, серед показників динаміки найвиразніше простежується закономірність у розташуванні абсолютних приростів. У другій половині ряду концептруються вищі їх величини, ніж у першій; вони мають стійкішу тенденцію до зростання. Для визначення істотності різниці середніх приростів обчислюють середню випадкову помилку для кожної половини ряду динаміки за формулою $m = \sigma : \sqrt{n}$. Середню випадкову помилку різниці знаходять, як суму таких помилок двох періодів ($m_p = \sqrt{m_1 + m_2}$). Порівнюючи різницю середніх абсолютних приростів ($A_2 - A_1$) із середньої) випадковою помилкою, одержують показник нормованого відхилення *(t)*.

Згідно з цією статистичною характеристикою і числом ступенів вільності за таблицею значень інтеграла ймовірностей Ст'юдента знаходять імовірність істотності різниці середніх абсолютних приростів. Аналогічні розрахунки виконують і для інших характеристик рядів динаміки – темпів росту і прискорень. У нашому прикладі інтеграл імовірності при нормованому відхиленні t=0,259 і 17 ступенях вільності варіація приростів урожайності зернових культур становить 0,616 (додаток 1). Таким чином, величина знайденого показника ймовірності менша за 0,9, що дає підставу стверджувати про можливість використання для екстраполяції врожайності зернових культур рівняння прямої лінії. Якщо прийняти гіпотезу про сталість закономірності для визначення врожайності зернових, то фактичні її рівні добре апроксимуються рівнянням $\bar{y}_t$ = 10,6+0,554*t*. Екстраполяція цієї лінії на наступні 8-10 років дає показник 28-29 ц/га.

Таблиця 70

Статистичні характеристики динаміки врожайності зернових культур

№ п.п. року	Середня врожайність, ц/га	П'ятирічна середня ковзна врожайність, ц/га	Показники динаміки середньої ковзної врожайності		
			абсолютний приріст, ц/га	темп росту, %	прискорення, %
1	9,1	–	–	–	–
2	10,3	–	–	–	–
3	16,5	10,6	–	–	–
4	13,7	13,1	2,5	123,6	–
5	3,6	14,2	1,1	108,4	-15,2
6	21,3	14,0	-0,2	98,6	-9,8
7	15,8	15,7	1,7	112,1	13,5
8	15,4	17,9	2,2	114,0	1,9
9	22,4	16,7	-1,2	93,3	-20,7
10	14,5	17,2	0,5	103,0	9,7
11	15,2	17,5	0,3	101,7	1,3
12	18,6	15,1	-2,4	122,4	20,7
13	16,9	15,3	0,2	86,3	-36,1
14	10,5	15,6	0,3	102,0	15,7
15	15,2	16,8	1,2	107,7	5,7
16	17,0	18,1	1,3	107,7	0
17	24,4	20,0	1,9	110,5	2,8
18	23,4	20,8	0,8	104,0	-6,5
19	20,2	22,2	1,4	106,7	2,7
20	19,0	23,6	1,4	106,3	-0,4
21	24,1	22,4	-1,2	94,6	-11,4
22	31,4	–	–	–	–
23	17,1	–	–	–	–

Існування нерегульованих факторів (наприклад, кліматичні умови) потребує обґрунтування форми тренда з урахуванням їх дії. Для розглядуваного прикладу найкращими формами тренда можуть бути експонента і парабола. Тут необхідний поглиблений попередній аналіз досліджуваного періоду, який ґрунтується на знанні агроекономічних факторів формування рівнів урожайності.

В таблиці 71 наведено статистичні характеристики рядів динаміки для різних культур, вирощуваних у степовій зоні України, і відповідно до них рекомендований тип тренда, який дасть найбільш об'єктивні результати розрахунків показників динаміки на прогнозований період. Безумовно, для більш точного статистичного прогнозування слід застосовувати множинні кореляційні моделі динамік», в яких би враховувалася дія факторів. Поки що теоретична база багатофакторного регресійного аналізу рядів динаміки розроблена і висвітлена недостатньо. Розглянемо основні методичні особливості його здійснення при вивченні економічних явищ.

Таблиця 71

Статистичні характеристики ряду динаміки врожайності (23 роки)

Культура	Статистична оцінка показників динаміки*			Тип тренда
	випадкова помилка різниць середніх (m)	нормоване відхилення (t)	значення інтеграла ймовірності $S_{(t)}$	
Озимі зернові	1,43	0,26	0,616	$\bar{y}_t = a_0 + a_1 t$
Кукурудза	3,21	0,53	0,688	$\bar{y}_t = a_0 a_1 t$
Картопля	1,70	1,29	0,894	$\bar{y}_t = a_0 + a_1 t$
Овес	2,63	1,30	0,894	$\bar{y}_t = a_0 a_1 t$
Овочі	1,72	0,16	0,578	$\bar{y}_t = a_0 + a_1 t$
Кормові коренеплоди	2,63	0,61	0,722	$\bar{y}_t = a_0 + a_1 t$
Силосні	2,14	0,61	0,722	$\bar{y}_t = a_0 + a_1 t$
Однорічні трави на сіно	0,79	0,50	0,688	$\bar{y}_t = a_0 + a_1 t$
Багаторічні трави на сіно	2,19	0,14	0,616	$\bar{y}_t = a_0 a_1 t$
Природні сіножаті	0,54	0,55	0,722	$\bar{y}_t = a_0 + a_1 t$

* Залежно від типу тренда визначали істотність різниць абсолютних приростів або темпів росту чи прискорень

Багатофакторні кореляційні моделі динаміки економічних явищ можуть бути побудовані на інформації різних ієрархічних рівнів і за неоднаковий період часу. Для таких моделей використовують ряди динаміки, що характеризують середні величини досліджуваних показників: 1) по країні в цілому; 2) по окремих галузях народного господарства; 3) по окремих галузях народного господарства за певний період часу, який приймається за одиницю виміру (наприклад, рік). Крім того, багатофакторна модель може будуватися на інформації, яка характеризує динаміку явища на кожному досліджуваному об'єкті (господарство, бригада, ферма), а також на просторовій інформації).

При побудові багатофакторних кореляційних моделей економічних явищ виникають дві математичні проблеми: автокореляція та мультиколінеарність.

Автокореляція у відхиленнях від трендів або регресійної моделі виникає з таких причин: 1) у моделі не враховано істотного фактора; 2) у моделі не враховано кілька неістотних факторів, дія яких збігається за напрямом і фазою; 3) неправильно обрано форму зв'язку залежної та незалежної змінних; 4) через особливість внутрішньої структури випадкової компоненти.

Для виявлення наявності автокореляції у відхиленнях від трендів або регресійної моделі розраховують критерій Дурбіна-Уотсона, який обчислюють за формулою:

$$d = \frac{\sum_{t=1}^{n} \left(\varepsilon_{t+1} - \varepsilon_t \right)^2}{\sum_{t=1}^{n} \varepsilon_t},$$

де ε_t – випадкові відхилення від тренда або регресійної моделі.

Розрахована таким чином величина d порівнюється з теоретичними її значеннями за стандартною математичною таблицею (додаток 14), яка містить відповідні нижні (d_1) та верхні (d_2) границі критерію Дурбіна-Уотсона, а також число змінних у кореляційній динамічній моделі (V_i) і довжину ряду динаміки (n).

При порівнянні фактичних і теоретичних величин критерію можливі три випадки: 1) $d < d_1$; 2) $d > d_2$; 3) $d_1 \le d \le d_2$. Дамо їм пояснення: 1 – гіпотеза про відсутність автокореляції у відхиленнях не приймається (відкидається); 2 – гіпотеза про відсутність автокореляції приймається; 3 – виникає потреба у подальших дослідженнях (наприклад, збільшення ряду динаміки).

Значення критерію Дурбіна-Уотсона коливається в межах $0 \leq d$ □ 4, при цьому величини їх різні для додатних та від'ємних коефіцієнтів. Для перевірки від'ємних автокореляцій обчислюють величину $(4 – d)$ й порівнюють її і за схемою, аналогічною у випадку додатної автокореляції.

Для зменшення автокореляції, крім виключення тренда, використовують й інший прийом: включення у множинну кореляційну моделі показника часу як аргументу (фактора), адже множинна регресія з відхиленнями від лінійних тенденцій еквівалентна прямому введенню часу в рівняння регресії (теорема Фриша і Воу).

До особливостей кореляційного аналізу в рядах динаміки слід віднести також мультиколінеарність, тобто наявність сильної кореляції між незалежними змінними, яка може існувати поза залежністю між результативною та факторними ознаками. Наявність мультиколінеарності в кореляційних моделях становить досить серйозну загрозу і для одержання об'єктивних оцінок взаємозв'язків, у зв'язку з чим ускладнюється сам процес здійснення аналізу. Пояснюється це такими причинами: 1) важко виділити найбільш істотні фактори, оскільки правило β-коефіцієнтів дійсне лише для некорельованих (або слабокорельованих) факторів; 2) викривлюється зміст коефіцієнтів регресії; 3) ускладнюється обчислювальний процес.

Розв'язання питань мультиколінеарності в кореляційному аналізі рядів динаміки повинне здійснюватися за такими етапами: 1) визначення факту наявності мультиколінеарності; 2) визначення області мультиколінеарності на множині факторів; 3) вимірювання ступеня мультиколінеарності; 4) з'ясування її причин; 5) розробка заходів щодо усунення мультиколінеарності.

Слід пам'ятати, що математична природа регресії виключає наявність лінійного зв'язку між незалежними змінними. Для економічних явищ ця обставина вважається не характерною, адже між факторами існують лінійні співвідношення, які знаходять своє відображення (в найпростішому випадку) у великих значеннях коефіцієнтів простої кореляції. Як вже відзначалося, фактори вважаються мультиколінеарними, якщо абсолютна величина парного І коефіцієнта кореляції перевищує 0,8.

Назвемо шляхи усунення (зменшення) в кореляційному аналізі мультиколінеарності в рядах динаміки: 1) побудова рівнянь регресії за відхиленнями від трендів; 2) одержання та залучення додаткової статистичної інформації; 3) перетворення множини незалежних змінних у кілька ортогональних (незалежних) множин із наступним використанням методів багатомірного статистичного аналізу (факторного, кластерного та ін.); 4) виключення з моделі одного або кількох лінійно зв'язаних факторів.

У спеціальній літературі рекомендується при багатофакторному регресійному аналізі в динаміці використовувати найпростіші – прямолінійні – залежності. Це пояснюється тим, що при складних дійсні залежності між явищами часто перекручуються. Вони також погано піддаються економічній інтерпретації.

ТЕМА 10. ІНДЕКСНИЙ МЕТОД

§ 10.1. Загальне поняття статистичних індексів. Основи індексного методу

В аналітичній роботі зі статистичними даними часто оперують різнорідними елементами. Наприклад, при аналізі сукупної зміни собівартості продукції рослинництва за певний проміжок часу мають справу з різними видами продукції – зерном, овочами, баштанними і т.д.; в аналізі сукупних змін затрат праці на виробництво картоплі – із різними видами польових робіт (оранка, боронування, культивація та ін.) Об’єднання різних елементів в одну сукупність називають **агрегатом.** Для аналізу змін, що відбуваються в таких агрегатах, найкращим прийомом вважається розрахунок індексів.

Статистичні індекси – це відносні величини, які одержують у результаті порівняння складних економічних явищ, утворених з різнорідних елементів, що не підлягають безпосередньому підсумовуванню.

Англійський термін «index number» означає «число – показник». Статистична практика широко використовує індекси при вивченні економічних явищ (хоча деякі економісти виявляються не підготовленими для такої роботи). Знання методології побудови індексів (далі будемо вживати термін «індексний комплекс») значно розширює аналітичні можливості дослідника, збагачує результативну інформацію досліджень.

За допомогою індексів можна характеризувати зміну як у часі, так і в просторі найрізноманітніших показників : обсягів виробленої продукції, посівних площ, урожайності, цін, вартості та собівартості, продуктивності праці і т.д. Їх поділяють на дві групи: до першої належать об'ємні (сумарні) показники (наприклад, розмір посівних площ, кількість худоби, обсяг продукції та ін.), які виражаються абсолютними величинами; до другої – показники, розраховані на певну одиницю (наприклад, урожайність, ціни, собівартість, продуктивність праці і т.д.). Останні умовно можна назвати якісними показниками, і виражаються вони у вигляді середніх величин. Зазначена особливість зумовлює поділ індексів на індекси **кількісних** та **індекси якісних показників.**

За допомогою статистичних індексів можна відображувати зміну у часі і просторі як окремих простих показників (наприклад, обсяг виробництва зерна, молока, м'яса і т.д.), так і однойменних показників по складних сукупностях (наприклад, зміна обсягу виробництва продукції рослинництва, тваринництва, по господарству в цілому і т.д.).

Класифікують індекси за ступенем охоплення елементів досліджуваного явища та способом побудови. За *ступенем охоплення елементів явища* індекси поділяють на індивідуальні й загальні.

Індекси, які відображують співвідношення простих одиничних показників, називають **індивідуальними,** а індекси, що характеризують зміну певного показника в цілому по будь-якій складній сукупності, - **загальними.** Останні в свою чергу розглядаються за широтою сукупності. Так, зведені індекси, що охоплюють всю сукупність досліджуваних явищ називають **тотальними** (або загальними), а індекси, які охоплюють частину (групу) елементів сукупності, - **груповими** (або субіндексами). Наприклад, індекс продукції рослинництва та індекс продукції тваринництва є груповими щодо тотального індексу продукції сільського господарства.

За способом побудови загальні індекси поділяють на агрегатні, середні із індивідуальних та середнього рівня.

Агрегатний індекс розраховують шляхом співвідношення двох сум. При цьому знаходять співмірник для різних елементів складного явища й додають елементи у звітному та базисному періодах, одержуючи відповідні суми.

Середній індекс визначають з індивідуальних індексів окремих елементів.

Індекс середнього рівня знаходять як співвідношення середніх величин поточного і базисного періодів.

Згідно із теоретичними концепціями агрегатні індекси вважаються основною формою економічних індексів, а середні із індивідуальних індексів – похідними, які одержують у результаті перетворення агрегатних індексів.

Обчислення загальних індексів, що дають змогу співвіднести між собою показники за складними сукупностями, являє собою особливий прийом дослідження, який називається **індексним методом.** За його допомогою можна не тільки вивчати динаміку показників, а й вимірювати вплив окремих факторів на динаміку складного показника. При цьому залежно від завдань аналізу можна фактори вивчати ізольовано, абстрагуючись від дії інших , або розглядати їх взаємопов’язано.

За допомогою індексного методу вирішуються такі завдання:

1) характеристика загальної зміни складного економічного явища чи окремих його елементів (складових); 2) виділення впливу одного з факторів шляхом елімінування впливу інших; 3) відокремлення впливу зміни структури явища на зміну індексованої величини.

Індексний метод має свою термінологію та символіку. Її дотримання є обов’язковою умовою в індексному аналізі. Розглянемо це питання дещо детальніше.

Для побудови статистичного індексу необхідно мати вихідну інформацію, як мінімум, за два періоди. Один з таких періодів називається базисним, другий – поточним. **Базисний -** це період, з яким порівнюють досліджувані явища, **поточний –** період, що порівнюється. Якщо досліджуються дані за кілька періодів, то один з них (як правило, початковий) буде базисним, а решта – поточними, або звітними.

Основними об’єктами для побудови індексних комплексів є кількість і ціна. Кількість позначають латинською літерою q (від лат. «quantitas»); ціну – латинською p (від лат. «pretium»). Важливе значення має **підписна нумерація.** За її допомогою позначається період, до якого належать дані. Так, якщо йдеться про кількість продукції за базисний період, то йому відповідає позначення q_0. Відповідно кількість продукції за перший поточний період позначається q_1, за другий поточний – q_2 і т.д. Аналогічні позначення вводять і для ціни.

Такі об'єкти для побудови індексних комплексів, як посівні площі, урожайність, затрати робочого часу на одиницю продукції та собівартість її одиниці, позначають відповідно символікою: П, *y,t,z.*

Виходячи з прийнятих позначень, для різних показників записують індивідуальні індекси, які позначають через *i*. Так, індивідуальний індекс обсягу виражається як $i = q_1 : q_0$, цін – $i = p_1 : p_0$, собівартості одиниці продукції – $i = z_1 : z_0$. Загальні індекси позначаються символом I і розраховуються дещо складніше.

У теорії індексів показник, зміну якого характеризує індекс, називають **індексованою величиною,** а пов'язану з нею величну, що використовують як постійну, **- елімінованою величиною,** або вагою. Остання відіграє роль сумірника. Використання зазначених двох видів величин вважається особливістю індексного методу аналізу.

Слід відзначити, що при побудові статистичних індексів насамперед необхідно вирішити такі питання: 1) який набір різнорідних елементів досліджуватиметься; 2) які показники виступатимуть індексованими величинами; 3) які величини виступатимуть сумірниками (вагами). При цьому встановлюють, які досліджувані показники при побудові індексів вважаються базисними, а які – поточними.

§ 10.2. Загальні індекси. Агрегатний індекс як основна форма індексу. Середні арифметичні й гармонійні індекси

Щоб розрахувати загальний індекс, необхідно подолати несумірність окремих елементів досліджуваної сукупності. Це досягається шляхом введення в індекс сумірника (ваги). Побудова формули загального індексу – одне з головних питань теорії індексів.

Агрегатний індекс вважається основною формою загального індексу. Його застосовують для вивчення складних суспільних явищ, які містять у собі різнойменні елементи. Найбільш типовим загальним індексом кількісних показників є *індекс фізичного обсягу.* Його можна побудувати двома способами: як агрегатний і як середній із індивідуальних. Наприклад, є дані про виробництво різних видів продукції в межах одного господарства за два періоди. Необхідно за допомогою загального індексу характеризувати зміни обсягу всієї продукції. Знаходимо загальний сумірник, який дає змогу виразити у співмірному вигляді загальний обсяг продукції в базисному та звітному періодах. Таким сумірником можуть бути: ціна, собівартість одиниці продукції, затрати праці на одиницю продукції і т.д.

Використовуючи прийняту символіку, виразимо вартість продукції в базисному і звітному періодах як q_0p_0 і q_1p_1. Порівнянням цих двох показників одержуємо агрегатний індекс вартості :

I= Σ q_1p_1: Σ q_0p_0. Оскільки вартість залежить від кількості продукції та цін, індекс може відобразити зміну обсягу виробництва продукції лише за умови постійності цін на окремі її види.

Побудований за цим принципом індекс називають **індексом фізичного обсягу.** Його формула має вигляд : $I_{фо} = \sum q_1 p : \sum q_0 p$, де q_0, q_1 – кількість виробленої продукції в базисному і звітному періодах ; p – ціни, зіставні для двох періодів.

Таким чином, **агрегатним індексом** називається загальний індекс, одержаний шляхом зіставлення підсумків, які виражають величину складного показника у звітному та базисному періодах, за допомогою сумірників (незмінних). Сам спосіб обчислення загального індексу називають **агрегатним.** Порівнювані суми в агрегатному індексі різняться між собою за індексованими величинами, сумірники тут незмінні.

Таблиця 72

Макет розрахунку індексу фізичного обсягу

Вид продукції	Обсяг виробництва, ц		Ціна одиниці продукції, грн.		Вартість продукції, тис. грн.	
	базисний період	звітний період	базисний період	звітний період	базисний період	звітний період
	q_0	q_1	p_0	p_1	q_0p_0	q_1p_0
...	...	...	...	...	...	...
Всього	-	-	-	-	130000	162000

Технічний бік розрахунків ілюструє макет, наведений у таблиці 72. Індекс фізичного обсягу одержимо порівнянням вартості продукції звітного і базисного років:

$$I_{фо} = \frac{\sum q_1 p}{\sum q_0 p} = \frac{162000}{130000} = 1.246$$ або 124,6 %.

Отже, загальний обсяг виробництва продукції збільшився у звітному році порівняно з базисним на 24,6 %. Різниця між чисельником і знаменником формули характеризує абсолютну зміну обсягу виробництва в поточному періоді за рахунок його збільшення. У нашому прикладі ця величина дорівнює 32000 грн. ($\sum q_1 p - \sum q_0 p$).

Загальний індекс може бути розрахований і як **середній із індивідуальних.** У такому випадку визначають індивідуальні індекси обсягу по окремих видах продукції *i*= q_1: q_0. З одержаних індивідуальних індексів розраховують середній.

У статистичній практиці середні індекси визначають у формі середньої арифметичної і середньої гармонійної. При цьому кожна з обраних форм повинна прийматися як середня зважена, тобто : $\bar{I} = \sum if : \sum f$; $\bar{I} = \sum M : \sum (M : i)$, де *i*- індивідуальні індекси обсягу; f і M – ваги відповідно в середньому арифметичному та середньому гармонійному індексах.

При визначенні ваг середнього арифметичного і середнього гармонійного індексів виходять із тотожності їх агрегатного індексу. Так, при обчисленні середнього арифметичного індексу повинна виконуватись умова :

$$\frac{\sum if}{\sum f} = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

Остання матиме місце при f = $q_0 p_0$. Дійсно, $\bar{I} = \sum i q_0 p_0 : \sum q_0 p_0 = \sum (q_1 : q_0) q_0 p_0 : \sum q_0 p_0$.

Отже, загальний індекс у формі середньоарифметичного матиме вигляд : $\bar{I} = \sum i q_0 p_0 : \sum q_0 p_0$.

Розраховуючи (аналогічно) ваги середньогармонійного індексу, слід пам'ятати умову $\sum M : \sum (M : i) = \sum q_1 p_0 : \sum q_0 p_0$. Така рівність буде дотримана, якщо M=$q_1 p_0$. Тоді $\bar{I} = \sum q_1 p_0 : \sum (q_0 p_0 : i) = \sum q_1 p_0 : \sum (q_1 p_0 : q_1) q_0 = \sum q_1 p_0 : \sum q_0 p_0$, тобто середньо - гармонійний індекс фізичного обсягу можна записати у вигляді : $\bar{I} = \sum q_1 p_0 : \sum (q_1 p_0 : i)$.

Вибір форми обчислення середнього індексу зумовлений насамперед наявністю в розпорядженні дослідника вихідної інформації поряд з індивідуальними індексами. Так, при наявності даних про вартість продукції в порівняльних цінах у базисному періоді загальний індекс із індивідуальних розраховують як середній арифметичний. Розглянемо послідовність таких дій на базі даних таблиці 73.

Таблиця 73

Вихідні і розрахункові дані для обчислення середнього арифметичного індексу

Вид продукції	Індивідуальний індекс обсягу	Вартість продукції у базисному періоді, тис. грн.
	$i = q_1 : q_0$	q_0p_0
Зерно	1,16	36200
Овочі	1,27	17900
Молоко	0,87	36700

$$\bar{I} = \frac{\sum iq_0p_0}{\sum q_0p_0} = \frac{1{,}16 \cdot 36200 + 1{,}27 \cdot 17900 + 0{,}87 \cdot 36700}{36200 + 17900 + 36700} = 1{,}064,$$

або 106,4 %.

Якщо в розпорядженні є дані про вартість продукції у звітному періоді в базисних цінах (табл.74), то загальний індекс визначають за принципом гармонійної середньої. Так, у нашому прикладі маємо:

$$\bar{I} = \frac{\sum iq_1p_0}{\frac{\sum q_1p_0}{i}} = \frac{17900 + 15500 + 33200}{\frac{17900}{1{,}09} + \frac{15500}{1{,}28} + \frac{33200}{1{,}00}} = 1{,}079,$$

або 107,9 %.

Таблиця 74

Вихідні і розрахункові дані для обчислення середнього гармонійного індексу

Вид продукції	Індивідуальний індекс обсягу	Вартість продукції у звітному періоді в цінах базисного , тис. грн.
	$i = q_1 : q_0$	q_1p_0
Зерно	1,09	17900
Овочі	1,29	15500
Картопля	1,00	33200

Між розглянутими вище видами індексів існує певне співвідношення, згідно з яким середній індекс повинен дорівнювати агрегатному.

§ 10.3. Система індексів для характеристики динаміки складного явища

Із розглянутого вище зрозуміло, що явища, динаміка яких вимірюється індексами, складаються з різнорідних елементів. Це зумовлює неможливість вимірювання рівнів таких явищ. Із цього випливає необхідність їх вимірювання в динамічному розрізі. Тому при обчисленні індексу завжди мають справу з двома рядами величин, які характеризують базисний і поточний періоди.

Залежно від періоду часу, який береться за основу при побудові індексів, останні можуть бути базисними та ланцюговими. **Базисними** називають індекси, які мають один і той самий період часу, взятий за основу розрахунків, тобто постійну базу порівняння. Наприклад, якщо при розрахунках індексів за 2001 – 2004 рр. за базу порівняння взяти 2001 р., то такі індекси називатимуться базисними. Якщо ж при обчисленні індексів база порівняння змінюватиметься і за таку базу братимуть період, який є попереднім щодо обчислюваного індексу, то останній називатиметься **ланцюговим.** Як базисний, так і ланцюговий індекси дають кількісну характеристику темпів розвитку явища.

Розглянемо приклад (табл.75). При обчисленні базисних індексів за базу порівняння прийнято 2001 р. Останній рядок таблиці показує: а) по графі базисних індексів – в 2004 р. порівняно з 2001 – м середня врожайність збільшилася на 1,5 % (індекс 1,015) ; б) по графі ланцюгових індексів – в 2004 р. порівняно з 1994 – м середня урожайність зменшиласЬ на 1,8 % (індекс 0,982).

Між базисними і ланцюговими індексами існує певний взаємозв'язок: добуток ланцюгових індексів дорівнює базисному індексу останнього періоду 1,018×1,015×0,982×=1,015.

Таблиця 75

Динаміка врожайності зернових культур у господарстві

Рік	Урожайність, ц/га	Індекси	
		базисні	ланцюгові
2001	33,1	-	-
2002	33,7	1,018	1,018
2003	34,2	1,033	1,015
2004	33,6	1,015	0,982

Щоб довести правильність цього рівняння, перемножимо ті дроби, на підставі яких були обчислені ланцюгові індекси, та здійснимо відповідні скорочення:

$$\frac{33{,}7}{33{,}1}\times\frac{34{,}2}{33{,}7}\times\frac{33{,}6}{34{,}2}=1{,}015.$$

Отже, одержано базисний індекс 2004р. У результаті скорочення залишаються лише знаменник першої ланки і чисельник останньої, що дає тотожність з базисним індексом : $i = \frac{33,6}{33,1} = 1,015$.

Зазначена взаємозалежність на прикладі індивідуальних індексів поширюється і на співвідношення агрегатних. Але таке співвідношення має місце за умови, що зважування, здійснюється лише постійними сумірниками (вагами)наприклад, базисного періоду.

Розглянемо зазначену залежність на прикладі співвідношення індексів фізичного обсягу:

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_2 p_0}{\sum q_1 p_0} \cdot \frac{\sum q_3 p_0}{\sum q_2 p_0} = \frac{\sum q_3 p_0}{\sum q_0 p_0}.$$

Як бачимо, знаменник другої ланки скорочується з чисельником першої, знаменник третьої ланки – із чисельником другої і залишаються чисельник третьої ланки та знаменник першої, що тотожне базисному індексу.

Слід відзначити, що при інтерпретації індексів, якими вимірюють динаміку явищ, іноді застосовують термін «пункт». Під останнім розуміють одиницю, якщо база порівняння при обчисленні індексу виражена у вигляді 100 %. Наприклад, якщо індекс обсягу виробництва продукції підвищився з 125 % у 2003 р. до 135 % у 2004 – му при базі порівняння 2000 р. 100 %, то можна сказати, що індекс збільшився на 10 пунктів.

Індекси з постійними і змінними вагами. Як уже згадувалося, взаємозв'язок між базисними і ланцюговими індексами (добуток ланцюгових дорівнює базисному) є безумовним лише для індивідуальних індексів. Для загальних такий взаємозв'язок не порушується, якщо ряд загальних індексів розраховано з постійною вагою.

Приклад. Маємо дані по підприємству про виробництво продукції та про ціни на неї за чотири роки:

2001	2002	2003	2004
q_1/p_1	q_2/p_2	q_3/p_3	q_4/p_4

При обчисленні базисних і ланцюгових індексів фізичного обсягу можна по-різному вирішити питання про елеміновані величини (ваги). Так, при визначенні ланцюгових індексів фізичного обсягу продукцію всіх періодів можна оцінити в одних і тих самих цінах (наприклад, у цінах 2001 р.) такі індекси мають вигляд:

$$I_{2/1}=\frac{\sum q_2p_1}{\sum q_1p_1};\quad I_{3/2}=\frac{\sum q_3p_1}{\sum q_2p_1};\quad I_{4/3}=\frac{\sum q_4p_1}{\sum q_3p_1}.$$

Оскільки розраховані таким чином індекси мають однакові ваги (p_1) , то представляють ряд **індексів з постійними вагами.** У цьому випадку можна перейти від ланцюгових індексів до базисного (і навпаки):

$$\frac{\sum q_2p_1}{\sum q_1p_1}\cdot\frac{\sum q_3p_1}{\sum q_2p_1}\cdot\frac{\sum q_4p_1}{\sum q_3p_1}=\frac{\sum q_4p_1}{\sum q_1p_1}.$$

Але при побудові ряду ланцюгових індексів можна було б піти іншим шляхом : для кожного періоду побудувати індекс фізичного обсягу за цінами попереднього періоду:

$$I_{2/1}=\frac{\sum q_2p_1}{\sum q_1p_1};\quad I_{3/2}=\frac{\sum q_3p_2}{\sum q_2p_2};\quad I_{4/3}=\frac{\sum q_4p_3}{\sum q_3p_3}.$$

Ці індекси побудовані за різними вимірниками (вагами), тобто вони є **індексами зі змінними вагами.** Для таких індексів перехід від ланцюгових до базисних (і навпаки) неможливий.

Відзначаючи позитивну особливість індексів із постійними вагами, що зумовлює перехід від ланцюгових індексів до базисних (і навпаки), слід вказати й на недоліки деяких видів економічних індексів. Так, чи можна вважати доцільною побудову індексів цін із постійними вагами? І дійсно, який сенс розрахунку індексу цін у четвертому періоді (порівняно з третім) за продукцією першого періоду (q_1)? З практичного боку для цін значно більший інтерес становлять індекси зі змінними вагами, хоча їм і не притаманна вказана вище взаємозалежність між ланцюговими та базисними індексами.

Індекси змінного і постійного складу. У розглянутих вище прикладах йшлося про випадки, коли для сукупності невимірних показників у натуральному вигляді визначалася середня зміна індексованих величин. При вивченні динаміки якісних показників часто треба визначати зміну середньої величини індексованого показника для будь – якої однорідної сукупності (наприклад, середньої врожайності, середньої трудоємності надою, середньої собівартості і т.д.).

У загальному вигляді динаміку таких середніх показників можна представити виразом ($x_1 : x_0$), який являє собою **середній індекс.**

Відносну величину, що характеризує динаміку двох середніх показників для однорідної сукупності, в статистиці називають **індексом змінного складу.**

Для якісних показників, наприклад урожайності і цін, індекси змінного складу легко записати у вигляді таких відношень:

$$I_y = \bar{y}_1 : \bar{y}_0 = \frac{\sum y_1\Pi_1}{\sum \Pi_1} : \frac{\sum y_0\Pi_0}{\sum \Pi_0};$$

$$I_p = \bar{p}_1 : \bar{p}_0 = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0}.$$

Назва індексу змінного складу зумовлена тим, що середні величини, динаміку яких вони відображують, можуть змінюватися залежно від змін кожної окремої одиниці досліджуваного явища та від змін його структури. Наприклад, збільшення середньої врожайності зернових культур залежить від підвищення врожайності кожної окремої культури (вівса, ячменю, гречки і т.д.) та від збільшення питомої ваги в загальній площі зернових найбільш урожайних культур.

Таким чином, індекс змінного складу характеризує спільний вплив зазначених вище факторів. У такому індексі знаходить прояв зміна обох величин – кількісних (Π_0, Π_1) т а якісних (y_0 і y_1).

У загальному вигляді формула індексу змінного складу така:

$$I_{зм.скл.} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} = \bar{x}_1 : \bar{x}_0,$$

де $\bar{x}$- осереднювана ознака ; f – вага (питома вага) досліджуваного явища.

Якщо дослідження має на меті виключити вплив змін структури сукупності на динаміку середніх показників розраховують середні для двох періодів по одній і тій самій структурі, що, як правило, фіксується по звітному періоду. Індекс, який відображує динаміку середніх величин при фіксованій структурі явища, називається **індексом постійного (фіксованого) складу.** Він характеризує вплив лише індексованої величини. Його структурна формула має вигляд:

$$I_{ф.скл.} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1}$$ або

$$I_{ф.скл.} = \frac{\sum q_1 p_1}{\sum q_1} : \frac{\sum q_0 p_1}{\sum q_1}$$

При скороченні в наведеній формулі на Σf_1 одержуємо вже відому формулу агрегатного індексу : $I = \sum x_1 f_1 : \sum x_0 f_{1\cdot 1}$.

У даному індексі вплив структурного фактора виключено. Прикладом таких індексів є індекси фізичного обсягу ($\frac{\sum q_1 p_0}{\sum q_0 p_0}$), індекси цін ($\frac{\sum p_1 q_1}{\sum p_0 q_1}$), індекси собівартості ($\frac{\sum Z_1 q_1}{\sum Z_0 q_1}$) та ін.

Відношенням індексів змінного складу до індексу фіксованого складу одержують **індекс структури :** $I_{стр.} = (\sum x_0 f_1 : \sum f_1) : (\sum x_0 f_0 : \sum f_0)$, тобто $I_{стр.} = I_{ф.скл.} : I_{зм.скл.}$. Таким чином завжди можна обчислити один з індексів, якщо відомі два інших. Даний індекс характеризує **вплив змін структури на зміну середньої величини.**

Таблиця 76

Вихідні і розрахункові дані для обчислення індексів змінного та фіксованого складу

Культура	Площа, га		Урожайність, ц/га		Валовий збір, ц		
	базисний період	звітний період	базисний період	звітний період	базисний період	звітний період	умовний
	$П_0$	$П_1$	y_0	y_1	y_0 $П_0$	y_1 $П_1$	y_0 $П_1$
Пшениця	500	590	30	35	15000	20650	17700
Жито	200	250	12	15	2400	3750	3000
Разом	700	840	-	-	17400	24400	20700

Приклад. Розглянемо розрахунку зазначених вище індексів за даними таблиці 76. Середня врожайність базисного і звітного періодів становить:

$$\overline{y_0} = \frac{\sum y_0 П_0}{\sum П_0} = \frac{17400}{700} = 24{,}9; \quad \overline{y_1} = \frac{\sum y_1 П_1}{\sum П_1} = \frac{24440}{840} = 29{,}0.$$

Обчислимо індекси врожайності змінного і постійного складу та їх співвідношення:

$$I_{зм.скл.} = \frac{\sum y_1 П_1}{\sum П_1} : \frac{\sum y_0 П_0}{\sum П_0} = \frac{24400}{840} : \frac{17400}{700} = 1{,}1647;$$

$$I_{ф..скл.} = \frac{\sum y_1 П_1}{\sum П_1} : \frac{\sum y_0 П_1}{\sum П_1} = \frac{24400}{840} : \frac{20700}{840} = 1{,}1789;$$

$$I_{стр.} = \frac{\sum y_0 П_1}{\sum П_1} : \frac{\sum y_0 П_0}{\sum П_0} = \frac{20700}{840} : \frac{17400}{700} = 0{,}9880;$$

$$I_{стр..} = I_{зм.скл.} : I_{ф.скл.} = 0{,}9880 = 1{,}1647 : 1{,}1789;$$

$$I_{зм.скл.} = I_{ф.скл.} \cdot I_{стр.} = 1{,}1647 = 1{,}1789 \cdot 0{,}9880.$$

Прокоментуємо результати розрахунків. Індекс урожайності змінного складу відображує динаміку середньої врожайності. У нашому випадку середня врожайність зернових культур підвищилася на 16,5 % як за рахунок збільшення врожайності окремих культур, так і за рахунок змін структури посівних площ. Перш ніж визначити вплив на динаміку врожайності змін структури посівних площ ($I_{стр.}$), розраховують індекс урожайності фіксованого складу ($I_{ф.скл}$). Він виключає вплив структури посівних площ. Індекс дорівнює 1,1789 і показує, що середня врожайність у звітному періоді порівняно з базисним підвищилася на 17,9 % за рахунок зростання врожайності окремих культур.

Розбіжність в індексах змінного та постійного складу можна пояснити впливом змін структури посівних площ. Величина індексу структури 0,988 свідчить про те, що зменшення питомої ваги високоврожайних культур у звітному періоді (пшениці – від 71 до 70 %) зумовило деяке зниження показника середньої врожайності (на 1,2 %).

Аналогічно можна розрахувати індекси змінного та фіксованого складу і для інших якісних показників у випадках, коли йдеться про сукупності, для яких розраховується середня величина індексованого показника (наприклад, для показників цін, собівартості, продуктивності праці і т.д.).

§ 10.4. Види економічних індексів, їх взаємозв'язок

Індекси фізичного обсягу. В аналізі соціально – економічних явищ часто використовується **індекс фізичного обсягу.** Він широко застосовується в наукових дослідженнях і практичних розрахунках, які стосуються вивчення динаміки виробництва продукції або ступеня виконання плану. Обчислюють цей індекс, нагадаємо, за формулою агрегатного індексу.

За інформацією макету таблиці 77, знаходять величину індексу фізичного обсягу, який характеризує зміну обсягу виробництва продукції в поточному періоді щодо базисного або планового періоду. Суми трьох останніх граф робочої таблиці дають вихідну інформацію для розрахунку індексів за такими формулами: $I_{ф.о.} = \sum q_1 p_0 : \sum q_0 p_0$; $I_{ф.о.} = \sum q_1 p_0 : \sum q_{gk/} p_0$. Перший із наведених індексів характеризує динаміку виробництва продукції, другий – ступінь виконання плану.

При обчисленні індексу фізичного обсягу індексуються обсяги виробництва продукції, елімінуються – ціни. Причому за незмінні ваги приймаються ціни базисного періоду. Отже, чисельник першого індексу являє собою умовну вартість виробленої продукції, знаменник – фактичну вартість продукції базисного періоду. Якщо обчислена величина цього індексу становить, наприклад, 1,163, це означає, що в поточному періоді порівняно з базисним було вироблено продукції на 16,3 % більше. За аналогічною схемою розраховують та інтерпретують індекс фізичного обсягу, коли базою порівняння є планові показники. При відсутності необхідних даних індекс фізичного обсягу обчислюють за формулою середнього арифметичного індексу.

Таблиця 77

Макет таблиці для обчислення індексу фізичного обсягу

Вид продукції	Виробництво продукції, ц			Ціна, грн.	Вартість продукції, тис. грн.		
	базисний період	план	поточний період		базисний період	поточний період	
						план	фактично
	q_0	$q_{пл.}$	q_1	p_0	q_0p_0	$q_{пл.}p_0$	q_1p_0
Зерно							
Овочі							
Картопля							
і т.д.							
Всього	-	-	-	-	$\sum q_0p_0$	$\sum q_{пл.}p_0$	$\sum q_1p_0$

Індекси цін. Досить важливим завданням статистики є характеристика динаміки цін. Воно вирішується за допомогою **агрегатних індексів цін.** Для їх розрахунку береться вартість однієї й тієї ж продукції у поточних та базисних цінах. Обчислюють індекс за схемою аналогічною індексу фізичного обсягу з тією різницею, що індексуються тут показники цін (p_1, p_0), а елімінованою величиною (вагою) приймається кількість продукції поточного періоду (q_1). Одержані суми вартості продукції з робочої таблиці (вона будується аналогічно попередній) підставляють у формулу індексу цін $I_ц = \sum p_1q_1 : \sum p_0q_1$.

Отже, чисельник формули являє собою вартість продукції поточного періоду, знаменник – умовну вартість. Наприклад, якщо вивчається динаміка цін на продукцію, реалізовану на ринку, й обчислена величина індексу цін становить 0,841, це означає, що ціни на реалізовану продукцію в поточному періоді порівняно з базисним знизилися на 15,9 %.

Агрегатний індекс цін розраховують у випадках, коли наявні дані про реалізовану продукцію в поточному періоді у цінах поточного та базисного періодів. Якщо ж є інформація про зміни обсягів реалізації і цін по окремих видах продукції, то для характеристики динаміки цін в цілому по всіх видах продукції розраховують середній гармонійний індекс цін: $I_ц = \sum p_1 q_1 : \sum \frac{p_1 q_1}{i}$.

Таблиця 78

Вихідні і розрахункові дані для обчислення середнього гармонійного індексу цін

Вид продукції	Обсяг реалізації в поточному періоді, тис. грн.	Зміни (+,-) в поточному періоді порівняно з базисним, %	Обсяг реалізації в поточному періоді в цінах базисного
	$p_1 q_1$	$i \times 100 - 100$	$\sum \frac{p_1 q_1}{i} = \sum p_0 q_1$
Зерно	6018	-16	7164
Овочі	3227	+9	2961
Молоко	1639	-3	1690
Всього	10884	-	11815

Розрахунок цього виду індексу розглянемо за даними, наведеними в таблиці 78. Як бачимо, у вихідній інформації відсутні дані про вартість реалізованої продукції у поточному періоді за цінами базисного. Тому динаміку цін розкриває середній гармонійний індекс:

$$I_ц = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i}} = \frac{10884}{11815} = 0.921$$
.

Таким чином, ціни в поточному році порівняно з базисним на реалізовану продукцію в цілому знизилися в середньому на 7,9%.

Індекс продуктивності праці. Оскільки продуктивність праці вимірюється кількістю продукції, виробленою за одиницю часу, або затратами робочого часу на виробництво одиниці продукції, для визначення рівня продуктивності праці необхідно знати обсяг продукції та час, затрачений на її виробництво. Використовують також вартісні показники продуктивності праці, які одержують шляхом ділення вартості продукції на затрачений робочий час.

При вивченні динаміки продуктивності праці на виробництві окремих видів продукції розраховують індивідуальні індекси продуктивності праці. При обчисленні їх за сукупністю різнорідної продукції співвідношення затрат робочого часу на виробництво продукції в поточному періоді визначають за продуктивністю праці вбазисному періоді та фактичними затратами робочого часу на продукцію звітного періоду. Структурна формула цього співвідношення має вигляд : $I = \sum t_0 q_1 : \sum t_1 q_1$, де t_0, t_1- затрати робочого часу на одиницю продукції відповідно у базисному і звітному періодах; q_1- кількість продукції за звітний період.

Таблиця 79

Вихідні і розрахункові дані обчислення агрегатного індексу продуктивності праці (трудового)

Вид продукції	Кількість виробленої продукції в поточному періоді, ц	Затрати праці на 1 ц продукції, люд.-год.		Затрати робочого часу, люд.-год.	
		базисний період	поточний період	умовні	фактичні
	q_1	t_0	t_1	$t_0 q_1$	$t_1 q_1$
А					
Б					
В					
і т.д.					
Всього	-	-	-	$\sum t_0 q_1$	$\sum t_1 q_1$

Обернену величину індексу продуктивності праці називають індексом затрат робочого часу. Її розраховують за формулою : $I = \sum t_1 q_1 : \sum t_0 q_1$.

Наведені індекси називають **трудовими індексами продуктивності праці.** Для зручності їх розрахунку складають робочу таблицю (табл. 79). Підсумки по останніх двох графах підставляють у чисельник та знаменник формули агрегатного індексу продуктивності праці. У статистиці він має назву «трудовий індекс продуктивності праці фіксованого складу» .

На практиці широко застосовується вартісний метод вимірювання динаміки продуктивності праці. Такий методичний підхід дає змогу одержати узагальнюючу характеристику динаміки продуктивності праці шляхом обчислення **вартісного індексу продуктивності праці.** Формула цього агрегатного змінного індексу змінного складу має вигляд: $I = (\sum q_1 p : \sum t_1 q_1) : (\sum q_0 p : \sum t_0 q_0)$, де $\sum q_0 p, \sum q_1 p$ - вартість одержаної продукції у порівнянних цінах відповідно у базисному та звітному періодах; $\sum t_0 q_0, \sum t_1 q_1$- затрати робочого часу в базисному і звітному періодах.

Розрахунки вартісного індексу здійснюють на базі допоміжної робочої таблиці (табл. 80). Підставивши у чисельник і знаменник підсумкові дані відповідних граф, одержимо агрегатний індекс продуктивності праці (вартісний) змінного складу.

Таблиця 80

Вихідні і розрахункові дані для обчислення агрегатного індексу продуктивності праці (вартісного)

Вид продукції	Кількість виробленої продукції, ц		Затрати праці на всю продукцію, люд.-г		Ціна, грн.	Вартість валової продукції, грн.	
	базисний період	поточний період	базисний період	поточний період		базисний період	поточний період
	q_0	q_1	$t_0\ q_0$	$t_1\ q_1$	p	q_0p	q_1p
Зерно							
Овочі							
Картопля							
і т.д.							
Всього	-	-	$\sum t_0\ q_0$	$\sum t_1 q_1$	-	$\sum q_0$p	$\sum q_1$p

Між розглянутими вище видами індексів існує така залежність: вартісний індекс продуктивності праці дорівнює відношенню індексу фізичного обсягу до індексу затрат праці, а саме:

$$\frac{\sum q_1 p}{\sum t_1 q_1} : \frac{\sum q_0 p}{\sum t_0 q_0} = \frac{\sum q_1 p}{\sum q_0 p} : \frac{\sum t_1 q_1}{\sum t_0 q_0},$$

Індекс собівартості. На підставі даних про кількість виробленої в підприємстві за два роки продукції та собівартість її одиниці можна обчислити середню зміну собівартості одиниці продукції по галузі (рослинництво, тваринництво) та підприємству в цілому. У даному випадку індексується собівартість одиниці продукції, а елімінується – кількість продукції поточного періоду. Розраховують індекс собівартості за схемою агрегатного загального індексу фіксованого складу $I_{соб..} = \sum Z_1 q_1 : \sum Z_0 q_1$, де $z_{0,}$ z_1 – собівартість одиниці продукції базисного і поточного періодів; q_1- кількість продукції поточного періоду.

Індекс загальних витрат обчислюють за формулою агрегатного індексу змінного складу : $I_{заг.витр...} = \sum Z_1 q_1 : \sum Z_0 q_0$. Перший із наведених індексів характеризує динаміку змін собівартості одиниці продукції в поточному періоді щодо базисного, другий – динаміку змін загальних витрат. Обчислюють ці індекси а підставі даних попередніх розрахунків, як показано в таблиці 81.

Таблиця 81

Вихідні і розрахункові дані для обчислення індексів собівартості та витрат

<table>
<tr><td rowspan="2">Вид продукції</td><td colspan="2">Кількість виробленої продукції, ц</td><td colspan="2">Собівартість одиниці продукції, грн.</td><td colspan="3">Витрати виробництва, грн.</td></tr>
<tr><td>базисний період</td><td>поточний період</td><td>базисний період</td><td>поточний період</td><td>базисний період</td><td>поточний період</td><td>умовні</td></tr>
<tr><td></td><td>q_0</td><td>q_1</td><td>z_0</td><td>z_1</td><td>$q_0 z_0$</td><td>$q_1 z_1$</td><td>$q_1 z_0$</td></tr>
<tr><td>А</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>В</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>В</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>і т.д.</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Всього</td><td>-</td><td>-</td><td>-</td><td>-</td><td>$\Sigma q_0 z_0$</td><td>$\Sigma q_1 z_1$</td><td>$\Sigma q_1 z_0$</td></tr>
</table>

Вибір ваги (сумірника) індексу. Слід відзначити, що в теорії індексного методу дискусійним вважається питання вибору елімінованої величини (ваги) при побудові агрегатних індексів. У статистичній практиці існує такий методичний підхід до побудови індексних комплексів : якщо результативний показник являє собою добуток об’ємного (кількісного) і якісного показників, то при визначенні впливу першого на результативний - якісний показник слід фіксувати на рівні базисного періоду. Наприклад, при побудові агрегатного індексу $\Sigma П_1 y_0$: $\Sigma П_0 у_0$ (де $П_0$, $П_1$- площі зернових у базисний та звітний періоди ; y_0 - урожайність) індексуються об’ємні показники ($П_0$, $П_1$), а елімінується якісний показник (y_0), тобто урожайність у базисному періоді. Економічний зміст одержаного результату зводиться до висновку: як змінився валовий збір зернових культур за рахунок кількісних змін посівних площ у поточному періоді.

Якщо ж визначається вплив якісного показника, то об’ємний (кількісний) фіксується на рівні поточного періоду. Так, при обчисленні агрегатного індексу $\Sigma\ y_1 П_1$: $\Sigma\ у_0 П_1$ індексуються показники урожайності ($у_0$,$у_1$), а вагою виступає показник площі ($П_1$) у поточному періоді. Одержаний результат характеризує зміну валового збору (однорідних культур) у поточному році за рахунок змін рівня врожайності. Аналогічний методичний підхід використовують при побудові агрегатних індексів цін ($\Sigma p_1 q_1 : \Sigma p_0 q_1$), собівартості ($\Sigma z_1 q_1 : \Sigma z_0 q_1$), продуктивності праці ($\Sigma t_0 q_1 : \Sigma t_1 q_1$) та ін.

При обчисленні зазначених вище індексів якісних показників виходять із суті останніх як економічної категорії, а не з математичного оформлення індексних комплексів. Наприклад, індекс цін повинний дати відповідь, на скільки та чи інша кількість продукції, оцінена за цінами попереднього і звітного періодів, змінила б свою грошову вартість. Отже, якщо ми вивчаємо реальну зміну цін, то нас повинна цікавити зміна їх на продукцію, яку одержуємо зараз, тобто в поточному періоді (q_1). З аналогічних міркувань виходять при виборі ваги для розрахунку індексів урожайності, собівартості, продуктивності праці тощо.

При обчисленні індексів кількісних показників, наприклад індексу фізичного обсягу, як уже згадувалося, сумірником береться ціна базисного періоду. Це зумовлено тим, що більш точно відобразити зміну кількості виробленої продукції можна лише, якщо виходити з передбачення, що в поточному періоді змінилася лише кількість виробленої продукції, а ціни залишилися на рівні базисного періоду.

Викладений порядок розрахунку індексів є загальноприйнятим, але залежно від завдань дослідження можуть бути певні відхилення від нього. Особливо це стосується розрахунку конкретних економічних індексів у статистичних дослідженнях.

§ 10.5 . Взаємозв’язок статистичних індексів. Визначення впливу окремих факторів

Уже відзначалося, що суспільно – економічні явища і процеси перебувають у взаємозалежності та взаємозумовленості. Тому значна частина статистичних показників взаємопов’язані. Наприклад, валовий збір є добутком показників урожайності на площу; виробничі витрати – добутком обсягу виробництва на собівартість; товарооборот - добутком кількості реалізованої продукції на ціну і т.д. Аналогічна взаємозалежність існує і між економічними індексами. Так, індекс валового збору дорівнює добутку індексу врожайності на індекс посівних площ; індекс товарообороту – добутку індексу фізичного обсягу на індекс цін і т.д.

Нижче наведено схему взаємозв’язків індексів, які найчастіше зустрічаються в економічних дослідженнях та аналізі.

Зміну валового збору однорідної групи культур визначають за допомогою наведених індексів та їх співвідношень:

Індекс валового збору - $I_{в.з.} = \frac{\sum П_1 y_1}{\sum П_0 y_0}$. Індекс урожайності – $I_{ур} = \frac{\Sigma П_1 y_1}{\Sigma П_1 y_0}$. Індекс розміру і структури посівів - $I_{рсп} = \frac{\Sigma П_1 y_0}{\Sigma П_0 y_0}$. Індекс розміру посівної площі - $I_{п.п} = \frac{\Sigma П_1}{\Sigma П_0}$. Індекс урожайності зі «строкатого» гектара - $I_{усг} = \frac{\Sigma П_1 y_1}{\Sigma П_1} : \frac{\Sigma П_0 y_0}{\Sigma П_0} = \frac{\overline{y_1}}{\overline{y_0}}$. Взаємозв'язок індексів - $I_{в.з.} = I_{ур} \times I_{рсп}$; $I_{в.з.} = I_{п.п} \times I_{усг}$.

Ступінь впливу змін у структурі посівних площ на валовий збір знаходять за таким співвідношенням індексів :

$\frac{\Sigma П_1 y_0}{\Sigma П_0 y_0} : \frac{\Sigma П_1}{\Sigma П_0}$ або $\frac{\overline{y_1}}{y_0} : \frac{\sum П_1 y_1}{\sum П_1 y_0}$.

Зміни в рівні товарообороту продукції можна визначити добутком індексу фізичного обсягу на індекс цін:

$$\frac{\Sigma q_1 p_1}{\Sigma q_0 p_0} = \frac{\Sigma q_1 p_0}{\Sigma q_0 q_0} \cdot \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0}.$$

Аналогічний взаємозв'язок індексів знаходимо при вивченні змін витрат виробництва, а саме: добуток індексу фізичного обсягу на індекс собівартості :

$$\frac{\Sigma q_1 z_1}{\Sigma q_0 z_0} = \frac{\Sigma q_1 z_0}{\Sigma q_0 z_0} \cdot \frac{\Sigma q_1 z_1}{\Sigma q_1 z_0}.$$

При вивченні змін у рівнях продуктивності праці знаходять співвідношення індексу фізичного обсягу та індексу затрат праці :

$$\left(\frac{\Sigma q_1 p_0}{\Sigma q_1 t_1} : \frac{\Sigma q_0 p_0}{\Sigma q_0 p_0} \right) = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} : \frac{\Sigma q_1 t_1}{\Sigma q_0 t_0}.$$

Існує також взаємозв'язок між індексами фонду заробітної плати, середньої зарплати та чисельності працівників . Так , добуток індексів середньої заробітної плати $(I_{сз})$ і чисельності працівників $(I_{п})$ дає індекс фонду заробітної плати $(I_{фз})$, тобто $I_{фз} = I_{сз} \times I_{п}$.

Слід відзначити , що існує взаємозв'язок і між індексами змінного складу (вони відображують зміни середніх рівнів якісних показників), індексами структурних зрушень та індексами фіксованого складу: $I_{з.с.} = I_{с.з.} \times I_{ф.с.}$.

Взаємозв’язки між наведеними вище індексами дають змогу дослідити вплив структурного фактора і зміну самої індексованої величини на зміну (у часі) середніх рівнів досліджуваного показника

Отже, при дослідженні динаміки показників соціально-економічних явищ можна використовувати широке коло статистичних індексів, різних за будовою і змістом, але взаємопов’язаних між собою та доповнюючих один одного, тобто їх систему. Систему взаємозалежних індексів використовують, зокрема, при вивченні ролі окремих факторів у загальній динаміці явищ, а також при обчисленні за двома відомими показниками третього, невідомого.

Характеристика дії фактора може бути одержана як у відносному, так і в абсолютному вираженні.

Загальний вигляд системи двофакторних взаємозалежних індексів можна представити у такому поєднанні:

$$I_{yx}= I_y I_x \text{ або } \frac{\sum y_1x_1}{\sum y_0x_0}=\frac{\sum y_1x_0}{\sum y_0x_0}\cdot\frac{\sum y_1x_1}{\sum y_1x_0}.$$

За даними таблиці 76 знайдемо зазначену взаємозалежність індексів:

$$I_{в.з.}=I_{рсп.}\cdot I_{ур.}=\frac{\sum П_1y_0}{\sum П_0y_0}\cdot\frac{\sum П_1y_1}{\sum П_1y_0}=\frac{20700}{17400}\cdot\frac{24400}{20700}=1{,}190\cdot1{,}179=1{,}40;$$

$$I_{в.з..}=\frac{\sum П_1y_1}{\sum П_0y_0}=\frac{24400}{17400}=1{,}40.$$

Висновок: валовий збір зернових культур у цілому збільшився на 40 %, у тому числі за рахунок збільшення розмірів посівних площ – на 19,0 %, за рахунок підвищення врожайності – на 17,9 %.

У розглянутому прикладі можна визначити абсолютний приріст результативного показника за рахунок змін кожного із факторних показників обчисленням різниць між чисельником та знаменником. Зокрема, загальний абсолютний приріст становить : $\Delta_{yx}=y_1x_1-y_0x_0$. Розклавши його за факторами, маємо: $\Delta_y=y_1x_0-y_0x_0=x_0(y_1-y_0)$; $\Delta_x=y_1x_1-y_1x_0=y_1(x_1-x_0)$. При такому методі розкладання абсолютного приросту одержимо $\Delta_{yx}=\Delta_y+\Delta_x$.

Для системи взаємозалежних індексів у двофакторних комплексах розкладання абсолютного приросту має вигляд: $\Delta_{yx}=\sum y_1x_1-\sum y_0x_0$, у тому числі за факторами : $\Delta_{yx}=\sum y_1x_0-\sum y_0x_0$; $\Delta_x=y_1x_1-\sum y_1x_0$.

За даними розглянутого вище прикладу валовий збір у поточному періоді порівняно з базисним збільшився на 7000 ц (24400 – 17400). Розраховуємо вплив кожного із факторів на визначений розмір приросту. Так, за рахунок збільшення посівних площ валовий збір підвищився на 3300 ц (20700 – 17400), а від підвищення урожайності – на 3700 (24400-20700), тобто 3300+3700 = 7000.

Аналогічний принцип розкладання абсолютного приросту на складові при вивченні дії трьох і більше факторів.

§ 10.6. Територіальні індекси, особливості їх обчислення

Як уже згадувалося, статистичні індекси використовують не тільки для дослідження змін явищ і процесів у часі, а й для характеристики змін рівнів соціально – економічних явищ у просторі. Зокрема, статистика широко застосовує метод порівняння показників у розрізі підприємств, міст, економічних районів, областей, країн. Узагальнюючі показники, які характеризують співвідношення рівнів складних економічних явищ у просторі, тобто в розрізі територій і об'єктів, називаються **територіальними індексами.** Побудова територіальних індексів має свої особливості порівняно з індексами, які характеризують динаміку явищ. Щодо обчислення індивідуальних територіальних індексів, то тут ніяких труднощів не виникає, адже обчислюються звичайні відносні величини порівняння. Наприклад, якщо порівнюється середньорічний надій від однієї корови у двох районах, в першому з яких він дорівнював 4700 кг, а в другому – 4300 кг, то зіставленням першого показника з другим одержуємо територіальний індекс 1,09 . По суті, одержано індивідуальний індекс, який показує, що надій у першому районі в 1,09 раза вищий, ніж у другому.

При побудові територіальних індексів (індивідуальних і загальних) за базу порівняння можуть бути прийняті показники будь – якої із порівнюваних територій. Проте інакше вирішується питання вибору елімінованої величини (ваги), тобто території чи об'єкта, на рівні якого слід зафіксувати вагу індексу. Загальноприйнята схема вибору ваги, розглянута вище, для обчислення індексів динаміки явищ не прийнятна. Адже в територіальних індексах порівнюються явища, які істотно різняться між собою за структурою, а тому ваги тут повинні певною мірою нейтралізувати структурні зрушення. Отже, для територіальних індексів методично правильним вважається вибір «стандартизованої» ваги, яка розрахована на базі, наприклад, районної галузевої та інших структур. Крім того, вибір бази порівняння при обчисленні територіальних індексів слід економічно обґрунтовувати. Наприклад, якщо порівнюються рівень заробітної плати тваринників спеціалізованих господарств з однаковими умовами виробництва, то за базу слід узяти передове господарство з найвищим рівнем оплати праці тваринників.

Щодо індексованої величини, то в територіальних індексах нею може бути кількісний чи якісний показник. Для останніх індексні комплекси розраховують за схемою індексів змінного та фіксованого складу.

При виборі ваги при побудові територіальних індексів слід виходити насамперед із завдань дослідження та його мети.

Досить просто вирішується питання сумірників при побудові індексів фізичного обсягу. Наприклад, при порівнянні обсягів виробництва продукції у двох областях використовують єдині порівнянні ціни (p). Тоді для областей А і Б територіальний індекс обсягу має вигляд : $I_{ф.о..} = \sum q_A p : \sum q_Б p.$

Але якщо просторове порівняння двох областей здійснюють із метою визначення, наприклад, різниці у показниках кормомісткості продукції тваринництва (витрати кормів на одиницю продукції), то за вагу слід брати кормомісткість кожного виду тваринницької продукції. Ця вага повинна бути зафіксована на одному й тому ж рівні для обох областей при індексуванні об'ємних показників. Який же показник слід обрати за фіксовану величину? Тут можливі варіанти: а) кормомісткість в області, яка порівнюється: б) кормомісткість в області, яка є базою порівняння; в) середній показник кормоміскості для обох областей; г) осереднений показник для ширшої території.

При побудові територіальних індексів якісних показників також маємо кілька варіантів вирішення питання вибору сумірника (ваги): а) обсяг продукції в області, яка порівнюється ; б) обсяг продукції в області, яка є базою порівняння ; в) сумарний обсяг продукції двох областей; г) будь – який інший об'ємний показник.

Зрозуміло, що індекси, обчислені за наведеними вище варіантами вибору ваг, дадуть різні відповіді. З практичного боку більш об'єктивною слід вважати відповідь за варіантом «а». Наприклад, при вивченні різниці в цінах реалізації продукції територіальний індекс по областях А і Б для цього варіанта матиме вигляд : $I_у = \sum q_А p_А : \sum q_А p_Б$.

Як бачимо, чисельник цього індексу містить показник фактично, тобто реального, розміру грошової виручки в області А, знаменник – умовний показник виручки при обсязі продукції області А, якщо продукція реалізовувалася за цінами області Б. Отже, додатна різниця між чисельником та знаменником характеризує втрати покупців області А, а від'ємна, - навпаки, виграш за рахунок різниці в цінах порівняно з областю Б.

За результатами дискусії з питань методології побудови територіальних індексів найбільш прийнятним слід вважати таке розв'язання питань вибору сумірників (ваги). При побудові індексних комплексів із показників інтенсивності доцільно обирати сумірником: 1) екстенсивний показник, що стосується територій, на якій інтенсивний показник є більш економічним; 2) загальну середню величину екстенсивного показника на порівнюваних територіях; 3) стандартну величину екстенсивного показника.

Якщо територіальні індекси будуються для екстенсивних показників, вагами можуть бути прийняті середні величини інтенсивного показника: 1)встановлені для території, прийнятої за стандарт; 2) розраховані для території, по якій здійснюється порівняння.

Зауважимо, що ми розглянули лише загальні риси методологічного підходу і принципів побудови територіальних індексів. Вирішення цих питань можливе після досконалого вивчення конкретного об'єкта дослідження та економічної природи явища чи процесу.

МОДУЛЬ 5

ТЕМА 11. ВИБІРКОВИЙ МЕТОД

§ 11.1. Загальне поняття вибіркового методу статистичного спостереження

Щоб вивчити будь - яку сукупність (а таке завдання вирішується за допомогою методів статистики), треба її охарактеризувати різного роду зведеними ознаками. Останні можуть бути досить різноманітні. Одні з них можуть характеризувати середні розміри ознак досліджуваного явища в сукупності, інші – характеризувати структуру сукупності, тобто частку будь – якого явища в сукупності спостережень. Зведені ознаки при всій їх різноманітності мають ту загальну властивість, що всі вони характеризують не окремі ознаки, а сукупність в цілому або певні її частини. У цьому зв'язку масове спостереження, тобто реєстрація кожної конкретної одиниці, має зміст лише як проміжний етап для одержання зведених ознак.

Статистиків давно цікавить питання, як спростити цей проміжний етап, тобто як перейти від вичерпної реєстрації ознак, що входять до досліджуваної сукупності, до часткової їх реєстрації. Мова йде про перехід від суцільного спостереження до несу цільного.

Як уже відомо з попереднього розгляду питання статистичного спостереження, залучення тих чи інших об'єктів (одиниць спостереження) для дослідження можна здійснювати двома шляхами: вивчати всі одиниці спостереження всього масиву або лише їх частину, відібрану за певними науковими принципами. У першому випадку здійснюється суцільне спостереження, в другому – несуцільне. Вибіркове спостереження є одним з видів несуцільного спостереження.

Вибірковими даними користуються досить широко в різних сферах людської діяльності. Наприклад, для оцінки якості зерна або молока немає необхідності в обстеженні всього обсягу продукції, досить лише взяти певну кількість проб. Незначна кількість дослідів виявляється достатньо, наприклад, для встановлення зараження зерна шкідниками, встановлення якості борошна, олійності соняшнику і т.ін. У галузі аграрної економіки до вибіркового методу вдаються при вивченні рівня і складу харчування різних груп населення, продуктивності праці, використання робочого часу тощо.

У сільському господарстві вибіркове спостереження застосовують для встановлення втрат урожаю при збиранні, засміченості посівів, якості продукції, продуктивності праці, для контрольних перевірок перепису худоби і т. ін. Цей вид спостереження одержав значне поширення у зв’язку з вивченням соціальних аспектів суспільного життя, зокрема, у дослідженні рівнів споживання і рівня добробуту населення. Так, в країні постійно проводяться бюджетні обстеження сімей працівників сільськогосподарських підприємств, здійснюються одноразові обстеження житлових умов сімей робітників і службовців, їх заробітної плати тощо. У статистичній практиці застосовується і вибіркова розробка статистичної інформації , зокрема, вибірково здійснюється розробка річних звітів підприємств для поглибленого вивчення продуктивності праці і собівартості виробництва продукції.

Приклад. Вивчають статистичну сукупність з десяти груп корів (кожна з них має однакову кількість голів), які характеризуються даними середньодобових надоїв від однієї корови (табл. 82).

Таблиця 82

Дані середньодобових надоїв від однієї корови

Номер групи	1	2	3	4	5	6	7	8	9	10
Надій, кг	30,6	29,3	30,1	32,5	28,3	29,5	32,7	33,3	24,7	25,9

Розрахунки показують, що середньодобовий надій від однієї корови по всіх групах у цілому становить 29,7 кг. Якщо взяти продуктивність тварин лише про трьох групах, наприклад, 1-й, 4-й і 10-й, то середній надій тут теж становитиме 29,7 кг. Отже, частина сукупності, представленої трьома групами корів, може замінити всю досліджувану сукупність поголів’я тварин. Наведений приклад свідчить, що несуцільне спостереження дає такі ж самі результати, що і суцільне. Але цього можна досягти при наявності певних умов.

Статистичною теорією і практикою розроблені наукові принцип вирішення проблеми відбору часткової сукупності через різні форми і способи. Ці питання вирішуються застосуванням вибіркового методу.

Вибіркове спостереження є одним із видів несуцільного спостереження (монографічного, основного масиву, анкетування). **Вибірковим** називається таке спостереження, при якому обстеженню підлягає лише частина одиниць сукупності, відібраних на основі науково розроблених принципів, які забезпечують одержання достатніх даних для характеристики всієї сукупності.

В основі теорії вибіркового спостереження лежать теореми закону великих чисел. Ними зумовлюються вирішення двох взаємопов'язаних і важливих у практичному відношенні питань вибіркового обстеження: розрахунок обсягу вибіркової сукупності при заданій точності дослідження і визначення помилки обстежень при заданому обсязі сукупності.

Мета вибіркового методу полягає в тому, щоб відібравши з генеральної сукупності певну кількість одиниць, дослідити їх і на цій основі оцінити невідомі параметри генеральної сукупності.

Теоретичні аспекти вибіркового методу деталізовані в навчальних курсах математичної статистики, загальної теорії статистики, спеціальних монографіях і в математичній літературі.

У даному розділі треба насамперед визначитися відносно термінології, яка прийнята в математичній теорії вибіркового методу, а також ввести деякі поняття і характеристики, зумовлені його початковою «присутністю» при розрахунках тих чи інших оцінок. У подальших міркуваннях будуть використовуватися терміни: вибірковий метод, вибірковий спосіб спостереження, вибіркове спостереження, вибірка – всі вони несуть однакове змістовне навантаження.

Загальна чисельність одиниць, із яких здійснюється відбір, називається **генеральною сукупністю.** Всі її характеристики мають назву генеральних (середня, частка, дисперсія тощо). Сукупність одиниць, відібраних на основі науково розроблених принципів, називається **вибірковою сукупністю.** Всі характеристики вибірки (відповідні узагальнюючі показники) називають **вибірковими** (середня, частка, дисперсія та ін.). Під терміном «вибірка» розуміють групу об'єктів, які відрізняються трьома особливостями: а) частина генеральної сукупності; б) частина генеральної сукупності, відібрана у випадковому порядку, певним чином; в) частка генеральної сукупності, яка досліджується для характеристики всієї генеральної сукупності.

Завдання вибірки - дати вірне уявлення про характер генеральної сукупності, оцінити значення її параметрів.

У практичних розрахунках інколи використовуються вибірки невеликі за обсягами (20-30 одиниць спостереження). Такі вибіркові сукупності в статистиці називаються **малими вибірками.** Вони мають принципові методичні особливості у своєму використанні. На відміну від малих вибірок, великі (звичайні) вибірки, які ґрунтуються на нормальному розподілі імовірностей, дають значно точніші результати.

Основні положення вибіркового методу, які будуть викладені нижче, розглядаються виходячи з передбачення, що маємо справу з вибірками досить великого обсягу.

Одержана в результаті статистичного спостереження вибіркова емпірична сукупність являє інтерес не сама по собі, а лише у тій мірі, в якій вивчення її дозволяє одержати інформацію про власності генеральної сукупності. Щоб властивості вибірки достатньо відображували властивості генеральної сукупності, вона повинна бути представничою – репрезентативною. Тобто, вибірка за певних умов стає більш чи менш точним відображенням усієї генеральної сукупності.

Властивість вибіркової сукупності відтворювати генеральну сукупність одержала назву **репрезентативність,** що означає представництво з певною точністю і надійністю. Репрезентативність вибіркових даних може бути виражена у достатньому або недостатньому ступені. У першому випадку у вибірці одержують вірогідні оцінки генеральних параметрів, у другому – невірогідні. Слід пам'ятати, що одержані невірогідні оцінки вибірки не применшують значення вибіркових показників для характеристики самої вибірки. Вірогідні оцінки значно розширюють сферу застосування результатів, одержаних при вибірковому обстеженні.

Теорія вибіркового методу завжди розвивались якраз у напрямку вирішення основної задачі – одержання надійних методів оцінки результатів вибірки. Зумовлено це тим, що несуцільному спостереженню притаманні деякі неточності по відношенню до суцільного спостереження. Розміри їх залежать від того, наскільки точно належна спостереженню частина сукупності відтворює (репрезентує) всю сукупність. Тому неточності несуцільного спостереження називають **помилками репрезентативності.** Останні є предметом окремого детального вивчення. Тут лише зазначимо, що у зв'язку з розміром можливих помилок репрезентативності, питання про переваги суцільного і несуцільного видів спостереження в кожному конкретному випадку вирішаються по-різному.

Якщо розмір помилок репрезентативності відносно невеликий, то ними нехтують, враховуючи те, що помилки ці компенсуються економією матеріально–грошових засобів. У випадках, коли помилки репрезентативності досить значні, перевагу віддають суцільному спостереженню. У такому разі вирішується питання у випадках, коли маємо справу лише з помилками репрезентативності. Але існують і інші помилки (див. 3.5. частина 3), розмір яких треба враховувати при характеристиках генеральної сукупності по вибіркових даних. Тому перед теорією вибіркового методу стоїть завдання визначити можливий розмір і границі цих помилок.

Вибірковий метод відрізняється від інших методів несуцільного спостереження (монографічного, основного масиву, анкетування) двома ознаками: 1) заздалегідь встановлюється кількість одиниць або частина генеральної сукупності, яка буде обстежена; 2) заздалегідь визначається порядок відбору одиниць спостереження, при якому вибіркова сукупність у достатній мірі репрезентувала б генеральну сукупність.

Перевагу вибіркового спостереження перед суцільним зумовлюють такі основні фактори:

1. Економія часу і засобів у результаті скорочення обсягу роботи. Наприклад, якщо вибірці підлягає 1 або 5 % загальної кількості одиниць, то обсяг робіт скорочується порівняно до суцільного обстеження відповідно в 100 і 20 разів. Особливо це важливо, коли зведення потрібні терміново.

2. Зведення до мінімуму пошкодження або навіть знищення досліджуваних об'єктів. Наприклад, визначення ступеня достигання цукрових буряків шляхом його вибіркового копання або визначення схожості насіння сільськогосподарських культур і т. ін.

3. Необхідність детального обстеження кожної одиниці спостереження при неможливості охоплення всіх одиниць. Наприклад, вивчення бюджету сімей фермерських господарств. У даному випадку необхідно, щоб облік здійснювався по окремих сім'ях, записи про доходи і витрати велися щоденно.

4. Досягнення вищої точності результатів обстеження завдяки скороченню помилок, які виникають при реєстрації. Якщо загальний обсяг роботи менший, можна залучити кваліфікованіший персонал, краще його підгодовувати, більш ретельно контролювати проведення обстежень і обробку їх результатів. Тому вибіркове обстеження може дати більш вірогідні результати, ніж відповідне суцільне обстеження.

5. Більш широка сфера застосування. Деякі види обстежень потребують для збору зведень досить кваліфікованого персоналу і спеціального обладнання. Як правило, і перше і останнє буває обмеженим. У таких випадках суцільне обстеження виявляється

неможливим: доводиться одержувати зведення або вибірковим шляхом, або не одержувати їх зовсім. Таким чином, перевага вибіркового спостереження тут очевидна. З іншого боку, якщо необхідно одержати точну інформацію про дрібні підрозділи вихідної сукупності, то необхідний для цього обсяг вибірки може виявитись настільки великим, що кращим буде суцільне спостереження.

Вибіркове обстеження пов'язано з певним ризиком. Але перевага його, як наукового методу дослідження, полягає в т ому, що воно дозволяє обчислити величину цього ризику. Використовуючи вибірковий метод, дослідник вирішує (залежно від поставлених завдань), на який ризик він повинен піти, тобто який рівень можливої помилки в результатах дослідження не буде істотно впливати на об'єктивність висновків. До речі, ризик слід вважати не слабкою, а сильною стороною вибіркового методу. Було б невірним вважати, що суцільне спостереження позбавлене ризику. З ряду причин результати суцільного спостереження можуть виявитися менш надійними, ніж результати вибіркового обстеження. На жаль, людський фактор інколи відіграє негативну роль у дослідженнях, оскільки вони пов'язані з людським судженням. Відомий факт, що при потоковому виробництві продукції контролерам (відділам технічного контролю) притаманне на вимірювальних приладах (при ідеальній їх роботі) читати очікувані показники замість дійсних. Якщо позбавитися ілюзій про надійність суцільного спостереження, то акції вибіркового спостереження переконливо зростають. До того ж у ряді випадків здійснення суцільного спостереження виявляється неможливим. Досить лише уявити чи стане тракторний завод випробувати кожен двигун чи інший його агрегат, піддаючи навантаженню, яке дорівнює звичайному терміну служби, а перевірити ракети чи, наприклад, водневі бомби означало б просто знищити їх. Не можна також уявити суцільного дослідження, завдання якого пов'язане з обстеженням посівної площі на предмет схожості насіння, адже така операція призвела б до повного знищення рослин на площі.

Коротка історична довідка. Вибіркове статистичне спостереження здійснювалося в різних країнах уже в XIX столітті. В Росії у 70-х роках цього століття вибіркове спостереження проводилося з метою вивчення господарських умов станиць Терського козацького війська після укладання миру на Кавказі. Потім вибіркове спостереження почали використовувати земські статистики. Початок покладено Воронезьким земством у зв'язку з вивченням питання про бюджети селянських господарств у середині 80-х років XIX століття. Такі спостереження ще не можна віднести до класичної форми вибіркового статистичного спостереження. В обох випадках обстеження зводилося до відбору типових об'єктів (станиці, поселення, селянські господарства).

Перші спроби використання вибіркового обстеження як наукового методу були зроблені у сфері соціальних досліджень (1899р., американець Раунтрі). У галузі економіки застосування вибіркового методу спостерігаємо значно пізніше.

Історія статистичної науки пов'язує початок розвитку наукової теорії вибіркового методу з ім'ям А.Кієра – директора норвезького статистичного бюро. Проте до створення теорії вибіркового методу на базі теорії ймовірностей він непричетний.

Лише у 1901 р. відомий статистик. В. Борткевич зробив акцент, що теоретичною основою вибіркового методу повинно служити обчислення ймовірностей. Така задача була вирішена у 1906 р. англійським статистиком А.Боулі на базі математичних розробок К.Пірсона і Р.Еджворта. Саме А.Боулі показав, як визначити випадкові помилки вибіркового обстеження. Вважається, що з цього моменту одержала розвиток теорія вибіркового обстеження, яка постійно удосконалюється.

§ 11.2. Теоретичні основи вибіркового методу

Кожна досліджувана сукупність залежить від дії певних суб'єктивних факторів, котрі зумовлюють коливання результатів досліджень. Кожна окрема одиниця спостереження звичайно належить деякій гіпотетичний необмеженій сукупності, яка підпорядкована певному закону розподілу сукупності. Відібрана одиниця спостереження дає додаткову інформацію про розташування і вид розподілу.

Отже, мета вибіркового обстеження полягає в тому, щоб мати можливість зробити певні ствердження про розподіл сукупності. На підставі інформації, одержаної з вибіркової сукупності, можна зробити висновки про величину параметрів досліджуваної сукупності, тобто дати їм оцінку. Як правило, потребують оцінювання такі параметри розподілу, як середня, дисперсія, коефіцієнти асиметрії і ексцесу. Щоб статистичні оцінки і висновки, одержані по вибірковій сукупності, були вірними, вибірка повинна репрезентувати всю сукупність.

Теоретичні аспекти вибіркового спостереження зводяться до таких питань: 1) встановлення репрезентативності вибіркових сукупностей; 2) вивчення видів вибірок, які можна одержати з даної сукупності; 3) ефект змін обсягу вибірки; 4) визначення найкращих способів оцінювання параметрів сукупності; 5) встановлення вірогідності одержаних результатів розрахунків (висновків).

Особливого значення у вибірковому спостереженні набуває питання про репрезентативність вибірки. Якщо остання не буде представляти (репрезентувати) всю досліджувану сукупність, одержані висновки можуть виявитися досить суб'єктивними і навіть невірними.

Щоб уявити, наскільки висновки, зроблені по результатах вибіркового обстеження, можуть відрізнятися від реальної дійсності, розглянемо такий гіпотетичний приклад. Припустимо, космічний корабель з іншої планети сів на земну кулю в зоні пустелі. Здійснивши статистичні спостереження за рослинним і тваринним світом, атмосферою, екіпаж корабля по результатах аналізу робить висновки: земля вкрита піском атмосфера її щільна, з високою температурою повітря, фауна і флора збіднілі і складаються з кількох видів тварин і рослин і т. ін.

Незнання теорії вибіркового спостереження небезпечно у справі його застосування в будь – якій галузі досліджень чи експериментів. Яким би не був план одержання вибірки, першою і самою головною його умовою повинна бути вимога, щоб вибірка давала по можливості найбільш вірну картину всієї сукупності, з якої здійснюється відбір одиниць спостереження. Причому всі одиниці вибірки мають бути одержані на основі випадкового відбору за певними об'єктивними правилами (про що йтиметься пізніше).

Щоб забезпечити достатню точність характеристики генеральної сукупності, необхідно організувати правильний відбір одиниць спостереження із цієї сукупності. Теорія і практика статистики передбачають кілька систем відбору одиниць у вибіркову сукупність. В їх основу покладено науковий принцип – забезпечити максимальну можливість вибору будь – якої одиниці з генеральної сукупності при дотриманні принципу випадковості. Тобто, для одержання правильної, не викривленої характеристики генеральної сукупності необхідно намагатися забезпечити можливість відбору у вибіркову сукупність одиниць спостереження з будь – якої частини генеральної сукупності. Зазначена вимога є основною і повинна виконуватися тим точніше, чим більше варіює досліджувана ознака. Ці й інші завдання вирішує теорія вибіркового методу.

В еволюції вибіркового методу чітко простежуються дві лінії його розвитку. Перша з них спрямована на вирішення завдань теоретичного плану, другій – властиві завдання практичного використання вибіркового методу в економічних дослідженнях.

Теоретичні аспекти вибіркового методу пов'язані з розвитком його у напрямі розробок питань одержання надійних методів оцінки результатів вибірки. Так, відокремлюючи статистичні характеристики генеральної сукупності (середня арифметична, частка, дисперсія і т. ін.) від аналогічних вибіркових характеристик, теорія становить і вирішує завдання встановити межі відхилень показників генеральної сукупності від показників вибіркової сукупності. Отже, теорія вибіркового методу вирішує проблему можливих помилок (їх розмір і межі), які мають місце у випадках, коли про характеристики генеральної сукупності судять на підставі відповідних характеристик, одержаних при вибірковому обстеженні. І якщо виникає потреба у встановленні причин, які зумовлюють розміри помилок, теорія вибіркового методу орієнтує на надійніші прийоми організації вибіркових досліджень.

Слід зазначити, що перші теоретичні дослідження щодо вибіркового методу відбувалися не в сфері економічних досліджень, а при вирішенні завдань, пов'язаних з визначенням очікуваних результатів азартних ігор і при обробці даних астрономічних спостережень. Як у першому, так і у другому випадках виникало питання про так звані випадкові помилки спостереження. Тому теорія вибіркового методу за таких умов ґрунтувалася винятково на принципі випадкового відбору.

Але проблеми, які виникали при вирішенні практичних завдань у господарській діяльності, мали іншу природу. І вирішення їх виявилось можливим лише при застосуванні вибіркового методу. Це стосувалося необхідності визначення урожаю по результатах пробних обмолотів у сільському господарстві. Аналогічна потреба виникала у торгівлі, коли якість всієї партії товару визначалась на підставі окремих проб. Розвиток державності зумовлював необхідність розробки бюджетних обстежень. Такі завдання за своєю природою мали інший зміст, явно відрізняючись від тих, на базі яких виникла теорія випадкового відбору. Тут виявилися відсутніми і «гра випадку» (характерна для азартних ігор), і необхідність багаторазових вимірів одного об’єкта (характерна для астрономічних досліджень). Вирішення нових практичних завдань вимагало одержання інформації про всю сукупність на підставі певної її частини. Тобто, постало завдання забезпечити таку систему відбору, при якій вибіркова сукупність досліджуваних об’єктів змогла б забезпечити одержання найбільш точної і повної інформації про всю сукупність.

В основу нового підходу до вибіркового спостереження було покладено метод планомірного відбору об’єктів . Але розбіжності між теорією і практикою (а такий підхід тривалий час не мав належної теоретичної розробки) призводили до того, що розроблені практикою методи вибірки вибраковувалися лише на тих підставах, що теорія не могла оцінити, наскільки пропонований метод надійніше того, для якого відомі закони, що визначають помилки репрезентативності. Останні виявилися меншими за тих, які одержувалися при забезпеченні надійного методу оцінки результатів вибірки пропонованої теорією. Так, практикою вибіркового методу було створено схему безповторного відбору, а в подальшому доведено ефективність прийомів механічного і типового відбору (про них мова йтиме пізніше).

Внутрішні протиріччя між теорією і практикою вибіркового методу стиралися у міру того, як практика вибіркового методу накопичувала багатий досвід (це стосувалось досліджень і в економці) застосування планомірних методів відбору, розвивались статистична теорія, а положення теорії ймовірностей поширилося на залежні явища. Переваги перед методом випадкового відбору стали очевидними.

При цьому принцип випадковості не порушується. Як і при випадковому відборі одиниць спостереження, так і при механічному і типовому прийомах відбору попадання тієї чи іншої одиниці у вибірку залежить від випадковості.

Кожна одиниця спостереження повинна мати рівну з іншими можливість бути відібраною. Саме на цьому ґрунтується математична теорія вибіркового методу. Покладений у її основу принцип випадковості відбору дозволяє розглядати кожну ознаку, що відбирається, як випадкову величину.

Відомо, що випадковість має своєї закони, які проявляються лише тоді, коли нагромаджено велику кількість спостережень.

Йдеться про закон великих чисел. Як зазначалося вище, теорія вибіркового методу заснована саме на законі великих чисел. Суть його зводиться до загального принципу - сукупна дія великої кількості випадкових факторів призводить до результату, майже не залежному від випадку. Цей закон є формою прояву необхідності через випадковість і належить лише до випадкових індивідуальних відмінностей, випадкової варіації. У соціально - економічній статистиці закон великих чисел розглядається як загальний принцип, через який кількісні закономірності, притаманні масовим суспільним явищам, вразливо проявляється тільки в досить великій кількості спостережень. Характерною особливістю цього закону є та, що закономірності проявляються у ньому лише як тенденція в середньому. Ця тенденція майже не залежить від випадку, проте намагається до відображення кількісної закономірності, не даючи ніколи абсолютно точного її відтворення.

Надаючи особливого значення закону великих чисел у вибірковому методі, слід пам'ятати, що не всі способи відбору, засновані на цьому законі. Мова йде про способи відбору, в основу яких покладено планомірно організований відбір, його результати виявляються точнішими, ніж при випадковій вибірці.

Результатом факту випадковості потрапляння окремих одиниць спостереження у вибіркову сукупність є відхилення вибіркової характеристики від відповідної генеральної характеристики. Величина зазначеного відхилення вважається помилкою вибірки. Така помилка є результатом факту випадковості потрапляння окремих одиниць спостереження у вибіркову сукупність. Розмір її може бути значний, малий або близький до нуля.

При збільшенні числа одиниць спостереження, як правило, зменшується різниця між вибірковими і генеральними характеристиками.

Теорія вибіркового методу дає змогу розраховувати можливі розміри помилок вибірки, а з іншого боку, залежно від конкретних завдань дослідження з врахуванням допустимої помилки - передбачити необхідний обсяг одиниць вибіркової сукупності.

§ 11.3. Способи відбору у вибіркову сукупність

Способи відбору одиниць з досліджуваної генеральної сукупності з метою утворення вибіркової сукупності можуть бути різні. Залежно від того, як поставлена вибірка, як організований відбір із загальної маси, розрізняють кілька варіантів утворення вибіркової сукупності. Щоб одержана вибіркова сукупність мала об'єктивну гарантію репрезентативності, застосування того чи іншого способу відбору повинно бути науково обґрунтованим. Основний принцип правильності відбору – строго об'єктивний підхід до відбору одиниць спостереження. Якщо цей принцип порушується і одиниці відбираються суб'єктивно, то результати такого спостереження не можна поширювати на генеральну сукупність, вони можуть бути застосовані лише щодо тієї частини, яка підлягала спостереженню.

Назва вибіркової сукупності зумовлюється способом відбору одиниць, тобто схемою її формування. За таких принципів утворюється класифікація видів вибірки. Але слід зазначити, що в навчальній і науковій літературі в цій справі існують деякі розбіжності.

При розгляді питань способів відбору (видів вибірок) будемо притримуватися традиційної їх класифікації : 1) власне випадковий відбір; 2) механічний відбір; 3) типовий відбір ; 4) серійний відбір ; 5) комбінований відбір.

Власне випадковий спосіб відбору – це такий спосіб формування вибіркової сукупності, коли відбір одиниць з генеральної сукупності здійснюється у випадковому порядку. Випадковість відбору полягає у дотриманні принципу однакової можливості для всіх одиниць генеральної сукупності потрапити у вибірку.

Для суворого забезпечення принципу випадковості самого процесу вибірки використовують різні технічні засоби (наприклад, урни з пронумерованими картками чи предметами), які є ніби моделлю генеральної сукупності. Найнадійнішим способом відбору вважається використання спеціальних таблиць випадкових чисел.

Випадковий відбір часто поєднують з іншими способами відбору. У разі використання його як самостійного способу він має назву **власне випадковий відбір.** Від назви способу відбору походить і назва вибірки – **власне випадкова.**

Випадкова вибірка може бути організована або за схемою **повторного відбору** або по схемі **безповторного відбору.** Зазначені схеми відбору дають однакові результати лише у разі нескінченої генеральної сукупності. При умові скінченності генеральної сукупності результати вибірок будуть різні. Особливість названих схем відбору полягає у такому.

При **повторному відборі** кожна одиниця бере участь у вибірці стільки разів, скільки відбирається одиниць, тобто після реєстрації вона повертається у генеральну сукупність і в подальшому може знов потрапити у вибіркову сукупність. За таких умов генеральна сукупність залишається незмінною, і тому для всіх одиниць сукупності забезпечується рівна ймовірність потрапити у вибірку.

При **безповторному** відборі кожна відібрана одиниця у подальшому відборі не бере участі, тобто не повертається у генеральну сукупність. Але це означає, що чисельність генеральної сукупності буде змінною після кожної операції відбору. У зв'язку з цим ймовірність потрапити у вибірку решти одиниць підвищується, а тому середня помилка вибірки тут буде менша, ніж при повторному способі відбору.

Особливістю власне випадкової повторної вибірки є та, що при її організації найбільш послідовно здійснюються принципи випадкового відбору. Цей вид вибірки подібний до багаторазових вимірів однієї й тієї ж величини. До речі, чисельність вибірки при застосуванні схеми повторного відбору може досягти обсягу, який перевищує генеральну сукупність (окремі одиниці враховуються у вибірці не один, а кілька разів). Тому власне випадкова вибірка зі схемою повторного відбору практичного значення не має і в дослідженні соціально – економічних явищ не використовується. У теорії вибіркового методу вона розглядається лише тому, що математична схема її найбільш проста і безперечна. Досить лише нагадати, що точність вибірки забезпечується положеннями теорії ймовірностей, зокрема законом великих чисел, який для даного випадку може бути сформульований так: при достатньо великому обсязі вибірки можна очікувати з ймовірністю близькою до одиниці, що вибіркова середня буде незначно відрізняти від генеральної середньої.

Безповторний спосіб відбору хоча і порушує принципи рівної ймовірності потрапити у вибірку, проте він не погіршує, а покращує результати відбору. Практично доведено, що при інших рівних умовах середня помилка безповторної вибірки менша, ніж при повторній, і потребує меншого обсягу одиниць для обстеження. Таким чином, практика внесла поправку в теорію випадкового відбору, забезпечивши високу точність вибіркового спостереження.

Власне випадкова вибірка знаходить практичне використання у польових дослідах і при дослідженні окремих соціальних явищ в аграрному секторі.

Механічний відбір. Механічним називається відбір, при якому генеральна сукупність поділяється на рівні частини відповідно до природного розташування її одиниць (географічного, просторового, алфавітного тощо) і з кожної частини обстежується одна одиниця. Тобто одиниці відбирають через рівні проміжки у порядку розташування їх сукупності. Наприклад, кожна п'ята одиниця при 20% - му відборі, кожна десята одиниця при 10 % -му відборі і т. ін. Якщо відбір здійснюється на земельній території, проби беруть у шаховому порядку. Проміжок між відібраними одиницями визначається залежно від прийнятої пропорції відбору. Цей проміжок розраховується як частка від ділення чисельності сукупності на обсяг вибірки. Наприклад, потрібно відібрати для обстеження 30 деталей із загальної чисельності 611 штук. Проміжок відбору (шаг) тут становитиме $\frac{611}{30} = 20{,}367$, тобто у вибірку потрібно включити кожну 20 одиницю, враховуючи «початкове число» відбору, за яке приймається будь –яка одиниця перших двадцяти одиниць. Як правило, першою одиницею відбирається та, яка знаходиться посередині проміжку (у даному випадку десята від початку відрахування).

Механічний спосіб забезпечує рівномірність відбору одиниць з усіх частин сукупності, тобто їх пропорційне представництво, а отже, і найбільш високу репрезентативність обстеження.

Слід зазначити, що при механічному способі відбору відібрані одиниці не мають імовірнісного характеру. Випадкові помилки тут зумовлюються не способом відбору, а наявністю випадковості у розташуванні матеріалу досліджуваної сукупності.

Механічний відбір можна здійснювати й іншими шляхами. Наприклад, якщо потрібно відібрати певну кількість річних звітів підприємств окремого регіону, їх вибирають з ретельно перемішаної сукупності в алфавітному порядку назв окремих підприємств.

Науковими дослідженнями доведено, що механічний відбір дає найточніші результати порівняно до випадкової вибірки. Середня вибіркова виявляється тут більш наближеною до середньої генеральної, ніж при випадковому відборі. За таких саме причин і при постановці будь – якого експерименту принципу механічного розміщення одиниць віддається перевага щодо випадкового їх розміщення. На рисунку 24 показано відбір у формі латинського квадрата, який носить назву «Хід коня». Це один з варіантів механічного розміщення елементів досліду, який виявляється доцільнішим, ніж випадковий.

A	B	C	D	E
D	E	A	B	C
B	C	D	E	A
E	A	B	C	D
C	D	E	A	B

Рис. 24. Схема механічного способу відбору «Хід коня»

Необхідно відзначити, що перевага механічного відбору у точності відтворення досліджуваної характеристики має місце лише при умові, що в організації такого відбору не міститься моментів, які сприяють зміщенню як самої вибірки, так і оцінки її результатів.

Про залежність помилки вибірки від характеру розташування досліджуваної сукупності свідчать результати механічного відбору з сукупності, розташованої у ранжирований ряд. Одержана таким шляхом вибірка може мати помилку репрезентативності не випадкову, а систематичну. При цьому величина її залежить від того, що прийнято за «початкове число». Так, якщо у ранжированому ряду, взятому в порядку зростання з проміжком відбору, рівному десяти, відбирається кожна перша одиниця (перша, одинадцята, двадцять перша і т.д.), вибірка буде мати помилку у бік зменшення, тобто вибіркова середня буде менша за генеральну середню. Якщо за «початкове число» обрати середину проміжку, то помилки не буде зовсім.

Механічний спосіб відбору, як правило, застосовується щодо матеріалу, в розташуванні якого мають місце і елементи випадковості і елементи деякої упорядкованості.

Поширеність даного способу відбору у порівнянні з випадковою вибіркою зумовлюється простотою і гнучкістю його застосування. Практична статистика досить часто здійснює роботи вибіркового характеру саме на підставі механічного відбору. Досить відзначити, наприклад, таку важливу галузь вибіркових робіт, як статистика сімейних бюджетів населення. Вибіркова сукупність тут формується за схемою механічного відбору одиниць у межах їх типових груп. Так, із списку, в якому господарства розташовані за розмірами розподіленого доходу на 1 людино–годину. Роботи, відбираються бюджети селян. За таким же принципом відбираються із списку, складеного в порядку зростання (чи зменшення) заробітної плати, підприємства і сім'ї працівників.

Типовий відбір. При типовому відборі сукупність попередньо розбивається на більш однорідні групи. Суть його зводиться до типового районування досліджуваної сукупності на однорідні групи з наступним відбором за власне випадковим принципом або механічним. Цей тип відбору також обмежує принцип рівної ймовірності. На точність результатів вибірки тут позитивно впливає сам принцип районування, адже він «знімає» вплив міжгрупової варіації ознак, що зумовлює менший розмір квадратичної помилки.

При типовому підборі генеральна сукупність розчленовується на певну кількість однорідних груп (підсумковостей) з груповими численностями ($N_1, N_2, N_3, ..., N_m$). Кожна типова група як часткова генеральна сукупність має свої характеристики: середню ($\bar{x}_1, \bar{x}_2, \bar{x}_3, ..., \bar{x}_m$), дисперсію ($\sigma_1^2, \sigma_2^2, \sigma_3^2, ..., \sigma_m^2$), частку $(p_1, p_2, p_3, ... p_m)$, показник асиметрії ($A_1, A_2, A_3, ..., A_m$) тощо. Кожна із зазначених характеристик оцінюється за частковими вибірковими сукупностями ($n_1, n_2, n_3, ..., n_m$), які одержують від кожної типової групи. Останні, як правило, різні за чисельністю одиниць. При цьому, хоча варіація ознаки в середині типових груп знижується, вона все ж таки має місце і відрізняється за розміром між групами. Зазначені особливості потребують певних методичних підходів у вирішенні питання обсягу часткових вибірок для забезпечення достатньої репрезентативності останніх.

На практиці вибірку можна здійснювати двома шляхами: 1)за способом рівномірного відбору; 2) за способом відбору пропорційно – груповим чисельностям. Розглянемо їх.

При рівномірному способі відбору з кожної типової групи відбирається однакова чисельність одиниць ($n_1 = n_2 = n_3 =, ..., n_m$). Його застосовують лише при умові рівної чисельності кожної з типових груп.

При пропорційній схемі відбору чисельність часткових вибірок береться пропорційно чисельності генеральних сукупностей типових груп або пропорційно їх дисперсіям чи середнім квадратичним відхиленням. Інколи використовується комбінована схема, тобто чисельність береться пропорційно і чисельностям і дисперсіям одночасно.

Розглянемо приклад **пропорційної вибірки.** Треба відібрати 150 підприємств лісостепової зони (генеральна сукупність), яка розчленована на підзони (центральну, лівобережну і правобережну) з груповими чисельностями $N_1= 200$; $N_2= 400$; $N_3= 500$.

При вибірці, пропорційній груповим чисельностям, відбір передбачає, що у всіх групах повинно зберігатися незмінним співвідношення:

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} =,....,= \frac{n_m}{N_m},$$

Таке співвідношення забезпечується, якщо відбір здійснити пропорційно питомій вазі кожної групи у генеральній сукупності (табл. 83).

Таблиця 83

Вихідні і розрахункові дані для обчислення часткових вибірок за схемою пропорційно чисельності

Типові групи	Обсяг генеральної сукупності , Nj	Питома вага групи у генеральній сукупності, d	Необхідний обсяг вибірки по групі (при *n*- 150), $n_j=nd$
I	200	0,18	27
II	400	0,36	55
III	500	0,46	68
Всього	1100	1,00	150

Для кожної типової групи відбирається постійна частка одиниць сукупності:

$$K = \frac{n}{N} = \frac{150}{1100} = 0{,}136 .$$

Використавши цей постійний коефіцієнт, знаходимо чисельність кожної вибірки : $n_1 = 200 \cdot 0{,}136 = 327;$ $n_2 = 400 \times 0{,}136 = 55;$ $n_3 = 500 \cdot 0{,}136 = 68;$ $n = n_1 + n_2 + n_3 = 27 + 55 + 68 = 150$.

Таким чином для умов нашого прикладу у вибірку із 150 одиниць повинно потрапити з першої групи (підзони) 27 одиниць, з другої – 55, з третьої – 68 одиниць спостереження.

Розглянемо приклад відбору за схемою відбору, **пропорційному середньому квадратичному відхиленню** (табл. 84).

Таблиця 84

Вихідні і розрахункові дані для обчислення обсягів часткових вибірок за схемою відбору пропорційно середньому квадратичному відхиленню

Типові групи	Обсяг генеральної сукупності , Nj	Середнє квадратичне відхилення, σ_j	Відносний коефіцієнт по групі $K = \frac{\sigma_J}{\sum \sigma_j}$	Необхідний обсяг вибірки по групі (при *n*- 150), $n_j = n\frac{\sigma_j}{\sum \sigma_j}$
I	200	6	0,33	50
II	400	5	0,28	42
III	500	7	0,39	58
Всього	1100	18	1,00	150

Практично така необхідність виникає у зв’язку з наявністю варіації в середині груп, яка зумовлює різницю в групових дисперсіях.

Для даного випадку відбору повинна зберігатися рівність співвідношень:

$$\frac{n_1}{\sigma_1} = \frac{n_2}{\sigma_2} = \frac{n_3}{\sigma_3} = ... = \frac{n_m}{\sigma_m}.$$

Значення середніх квадратичних відхилень по типових групах визначають шляхом пробних досліджень або беруть з даних попередніх досліджень.

Припустимо, в нашому прикладі для трьох досліджуваних груп рівні їх становили: $\sigma_1 = 6$; $\sigma_2 = 5$; $\sigma_3 = 7$.

Результати розрахунку необхідності чисельності для кожної вибірки подано в таблиці 84.

Чисельності часткових вибірок визначаємо згідно із одержаним коефіцієнтом: $n_1 = 150 \cdot 0{,}33 = 50$; $n_2 = 150 \cdot 0{,}28 = 42$; $n_3 = 150 \cdot 0{,}39 = 58$; $n = 50 + 42 + 58 = 150$.

При розглянутій схемі формування групових вибіркових сукупностей для всіх груп зберігається постійне співвідношення:

$$K = \frac{150}{18} = 8{,}3; \quad K_j = \frac{50}{6} = \frac{42}{5} = \frac{58}{7} = 8{,}3.$$

Якщо відбір здійснювати пропорційно дисперсіям, то питома вага груп з великою варіацією ознак у вибірці різко зросте.

Розглянемо приклад для випадку **пропорційно – комбінованої схеми відбору.**

Необхідність такої схеми відбору пояснюється тим, що на практиці типові групи, як правило, різняться як за чисельністю, так і за варіацією ознак. Обидва ці фактори треба враховувати при формуванні сукупності. Групові вибірки в даному випадку відбирають таким чином, щоб залишилися незмінними такі співвідношення:

$$\frac{n_1}{N_1\sigma_1} = \frac{n_2}{N_2\sigma_2} = \frac{n_3}{N_3\sigma_3} = ... = \frac{n_m}{N_m\sigma_m}.$$

Чисельність одиниць спостереження для кожної часткової вибірки визначається за формулою: $n_j = n\frac{Nj\sigma j}{\sum Nj\sigma}$.

Робочі етапи розрахунків наведено в таблиці 85.

Таблиця 85

Вихідні і розрахункові дані для обчислення обсягів часткових вибірок за схемою пропорційно-комбінованого відбору

Типові групи	Обсяг генеральної сукупності , Nj	Середнє квадратичне відхилення, σ_j	Зважене середнє квадратичне відхилення, $\sigma_j N_j$	Відносний коефіцієнт по групі $\frac{\sigma_j N_j}{\sum \sigma_j N_j}$	Необхідний обсяг вибірки по групі (при *n*- 150), $n_j = n\frac{\sigma_j N_j}{\sum \sigma_j N_j}$
I	200	6	1200	0,18	27
II	400	5	2000	0,30	45
III	500	7	3500	0,52	78
Всього	1100	18	6700	1,00	150

Як бачимо, при такій схемі відбору для всіх груп зберігається незмінним співвідношення: $\frac{n_j}{N_j\sigma_j} = \frac{27}{1200} = \frac{45}{2000} = \frac{78}{3500} = 0,022$.

Зазначимо, якщо б відбір здійснювався не за схемою пропорційності, то чисельність для кожної вибірки була б однакова, тобто: $\frac{n}{3} = \frac{150}{3} = 50$.

Зведені дані про чисельність часткових вибірок при типовому відборі, одержані за різними схемами відбору наведено в таблиці 86.

Таблиця 86

Обсяг групових вибірок, одержаних при різних способах формування вибіркової сукупності типової вибірки

Типові групи	Принцип відбору			
	рівномірний	пропорційний		
		чисельності типових груп	середнім квадратичним відхиленням	зваженим середнім квадратичним відхиленням
I	50	27	50	27
II	50	55	42	45
III	50	68	58	78
Всього	150	150	150	150

Як свідчать її дані, одержані результати різняться між собою. Кожен з варіантів розрахунків забезпечує репрезентативність вибірки. Але найбільшу репрезентативність забезпечує вибірка пропорційна комбінованої, тобто та, яка організована пропорційно зваженим середнім квадратичним відхиленням. Пояснюється це тим, що при такій схемі відбору враховуються як обсяг типових груп, так і варіація ознаки. Оскільки такий спосіб потребує багато попередньої інформації про генеральну сукупність, на практиці частіше використовується вибірка, пропорційна обсягу сукупності типових груп.

Серійний відбір. Для розглянутих вище способів відбору (власне випадкового, механічного, типового) характерним є, що відбір одиниць з генеральної сукупності здійснюється в індивідуальному порядку. На практиці інколи виявляється доцільним проводити відбір не окремих одиниць, а цілих груп (серій, гнізд), котрі підлягають потім суцільному обстеженню. Групи (серії) відбирають за методом власне випадкової безповторної вибірки або способом механічного відбору.

Серійний відбір (в англійській термінології «cluster sampling» – груповий відбір; у сільськогосподарський статистиці – «гніздовий відбір») має практиці переваги, особливо у сільськогосподарській статистиці, де обстеження кількох груп підприємств, розташованих безпосередньо одне біля, одного менш утруднене, ніж обстеження такої ж чисельності окремих підприємств, розташованих по всій території району.

У сільському господарстві спосіб відбору застосовують при вивченні бюджету сімей робітників і службовців галузі, при контрольних обходах, які проводяться після обліку худоби і т. ін. З цією метою відбирають підприємства або їх групи, в яких проводять суцільне обстеження.

У статистичній практиці серійний відбір здійснюють у двох варіантах : 1) всі серії мають однакову кількість одиниць; 2) всі серії не однакові за обсягом. Більш поширеним вважається перший варіант.

Серійний спосіб відбору має свої переваги і недоліки перед відбором окремих одиниць. Перевагою вважається те, що його легко організовувати. Але те, що при серійному відборі значно порушується рівномірність розподілу відібраних одиниць у межах генеральної сукупності і більш висока помилка вибірка – є суттєвим недоліком цього способу. Виправляється вказаний недолік збільшенням чисельності вибірки.

Точність результатів серійної вибірки залежить від того, наскільки точно середні показники серій будуть репрезентувати генеральну сукупність. Чим менше вони будуть відхилятися від генеральної середньої , тим точнішими вважаються результати серійної вибірки. Якщо серії в генеральній сукупності мають однакову чисельність (перший варіант відбору), то загальна середня по всій вибірці розраховується як середня арифметична проста із серійних середніх. При неоднакових чисельностях серій визначаться середня арифметична зважена.

Багатоступенева вибірка. Значна частина великих вибіркових обстежень здійснюється не на підставі одного способу відбору, а комбінуванні (поєднанні) двох і більше способів, які утворюють **ступені відбору.**

Так, типовий відбір поєднують з кількома стадіями (ступенями) відбору. При цьому кожна стадія має свою одиницю відбору. Така вибірка називається **багатоступеневою** або **комбінованою.** Наприклад, при обстеженні бюджетів сімей робітників і службовців сільськогосподарських держпідприємств спочатку загальне число сімей, які підлягають обстеженню, розподіляють по держпідприємствах з різним напрямом спеціалізації і по областях. Це перша стадія відбору, яка забезпечує репрезентативність тих чи інших типів спеціалізації господарств і адміністративних областей. У даному випадку одиницею відбору є область і тип спеціалізації. На наступному етапі відбирають підприємства в межах кожного типу спеціалізації в області. Це друга стадія відбору. Здійснюється вона на підставі науково розроблених принципів її організації .

Розглянемо схематично випадок застосування двоступеневої вибірки, коли кожна одиниця містить одне й те ж число М елементів, *m* з яких відбирається із цієї одиниці. Схема двоступеневої вибірки при М= 16 і *m* = 4 зображена на рисунку 25.

Перевага двоступеневої вибірки в тім, що вона надає більшу свободу дій, ніж одноступенева. Коли М= *m,* двоступеневий відбір зводиться до одноступеневого. Але, якщо таке значення *m* не найкраще, беруть дещо менше значення *m,* яке виявляється результативним, тобто дає найкраще співвідношення між статистичною точністю і витратами, зумовленими обсягом вибіркових робіт.

Якщо досліджувані ознаки всередині однієї і тієї ж одиниці спостереження відрізняються незначно, то з міркувань точності величина *m,* повинна бути невеликою. З іншого боку, інколи витрати на спостереження всієї одиниці і деякі підвибірки з неї можуть бути однаковими. Наприклад, якщо одиницею спостереження при бюджетних обстеженнях служить домашнє господарство селян, то одна людина може дати точні зведення про всіх членів сім'ї цього господарства.

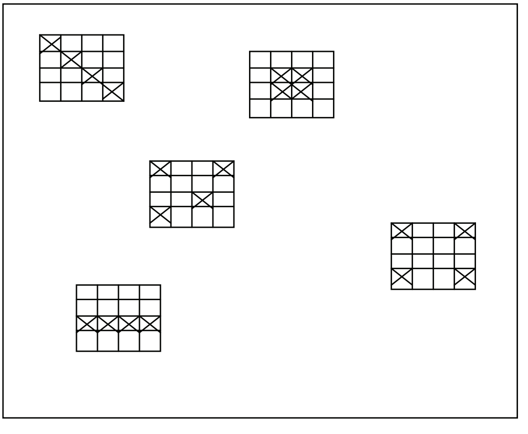


Рис. 25. Схематичне зображення двоступеневої вибірки
(N=256; n=5; M = 16; *m= 4*)

При правильному відборі ознак на кожному ступені відбору можна використати ті з них, які спостерігаються у вибірці великого об'єму як «контрольні», тобто коригують на їх підставі результати спостереження пов'язаних з ними ознак.

Багатоступенева вибірка застосовується при одержанні проб для лабораторних аналізів. Наприклад, при відборі пробних зразків продукції, яка надходить в різній тарі (бочках, ящиках, мішках тощо). Спочатку відбирають певну кількість одиниць продукції в тарі, а потім беруть проби в середині з відібраних видів тари (бочок, мішків тощо). Зокрема, якщо зерно одержують в мішках, спочатку відбирають певну кількість мішків, а потім з них беруть проби зерна щупом зверху, всередині і знизу кожного мішка.

В окремих випадках утворюють так звану середню пробу шляхом змішування окремих проб. Додатковими ступенями відбору тут будуть наступні проби, які беруть у вигляді деякої частини до вихідної (середньої) проби; з останньої беруть знов певну частину і так доти, доки розмір проби не зменшиться до норми, приданої для лабораторного аналізу.

§ 11.4. Помилки вибірки, їх визначення при різних способах відбору

Між характеристиками вибіркової сукупності і шуканими параметрами відповідних характеристик генеральної сукупності існують певні розбіжності. Їх називають **помилками спостереження.** Загальна величина помилки вибіркового спостереження зумовлюється можливістю виникнення двох видів помилок: помилки реєстрації і помилки репрезентативності.

Помилки реєстрації виникають внаслідок недостатнього рівня кваліфікації працівників, неточності підрахунків, недосконалості вимірювальних приладів і т.ін. Ймовірність виникнення помилок реєстрації при вибірковому обстеженні значно менша, ніж при суцільному, адже вибіркове здійснюється кваліфікованішими працівниками і організовується більш ретельно і конструктивно, ніж суцільне. При вибірковому спостереженні завдяки скороченню кількості досліджуваних одиниць значно зменшується можливість одержати помилки реєстрації. Спеціально підібрані і навчені спостережувачі не зацікавлені у викривленні спостережуваних даних, що також сприяє одержанню більш об'єктивної інформації про обстежувану сукупність об'єктів.

У той же час при вибірковому спостереженні виникають помилки, які не мають місця при суцільному обстеженні – **помилки репрезентативності** . Вони являють собою розбіжність між величиною одержаних по вибірці показників і величиною тих показників, котрі були б одержані при проведенні з однаковим рівнем точності суцільного спостереження.

Отже, помилка вибірки (помилка репрезентативності) – це абсолютна величина різниці між відповідними вибірковою і генеральною характеристиками: ($\tilde{x} - \bar{x}$) – помилка для середньої; (w-p) – помилка для частки (w, p - частка ознаки відповідно у вибірковій і генеральній сукупностях). Природа виникнення такої помилки полягає в тому, що вибіркова сукупність не точно відтворює генеральну сукупність.

Помилки репрезентативності можуть бути **випадковими і систематичними.** Так, при вибірковому викопуванні коренів цукрових буряків для визначення їх урожайності у вибіркову сукупність випадково можуть потрапити дещо кращі від середніх екземпляри. У цьому випадку може йти мова про випадкову помилку репрезентативності. У разі, якщо у вибірку будуть систематично відбиратися кращі екземпляри, то мова буде йти про систематичну помилку репрезентативності, яка зумовлена навмисним порушенням правил відбору.

Таким чином – систематичні помилки спрямовані в один бік і можуть виникати у зв'язку з особливостями прийнятої системи відбору і обробки даних спостереження або у зв'язку з порушенням встановлених правил і принципів відбору.

Випадкові помилки не мають певного напряму. Їх виникнення пояснюється недостатньо рівномірним представленням у вибірковій сукупності різних категорій одиниць генеральної сукупності. Оскільки розподіл одиниць спостереження вибіркової сукупності не зовсім точно відтворює розподіл одиниць генеральної сукупності, вибірка не може точно відображати генеральну сукупність, а отже, повністю усунути випадкові помилки неможливо, їх можна звести до незначних розмірів.

Питання визначення можливої і фактичної помилки вибірки має першочергове значення при організації і проведенні вибіркового обстеження. Її величина характеризує ступінь надійності одержаних результатів вибіркового обстеження і зумовлює об’єктивність оцінок параметрів генеральної сукупності. Як і сама вибіркова характеристика, помилка вибірки є випадковою величиною. Розмір випадкової помилки вибірки визначається згідно із граничними теоремами ймовірностей. Розрізняють середню і граничну помилку вибірки. Під **середньою (стандартною) помилкою вибірки** розуміють таке розходження між вибірковою і генеральною середньою ($\tilde{x}-\bar{x}$), яке не перевищує розмір середнього квадратичного відхилення ($\pm\sigma$). Максимально можливе розходження ($\tilde{x}-\bar{x}$) називають **граничною помилкою вибірки,** тобто – це максимум помилки при заданій імовірності її появи.

Існують дві формули середньої помилки вибірки. Одна з них використовується при вимірюванні середнього значення ознаки (наприклад, на підприємстві вибірково обстежується середній розмір зарплати працюючих), друга – коли вибірково вимірюється частка ознаки (наприклад, частка високооплачених працівників на підприємстві).

Коли вибірка здійснюється за принципом повторного відбору, то формули середньої помилки мають вигляд: для середньої - $m=\sqrt{\frac{\sigma^2}{n}}$ для частки - $m=\sqrt{\frac{w(1-w)}{n}}$.

Повторну вибірку використовують дуже рідко. Як правило, вибірка організовується за принципом безповторного відбору. Стосовно до цього принципу відбору в наведених вище формулах середньої помилки в підкореневий вираз вводиться додатковий множник ($1-\frac{n}{\mathrm{N}}$), де N - чисельність генеральної сукупності.

Отже, для безповторної вибірки формули середньої помилки набудуть вигляду:

а) при визначенні середнього значення ознаки - $m=\sqrt{\frac{\sigma^2}{n}(1-\frac{n}{\mathrm{N}})}$;

б) при визначенні частки ознаки - $m=\sqrt{\frac{w(1-w)}{n}\times(1-\frac{n}{N})}$.

Теорією ймовірностей доведено : твердження про те, що генеральні характеристики не відхиляються від вибіркових на величину більшу, ніж величина помилки вибірки (m), завжди має постійний ступінь імовірності 0,683. Імовірність твердження можна підвищити, подвоївши або потроївши середню помилку (2 *m;* 3 *m*). У цьому випадку ймовірність стверджень досягає рівнів 0,954 або 0,997, тобто з тисячі випадків відповідно в 954 і 997 випадках вибіркові характеристики будуть відрізнятися від генеральних на величину обчисленої помилки вибірки. У решти випадків (46 і 3) відхилення генеральних і вибіркових параметрів може виходити за межі обчисленої помилки.

Таким чином, щоб підвищити ймовірність ствердження, необхідно розширити межі відхилень шляхом збільшення середньої помилки в $\frac{|\tilde{x} - \bar{x}|}{\sigma}$ разів, де відношення різниці середніх до середнього квадратичного відхилення являє собою величину так званого **нормованого відхилення** (t).

Отже, з визначеною ймовірністю можна стверджувати, що відхилення генеральних і вибіркових характеристик не перевищать деякої величини – граничної помилки вибірки (Δ). Гранична помилка вибірки пов'язана з середньою помилкою рівнянням Δ= t *m,* де t-нормоване відхилення (коефіцієнт кратності, коефіцієнт довіри), яке залежить від рівня ймовірності.

Величина ймовірності задається залежно від мети і завдань дослідження. Ймовірність потрапляння помилки репрезентативності у межах $\pm t$ визначається за формулою інтеграла ймовірностей $p(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$.

Таблиця 87

Витяг із стандартних таблиць «Функція Лапласа»

t	1,00	1,96	2,00	2,50	2,58	3,00	3,30
p	0,683	0,950	0,954	0,997	0,990	0,667	0,999

Значення цього інтеграла міститься в стандартних математичних таблицях «Функція Лапласа» (див. додат. 5,). В таблиці 87 наведені рівні ймовірностей *p* для деяких цілих і дробових значень t.

Припустимо, що помилку вибірки треба оцінити з імовірністю 0,954. Це означає, що розбіжність між вибірковою і генеральною середньою не перевищить двох величин середньої помилки, тобто в 95,4 % випадків помилка репрезентативності не вийде за межі ± 2; при ймовірності 0,997 – за межі ± 3 і т. ін.

Для чисельно малих статистичних сукупностей не може бути застосована теорема Ляпунова, яка з'ясовує загальні умови, при здійсненні котрих розподіл суми незалежних випадкових величин прямує до нормального, оскільки значення вибіркової середньої ($\bar{x}$) тут занадто залежить від величини кожної випадкової змінної. Характер розподілу $\bar{x}$ в цих умовах буде істотно відрізнятися від нормованого розподілу, а довірчі інтервали і довірчі ймовірності (про них мова піде нижче) при малих вибірках можуть бути розраховані тільки за умов нормального розподілу досліджуваної ознаки. За розрахунками Ст'юдента, ймовірність того, що абсолютна величина різниці вибіркової і генеральної середньої буде менше граничної помилки вибірки ($\left|\tilde{x}-\bar{x}\right|\langle t\mu$) і являє собою функцію від нормативного відхилення (t) і чисельності вибірки (n). Формула цього доведення має вигляд: $p(\left|\tilde{x}-\bar{x}\right|\langle t\mu)=\int_{-t}^{+t}A(1+\frac{t^2}{n-1})^{-\frac{n^2}{2}}dt$,

де $A=\frac{\Gamma(\frac{n}{2})}{\sqrt{n(n-2\Gamma(\frac{n-1}{2})}}$, $\Gamma(\frac{n}{2})$ - гамма – функція.

У практичних розрахунках використовуються таблиці розподілу С'юдента S (t), в яких дано рівні ймовірностей для різних значень n і t (додат.2).

На основі теоретичних рівнів імовірностей розраховують фактичні їх рівні. При цьому розрахункова ймовірність (p) становитиме : p = $[S(t)-0.5]\cdot 2$. Із сказаного вище випливає, що після обчислення середньої помилки вибірки виникає питання обчислення граничної помилки репрезентативності (Δ) розмір її у вибірковому спостереженні може бути менший або більший від середньої помилки репрезентативності (μ). Згідно теореми Чебишева і Ляпунова, яка визначає ймовірність того, що гранична помилка вибірки не перевищить t разів взяту середню помилку вибірки (μ), вирішують питання про граничну помилку. Наведемо формулювання теореми Чебишева: з імовірністю, як завгодно близькою до одиниці, можна стверджувати, що при достатньо великій кількості незалежних спостережень вибіркова середня ($\tilde{x}$) буде як завгодно мало відрізнятися від генеральної середньої ($\tilde{x}$). Отже, гранична помилка вибірки (Δ) обчислюється з певною ймовірністю (p), якій відповідає t – разове значення середньої помилки (μ):Δ= tμ.

Межі середньої характеристики в генеральній сукупності становитимуть: для середньої - $\bar{x} = \tilde{x} \pm \Delta$;

для частки - $p = w \pm \Delta$.

У розгорнутому вигляді формули граничної помилки для повторної і безповторної схеми відбору наведені у наступному прикладі.

Приклад. Розглянемо конкретний приклад розрахунку граничної помилки вибірки при визначенні середньої характеристики у вибірковій сукупності і частки вибірки.

Умова. У сільськогосподарських підприємствах району площа зернових культур становить 20000 га. При 10 % - му безповторному відборі встановлено, що середня урожайність зернових в районі дорівнює 30 ц/га, середньоквадратичне відхилення урожайності становить 2 ц. Питома вага високоврожайних культур 60%. Потрібно визначити з ймовірністю 0,954 граничну помилку середньої врожайності зернових культур по вибірці і граничну помилку частки, тобто питомої ваги високоврожайних культур в загальній площі посіву.

Хід рішення. Встановлюємо чисельність вибіркової сукупності – 10% від 20000 га, вона дорівнює 2000 га. Маємо : $\tilde{x}$ = 30; N= 20000; *n*= 2000; σ = 2; w= 0,60; p= 0,954; t= 2.

$$\Delta_{\bar{x}} = t\sqrt{\frac{\sigma^2}{n}(1-\frac{n}{N})} = 2\sqrt{\frac{4}{2000}(1-\frac{2000}{20000})} = 0{,}08.$$

Отже, різниця між вибірковою середньою урожайністю і генеральною середньою буде не більша за 0,08 ц. Межі середньої генеральної урожайності в центнерах: $29{,}92 \le 30 \le 30{,}08$.

Гранична помилка для частки становить:

$$\Delta_w = t\sqrt{\frac{w(1-w)}{n}(1-\frac{n}{N})} = 2\sqrt{\frac{0{,}60(1-0{,}60)}{2000}(1-\frac{2000}{20000})} = 0{,}02.$$

Таким чином, помилка у визначенні частки високоврожайних культур у вибірковій сукупності не перевищить 2 %, тобто питома вага високоврожайних культур знаходиться у межах $58 \le 60 \le 62\%$.

Величина випадкової помилки репрезентативності залежить: 1) від способу формування (відбору) вибіркової сукупності; 2) від обсягу вибірки; 3) від ступеня варіації досліджуваної ознаки у генеральній сукупності.

А це означає, що для одержання мінімальної помилки необхідно дотримуватися таких математичних положень: 1) чим більший обсяг вибірки, тим повніше взаємопогашаються випадкові відхилення. Величина помилки вибірки обернено пропорційна кореню квадратному з чисельності вибірки. При збільшенні вибіркової сукупності у чотири рази помилка вибірки зменшується у два рази; 2) збільшення показника варіації досліджуваної ознаки зумовлює збільшення помилки вибірки, тобто величина останньої прямо пропорційна середньому квадрату відхилень.

Слід пам'ятати, що при вибірковому обстеженні відсутня інформація про розмір дисперсії, тому величина її приймається наближеним показником у вигляді вибіркового середнього квадрата відхилень.

Для кожного конкретного способу відбору у вибіркову сукупність величина помилки репрезентативності може бути визначена за відповідними формулами.

Повернемося до повторної і безповторної схеми відбору з генеральної сукупності. Оскільки при безповторному відборі чисельність генеральної сукупності зменшується (при повторному – вона залишається незмінною), після кожного відбору ймовірність потрапити у вибірку для одиниць, що залишаються, підвищується. Тому середня помилка тут буде меншою, ніж при повторному відборі.

Перетворення формули середньої помилки для середньої при повторному відборі $m_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$ у вигляд $m_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ дає підстави стверджувати, що середня квадратична помилка (середнє квадратичне відхилення вибіркової середньої від генеральної) прямо пропорційна варіації ознаки у генеральній сукупності і обернено пропорційна кореню квадратному з обсягу вибірки. Гранична помилка вибірки (Δ), як випадкова величина, може бути в кожному конкретному випадку менша, рівна або більша за середню помилку(m) . Ймовірність її величини при досить великій сукупності вибірки визначають за теоремою Ляпунова:

$$p(\Delta \le tm = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt = f(t)$$.

Значення інтеграла Лапласа (функція від t) містяться в стандартних математичних таблицях (додаток). За такими таблицями можна встановити, що

при t = $1 p(\Delta \le m) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt = 0{,}683$;

при t = $2 p(\Delta \le 2m) = \frac{1}{\sqrt{2\pi}} \int_{-2}^{+2} e^{-\frac{t^2}{2}} dt = 0{,}954$.

Наведені розрахунки свідчать про те, що практично неймовірно одержати помилку, більшу за 3 m, тобто більшу, ніж $3\frac{\sigma}{\sqrt{n}}$. Отже, практично вірогідно, що генеральна середня не вийде за границі:

$$\tilde{x} - 3\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \tilde{x} + 3\frac{\sigma}{\sqrt{n}}.$$

Як уже зазначалося вище, при безповторному способі відбору для середньої помилки вибірки вводять поправочний коефіцієнт $\frac{N-n}{N-1}$, де *n,* N- відповідно чисельність вибіркової і генеральної сукупностей. Для досить великих обсягів генеральної сукупності замість значення N- 1 вводять значення N, тоді формула набуває вигляду: $\frac{N-n}{N-1} \approx \frac{N-n}{N} = 1 - \frac{n}{N}$.

З врахуванням наведеної поправки дисперсія вибіркової середньої становить:

$$\sigma_{\tilde{x}}^2 = \frac{\sigma^2}{n}(1 - \frac{n}{N}).$$

Середні помилки середньої і частки у вибірковій сукупності для власне випадкового відбору наведено у згаданій вище таблиці.

Для механічного способу відбору помилка репрезентативності розраховується аналогічно формулам для власне випадкового відбору.

При типовому способі відбору розрахунок середньої помилки має деякі особливості. Розглянемо їх.

З викладеного вище зрозуміло, що середня помилка вибірки залежить від середнього квадрата відхилень (дисперсії) досліджуваної ознаки. Згідно правилу складання і розкладання дисперсій маємо: $\sigma_y^2 = \sigma_x^2 + \sigma_Z^2$, де σ_y^2- загальна дисперсія; σ_x^2- між групова дисперсія; σ_Z^2 - внутрішньогрупова дисперсія.

Для типової вибірки міжгрупова дисперсія вимірює варіацію групових середніх ($\tilde{x}_j$) відносно загальної середньої ($\tilde{x}$), тобто :

$$\sigma_x^2 = \frac{\sum(\tilde{x}_j - \tilde{x})^2 n_i}{\sum n_i}.$$

Цей вид дисперсії пояснює варіацію, викликану ознакою, покладеною в основу групувань при виділенні типових груп і не може розглядатися як помилка вибірки. Називають її систематичною дисперсією. Отже, при розрахунках середньої помилки вибірки цей вид дисперсії виключається.

Але кожна типова група має варіацію ознаки, викликану впливом різних неврахованих факторів - внутрішньогрупову (залишкову) варіацію. Остання розраховується як середня арифметична з групових дисперсій:

$$\sigma_Z^2 = \frac{\sum \sigma_j^2 N_j}{\sum N_j},$$

де σ_j^2 - групові вибіркові дисперсії.

Саме ця частина варіації залишається непоясненою і повинна розглядатися як помилка вибірки. Тобто формула середньої помилки типової вибірки має вигляд:

$$m_{тип} = \sqrt{\frac{\sigma_Z^2}{n}},$$

де *n* – загальний обсяг вибірки ($n = \sum n_j$).

За правилом складання і розкладання дисперсій маємо : $\sigma_Z^2 \langle \sigma_y^2$, тому середня помилка типової вибірки, як правило, менша за середню помилку при власне випадковій вибірці. Оскільки середня помилка типової вибірки дає точніші результати (висновки), її широко використовують в дослідженні економічних явищ.

Треба пам’ятати, що організація типової вибірки зумовлена, як правило, власне випадковим відбором. Адже відбір одиниць з кожної групи здійснюють власне випадковим методом. При цьому застосовується схема безповторного відбору. З цих причин до середньої помилки середньої чи частки при безповторній схемі відбору вводять поправку $1 - \frac{n}{N}$.

При серійному способі відбору по кожній відібраній серії розраховується значення дисперсії. Середня арифметична з цих дисперсій становить внутрішньосерійну (σ_{BC}^2), тобто залишкову дисперсію.

Варіація серійних середніх ($\tilde{x}_j$) навколо загальної вибіркової середньої $\tilde{x}$ характеризується міжсерійною дисперсією (σ_{MC}^2) . Структурна формула її має вигляд:

$$\sigma_{MC}^2 = \frac{\sum (\tilde{x}_j - \tilde{x})^2}{n_c},$$

де n_c - чисельність вибірки в серії.

Згідно з правилом розкладання дисперсії маємо $\sigma_y^2 = \sigma_{вс}^2 + \sigma_{мс}^2$.

Внутрішньосерійна дисперсія розраховується на основі даних суцільного спостереження відібраних серій. А це означає, що помилка репрезентативності залежить від міжсерійної дисперсії. Її розраховують за схемою безповторного відбору серій:

$$m = \sqrt{\frac{\sigma_{мс}^2}{n_c}(1 - \frac{n_c}{N})}.$$

Для зазначених вище способів відбору при розрахунках граничної помилки як для середньої, так і для частки, середню помилку множать на коефіцієнт довіри (t), величина якого залежить від рівня обраної ймовірності. Детально це питання викладено в § 4.4.

§ 11.5. Організація вибіркового спостереження

Здійснення вибіркового обстеження ґрунтується, насамперед, на знанні природи досліджуваних процесів та явищ і глибокому теоретичному аналізі. Вибіркове обстеження починають з копіткої підготовки роботи, яка передбачає вирішення таких питань: мета і об'єкт дослідження; програма та інструментарій обстеження; джерела і способи збирання необхідної інформації; підбір і підготовка кадрів, пробні обстеження і ряд інших питань.

Організація вибіркового спостереження з метою відтворення генеральної сукупності висуває ряд завдань, вирішення яких ґрунтується на теорії вибіркового методу. Розглянемо їх .

1. По-перше, це вирішення питання щодо встановлення чисельності вибіркової сукупності. Суть цього завдання полягає в тому, щоб знайти відповідь на запитання - скільки потрібно відібрати одиниць спостереження, щоб помилка вибірки з певним рівнем ймовірності не перевищувала встановлений розмір.

2. Друге завдання вибіркового спостереження має на меті оцінку показників, одержаних за вибірковими даними. Вирішення цього завдання полягає у визначенні граничної помилки вибіркової сукупності.

3. Третє завдання вибірки зводиться до встановлення ймовірностей здійснення певного розміру помилки. Для цього необхідно знати середню і граничну помилки вибірки, розрахувати нормоване відхилення, на підставі якого за стандартними таблицями інтеграла ймовірності визначається рівень ймовірності.

Визначення границь, в яких знаходяться характеристики всієї сукупності, ускладнюється у випадках, коли генеральна сукупність досліднику невідома. Адже при відомій генеральній сукупності завжди можна побудувати схему всіх можливих випадків вибірки з цієї сукупності. При невідомій генеральній сукупності таку схему побудувати неможливо і кількісні характеристики генеральної сукупності не можуть бути використані для оцінки одержаних вибіркових характеристик.

Для оцінки різниці між вибірковою і генеральною характеристиками застосовують так звані «прямі» теореми теорії ймовірності (теореми Бернулі, Муавра – Лапласа, Чебишева та ін.) лише у разі, коли генеральна характеристика - величина постійна, а вибіркова – випадкова змінна. Оскільки в нашому випадку дана умова не витримується, при вирішенні завдання одержання оцінки виходять з припущення, що невідома нам генеральна характеристика може мати ті чи інші значення з відповідними їм ймовірностями, тобто вона розглядається як випадкова змінна. Теорія ймовірностей для вирішення цього завдання застосовує так звані «обернені» теореми Бернулі, Лапласа, «другий закон великих чисел» та ін. Застосування їх поширюється лише на вибірки необмеженого обсягу і не може бути застосовано для оцінок точності кінцевих вибірок.

4. У зв’язку з тим, що в практиці вибіркового обстеження виникають помилки (чи похибки), розміри яких певною мірою зумовлені способом застосованого відбору та обсягом вибірки, постає питання оцінки того чи іншого способу відбору. Отже, виникає завдання оцінки точності способу, за яким здійснюється відбір у вибіркову сукупність. Це завдання вирішується шляхом з’ясування таких питань: чи відомі дані про всю сукупність; яким способом здійснюється вибірка ; якими будуть результати, тобто якими властивостями буде володіти вибіркова сукупність. Чи будуть ці властивості і характеристики вибіркової сукупності відрізнятися від відомих досліднику властивостей і характеристик генеральної сукупності? Наскільки великими будуть відхилення (помилки вибірки)? Чим вони зумовлюють і чи можуть бути зменшені і наскільки?

Вирішення цього завдання ґрунтується на всебічному аналізі властивостей досліджуваної сукупності і врахуванні властивостей застосованого способу відбору. Важлива роль тут належить теорії ймовірностей, адже мова йде про оцінку точності вибірки. Дослідник розглядає всі можливі варіанти вибірок, зіставляючи результати аналізу з характеристиками заданої генеральної сукупності. Застосовуються тут «прямі» теореми теорії ймовірностей.

Вирішення зазначених вище завдань вибіркового спостереження можливе лише за умов знання теорії вибіркового методу. Вони спрямовані на досягнення основної мети вибірки – відтворення генеральної сукупності. Їх вирішення потребує детального розгляду питань про помилки вибіркового обстеження, точкову і інтервальну оцінки, нормоване відхилення, довірчу ймовірність і довірчий інтервал. Цим питанням присвячуються окремі параграфи посібника.

Оскільки вибіркове спостереження відбирає, «вихоплює» факти з досліджуваної генеральної сукупності, при невдалій його організації можуть бути зареєстровані явища, не пов'язані одне з одним, перед організацією вибіркового обстеження стоїть завдання здійснити відбір таким чином, щоб існуючі взаємозв'язки між одиницями знайшли відображення у вибірковій сукупності. Таке завдання вирішується шляхом укрупнення одиниць відбору, переходу до серійної вибірки. Остання гарантує охоплення в цілому групи взаємно пов'язаних фактів. А це означає, що створена можливість для вивчення їх взаємодії і взаємозалежності.

В організації вибіркового спостереження мають місце випадки, коли поєднується відбір серій з відбором одиниць всередині останніх. Така вибірка називається комбінованою. Трапляються випадки організації вибіркового спостереження, при якій відбір здійснюється всередині однієї з пов'язаних між собою сукупностей, а з другої сукупності в обстеження включають ті одиниці, котрі свідомо пов'язані з відібраними одиницями першої сукупності.

В організації вибіркового спостереження виняткового значення набуває питання про визначення таких категорій обстеження, як одиниця відбору і одиниця спостереження. Під **одиницями відбору** розуміють частини сукупності, які включені у склад вибірки. **Одиницями спостереження** називають такі частини сукупності, ознаки яких підлягають спостереженню і реєструються окремо від інших в процесі обстеження. Одиниця відбору – це категорія, притаманна виключно вибірковому методу, у той час як одиниця спостереження є категорією загальностатистичною, яка має відношення і до суцільного і до несуцільного спостереження.

Відмінність понять одиниці відбору і одиниці спостереження особливо різко виявляється при серійній вибірці, при організації якої відбір здійснюється групами (серіями) одиниць, які складають досліджувану сукупність, а спостереженню підлягає кожна одиниця спостереження. В серійних вибірках одиниця спостереження є частиною одиниці відбору.

Як правило, у вибірковому обстеженні одиниця відбору і одиниця спостереження збігаються між собою. Але одиниця відбору не повинна бути меншою за одиницю спостереження. Наприклад, якщо одиницею спостереження є підприємство, то одиницями відбору не можуть бути відділки, бригади, ферми тощо. З іншого боку, при фіксованій чисельності вибірки або частки явища одиниця відбору не повинна значно перевищувати одиницю спостереження. Якщо, наприклад, при серійному відборі відокремлюється 60 одиниць з 1000, то краще відібрати 6 серій по 10 одиниць, ніж 3 серії по 20 одиниць.

Досить важливим питанням організації вибіркового обстеження є визначення схеми і способу відбору. При його вирішенні виходять перш за все з розміру помилки вибірки.

Наступним важливим питанням вважається вирішення завдання щодо чисельності вибіркової сукупності. Тут статистик повинен орієнтуватися стосовно граничних помилок розміру досліджуваного явища і ймовірності, з якою ці границі повинні бути гарантовані.

Організація вибірки передбачає попередню перевірку її репрезентативності. Із цією метою за даними поточного або минулого періоду розраховують кілька найбільш важливих показників для генеральної і для вибіркової сукупностей. Одержані показники зіставляють і таким чином знаходять різницю між ними. У соціально – економічних дослідженнях вважається задовільним відхилення вибіркових даних від даних по генеральній сукупності в межах 5 %. Якщо відхилення перебільшують зазначений рівень, вибірка вважається нерепрезентативною, і відбір повторюють. У випадках, коли повторний відбір не дає бажаних наслідків, збільшують чисельність вибіркової сукупності.

Організація вибіркового спостереження потребує врахування всіх можливих умов і обставин, в яких буде відбуватися обстеження. Якщо ж виникають нові обставини в ході обстеження – в організацію вибірки вносять відповідні корективи.

Викладене вище зумовлює здійснення вибіркового спостереження за такими етапами: 1) формування мети спостереження; 2) відмежування генеральної сукупності; 3) встановлення системи відбору одиниць для спостереження; 4) визначення числа одиниць, які підлягають відбору; 5) підготовка основи вибірки (відбору); 6)проведення статистичного спостереження, тобто планомірного, науково організованого збору даних про досліджуване явище (процес); 7) розрахунок вибіркових характеристик і їх помилок; 8) поширення вибіркових даних на генеральну сукупність. Перелічені вище етапи знаходять своє відображення у висвітленні теоретичних і практичних аспектів вибіркового методу.

При організації вибірки значення має визначення необхідної її чисельності (*n*). Остання залежить від варіації одиниць обстежуваної сукупності. Чим більша коливність, тим більшою повинна бути чисельність вибірки. Зворотний зв'язок існує між чисельністю вибірки її граничною помилкою. Прагнення отримати меншу помилку вимагає збільшення чисельності вибіркової сукупності.

Необхідна чисельність вибірки визначається на основі формул граничної помилки вибірки (Δ) із заданим рівнем ймовірності (р). Шляхом математичних перетворень отримують формули розрахунку чисельності вибірки (табл. 88).

Таблиця 88

Розрахунок необхідної чисельності вибірки

Спосіб відбору	Чисельність вибірки (*n*) при визначенні та оцінці параметра	
	середньої ($\bar{x}$)	частки (w)
Повторний	$\frac{t^2\sigma^2}{\Delta^2_{\bar{x}}}$	$\frac{t^2 w(1-w)}{\Delta^2_w}$
Безповторний	$\frac{t^2\sigma^2 \mathrm{N}}{\Delta^2_{\bar{x}}\mathrm{N} + t^2\sigma^2}$	$\frac{t^2 w(1-w)\mathrm{N}}{\Delta^2_w \mathrm{N} + t^2 w(1-w)}$

При вирішенні питання про визначення чисельності вибірки дослідник вимушений роботи припущення про генеральні характеристики або здійснювати для них пробну вибірку. Лише за таких умов буде об'єктивною величиною чисельність вибіркової сукупності, розрахована за розглянутими формулами.

Для розрахунку чисельності вибірки можна користуватися стандартними таблицями і номограмами, в яких для певних величин граничної помилки і довірчої ймовірності наводиться рекомендована чисельність одиниць вибіркового спостереження. Прикладом такої довідкової таблиці є таблиця 89. Останню можна перебудувати таким чином, що заданими величинами будуть чисельність вибірки і гранична помилка. У такому разі можна знайти рівень ймовірності або величини чисельності вибірки та рівня ймовірності, за якими знаходиться можлива гранична помилка (табл. 90).

Таблиця 89

Чисельність вибірки при заданих рівнях граничної помилки(Δ) і ймовірності (*P*)

P	Δ									
	0,10	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01
0,85	51	63	80	105	143	207	323	575	1295	5180
0,90	67	83	105	138	187	270	422	751	1090	6763
0,95	96	118	150	195	266	384	600	1067	2400	9603
0,99	165	204	359	338	460	663	1036	1843	4146	16587
0,997	220	271	344	449	611	880	1376	2446	5504	22018
0,999	270	334	422	552	751	1082	1691	3007	6767	27069

Таблиця 90

Рівень довірчої ймовірності при заданому числі спостережень (*n*) і граничній помилці(Δ)

n	Δ									
	0,10	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01
50	0,843	797	242	678	604	520	428	329	223	112
100	0,954	928	890	838	770	689	576	452	310	159
200	0,995	989	976	952	910	843	742	604	428	223
300	0,999	998	994	985	962	917	834	701	511	271
500				0,998	993	975	926	820	629	345
1000						0,998	989	942	794	473
2000								0,993	926	629
5000									0,995	843
10000										0,954

Одержані в результаті вибіркового спостереження статистичні характеристики з врахуванням їх точності можуть бути поширенні на генеральну сукупність. Пов’язану із цим обчислювальну роботу здійснюють двома способами, які вважаються найпростішими у статичній практиці: 1) спосіб прямого перерахунку; 2) спосіб поправочних коефіцієнтів.

Спосіб прямого перерахунку полягає у тому, що вибіркові характеристики (середня, частка) помножують на відповідний показник обсягу генеральної сукупності з врахуванням значення граничної помилки репрезентативності . Наприклад : вибірковим обстеженням 10 % приватних господарств при загальній їх кількості 2500 встановлено, що середня чисельність птиці на одне подвір’я становить 30 голів. Щоб визначити загальну чисельність поголів'я птиці в приватному секторі обстежуваного об'єкта, необхідно загальну кількість дворів перемножити на середню чисельність птиці в одному дворі (2500 · 30=75000).

Спосіб поправочних коефіцієнтів - застосовується для перевірки і уточнення даних суцільного спостереження. Із цією метою проводять контрольні вибіркові спостереження. Суть його полягає у тому, що по одних і тих же об'єктах зіставляють дані суцільного і контрольного вибіркового обстеження. У результаті такого зіставлення обчислюють поправочні коефіцієнти. Останні використовують для внесення відповідних поправок по даних суцільного спостереження.

Наприклад, у результаті суцільного обстеження фермерських господарств району встановлено, що кількість поголів'я свиней в них становить 7000 гол. Контрольне вибіркове обстеження 10 % фермерських господарств показало, що в них при суцільному обстеженні зареєстровано 800 голів, а при контрольному обстеженні 816 голів. Отже, при суцільному обстеженні було недорахована 16 голів, що становить 2 % від 800 голів. Поправочний коефіцієнт тут становить 102 % (1,02). Цей показник використовується для внесення поправок у результати суцільного спостереження. В нашому прикладі слід вважати загальну чисельність поголів'я свиней не 7000, а 7140 голів (7000 · 1,02).

При поширенні даних вибіркової сукупності на генеральну слід враховувати таку важливу обставину, як однорідність вибірки і всієї досліджуваної сукупності. Тому зазначеним способом треба користуватися з обережністю. Особливо це стосується дослідження соціально – економічних явищ і процесів, яким притаманна риса мінливості.

ТЕМА 12. ПОДАННЯ СТАТИСТИЧНИХ ДАНИХ: ТАБЛИЦІ, ГРАФІКИ, КАРТИ

§ 12.1. Статистичні таблиці, їх види і правила оформлення

Результати обробки статистичних даних оформляють у вигляді статистичних таблиць. **Статистична таблиця** – це форма раціонального і наочного викладення цифрових характеристик досліджуваних явищ і їх складових частин. За допомогою інформації у вигляді зведених статистичних таблиць створюється можливість характеристики явищ з погляду їх розміру, структури та динаміки розвитку. А стислість і наочність форм зображення даних у таблиці зумовлюють їх сприймання, розуміння й аналіз. Адже таблична форма дозволяє викласти матеріал найбільш зручно, компактно, наочно і раціонально.

До складових статистичної таблиці належать вертикальні графи, горизонтальні рядки, які, перетинаючись, утворюють клітини, а також відповідні заголовки. У сукупності зазначені елементи форми таблиці утворюють її макет (рис. 26).

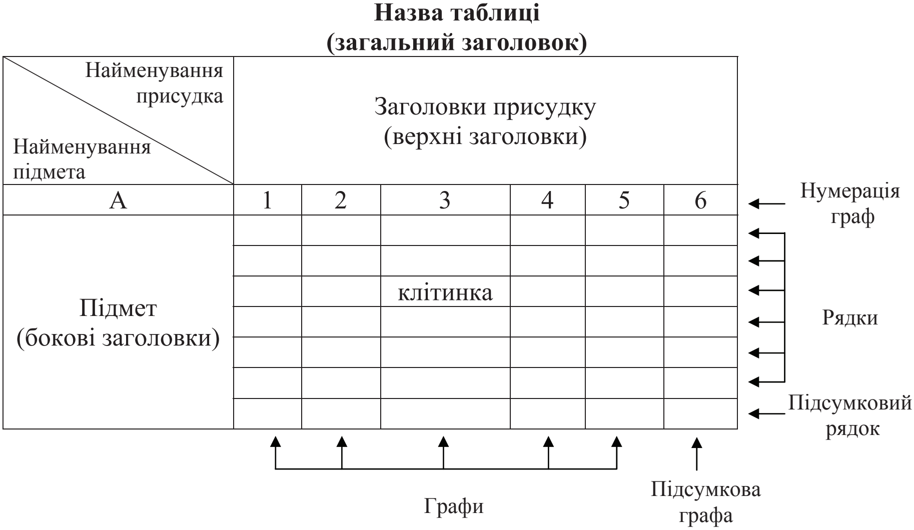


Рис.26.Схема макету статистичної таблиці

Статистична таблиця має загальний і внутрішні заголовки Перший з них у стислій формі відображує зміст таблиці. Останні – пояснюють змістовне навантаження відповідних рядків і граф.

Основними елементами статистичної таблиці є **підмет** і **присудок.**

Підметом таблиці є одиниці статистичної сукупності, або їх групи, які характеризуються показниками. Присудок таблиці – це система показників, які характеризують підмет.

Залежно від побудови (розробки) підмета статистичні таблиці поділяють на три групи: **прості, групові, комбінаційні**.

1. **Таблиця проста**, або перелікова, містить зведені показники щодо переліку одиниць спостереження, переліку хронологічних дат або територіальних підрозділів. Відповідно їх називають таблицями простими, або переліковими, хронологічними, або територіальними. Отже, в підметі простої таблиці перелічуються одиниці сукупності (підприємства, виробничі підрозділи, види продукції тощо) або одиниці часу (роки, квартали, місяці тощо).

2. **Таблиця групова** містить зведення про сукупність, розчленовану на окремі групи за однією ознакою. При цьому кожна група може бути охарактеризована рядом показників (табл. 91).

Таблиця 91

Групування підприємств за розміром посівної площі зернових культур

Групи підприємств за розміром посівної площі	Кількість підприємств у групі	Посівна площа, га	Урожайність, ц/га
І-до 500	3	460	36,3
ІІ – 500-700	5	680	39,6
ІІІ – 700 - 900	16	810	43,8
IV – 900 - 1100	4	1030	50,6
V –понад 1100	3	1270	53,2
Всього	31	-	-

3. **Таблиця комбінаційна** – у підметі такої таблиці містяться групи за двома і більше ознаками (табл. 92).

Таблиця 92

Вплив якості ґрунту та кількості внесених добрив на урожайність зернових культур

<table>
<tr><td rowspan="4">Показники</td><td colspan="6">Групи підприємств за якістю ґрунту, балів</td></tr>
<tr><td colspan="2">І –до 50</td><td colspan="2">ІІ -50-70</td><td colspan="2">ІІІ – понад 70</td></tr>
<tr><td colspan="6">підгрупи підприємств за кількістю внесених мінеральних добрив, ц діючої речовини</td></tr>
<tr><td>до2,5</td><td>понад 2,5</td><td>до 2,5</td><td>понад2,5</td><td>до 2,5</td><td>понад 2,5</td></tr>
<tr><td>Кількість підприємств у групі</td><td>3</td><td>6</td><td>4</td><td>7</td><td>6</td><td>3</td></tr>
<tr><td>Урожайність, ц/га</td><td>32,7</td><td>39,6</td><td>40,2</td><td>43,4</td><td>45,0</td><td>52,6</td></tr>
</table>

Статистичні таблиці можуть бути побудовані за принципом розробки їх присудка, а саме: з простою розробкою присудка та із складною розробкою присудка. У таблиці з простою розробкою присудка показники, які характеризують підмет, не пов’язані між собою (табл. 93), а у таблиці складною розробкою присудка такі показники пов'язані між собою (табл. 94).

Таблиця 93

Макет статистичної таблиці з простою розробкою присудка

Бригади підприємства	Кількість працюючих, чол.	У тому числі			
		чоловіки	жінки	з виробничим стажем, років	
				до 5	понад 5
1					
2					
3					

Таблиця 94

Макет статистичної таблиці з складною розробкою присудка

Бригади підприємства	Кількість працюючих, чол.	У тому числі з виробничим стажем, років			
		до 5		понад 5	
		чоловіки	жінки	чоловіки	жінки
1					
2					
3					

Вигляд статистичної таблиці залежить не від розробки присудка, а від побудови підмета.

При побудові комбінаційних таблиць найчастіше використовують комбінаційне групування за 2-3 ознаками. Якщо виділяти більшу кількість груп і підгруп, одержимо досить громіздке групування. Наприклад, якщо за першою ознакою виділити 2 групи, за другою – 3, за третьою– 4, то у підметі такого групування одержимо 24 рядки (2×3×4), не рахуючи підсумкових рядків по групах і загального підсумку. Тому з введенням кожної нової групувальної ознаки число рядків у кілька разів збільшується, що значно ускладнює читання такої таблиці і та аналіз.

Слід також враховувати ту обставину, що при комбінаційному групуванні за кількома ознаками відбувається велике подрібнення сукупності на групи і підгрупи. У результаті цього в одержаних підгрупах може виявитися недостатня кількість одиниць спостереження. У такому разі числові показники не будуть характеризувати типові властивості, адже вони не втрачають своєї індивідуальності.

Результати комбінаційного групування за великою кількістю ознак навіть при невеликій кількості інтервалів важко сприймаються, а сама таблиця втрачає свою найважливішу перевагу – наочність. Тому комбінаційні таблиці доцільно складати не більше як за трьома факторами та кількістю інтервалів не більше трьох (інколи чотирьох).

Уперше прості і групові таблиці були застосовані при вивченні статистичних даних у 1727 р. І. К. Кириловим, а комбінаційна таблиця була розроблена і запропонована у 1882 р. А. П. Шлікевичем,

Для досягнення найбільшої виразності статистичної таблиці необхідно при її оформленні дотримуватися певних правил. Розглянемо їх.

1. Форма статистичної таблиці повинна бути узгоджена із раніше існуючими таблицями для забезпечення можливості порівняння даних за ряд відрізків часу.

2. Назва таблиці (загальний заголовок) повинна коротко і точно характеризувати основний її зміст. Ця вимога рівною мірою стосується і назв підмета та присудка таблиці. Якщо загальний заголовок недостатньо докладно сформульований, то можна зробити примітки до нього.

3. У таблиці (в загальній назві або в назвах підмета і присудка) повинно бути вказано, якої території або якого періоду чи моменту часу стосуються наведені дані, а також характер цих даних (фактичні, нормативні, розрахункові і т. д.).

4. Показники таблиці повинні мати одиниці виміру. Якщо для всіх показників використовується одна одиниця виміру, її пишуть наприкінці заголовка таблиці, а якщо їх кілька – в кінці рядків або граф.

5. Усі числові значення даного показника зазначаються з однаковою точністю.

6. Якщо кількість показників підмета і присудка таблиці досить велика, то їх нумерують. При цьому графи, які містять перелік об'єктів або їх груп, позначають великими літерами алфавіту, а графи з показниками присудка – арабськими цифрами.

7. У кожній клітині таблиці слід писати відповідне -число або умовні знаки, які мають таке значення: ... (три крапки) – відсутність даних, або робиться запис «немає зведень»; – (тире) – замість нульового значення або якщо числове значення не має логічного

змісту чи неможливе зовсім; число 0,0 ставиться у тих випадках, коли величина показника не перевищує 0,05.

8. Таблиця повинна бути замкненою, тобто мати підсумки (в цілому по групах і підгрупах).

9. Якщо по тексту зустрічаються дві і більше таблиць, їх нумерують арабськими цифрами. Номер таблиці ставиться перед її назвою, слово «Таблиця» і знак «№» не пишуть.

При необхідності до таблиці дають примітки і зноски. Примітки даються у випадках необхідності додаткових пояснень щодо показників таблиці. У зносках вказують джерела зведень, наведених у таблиці, уточнюють дату тощо.

Грамотно складена статистична таблиця є важливим засобом викладення обробленого статистичного матеріалу та його аналізу.

§ 12.2 Графічний метод

12.2.1. Роль і значення графічного методу в наукових дослідженнях

Графічні методи вважаються досить важливим та ефективним знаряддям сучасної науки, вони надійно увійшли в методику наукових досліджень. Особливо велику роль ці методи відіграють у статистичних дослідженнях, де вивчаються складні взаємозв'язки соціально-економічних явищ і процесів в русі показників динаміки, а також складні переплетіння зв'язків у просторі .

Статистичні графіки використовують з метою узагальнення статистичних даних, їх аналізу та популяризації (останнє стосується неспеціалістів).

Що ж являє собою **статистичний графік?** Це спосіб наочного подання і викладення статистичних даних та їх співвідношень за допомогою геометричних знаків (сукупності крапок, ліній, поверхонь) та інших графічних засобів з метою їх узагальнення й аналізу. За допомогою графіків більш глибоко вивчають склад і динаміку явищ, а також взаємозв'язки між ними.

Застосування графічного методу при вивченні соціально-економічних явищ досить різнопланове. Так, його використовують для порівняння обсягів певних статистичних сукупностей та вивчення їх складу. Прикладом може бути графічне зображення складу спеціалістів певної галузі народного господарства за віком, статтю, фахом або обсягів і по-галузевого складу валової продукції сільського господарства тощо. У даному випадку роль графічного методу зводиться до наочного представлення співвідношення окремих елементів, які утворюють досліджувану статистичну сукупність, показу зміни обсягів і структури цих сукупностей.

Об'єктами графічних зображень можуть бути процеси відтворення, які розглядаються у демографічній та економічній статистиці. Особливу роль відіграє графічний метод при вивченні динаміки соціально економічних явищ, де використовують графічні характеристики рядів динаміки; у статистико-географічних дослідженнях, де статистичні дані зображують у вигляді статистичних карт. Побудовою останніх займається прикладна наука – економічна картографія, в якій тісно поєднуються статистичні і географічні аспекти дослідження явищ.

Специфічною особливістю графічних зображень є їх лаконічність, простота кодування інформації та однозначність тлумачення (за змістом) записів у символічній формі . До окремих особливостей статистичних графіків належать також їх виразність, універсальність (для них не існує мовних перешкод), доступність для огляду та ін.

Графічна мова вважається специфічною формою наукового мислення та узагальнення. Це особлива форма інформації, яка трактується в сучасних поняттях теорії пізнання як своєрідна знакова система.

Мова статистичних графіків належить до умовних символічних мов і має такі особливості:

1) двомірність графічних знаків, тобто домірність запису. Це основна ознака графічної мови як знакової системи, джерело інформації та пізнавальної сили. Так, у двомірному символічному записі «працює на інформацію» як послідовність розташування знаків у лінійному ряду, так і їх розташування в просторі. Це, безумовно, розширює інформаційні й пізнавальні можливості графічної мови;

2) безперервність виразу. У статистичних графіках відповідна інформація представлена не окремими дискретними знаками, а взаємопов'язаною системою, геометрично орієнтованою у просторі. Цим графічна мова відрізняється, наприклад, від мови математичних формул, яка зберігає дискретність знаків і лінійну (одномірну) послідовність їх виразу (і чергування);

3) відокремленість викладу. Статистичні графіки як знаряддя наукової інформації відособлюються від тексту взаємопов'язаної за змістом інформації, яка подається в усній або письмовій формі. Тимчасом як мова математики (фізики чи хімії), як правило пов'язана з такими формами подання інформації.

Своєрідність статистичного графіка як знакової системи полягає і в тому, що основним засобом передачі інформації при такому способі зображення є не знаки – коди, а знаки – образи. На відміну від перших, які є найпростішими умовними сигналами, знаки – образи являють собою складніше організовані системи сигналів, які зовнішньо відображають об'єкти за принципом їх схожості.

Предметом дослідження при визначенні статистичного графіка є статистичні дані про масові суспільні явища і процеси. Саме в цьому полягає відмінність статистичних графіків від графіків взагалі. Вони являють собою не просту ілюстрацію явищ, а дають нове знання про предмет дослідження, відображуючи ті розумові побудови, які вивчає статистична наука і практика.

12.2.2. Основні елементи статистичного графіка

Статистичний графік являє собою рисунок, який описує статистичні сукупності умовною мовою геометричних знаків тієї чи іншої форми крапок, ліній, площин, фігур та різних їх комбінацій. У більшості випадків статистичних графіків використовують не об'ємне зображення, яке є складним за побудовою, а площинне. Останнє досить різноманітне за формою і водночас має ті ж самі складові елементи. Розглянемо основні з них .

Поле графіка – це простір, в якому розміщуються геометричні або інші графічні знаки, що утворюють графік. Розмір поля графіка залежить від його призначення і характеризується розміром та пропорціями сторін. З погляду естетичних вимог і зорового сприйняття зображених даних рекомендується таке співвідношення сторін: від 1:1,3 до 1:1,5. Найзручнішим для візуального сприйняття вважається формат, сторони якого знаходяться у співвідношенні 1:2. Таке співвідношення одержують коли довша сторона прямокутника дорівнює діагоналі квадрата, побудованій на короткій стороні прямокутника. Ідеальні графіки прямокутної форми зі співвідношенням сторін 3:5, 5:8, 8:13 і т. д. Такі співвідношення сторін відомі під назвою «правило золотого перетину», згідно з яким висота прямокутника відноситься до його основи як основа до висоти плюс основа. Якщо статистичні графіки представлені у формі рівнобічного трикутника, то його основа повинна відноситися до висоти, як 1:3.

Слід відзначити, що розмір графіка повинен відповідати його призначенню.

Геометричні знаки (або графічні образи) – сукупність геометричних або графічних знаків для зображення статистичних даних. Насамперед це крапки, за допомогою яких наочно зображуються лічильні множини, тобто окремі елементи статистичної сукупності. Одна крапка може означати один випадок або будь –яку їх кількість (наприклад, одне підприємство, 400 кг, 6 км і т .д.). Геометричними знаками статистичних графіків можуть бути відрізки прямих ліній, що поєднують дві сусідні крапки у полі графіка. Змістове наповнення такого знака пов’язується з довжиною відрізка та кутом нахилу щодо осі абсцис. Довжина відрізків характеризує розмір явища, а кут – інтенсивність його розвитку у часі чи просторі. Відрізки з’єднані в один ланцюг, утворюють одна ламану лінію – криву графіка. Остання є досить поширеною формою знакової системи.

Значне місце в цій системі займають знаки у вигляді **площин** різних геометричних форм (квадрат, сектор, коло і т.п.). Їх використовують для порівняння явищ, які характеризуються абсолютними і відносними величинами.

Графічні зображення в статистиці можуть бути представлені і негеометричними знаками, зокрема силуетами чи малюнками. Наприклад, динаміку книжкової продукції на графіку можна зобразити у вигляді книжкових полиць, інфляційні процеси – у вигляді банкнотів тощо.

Просторові орієнтири у статистичних графіках використовують для визначення порядку розміщення геометричних знаків у полі графіка. Вони задаються системою координатних сіток контурних ліній, які ділять це поле на частини. Як правило, в статистиці використовується система прямокутників координат, але іноді може застосовуватися і полярна система (колові графіки).

Масштабні орієнтири визначаються системою масштабних шкал або спеціальними знаками для визначення розмірів графічних знаків.

Експлікація графіка являє собою словесне пояснення основних елементів графіка та його змісту. Вона включає : назву графіка, надписи вздовж масштабних шкал, окремі пояснювальні надписи, що розкривають зміст елементів графічного образу. Статистичний графік – це знакова модель, без експлікації його не можна зрозуміти, тобто перенести знання із формалізованої системи характеристики дійсності на саму дійсність.

12.2.3. Види статистичних графіків і способи їх побудови

Статистичні графіки за напрямом використання характеризуються значною різноманітністю. Їх наукова класифікація передбачає такі ознаки, як загальне призначення, види, форми і типи основних елементів. Традиційно теорія статистики розглядає класифікацію графіків за видами їх поля. За цим принципом графічні зображення поділяють на діаграми, картограми та картодіаграми.

Діаграми - це умовні зображення числових величин та їх співвідношень за допомогою геометричних знаків.

Картограми – зображення числових величин та їх співвідношень за допомогою нанесення умовної штриховки або розцвітки на карту – схему.

Картодіаграми - це поєднання діаграми із картою – схемою. При побудові діаграми встановлюється певний масштаб, тобто співвідношення між розмірами величин на графіку і дійсною величиною зображуваного явища в натурі.

Найбільш поширеним видом статистичних графіків є діаграми. Залежно від способу зображення статистичних даних вони можуть бути в одному вимірі, коли ці дані зображують у вигляді прямих ліній або смуг однакової ширини, і в двох вимірах (площині), на яких даних зображують за допомогою площ геометричних фігур (прямокутників, квадратів, кіл.).

До першого виду діаграм належать лінійні, стовпчикові, стрічкові та ін.; до другого – прямокутні (квадратні, «Знак Варвара»), колові, секторні, радіальні, фігурні.

Лінійна діаграма відображує розмір показника у формі ліній різної довжини, які утворюються в результаті з'єднання крапок у координатному полі. Одним із видів лінійних діаграм є лінійний графік виконання плану та обліково-плановий графік (рис. 27, 28).

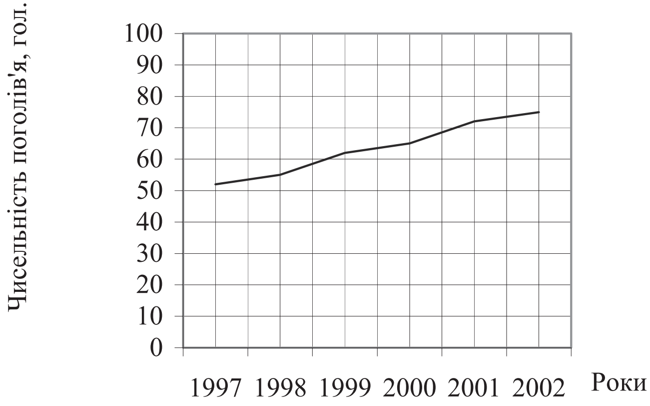


Рис. 27. Лінійний графік динаміки поголів’я коней у господарстві

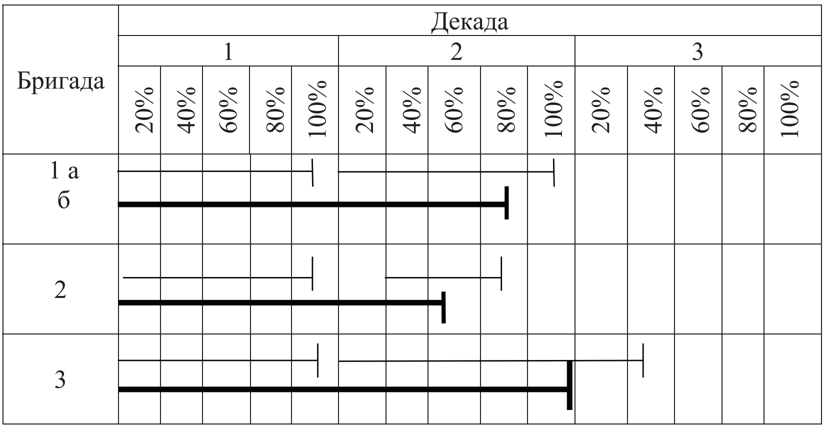


Рис. 28. Обліково-плановий графік виконання підприємством плану виробництвом продукції впродовж місяця: а – за декаду; б – наростаючим підсумком

Застосовують лінійні діаграми в основному для вивчення розвитку явищ у часі.

До будови лінійних діаграм ставлять такі вимоги:

1) діаграма повинна читатися по горизонталі зліва на право, по вертикалі – знизу вверх;

2) на осі ординат обов’язково позначається нульова величина. У випадках, коли дотримання цього правила пов’язане зі значним

зменшенням масштабу та погіршенням наочності, слід зробити розрив по всіх ординатах (при цьому нульова лінія зберігається.);

3) відрізки на осі абсцис повинні відповідати інтервалам (для рядів динаміки – періоду часу);

4) нульова лінія повинна різко відрізнятися від інших паралельних ліній ;

5) при побудові діаграми із застосуванням процентної шкали треба чітко виділити лінію, яка означає 100 %;

6) крива лінія діаграми повинна різко відрізнятися від ліній сітки;

7) цифрові показники розміщують на графіку таким чином, щоб їх можна було легко прочитати;

8) площа графіка повинна бути квадратною або прямокутною.

Стовпчикові діаграми. На цьому виді діаграми статистичні дані зображують у вигляді прямокутників (стовпчиків) однакової ширини. Розташовують їх вертикально чи горизонтально. Величину явищ характеризує висота стовпчика (рис. 29).

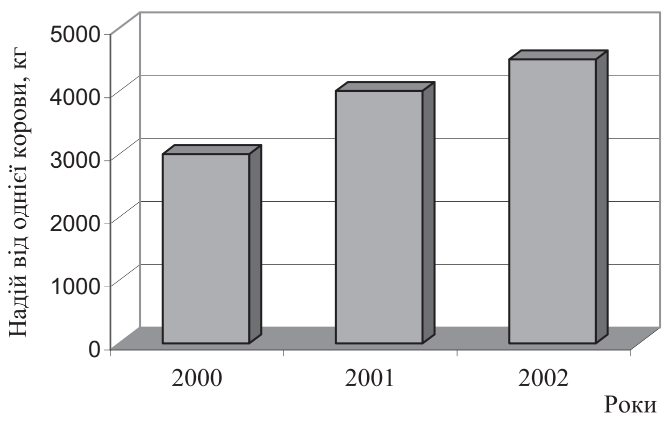


Рис. 29. Стовпчикова діаграма динаміки валового виробництва продукції підприємством

Стовпчикові діаграми застосовуються: 1) при порівнянні між собою різних явищ; 2) для зображення явищ у часі; 3) для відображення структури явищ.

Розглянемо основні правила побудови стовпчикових діаграм:

1) ширина стовпчиків та відстань між ними повинні бути однаковими;

2) стовпчики розташовують від меншого до більшого або навпаки (просторова модель);

3) в основі стовпчиків проводиться та виділяється базова лінія;

4) вказується назва і цифрові дані стовпчиків;

5) на шкалі повинні бути поділки, основні з яких позначаються цифрами;

6) вказують одиницю виміру.

Різновидом стовпчикової діаграми є **гістограма**, за допомогою якої зображуються варіаційні ряди розподілу.

Стрічкові діаграми. На відміну від стовпчикових, при побудові стрічкових діаграм прямокутники, якими зображують розмір явищ, розташовують не по вертикалі, а по горизонталі (рис. 33). Вимоги, що ставляться до побудови цього виду діаграм, аналогічні вимогам до стовпчикових діаграм.

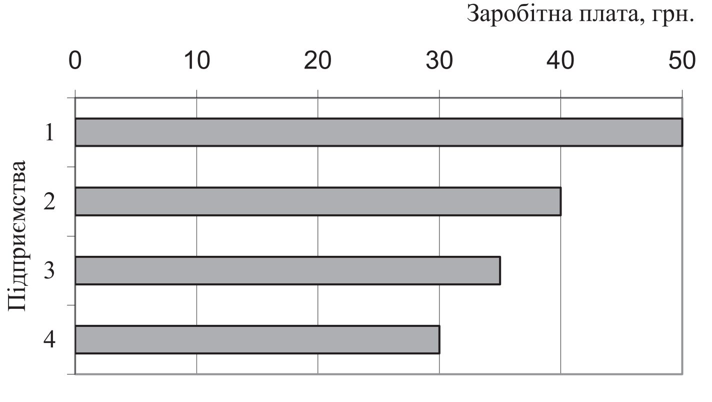


Рис. 30. Стрічкова діаграма денної заробітної плати на підприємствах

Секторні діаграми являють собою коло, поділене на сектори, величини яких відповідають (у пропорціях) зображуваним розмірам явищ. Секторні діаграми будують для відображення структури явищ (рис. 31).

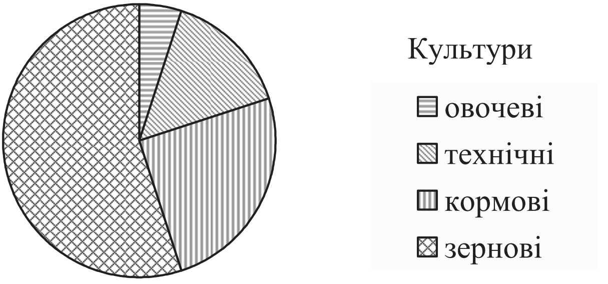


Рис. 31. Секторна діаграма структури посівних площ сільськогосподарського підприємства

Прямокутні діаграми. Цей вид діаграм величину досліджуваних явищ зображує у вигляді площ. Прямокутні діаграми застосовують для зображення явищ, які змінюються у часі, а також для порівняння різних величин у просторі.

До прямокутних діаграм належать квадратні діаграми та «Знак Варвара».

Квадратні діаграми використовують при порівнянні абсолютних величин. Для визначення сторони квадрата слід добути квадратний корінь із досліджуваних (діаграмованих) величин. За даними таблиці 95 проводимо відповідні розрахунки, прийнявши масштаб 30=1 см. Переводимо в масштабні одиниці показники , одержані після добування квадратного кореня із величин площ сільськогосподарських угідь: 81,2 : 30= 2,7 см; 76,8 : 30= 2,6 см; 72,8 : 30=2,4 см одержані числові значення приймаються за величину сторони квадрата (рис.32).

Таблиця 95

Вихідні і розрахункові дані для побудови квадратних та колових діаграм

Номер підприємства	Площа сільсько-господарський угідь, га	Квадратний корінь із розміру площі	Довжина радіуса, см, при масштабі 100=2 см
1	6600	81,2	$1{,}62 = (\frac{81{,}2 \times 2}{100})$
2	5900	76,8	1,54
3	5300	72,8	1,46

«Знак Варзара». Використовується для порівняння трьох пов’язаних між собою величин. Він являє собою прямокутник, в якому довжина відображує величину одного явища, ширина – іншого, а площа його характеризує добуток цих у двомасштабному порівнянні: один масштаб – для основи прямокутника, другий – для його висоти.

«Знаком Варзара» одночасно порівнюється, як уже згадувалося, три пов’язані між собою величини, тобто діаграмовий показник є добутком двох інших. Наприклад, якщо площа прямокутника діаграми ілюструє валовий збір, то одна його довжина – посівну площу, друга – висота - урожайність. Цей вид діаграми зображено на рисунку 33.

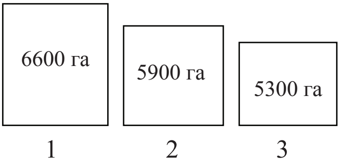


Рис. 32. Квадратна діаграма розмірів площ сільськогосподарських угідь підприємства

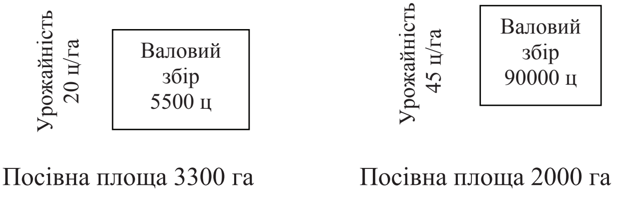


Рис.33. Прямокутна діаграма «Знак Варзара».

Колові діаграми своєю площею відображують величину досліджуваних явищ. Вони ґрунтуються на використанні площі кола для ілюстрації порівнюваних однорідних величин. При їх побудові береться до уваги, що площі кіл відносяться між собою як квадрати їх радіусів. Для визначення радіуса кола необхідно добути квадратний корінь із діаграмової величини; на цій основі накреслити його в певному масштабі й за його величиною описати коло. На рисунку 34 зображено колову діаграму за даними таблиці 95.

Радіальні діаграми. Цей вид діаграм застосовується для графічного зображення явищ, які змінюються в замкнуті календарні строки. В основу їх побудови покладено полярну систему координат, де за вісь абсцис приймається коло, за весь ординат – його радіуси.

Залежно від того, який зображується цикл діаграмованого явища – замкнутий чи продовжуваний (із періода в період), – розрізняють радіальні діаграми *замкнуті і спіральні.* Наприклад, якщо весь цикл зміни зображуваного явища охоплює річний період, радіальну діаграму будують за формою замкнутої.

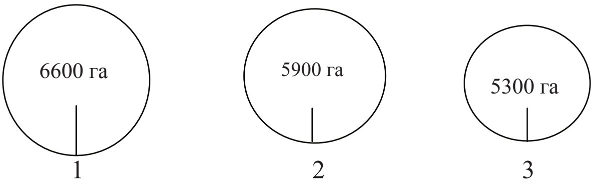


Рис. 34. Колова діаграма розмірів площ сільськогосподарських угідь підприємства

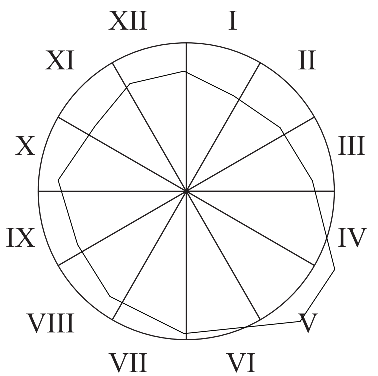
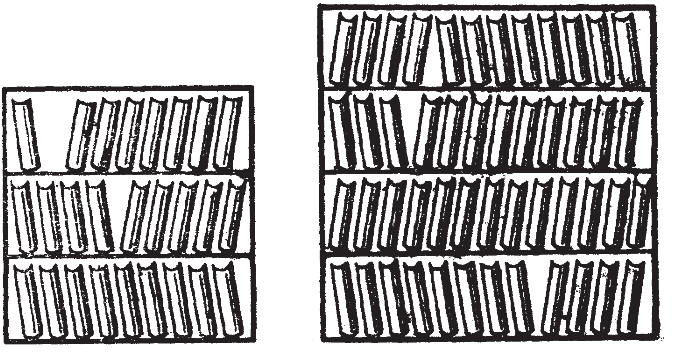


Рис. 35. Радіальна діаграма відпрацьованих людино-годин на підприємстві впродовж року

Якщо ж зміна явища вивчається впродовж циклу діаграмованого періоду (наприклад, грудень одного року сполучається з січнем другого року і т.д.), ряд динаміки зображується у вигляді суцільної кривої, яка візуально має вигляд спіралі.

При побудові радіальних діаграм початком відліку (полюсом) може бути центр або окружність. Якщо за полюс прийнято центр кола, то радіальну діаграму будують у такій послідовності: коло ділять на стільки частин, скільки періодів має діаграмований цикл (наприклад, рік – 12 міс.), і будують відповідно їм радіуси (у даному випадку – 12). Періоди розміщують за годинниковою стрілкою і на кожному радіусі у масштабному вимірі відкладають відрізки (від центра кола), пропорціональні розмірам явищ. Кінці відрізків на радіусах з'єднують, у результаті чого утворюється концентрична ламана лінія. Приклад замкнутої радіальної діаграми з початком відліку від центра кола наведено на рис. 35.

Метод фігур – знаків. Цей метод зображення діаграмованих явищ передбачає заміну геометричних фігур малюнками, які відповідають змісту статистичних даних (рис. 36). Тобто величина показника зображується за допомогою фігур (символів, рисунків): наприклад, поголів'я коней – у вигляді силуета коня, виробництво автомобілів – у вигляді малюнка автомобиля і т.п. Переваги такого виду діаграм перед геометричним – їх наочність та дохідливість. Символічне зображення робить діаграму виразнішою й привабливішою.



2000 р. 2004 р

Рис. 36. Динаміка книжкових видань з питань ринкової економіки в районній бібліотеці

Метод фігур – знаків (так званий віденський) має свої особливості і характеризується більш насиченим змістом, що має принципове значення й вимагає дотримання певних правил щодо побудови таких діаграм, а саме :

1) символи повинні бути зрозумілими самі по собі й не вимагати детальних пояснень. Як правило, вони зображують контур чи силует діаграмованих об'єктів;
2) забезпечувати однозначність трактування;
3) однозначність теми;

4) групувальні ознаки розташовують вертикально, а показники, які їх характеризують, – горизонтально;

5) зображення знаків – символів повинне відповідати принципам гарного малюнку;

6) виключними вважаються зайва деталізація та прикрашання;

7) стандартизація знаків – символів. Компонування діаграми повинне здійснюватися стандартизованими знаками – символами, виготовленими у друкарні і монтованими методом аплікації. Існують спеціальні зразки таких знаків;

8) обов'язковість назви діаграми і текстових позначень окремих сукупностей (груп), які зображується певною фігурою; масштабне позначення з вказівкою числового значення кожного знака – символа.

Напівлогарифмічні графіки. Цей вид статистичного графіка будується в системі координат. Числа, що характеризують діаграмоване явище, знаходяться у масштабі логарифмів. Логарифми точок розташовують на осі ординат, а дату явища (роки) – на осі абсцис (рис. 37).

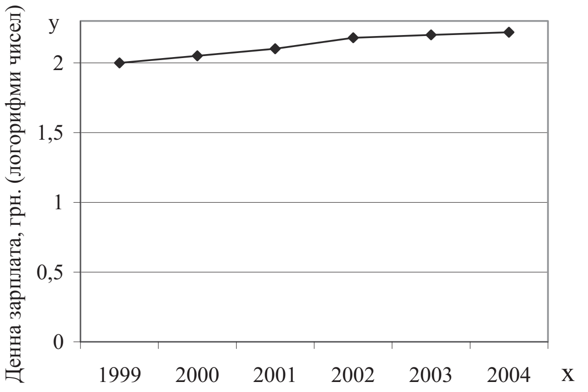


Рис. 37. Напівлогарифмічний графік динаміки показників денної заробітної плати на підприємстві

Картограми і картодіаграми. Картограми являють собою контурну географічну карту або схему, на якій штриховкою різної густоти, крапками або фарбами різного ступеня насиченості зображена порівняльна інтенсивність будь-якого показника в межах кожної одиниці нанесеного на карту територіального поділу. На картограмах, як правило, зображують явища, що характеризуються відносними або середніми величинами (наприклад, кількість працюючих пенсіонерів у загальній кількості працюючих за регіонами, меліорованість земель у процентах до загальної площі, середня заробітна плата на підприємствах по районах області і т.д.).

За способом зображення діаграмованих явищ розрізняють *картограми крапкові і фонові.*

У перших рівень явища показують за допомогою крапок, розташованих на контурній карті територіальної одиниці. Для наочності зображення щільності або частоти появи певної ознаки крапкою позначають одну або кілька одиниць сукупності.

На фонових картограмах штриховкою різної густоти або фарбою різного ступеня насиченості зображують інтенсивність будь-якого показника в межах територіальної одиниці. Один із випадків картограм зображено на рисунку 38.

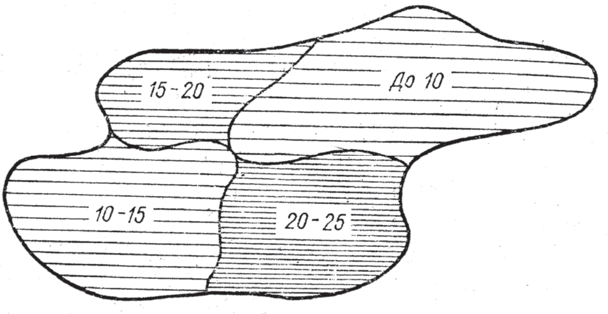


Рис. 38. Картограма щільності поголів'я корів на 100 га сільськогосподарських угідь у господарствах району

Якщо на контурну карту наносяться статистичні дані у вигляді діаграм, одержують **картодіаграму.** Яскравим її прикладом є географічна карта, на якій чисельність населення великих міст зображена у вигляді кіл різної величини.

Крім розглянутих способів графічного зображення досліджуваних явищ, існують і інші. Практичне їх використання при відображенні динаміки явищ, їх структури та взаємозв'язків розглянуто в попередніх розділах.

ПРОГРАМОВАНИЙ КОНТРОЛЬ ЗНАНЬ

МОДУЛЬ I

Тема 1. Методологічні засади статистики

1.1. Яке із зазначених нижче положень виходить за межі визначення терміну «Статистика»?
– Статистика – наука.
– Статистика – параметр.
* Статистика – кількісний бік окремих суспільних явищ.
– Статистика – процес збирання, зберігання і обробки даних про масові суспільні явища.

1.2. Яке з положень належить до визначення статистичної методології?
– Вивчення кількісного боку масових явищ.
– Своєрідний метод пізнання.
* Сукупність статистичних методів дослідження.
– Єдність статистичної теорії і практики.

1.3. Що собою являє статистична наука? Знайти правильну відповідь.
– Своєрідний метод пізнання.
* Статистика – самостійна суспільна наука, яка вивчає кількісний бік масових суспільних явищ у нерозривному зв'язку з їх якісним боком.
– Метод розробки принципів збирання і обробки даних.
– Вивчення взаємозв'язків і закономірностей розвитку явищ.

1.4. Яке з наведених положень належить до визначення загальної теорії статистики?
– Галузь статистики, яка вивчає кількісний бік масових явищ.
* Вивчення загальних правил і методів дослідження масових суспільних явищ.
– Галузь математичних знань.
– Розробляє раціональні прийоми систематизації обробки і аналізу даних статистичних спостережень.

1.5. Що вивчає теорія статистики?
– Кількісний бік масових явищ .і процесів, які відбуваються у народному господарстві.

* Загальні правила і методи статистичного дослідження.
– Взаємозв'язки між окремими одиницями суспільних явищ.
– Кількісний бік масових явищ у сфері виробництва.

1.6. Що вивчає соціальна статистика?
– Тенденції руху показників у сфері соціального життя.
– Стан і розвиток умов виробництва і умов соціального життя.
– Кількісний і якісний бік масових суспільних явищ, процесів.
* Кількісний і якісний бік масових суспільних явищ і процесів, які відбуваються у соціальному житті.

1.7. Що вивчає економічна статистика?
– Реєструє масові суспільні явища.
* Вивчає масові суспільні явища (спираючись на положення теорії статистики) у сфері матеріального виробництва.
– Вивчає загальні правила і методи дослідження масових явищ.
– Вивчає взаємозв'язки між масовими суспільними явищами і процесами.

1.8. Що вивчають галузеві статистики?
– Правила і основні принципи вивчення економіки галузей.
* Загальні положення про статистичні показники процесів виробництва в галузях народного господарства.
– Кількісний і якісний бік масових явищ у сфері виробництва.
– Вивчають показники процесу виробництва в галузях матеріального виробництва, сфері обігу; показники роботи галузей невиробничої сфери і т. ін.

1.9. Знайти правильне визначення статистичної сукупності.
– Ознаки, які відображують розміри чи обсяги явищ і процесів.
– Ознаки, які дають характеристику безпосередньо одиницям спостереження.
– Первинні елементи масових однорідних явищ.
* Маса однорідних елементів (явищ, фактів тощо), які мають єдину якісну основу.

1.10. Що являє собою одиниця сукупності?
– Множина реально існуючих у часі і просторі матеріальних предметів.
* Окремі первинні елементи або індивідуальні явища, які

складають статистичну сукупність.
– Варіюючі ознаки про масові явища і процеси.
– Вторинні ознаки досліджуваних явищ.

1.11. Що є предметом математичної статистики?
– Кількісний і якісний бік масових суспільних явищ і процесів.
– Кількісний бік масових явищ.
– Якісний і кількісний аналіз даних про масові явища.
* Математична теорія математико-статистичних методів не залежно від специфіки і галузі їх застосування.

1.12. Що є предметом статистики як суспільної науки?
– Кількісний аналіз окремих одиниць статистичної сукупності.
– Сукупність прийомів і методів дослідження суспільних явищ.
* Кількісний бік масових суспільних явищ у конкретних умовах простору і часу.
– Вивчення кількісних зв'язків соціально-економічних явищ.

1.13. Складові статистики як суспільної науки.
– Математична статистика, загальна теорія статистики.
– Математична статистика, загальна теорія статистики, економічна статистика.
* Математична статистика, загальна теорія статистики, економічна статистика, галузеві статистики.
– Загальна теорія статистики, економічна статистика, галузеві статистики.

1.14. Що входить в систему наукових статистичних дисциплін?
– Загальна теорія статистики, економічна статистика, галузеві статистики.
* Математична статистика, загальна теорія статистики, економічна статистика, галузеві статистики, галузеві статистики, статистичне моделювання, статистичне прогнозування.
– Економічна статистика, галузева статистика, статистичне моделювання.
– Галузеві статистики, загальна теорія статистики, статистичне моделювання.

1.15. Визначення математичної статистики як наукової дисципліни.
– Статистична методологія і математична теорія.

– Статистична теорія методологія і математична теорія.
* Галузь математичних знань.
– Принципи статистичної науки стосовно різних сторін суспільного життя.

1.16. Дати визначення предмету математичної статистики.
– Загальні властивості кількісних відносин соціально-економічних явищ.
– Показники, які характеризують масові суспільні явища.
– Кількісні характеристики процесів і явищ суспільного життя
* Формальна математична сторона статистичних методів дослідження байдужа до специфічної природи об'єктів, які вивчаються.

1.17. Теоретична база математичної статистики.
– Статистика методологія.
– Статистична теорія і статистика методологія.
* Теорія ймовірності.
– Чисто математична теорія.

1.18. Завдання математичної статистики.
– Вивчення кількісних сторін масових суспільних явищ.
* Встановлення законів розподілу, оцінка невідомих параметрів різних розподілів, перевірка статистичних гіпотез.
– Кількісна оцінка якісної сторони масових суспільних явищ.
– Збір, систематизація, обробка і аналіз даних про явища суспільного життя.

1.19. Визначення категорії «статистика сукупність».
– Сукупність статистичних показників різних за кількісними та якісними ознаками.
* Сукупність однорідних об'єктів чи явищ, об'єднаних за певними ознаками в єдине ціле.
– Статистичні характеристики масових даних, одержані в результаті статистичного спостереження.
– Середні величини, показники варіації, міри асиметрії.

Тема 2. Статистичне спостереження

2.1. Що являє собою статистичне спостереження?
– Збирання та аналіз даних про масові явища.

– Первинна обробка масових даних.
* Планомірний науково організований збір даних про явища і процеси суспільного життя шляхом реєстрації по заздалегідь розробленій програмі спостереження.
– Вивчення кількісних взаємозв'язків явищ по заздалегідь розробленій програмі спостереження.

2.2. Що є об'єктом статистичного спостереження?
* Сукупність суспільних явищ і процесів, які підлягають статистичному спостереженню.
– Сукупність масових суспільних явищ.
– Елементи явищ, які є носіями істотних ознак, що підлягають реєстрації.
– Одиниці суспільних явищ, які підлягають спостереженню.

2.3. Що є одиницею статистичного спостереження ?
* Первинний елемент об'єкта дослідження, який є носієм істотних ознак і властивостей, що підлягають реєстрації.
– Первинний елемент масового суспільного явища.
– Декілька елементів об'єкта статистичного спостереження.
– Елементи явищ суспільного життя.

2.4. До якого виду статистичного спостереження належить звітність сільськогосподарських підприємств перед органами державної статистики?
– Вибіркове.
– Монографічне.
* Суцільне.
– Обстеження основного масиву.

2.5. У якому документі статистичного спостереження формуються мета і завдання спостереження ?
– В інструкції.
– У статистичному формулярі.
* В організаційному плані.
– У програмі спостереження.

2.6. До якого виду статистичного спостереження належить обстеження бюджету сімей ?
– До суцільного.
– До монографічного.
– До анкетного обстеження.
* До вибіркового обстеження.

2.7. Які види спостережень розрізняють залежно від повноти охоплення статистичної сукупності?

* Суцільне і несуцільне спостереження.
– Спостереження основного масиву.
– Вибіркове спостереження.
– Монографічне спостереження.

2.8. Який вид спостереження називають вибірковим?

– Спостереження, при якому обстеженню підлягає більша половин статистичної сукупності, відібраної на основі науково розроблених принципів відбору.
* Спостереження, при якому обстеженню підлягає частина статистичної сукупності, відібраної на основі науково-розроблених принципів відбору.
– Спостереження, при якому обстеженню підлягає менша половина статистичної сукупності, відібраної на основі науково розроблених принципів відбору.
– Вид несуцільного спостереження, при якому обстежується не більше 10% сукупності.

2.9. Який вид спостереження називають обстеження основного масиву?

* Несуцільне обстеження, при якому з усієї сукупності одиниць відбирається така їх частина, в якій обсяг досліджуваної ознаки становить питому вагу більшу за 50% загального обсягу сукупності.
– Несуцільне обстеження, при якому з усієї сукупності відбираються типові одиниці спостереження.
– Вид несуцільного спостереження, яке організовується періодично впродовж господарського року.
– Обстеження, організоване за спеціальною програмою і зумовлене виробничою необхідністю.

2.10. Який вид спостереження називають монографічним?

– Різновидність суцільного спостереження.
– Спостереження, при якому здійснюється контроль інформації, одержаної при анкетуванні.
* Детальне вивчення окремих одиниць статистичної сукупності або їх груп, подібних у певному відношенні.
– Спостереження, при якому обстежуються однорідні об’єкти.

2.11. Яка мета монографічного спостереження.
- – Контроль даних вибіркового обстеження.
- * Виявлення тенденції розвитку прогресивних явищ і поширення передового досвіду.
- – Коригування даних при обстеженні серій однорідних об’єктів.
- – Виявлення помилок при несуцільних обстеженнях.

2.12. За якою ознакою поділяють статистичні спостереження на поточні періодичні й одноразові?
- – За вимогами до організаційних форм спостереження.
- * За часом проведення спостереження.
- – За часом надходження даних статистичної звітності від підприємств.
- – За організацією статистичної звітності.

2.13. Сутність поточного статистичного спостереження.
- * Безперервна реєстрація фактів явищ у міру їх виникнення.
- – Систематична реєстрація фактів і явищ.
- – Епізодична реєстрація фактів і явищ.
- – Одноразова реєстрація фактів і явищ.

2.14. Сутність періодичного спостереження.
- * Спостереження, яке повторюється через певні, заздалегідь установлені проміжки часу.
- – Спостереження, яке здійснюється за програмою монографічного обстеження.
- – Спостереження, яке здійснюється на підставі документів оперативно – технічного обліку.
- – Спостереження, яке проводять з метою вивчення явища у разі потреби.

2.15. Сутність одноразового спостереження.
- – Спостереження, яке здійснюється за вимогою директивних органів.
- – Спостереження за фактами та явищами шляхом їх безперервної реєстрації.
- * Спостереження, яке проводиться з метою вивчення якогось явища на певний момент часу.
- – Спостереження за фактами та явищами за певний період часу.

2.16. До якого виду статистичного спостереження за часом належить реєстрація народжень ?
- – До періодичного.
- * До поточного.
- – До одноразового.
- – До безпосереднього.

2.17. Які помилки визначають розбіжність між спостережуваним показником і дійсним його розміром ?
- – Помилки репрезентативності.
- – Помилки випадкові.
- * Помилки статистичного спостереження.
- – Помилки систематичні.

2.18. Які помилки спостереження називають помилками реєстрації ?
- * Помилки, які виникають внаслідок неправильного встановлення, фактів або неправильного їх запису у формуляр.
- – Помилки, які виникають внаслідок неправильного встановлення фактів.
- – Помилки, які виникають внаслідок невипадкових причин.
- – Помилки, які виникають внаслідок перекручення дійсності.

2.19. Що називають точністю статистичного спостереження ?
- – Арифметичний контроль даних спостереження.
- – Вірогідність одержання об'єктивної інформації за даними спостереження.
- * Ступінь відповідності величини ознаки, встановленої за даними спостереження, дійсної величини.
- – Розбіжність між величиною показника, встановленою за допомогою спостереження і дійсним його розміром.

Тема 3. Зведення і групування статистичних даних

3.1. Що називають статистичним групуванням?
- – Зведення результатів обчислення у статистичних таблицях;
- – Раціональну форму викладення результатів обстеження явищ;
- – Побудову варіаційного ряду;
- * Розподіл статистичної сукупності на частини (групи) за рядом характерних для них ознак.

3.2. Яким видом групувань вирішується завдання вивчення причинно-наслідкових зв’язків між досліджуваними ознаками?
– Комбінаційним;
– Структурним;
* Аналітичним;
– Типологічним.

3.3. Яка статистична таблиця називається комбінаційною?
– Підмет містить одну або більше ознак;
* Підмет містить групи за двома і більше ознаками;
– Підмет містить групи одиниць спостереження;
– Підмет містить групи одиниць спостереження за однією ознакою.

3.4. Які за видом графіки форм розподілу не вивчає математична статистика?
– Одновершинні;
* Багатовершинні;
– Помірноасиметричні;
– Крайньоасиметричні .

3.5. За допомогою якого виду графіків рядів розподілу зображуються інтервальні варіаційні ряди?
– Полігон;
* Гістограма;
– Кумулята;
– Огіва.

3.6. Яка відносна величина характеризує співвідношення між складовими частинами цілого?
* Відносна величина координації;
– Відносна величина структури;
– Відносна величина порівняння;
– Відносна величина інтенсивності.

3.7. Яка з наведених відповідей виходить за межі вимог до статистичних показників?
– Повнота вихідних даних;
– Порівнюваність;
– Вірогідність;
* Ефективність.

3.8. Що є статистичною характеристикою центра розподілу у ряді розподілу?
* Середня арифметична;

– Дисперсія;
– Мода;
– Медіана.

МОДУЛЬ 2

Тема 4. Узагальнюючі статистичні показники

4.1. Знайти правильну відповідь до визначення абсолютних показників.
 * Показники, які відображують розмір кількісних ознак досліджуваних явищ.
 – Показники, які відображують розміри кількісних ознак окремих одиниць сукупності.
 – Показники, які відображують кількісні ознаки певної сукупності.
 – Показники, які відображують кількісні і якісні ознаки досліджуваних явищ.

4.2. Знайти неправильну відповідь на запитання: «В яких вимірниках (одиницях виміру) статистика застосовує абсолютні показники?»
 * У прямих і непрямих.
 – У натуральних і умовно-натуральних.
 – У вартісних і трудових.
 – У комбінованих.

4.3. При обчисленні відносних величин, що виступає базою порівняння у формулі співвідношення абсолютних показників?
 – Чисельник.
 * Знаменник.
 – 100 %.
 – Звітна величина.

4.4. В яких одиницях виражаються відносні показники, коли базова величина приймається за 1000?
 – У процентах.
 – У коефіцієнтах.
 * У проміле.
 – У продециміле.

4.5. Які з перелічених величин характеризують відношення між однойменними показниками?

– Відносні величини інтенсивності.
– Відносні величини координації.
* Відносні величини структури.
– Інтегровані відносні величини.

4.6. Які з перелічених величин характеризують відношення між різнойменними показниками?
– Відносні величини виконання плану.
– Відносні величини структури.
– Відносні величини динаміки.
* Відносні величини інтенсивності і відносні величини координації.

4.7. Яка відносна величина характеризує відношення планового показника до іншої величини, прийнятої за базу порівняння?
– Відносна величина виконання плану.
– Відносна величина порівняння.
– Відносна величина координації.
* Відносна величина виконання планового завдання.

4.8. Яка відносна величина характеризує зміну явищ і процесів у часі?
– Відносна величина структури.
– Відносна величина порівняння.
* Відносна величина динаміки.
– Відносна величина інтенсивності.

4.9. Яка відносна величина характеризує співвідношення між складовими частинами цілого?
* Відносна величина координації.
– Відносна величина структури.
– Відносна величина порівняння.
– Відносна величина інтенсивності.

4.10. Яка величина характеризує склад того чи іншого суспільного явища?
– Відносна величина порівняння.
* Відносна величина структури.
– Відносна величина координації.
– Відносна величина динаміки.

4.11. В якому з наведених прикладів обчислена відносна величина координації?
– Кількість автомобілів на початок року в одному підприємстві по відношенню до іншого підприємства

становить 86 %.
– Щільність поголів'я корів на 100 га сільськогосподарських угідь у становить 27 голів.
* На 100 робітників підприємства припадає 70 жінок.
– Питома вага зернових культур у загальній площі посівустановить 36 %.

4.12. Яка з наведених відповідей виходить за межі вимог до статистичних показників?
– Повнота вихідних даних.
– Порівнюваність.
– Вірогідність.
* Ефективність.

4.13. До якого виду вимірників абсолютних величин належить показник обсягу виробництва валової продукції по підприємству?
– До трудових.
– До натуральних.
– До умовно-натуральних.
* До вартісних.

4.14. До якого виду вимірників абсолютних величин належить показник обсягу витрат кормів у кормових одиницях?
– До вартісних.
– До натуральних.
* До умовно-натуральних.
– До трудових.

4.15. До якого виду відносних величин належить показник виходу телят на 100 корів?
* Інтенсивності.
– Структури.
– Порівняння.
– Координації.

4.16. Яка з середніх належить до степеневої середньої ?
– Геометрична;
– Арифметична;
– Гармонійна;
* Квадратична.

4.17. Що станеться із середньою арифметичною величиною, якщо до кожної варіанти ряду розподілу додати або відняти одну і ту ж величину?

– Не зміниться;
– Збільшиться ;
– Зменшиться;
* Збільшиться або зменшиться на цю ж величину.

4.18. Різні види середніх, розраховані для одного й того ж варіаційного ряду, різняться між собою. У якій відповіді простежується послідовне їх зростання?
* Гармонійна, геометрична, арифметична, квадратична;
– Геометрична, гармонійна, арифметична, квадратична;
– Квадратична, геометрична, гармонійна, арифметична;
– Арифметична, геометрична, гармонійна, квадратична.

4.19. Щоб середня величина була дійсно типовою, яких необхідно дотримуватись вимог при її обчислені?
– Середня якомога менше повинна підлягати дії випадкових коливань;
* Сукупність об’єктів повинна бути якісно однорідною;
– Середня повинна обчислюватись за всім колом явищ;
– Чисельність сукупності повинна бути достатньо великою.

Тема 5. Аналіз рядів розподілу

5.1. Яка з перелічених відповідей виходить за межі видів рядів розподілу?
– Атрибутивні, варіаційні.
– Дискретні.
– Інтервальні.
* Структурні.

5.2. За допомогою якого виду графіків рядів розподілу зображуються дискретні варіаційні ряди?
* Полігон .
– Гістограма.
– Кумулята.
– Огіва.

5.3. Як класифікуються ряди розподілів за формами їх графіків?
– Гістограма.
– Кумулята.
* Одновершинні і багато вершинні.

– Гостровершинні і похиловершинні.

5.4. Які за видом графіків форм розподілу не вивчає математична статистика?

– Одновершинні.
* Багатовершинні.
– Помірноасиметричні.
– Крайньоасиметричні .

5.5. Назвати складові елементи статистичних рядів розподілу.

– Варіанта, частість.
– Частота, частість.
– Частість.
* Варіанта і частота.

5.6. За допомогою якого виду графіків рядів розподілу зображуються інтервальні варіаційні ряди?

– Полігон .
* Гістограма.
– Кумулята.
– Огіва.

5.7. За які межі не повинно виходити число інтервалів при визначенні їх кількості через корень квадратний з обсягу вибірки?

– 3 – 15.
– 5 – 30.
* 5 – 20.
– 10 – 20.

5.8. На яку кількість інтервалів розподіляють статистичну сукупність при невеликій її кількості (до 30 одиниць спостереження?)

* Три.
– Чотири.
– П'ять.
– Сім.

5.9. Який показник характеризує абсолютну міру варіації ознаки в статистичній сукупності?

– Розмах варіації;
* Середнє квадратичне відхилення;
– Середній квадрат відхилення;
– Коефіцієнт варіації.

5.10. Назвати практичну значимість центрального моменту третього порядку
- – Характеризує міру варіації;
- – Характеризує однорідність сукупності;
- * Використовується для характеристики асиметрії розподілу;
- – Використовується для характеристики гостровершинності розподілу.

5.11. Назвати практичну значимість центрального моменту четвертого порядку.
- – Характеризує міру варіації;
- – Характеризує однорідність сукупності;
- – Використовується для характеристики асиметрії розподілу;
- * Використовується для характеристики гостровершинності розподілу.

Тема 6. Аналіз подібності розподілів

6.1. Знайти правильне визначення статистичної оцінки.
- – Узагальнююча характеристика.
- – Будь - який вид середньої величини.
- * Метод суджень про числові значення параметрів розподілу генеральної сукупності по вибіркових даних .
- – Метод суджень про результати одержаних вибіркових характеристик на підставі довірчої ймовірності.

6.2. Як називається в статистиці наближене значення параметра генеральної сукупності, одержане за результатами вибірки?
- – Довірчий інтервал.
- – Одиниця вибірки.
- – Вибіркова характеристика.
- * Статистична оцінка.

6.3. Якими властивостями повинна бути наділена статистична оцінка, щоб вона була максимально наближена до генеральної характеристики?
- * Незміщеність, ефективність, спроможність, достатність .
- – Незміщеність, ефективність, вірогідність.
- – Незміщеність, істотність.
- – Вірогідність.

6.4. Як називається статистична оцінка середньої, якщо вибіркове її значення відповідає генеральному значенню?

– Ефективна.
– Спроможна.
* Незміщена.
– Достатня.

6.5. Яка статистична оцінка зумовлює повноту охоплення всієї вибіркової інформації, тобто є вичерпною?
– Ефективна.
– Спроможна.
– Незміщена.
* Достатня.

6.6. Яка з зазначених нижче оцінок не є точковою оцінкою?
– Середня арифметична.
– Середня квадратичне відхилення.
* Границі інтервалу генеральної середньої.
– Кількість елементів в групі генеральної сукупності.

6.7. Як називається доведена ймовірність того, що помилка вибірки не перевищить деяку задану величину ?
– Поріг імовірності.
* Довірча ймовірність.
– Рівень істотності.
– Рівень вірогідності.

6.8. Як називаються границі, в яких із заданою ймовірністю може знаходитися генеральна характеристика?
– Істотні інтервали.
– Інтервальна різниця.
– Розмах варіації.
* Довірчі інтервали.

6.9. В яких з перелічених випадків використовується наведена формула:

$$P(t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt \text{ ?}$$

* При відсутності стандартних таблиць інтервалу ймовірностей.
– При визначенні граничної помилки.
– При визначенні рівня ймовірності, середня генеральної.
– При визначенні рівня ймовірності, коли невідоме нормоване відхилення

6.10. Як називається довірчий інтервал, коли розраховується лише значення ознаки, які перевищують (або не перевищують) значення шуканого параметра?
 * Односторонній довірчий інтервал.
 – Двосторонній довірчий інтервал.
 – Довірчий інтервал.
 – Інтервальна різниця.

6.11. Які закони розподілу вважаються класичними по відношенню до інших ?
 – Біноміальний, нормальний, Ст'юдента.
 * Біноміальний, нормальний, Пуассоновий.
 – Нормальний, Пірсона.
 – Нормальний, Ст'юдента, Пірсона.

6.12. На якому законі ґрунтується переважна більшість статистичних методів дослідження?
 – Фішера-Спедекора.
 – Ст'юдента.
 – Пірсона.
 * На нормальному.

6.13. Як називається теоретичний розподіл, до якого прямує емпіричний розподіл при $n \to \infty$?
 – Класичний розподіл
 – Умовний розподіл.
 * Закон розподілу.
 – Стандартний розподіл.

6.14. Яким вченим відкритий закон нормального розподілу?
 – Бернулі
 – Фішером.
 – Ст'юдентом.
 * Гауссом.

6.15. Яким вченим зроблено відчутний теоретичний вклад у розробку нормального закону?
 – Пуассоном.
 * Лапласом.
 – Пірсоном.
 – Фішером.

6.16. Якими математичними параметрами визначається нормальний розподіл?
 * $\bar{x}, \sigma$.

– x_i σ^2 .
– σ. t.
– x_it.

6.17. Які статистичні характеристики зумовлюють форму і положення нормальної кривої?
– Середня.
– Середнє квадратичне відхилення.
* Середня і середнє квадратичне відхилення.
– Дисперсія і середнє лінійне відхилення.

6.18. При якому розподілі середня арифметична, мода і медіана будуть рівні між собою?
* При нормальному симетричному.
– При помірно асиметричному.
– При асиметричному.
– При крайньоасиметричному.

6.19. Яке спостерігається співвідношення між середньою арифметичною, модою і медіаною при симетричному нормальному розподілі?
– Середня арифметична більша за моду і медіану.
– Середня арифметична менша за моду, більша за медіану.
* Середня арифметична, мода і медіана рівні між собою.
– Середня арифметична менша за моду і медіану.

6.20. Як називається крива нормального розподілу, коли $\bar{x}=0$ і $\sigma=1$?
– Теоретичною.
– Канонічною.
– Логарифмічною.
* Нормованою.

6.21. При обчислені теоретичних частот, яку кількість нечисленних частот прийнято об'єднувати?
– До 7.
– До 6.
* До 5.
– До 4.

6.22. При якому абсолютному розмірі відношення коефіцієнта асиметрії до своєї середньоквадратичної помилки робиться висновок про невідповідність емпіричного розподілу характеру нормального розподілу?
– >2.
* >3.

– >1.
– >0,5.

6.23. Яким правилом користуються на практиці при дослідженні сукупності на предмет її узгодження з нормальним законом?
– Правилом складання дисперсії.
* Правилом 3 сигм.
– Правилом золотого перетину.
– Правилом розкладання дисперсії.

6.24. В яких сферах людської діяльності зустрічаються розподіли, близькі до нормального, найрідше?
– У техніці.
– У біології.
* В економіці.
– В астрономії.

6.25. Яке з названих нижче положень виходить за межі аспектів застосування нормального розподілу?
– Визначення ймовірності конкретного значення ознаки.
– Оцінка статистичних параметрів.
– При визначенні довірчого інтервалу.
* При визначенні чисельності вибірки.

6.26. Від яких статистичних характеристик залежить імовірність значення t в сукупності з розподілом Ст’юдента ?
– $\bar{x}$, t.
– п, $\bar{x}$.
* n, t.
– v, $\bar{x}$.

6.27. При яких умовах розподіл Ст’юдента наближається до нормального?
– При зменшенні чисельності вибірки.
* При збільшенні чисельності вибірки.
– При збільшенні середнього квадратичного відхилення.
– При п > 15.

6.28. Як називається критерій, розроблений К.Пірсоном для з’ясування відповідності певного закону розподілу вибраного для відображення досліджуваного ряду розподілу ?
– Критерій Фішера.
* Хі-квадрат критерій.
– Критерій Ст’юдента.
– Критерій Бартлета.

6.29. При яких змінах чисельності вибірки розподіл Хі –квадрат переходить у нормальний?

* При збільшенні.

– При зменшенні.

– При збільшенні або зменшенні.

– При п > 15.

6.30. За якою з наведених формул розраховується Хі-квадрат критерій?

* $\chi^2 = \sum \frac{(n_i - n_T)^2}{n_T}$.

– $\chi^2 = \sum \frac{(n_i - n_T)}{n_T}$.

– $\chi^2 = \sum \frac{(n_i - n_T)^2}{n_i}$.

– $\chi^2 = \sum \frac{(n_i - n_T)}{n_i}$.

6.31. Як називають критерій розподілу, для визначення якого знаходиться співвідношення факторної і залишкової дисперсії?

– Пірсона.

– Ст'юдента.

* Фішера.

– Лапласа.

6.32. Як називають кількість одиниць спостереження, здатних приймати будь-які (вільні) значення, що не змінюють середньої величини, тобто, загальної їх характеристики?

– Вибіркова сукупність.

– Мала вибірка.

* Число ступенів вільності.

– Велика вибірка.

6.33. Чи вирішує мала вибірка типові завдання: оцінка середньої ; визначення довірчих інтервалів генеральної середньої; оцінка різниць двох вибіркових середніх; оцінка середньої різниці?

* Вирішує.

– Не вирішує.

6.34. Для якої за обсягом вибірки розподіл Ст'юдента вважається точним?

– Для малої.

– Для великої.

* Для будь-якої.

– Для вибірки з п >20.

6.35. В яких вибірках підлягає статистичній оцінці різниця середніх?
* У малих незалежних.
– У малих.
– У великих.
– У залежних.

МОДУЛЬ 3

Тема 7. Статистичні методи вимірювання взаємозв'язків

7.1. У чому полягає головне завдання дисперсійного аналізу ?
– Статистичне вивчення варіації середньої величини ознаки.
– Визначення вибіркової дисперсії.
* Статистичне виявлення впливу факторів, які зумовлюють мінливість ознаки.
– Обчислення факторної дисперсії.

7.2. Які відповіді виходять за межі завдань дисперсійного методу, коли він виконує не допоміжні, а самостійні функції в аналізі економічних явищ?
– Кількісне вимірювання сили впливу факторних ознак та їх сполучень на результативну ознаку.
– Визначення вірогідності впливу та його двірчих границь.
– Аналіз окремих середніх та статистична оцінка їх різниці.
* Визначення вірогідності впливу факторних ознак, результати групувань.

7.3. Яка з наведених формул визначає ступінь впливу факторних ознак на результативну ?
* $\eta^2 = C_x : C_y$.
– $\eta^2 = C_y : C_x$.
– $\eta^2 = C_Z : C_y$.
– $\eta^2 = C_Z : C_x$.

7.4. Що означає наведене рівняння : $C_x = C_A + C_B + C_C + C_{AB} + ... + C_{ABC}$?
– Дисперсію дії факторів.
– Дисперсію дії поєднання факторів.
* Дисперсію дії факторів та їх поєднань.
– Дисперсію результативної ознаки.

7.5. Які відповіді виходять за межі етапів дисперсійного аналізу?
– Обробка дисперсійного комплексу для одержання

загальної, фактичної та залишкової дисперсій.
– Знаходження питомої ваги факторної і залишкової дисперсій у загальній дисперсії та їх коригування на число ступенів вільності.
– Оцінка вірогідності впливу факторів та їх поєднань
* Оцінка вірогідності залишкової дисперсії.

7.6. Яка з наведених формул визначає величину загальної дисперсії?
_ $C_x = C_y - C_Z$.
* $C_y = C_x + C_Z$.
_ $C_Z = C_y - C_x$.
_ $\eta_x^2 = C_x : C_y$.

7.7. Яка з наведених формул виходить за межі обчислення дисперсій в одно факторному статистичному комплексі?
_ $C_y = \sum V^2 - \frac{\left(\sum V\right)^2}{n}$.
_ $C_x = \sum h - \frac{\left(\sum V\right)^2}{n}$.
_ $C_Z = \sum V^2 - \sum h$.
* $F = C_x : C_Z$.

7.8. Яку статистичну характеристику визначає формула $K = \frac{C_x}{C'_x}$?
– Девіата .
– Коефіцієнт окремого визначення.
* Поправочний коефіцієнт для розкладу сумарної дисперсії.
– Кінцева дисперсійна структура.

7.9. Яка з наведених формул визначає ступінь впливу неврахованих факторів на результативну ознаку ?
_ $\eta^2 = C_x : C_y$.
* $\eta^2 = C_Z : C_y$.
_ $\eta^2 = C_y : C_Z$.
_ $\eta^2 = C_y : C_x$.

7.10. Що означає вираз $\eta_{AB}{}^2 = C_{AB} : C_y = 16{,}6\%$?
– Ступінь впливу факторної ознаки на результативну.
– Ступінь впливу факторів та їх поєднань на результативну ознаку.
– Кількісну залежність результативної ознаки від двох

факторів.
- * Ступінь впливу поєднань двох факторів на результативну ознаку.

7.11. За яким критерієм визначається вірогідність дії досліджуваних факторів у дисперсійному комплексі ?
- * Критерій Фішера.
- – Критерій Ст’юдента.
- – Критерій Пірсона .
- – Критерій Лапласа.

7.12. Як називають стандартні відношення девіат, які знаходять за математичними таблицями ?
- – Хі – квадрат критерій.
- * F – критерій.
- – t – критерій.
- – λ – критерій.

7.13. Як у дисперсійному аналізі називається математична модель, в якій досліджується дія трьох факторів?
- – Факторний дисперсійний комплекс.
- – Багатофакторний дисперсійний комплекс.
- * Трифакторний дисперсійний комплекс.
- – Дисперсійна модель.

7.14. Чому дорівнює сума окремих ступенів вільності у двофакторному дисперсійному комплексі?
- – Числу одиниць спостережень.
- – Числу одиниць спостережень, зменшеному на одиницю.
- * Числу ступенів вільності, зменшеному на дві одиниці.
- – Числу ступенів вільності для загальної дисперсії.

7.15. В яких випадках вважається вірогідним досліджуваний вплив факторів на результативну ознаку?
- – FT>Fp.
- * FT<Fp.
- – FT=Fp.
- – FT≠Fp.

7.16. Які відповіді виходять за межі характеристики дисперсійного методу з боку виконуваних ним допоміжних функцій ?
- – Оцінка результатів групувань.
- – Оцінка істотності коефіцієнт кореляції та різниці середніх.
- * Оцінка характеру розподілу вибірки.
- – Оцінка лінійної множинної регресії.

7.17. Який зв'язок називається кореляційним?
- – Повний зв'язок між ознаками.
- – Повний зв'язок між двома і більше ознаками.
- * Неповний зв'язок між ознаками, який проявляється при спостереженні масових даних.
- – Неповний зв'язок між ознаками, встановлений на підставі одиничного спостереження.

7.18. Яка з відповідей виходить за межі правильного визначення поняття «кореляція»?
- – Зміна середньої величини однієї ознаки залежно від значення іншої.
- – Залежність між випадковими величинами, яка не має функціонального характеру.
- – Неповна залежність між ознаками.
- * Визначення форми зв'язку.

7.19. Що являє собою поняття «регресія»?
- – Тіснота зв'язку.
- – Математичне очікування змінної величини, зумовлене зміною випадкової.
- * Лінія, вид залежності середньої величини результативної ознаки від факторної.
- – Вид пропорціональної залежності двох змінних.

7.20. Пояснити поняття «стохастичний зв'язок».
- – Вид кореляційного зв'язку
- – Форма кореляційного зв'язку.
- – Тип зв'язку між випадковими величинами.
- * Зв'язок між випадковими величинами, при якому зміна однієї з них зумовлює зміну закону розподілу інших.

7.21. Дати визначення поняттю «форма кореляційного зв'язку».
- * Тип аналітичної формули; яка відображує залежність між досліджуваними ознаками.
- – Аналітичне рівняння зв'язку.
- – Кутовий коефіцієнт у прямолінійному рівнянні зв'язку.
- – Вид дослідження взаємозалежностей між ознаками.

7.22. Яка з відповідей виходить за межі визначення завдань кореляційного аналізу?
- * Визначення ступеня відокремленого спільного впливу факторів на результативну ознаку.
- – Оцінка параметрів нормально розподіленої генеральної

сукупності (середніх, дисперсій, коефіцієнтів кореляції).
- Перевірка істотності оцінюваних параметрів.
- Виявлення структури взаємозалежності ознак.

7.23. Дати визначення поняттю «мультиколінеарність».
- Кореляційна залежність між досліджуваними ознаками.
- Теоретично не доведений кореляційний зв'язок між факторами.
* Кореляція між факторами.
- Комбінація кількох рядів розподілу з метою побудови кореляційної моделі.

7.24. Що означає поняття «лінеаризація»?
- Аналітичне вирівнювання досліджуваних зв'язків за математичними формулами.
- Виявлення кореляційних зв'язків шляхом виключення факторів, лінійно пов'язаних між собою.
- Перехід від лінійного зв'язку до нелінійного.
* Перехід від нелінійного зв'язку до лінійного.

7.25. Яка з відповідей виходить за межі визначення переваг кореляційно -регресійного методу перед методом статистичних групувань?
- Елімінування випадкових коливань досліджуваних ознак.
- Можливість одночасного вивчення зв'язків між кількома ознаками.
- Одержання показників тісноти зв'язку та оцінка параметрів генеральної сукупності за даними вибірки.
* Раціональне й наочне викладення цифрових характеристик досліджуваних явищ.

7.26. Як називається кореляція, коли ознака розглядається як результат дії двох і більше факторів?
- Прямолінійною.
- Криволінійною.
- Простою.
* Множинною.

7.27. Як називається кореляційний зв'язок, при якому значення результативної ознаки змінюється в протилежному напрямі щодо факторної?
- Криволінійний.
* Обернений.
- Прямий.

– Прямолінійний.

7.28. Які з перелічених етапів роботи не стосуються кореляційного аналізу?

– Математично-економічне моделювання.
– Знаходження параметрів кореляційного рівняння.
* Визначення кореляції атрибутивних ознак.
– Оцінка й аналіз одержаних результатів.

7.29. Який можна зробити висновок про характер кореляційного зв'язку, якщо величина одержаного коефіцієнта кореляції становить -0,816?

– Зв'язок прямий.
* Зв'язок обернений.
– Зв'язок криволінійний.
– Зв'язок прямолінійний.

7.30. Який вид залежності характеризує взаємозв'язок наведених нижче параметрів $\overline{yx} = \overline{y} \cdot \overline{x} = \sigma_y \sigma_x$?

– Прямий зв'язок між ознаками.
– Обернений зв'язок між ознаками.
– Відсутність лінійного зв'язку.
* Повну лінійну функціональну залежність.

7.31. Яка схема рішення вважається більш прийнятною, якщо обчислювальні операції кореляційного аналізу виконуються за допомогою засобів малої механізації?

– Алгоритми Таусса.
– Метод послідовних наближень Зейделя.
– Правило Сарруса.
* Схема Чебишева поліномах Дулітля.

7.32. Дати визначення показника коефіцієнта кореляції.

– Вимірник тісноти зв'язку при простій кореляційній залежності.
– Параметр рівняння регресії.
– Вимірник тісноти кореляційного зв'язку.
* Вимірник тісноти зв'язку при простій прямолінійній залежності.

7.33. Яка статистична характеристика визначається за формулою $r = (\overline{yx} - \overline{y} \cdot \overline{x}) : \sigma_y \sigma_x$?

– Індекс кореляції.
* Коефіцієнт кореляції.

– Частковий коефіцієнт кореляції.
– Кореляційне відношення.

7.34. Розраховано коефіцієнт регресії врожайності вівса (ц/ га) і собівартості його виробництва (грн.). В яких одиницях виміру інтерпретується цей коефіцієнт?
– Гривні з розрахунку на 1гектар.
– Центнери з розрахунку на 1 гектар.
* Гривні з розрахунку на 1 центнер.
– Центнери з розрахунку на 1 гривню.

7.35. В яких випадках можна одержати коефіцієнт кореляції з від'ємним знаком?
– При множинному лінійному кореляційному зв'язку.
– При множинному криволінійному кореляційному зв'язку.
* При парному лінійному зв'язку.
– При парному криволінійному зв'язку.

7.36. Яка з наведених формул використовується для визначення коефіцієнта множинної кореляції?
* $\sqrt{\sigma_x^2 : \sigma_x^2}$.
– $\sqrt{\sigma_{yx}^2 : \sigma_y^2}$.
– $\sqrt{1 - \frac{\sigma_x^2}{\sigma_y^2}}$.
– $\sigma_{\overline{yx}} : \sigma_y$

7.37. Що характеризує числове значення коефіцієнта множинної детермінації 0,858?
* Варіація результативної ознаки на 85,8 % зумовлена варіацією досліджуваних факторів.
– На невраховані фактори припадає 85,8 % питомої ваги впливу на зміну результативної ознаки.
– Варіація результативної ознаки на 85,8 % зумовлена варіацією всіх можливих факторів.
– Не інтерпретується.

7.38. Що означає наведена формула: $t = \left(r\sqrt{n-2}\right) : \left(\sqrt{1-r^2}\right)$?
– Істотність параметрів рівняння регресії.
– Довірчий інтервал коефіцієнта кореляції.
– Нормоване відхилення.
* Істотність коефіцієнта кореляції.

7.39. За якою формулою перевіряється істотність коефіцієнта множинної кореляції?

– $\left(r\sqrt{n-2}\right):\left(\sqrt{1-r^2}\right)$.

* $R : S_R$.

– $\left(1-R^2\right):\left(\sqrt{n-1}\right)$.

– $\left(1-r^2\right)\left(n-p\right)$.

7.40. В яких випадках вважається істотним коефіцієнт кореляції?

– $t_T \rangle t_p$.

– $t_p \rangle t_T$.

– $t_T = t_p$.

* $t_p \geq t_T$

7.41. Для параметрів рівняння регресії необхідно встановити довірчі границі випадкових коливань. Яка з наведених формул буде використана з цією метою?

– $R : S_R$.

* $\left(\sum y^2 - a_0\sum y - a_1\sum xy\right):\left(n-2\right)$

– $\sigma_{\overline{yx}} : \sigma_y$.

– $\left(1-R^2\right):\left(\sqrt{n-1}\right)$.

7.42. Чому у формулі 2 (п. 1.25) за знаменник (дільник) взято число ступенів вільності?

– Щоб одержати більш точну статистичну характеристику.

– Щоб одержати вірогідну оцінку показників генеральної сукупності.

* Щоб одержати незмінену оцінку показників генеральної сукупності.

– Щоб одержати ефективну оцінку показників генеральної сукупності.

7.43. Обчисленням стандартної помилки рівняння регресії встановлено, що довірчий інтервал коливається у значних межах. Чим пояснюється така неточність передбачень теоретичного рівня результативної ознаки?

* Досліджувана залежність є кореляційною, а не функціональною.

– Залежність ознак досліджувалася з низьким рівнем імовірності.

– Одержані результати кореляційного аналізу є вірогідними, а не точними.

– Стандартна помилка рівняння регресії виявилася невірогідною.

7.44. Що характеризує частковий коефіцієнт кореляції у випадку дослідження впливу двох факторних ознак?

– Тісноту лінійного зв'язку між результативною і даною факторною ознакою.
– Тісноту лінійного зв'язку між двома факторними ознаками.
– Тісноту зв'язку між результативною і факторною ознакою.
* Тісноту лінійного зв'язку результативної ознаки з однією із факторних при виключенні дії іншої факторної ознаки.

7.45. Яка з перелічених відповідей виходить за межі вимог до побудови кореляційно - регресійних моделей аграрно-економічних явищ?

– Фактори-аргументи повинні відображувати об'єктивні особливості сільськогосподарських підприємств.
– Залежна і жодна з незалежних змінних не повинні перебувати у функціональній залежності від іншої або їх групи.
– Кількість включених у модель факторів повинна бути не дуже великою.
* Величина коефіцієнта кореляції між включеними у модель факторами не повинна перевищувати 0,9.

7.46. Які з перелічених статистичних характеристик не стосуються непараметричних критеріїв кореляційних зв'язків?

* Коефіцієнт кореляції.
– Коефіцієнт кореляції рангів.
– Коефіцієнт асоціації.
– Критерій знаків.

7.47. Який непараметричний критерій розраховують за формулою $p = 1\frac{\sigma\sum d^2}{n\left(n^2-1\right)}$?

– Коефіцієнт асоціації.
– Коефіцієнт Фехнера.
* Ранговий коефіцієнт кореляції Спірмена.
– Коефіцієнт контингенції.

7.48. Між якими непараметричними критеріями існує залежність, яку визначає відношення $p = r_{yx} : 2\sin\frac{\pi}{\sigma}$?

– Коефіцієнт кореляції та коефіцієнт асоціації.

– Коефіцієнт кореляції рангів і коефіцієнт контингенції.
– Коефіцієнт Фехнера та коефіцієнт Спірмена.
* Лінійний коефіцієнт кореляції і коефіцієнт кореляції рангів.

7.49. Який непараметричний критерій розраховують за формулою $K_{Ф} = (n_a - n_b):(n_a + n_b)$?
* Критерій знаків (коефіцієнт Фехнера).
– Коефіцієнт асоціації.
– Коефіцієнт контингенції.
– Рантовий коефіцієнт кореляції Спірмена.

7.50. Що розуміють у кореляції рядів динаміки під поняттям «тренд»?
* Зміна, яка визначає загальний напрям розвитку, основну тенденцію ряду динаміки.
– Непараметричний критерій.
– Наявність автокореляції у рядах динаміки.
– Специфічна структура випадкової компоненти у ряді динаміки.

МОДУЛЬ 4

Тема 8. Аналіз інтенсивності динаміки

8.1. Яка з відповідей виходить за межі вимог до побудови рядів динаміки?
– Вірогідність, точність, наукова обґрунтованість.
– Порівнянність за змістом.
– Порівнянність за територією.
* Порівнянність моментних і періодичних рядів.

8.2. Яка з відповідей виходить за межі дискретних рядів динаміки?
– Моментні ряди.
* Інтервальні ряди.
– Неперервні ряди.
– Ряди середніх.

8.3. До яких рядів динаміки належать показники, одержані через певні проміжки часу?
– Моментні.
– Інтервальні.
* Дискретні.

– Неперервні.

8.4. До яких рядів динаміки належать показники, що характеризують розміри явищ за певні проміжки часу?

– Дискретні.
– Моментні.
* Інтервальні.
– Ряди середніх.

8.5. До якого виду динаміки належать показники поголів'я худоби на початок кожного місяця року?

* Моментні.
– Інтервальні.
– Ряди середніх.
– Неперервні.

8.6. З яким видом середньої розраховують середньорічну кількість худоби, якщо відома її чисельність на початок кожного місяця року?

– Арифметична.
* Хронологічна.
– Гармонійна.
– Геометрична.

8.7. За яким видом середньої визначають середньорічний рівень виробництва продукції, якщо відомі щорічні обсяги її виробництва за 6 років?

* Арифметична.
– Хронологічна.
– Гармонійна.
– Геометрична.

8.8. За допомогою яких статистичних характеристик визначають варіацію рядів динаміки навколо середньої?

– Розмах варіації.
– Середнє лінійне відхилення.
* Середнє квадратичне відхилення і коефіцієнт варіації.
– Дисперсія та коефіцієнт осциляції.

8.9. Який аналітичний показник ряду динаміки характеризує абсолютну величину розміру змін явища?

– Коефіцієнт зростання.
– Темп приросту.
– Абсолютне значення 1% приросту.
* Абсолютний приріст.

Тема 9. Аналіз тенденцій розвитку та коливань

9.1. За яким видом середніх розраховують середній коефіцієнт зростання?
   – Арифметична.
   * Геометрична.
   – Квадратична.
   – Хронологічна.

9.2. Яка кількісна статистична характеристика ряду динаміки визначає тенденцію розвитку явища?
   – Автоковаріація.
   – Автокореляція.
   * Тренд.
   – Регресія.

9.3. Який вид тенденції розвитку явища характеризує тенденцію змін зв'язку між окремими рівнями ряду?
   – Тенденція середнього рівня.
   – Тенденція дисперсії.
   * Тенденція автокореляції.
   – Тенденція в русі показників приросту.

9.4. У чому полягає суть завдання щодо використання прийомів обробки рядів динаміки з метою виявлення головної тенденції розвитку явища?
   – Встановлення характеру дії основних причин, що визначають динаміку явища.
   – Елімінування дії випадкових, другорядних причин, що визначають динаміку явища.
   * Елімінування дії випадкових причин та встановлення характеру дії основних причин, що визначають динаміку явища.
   – Побудова математичних функцій динаміки.

9.5. В яких випадках використовують прийом змикання рядів динаміки?
   * При непорівнянності рівнів рядів динаміки.
   – При виявленні закономірності розвитку явища.
   – При виявленні характеру головної тенденції динаміки.
   – При виявлені типу загальної тенденції динаміки.

9.6. Що розуміють під загальною тенденцією динаміки?
   – Тенденція в русі показників динаміки.

– Тенденція до зростання рівня явища.
– Тенденція до зростання або зниження рівнів ряду.
* Тенденція до зростання, стабільності або зниження рівня даного явища.

9.7. Яка з відповідей виходить за межі типів динаміки?
– Абсолютні прирости зростають.
– Абсолютні прирости стабільні.
– Темни зростання стабільні; темпи зростання збільшуються.
* Темпи зростання зменшуються.

9.8. Які з прийомів виявлення загальної тенденції розвитку і характеру динаміки слід використовувати, коли рівні ряду динамки значно варіюють?
* Згладжування шляхом укрупнення інтервалів, згладжування за допомогою ковзної середньої.
– Побудова графіків рядів динаміки.
– Змикання рядів динаміки.
– Визначення автокореляції у рядах динаміки.

9.9. З метою встановлення тенденції розвитку явища дослідником виділено певний етап його розвитку й обрано тип аналітичної функції $\overline{y}_t = a_0 + a_1 t + a_2 t^2$. Який спосіб обробки рядів динаміки використано в даному разі?
– Вирівнювання шляхом укрупнення інтервалів.
– Вирівнювання способом ковзної середньої.
* Аналітичне вирівнювання.
– Побудова математичних функцій динаміки.

9.10. Який тип аналітичної функції використовують для вирівнювання ряду динаміки у випадках, коли абсолютні прирости рівномірно збільшуються?
– Рівняння прямої.
* Рівняння параболи.
– Рівняння показової функції.
– Ряд Фур'є.

9.11. Яку з наведених математичних функцій використовують для вирівнювання рядку динамки, якщо коефіцієнти зростання (ланцюгові) стабільні?
– $\overline{y}_t = a_0 + a_1 t$.
– $\overline{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$.
* $\overline{y}_t = a_0 a_1 t$.

– $\bar{y}_t = a_0 + \sum_{\kappa-1}^{m}(a_\kappa \cos \kappa t + b_\kappa \sin \kappa t^2$.

9.12. Яка з відповідей виходить за межі способів визначення сезонних коливань у рядах динаміки?
- – Розраховується середня арифметична ряду, з якою порівнюється щомісячні рівні.
- – Розраховуються індекси сезонності за способом,
- – Розраховуються індекси сезонності за способом, вказаним у п. 1; потім знаходиться середня арифметична з індексів.
- – Розраховується відношення фактичних щомісячних рівнів до ковзної середньої, розрахованої за 12 міс. На підставі цих співвідношень за ряд років розраховується середня арифметична для кожного місяця.
- * Розраховується середня арифметична ряду. З нею порівнюються середні ковзні тримісячні рівні.

9.13. Що розуміють у кореляції рядів динаміки під поняттям «тренд»?
- * Зміна, яка визначає загальний напрям розвитку, основну тенденцію рядів динаміки.
- – Непараметричний критерій.
- – Наявність автокореляції у рядах динаміки.
- – Специфічна структура випадкової компоненти у ряді динаміки.

9.14. Назвати критерій, який використовується для виявлення наявності автокореляції у відхиленнях від тренда.
- – Критерій Пірсона.
- – Критерій Стьюдента.
- – Критерій Дурбіна.
- * Критерій Дурбіна-Уотсона.

9.15. Що означає термін «мультиколінеарність» в рядах динаміки?
- – Наявність функціонального зв’язку.
- – Ступінь тісноти.
- – Криволінійний зв’язок.
- * Наявність сильної кореляції між незалежними змінними.

Тема 10. Індексний метод

10.1. Як називають в індексному аналізі об'єднання різнорідних елементів в одну сукупність?
- – Індексний комплекс.

– Модель індексного аналізу.
* Агрегат.
– Агрегатний індексний комплекс.

10.2. Яка з відповідей дає визначення статистичного індексу?
– Показник.
– Відносна величина.
– Комплексний показник.
* Відносна величина, одержана в результаті порівняння складних економічних явищ, що не підлягають безпосередньому підсумовуванню.

10.3. Які індекси відображують співвідношення простих одиничних показників?
– Тотальні.
– Субіндекси.
* Індивідуальні.
– Загальні.

10.4. Як називається в теорії індексів показник, зміну якого характеризує індекс?
– Сумірник.
* Індексована величина.
– Елімінована величина.
– Середня величина.

10.5. Як називається в індексному комплексі постійна величина, пов'язана з індексованою?
* Сумірник (вага).
– Порівнювана величина.
– Константа.
– Середня величина.

10.6. Як класифікуються індекси за способом побудови?
– Агрегатні, тотальні, середні.
* Агрегатні, середні із індивідуальних, середнього рівня.
– Агрегатні, групові, індивідуальні.
– Агрегатні, середнього рівня, індивідуальні.

10.7. Як класифікуються індекси за ступенем охоплення елементів явищ?
* Індивідуальні, загальні.
– Індивідуальні, агрегатні.
– Загальні, тотальні.
– Групові, індивідуальні.

10.8. Якими способами можна побудувати індекс фізичного обсягу?
- * Як агрегатний і як середній із індивідуальних.
- – Як загальний і як індивідуальний.
- – Як тотальний.
- – Як груповий.

10.9. Як називається індекс, одержаний за рівнянням: $\sum q_1 p : \sum q_0 p$?
- – Агрегатний індекс вартості.
- * Агрегатний індекс фізичного обсягу.
- – Індекс вартості.
- – Індекс цін.

10.10. За якою формою середньої розраховують середні індекси?
- * Арифметичної, гармонійної.
- – Арифметичної.
- – Гармонійної.
- – Структурної.

10.11. Яку форму індексу використовують в аналізі, якщо вихідні дані несуть інформацію про вартість продукції звітного періоду в базисних цінах?
- – Середню арифметичну.
- * Середню гармонійну.
- – Середню арифметичну або середню гармонійну.
- – Будь-яку середню форму.

10.12. Яка форма індексу буде використана в розрахунках, якщо в розпорядженні дослідника є дані: а) індивідуальні індекси обсягу; б) вартість продукції у базисному році? Треба визначити індекс фізичного обсягу.
- – Середній гармонійний.
- – Агрегатний.
- – Середній з індивідуальних.
- * Середній арифметичний.

10.13. Як класифікують індекси залежно від періоду часу, взятого за основу порівняння?
- – Періодичні.
- – Базисні.
- – Ланцюгові.
- * Базисні та ланцюгові.

10.14. Який взаємозв'язок існує між базисними і ланцюговими індексами?
- – Прямий.

– Обернений.
– Добуток базисних індексів дорівнює ланцюговому останнього періоду.
* Добуток ланцюгових індексів дорівнює базисному останнього періоду.

10.15. Який термін використовують при інтерпретації індексів, якщо за базу порівняння при обчисленні береться 100 %?
– Процент.
* Пункт.
– Проміле.
– Продециміле.

10.16. Знайдіть правильну відповідь за такими результатами розрахунків: у 2004 р. індекс цін щодо 2003 р. підвищився від 115 до 120%. За базу порівняння (100 %) взяте 2003р.
– Індекс збільшився на 5 %.
– Індекс збільшився на 5 одиниць свого виміру.
– Індекс збільшився на 0,05.
* Індекс збільшився на 5 пунктів.

10.17. Як називається індекс, представлений відносною величною, що характеризує динаміку двох середніх показників?
* Індекс змінного складу.
– Індекс фіксованого складу.
– Індекс з постійною вагою.
– Індекс із змінною вагою.

10.18. Яку статистичну характеристику одержують відношенням індексу змінного складу до індексу фіксованого складу?
– Індекс середнього рівня.
* Індекс структури.
– Індекс з постійною вагою.
– Середній індекс.

10.19. Яку статистичну характеристику одержують добутком індексу структури та індексу фіксованого складу?
* Індекс змінного складу.
– Індекс із змінними вагами.
– Індекс із постійними вагами.
– Середній індекс.

10.20. Яка відповідь відображує основні види економічних індексів?
– Індекси середнього рівня.
* Індекси продуктивності праці, індекси фізичного обсягу,

індекси цін, індекси собівартості.
- Індекси структури.
- Індекси товарообороту.

10.21. Як називаються індекси, що характеризують співвідношення рівнів явищ у просторі?
- Загальні.
- Тотальні.
* Територіальні.
- Субіндекси.

МОДУЛЬ 5

Тема 11. Вибірковий метод

11.1. Як називається вид статистичного спостереження, при якому обстеженню підлягає лише частина одиниць сукупності, відібраних на основі науково розроблених принципів?
* Вибіркове.
- Суцільне.
- Обстеження основного масиву.
- Анкетування.

11.2. Який використовують спосіб відбору у вибіркову сукупність, якщо відбір одиниць з генеральної сукупності здійснюють через рівні проміжки?
- Типовий.
- Власне випадковий.
* Механічний.
- Серійний.

11.3. Який вид статистичного спостереження застосовують для одержання характеристик генеральної сукупності при умові, що затрати праці і засобів на збір інформації повинні бути мінімальними?
- Обстеження основного масиву.
- Суцільне.
- Несуцільне.
* Вибіркове.

11.4. З перелічених нижче вибіркових сукупностей, які вибірки вважаються малими за обсягом одиниць спостереження?
– До 50.
– До 70.
* До 30.
– До 100.

11.5. Як називають властивість вибіркової сукупності відтворювати генеральну сукупність?
– Ідентичність.
– Типовість.
* Репрезентативність.
– Уніфікованість.

11.6. У скільки разів скорочується обсяг робіт порівняно до суцільного спостереження, якщо вибірці підлягає 5 % одиниць загальної кількості?
* У 20 разів.
– У 25 разів.
– У 10 разів.
– У 15 разів.

11.7. Як називається помилка вибірки, одержана за формулою $\frac{\sigma}{\sqrt{n}}$?
– Гранична.
* Середня.
– Випадкова.
– Систематична.

11.8. Як називається помилка вибірки, одержана за формулою $t\sqrt{\frac{\sigma^2}{n}}$?
* Гранична.
– Середня.
– Випадкова.
– Систематична.

11.9. Величина якої помилки вибірки характеризує середнє квадратичне відхилення всіх можливих вибіркових середніх від генеральної середньої?
– Граничної.
– Випадкової.
– Систематичної.
* Середньої.

11.10. Який спосіб відбору потребує попередньої градації генеральної сукупності на якісно відмінні групи?
* Типовий.
– Серійний.
– Власне випадковий.
– Механічний.

11.11. За якою формулою визначається гранична помилка середньої при безповторному відборі?
– $m = \frac{\sigma}{\sqrt{n}}$.
– $m = \sqrt{\frac{\sigma}{n}\left(1 - \frac{n}{\mathrm{N}}\right)}$.
* $\Delta = t\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{\mathrm{N}}\right)}$.
– $\Delta = t\frac{\sigma}{\sqrt{n}}$.

11.12. Як зміниться середня помилка вибірки при повторному відборі, якщо чисельність вибірки збільшити у 4 рази?
– Не зміниться.
– Збільшиться у 4 рази.
* Зменшиться у 2 рази.
– Зменшиться у 4 рази.

11.13. За якою формулою визначається середня помилка частки ознаки, якщо обстеження здійснено за принципом безповторного відбору?
– $m = \sqrt{\frac{w(1 - w)}{n}}$.
* $m = \sqrt{\frac{w(1 - w)}{n}\left(1 - \frac{n}{\mathrm{N}}\right)}$.
– $\Delta = t\sqrt{\frac{\sigma}{n}\left(1 - \frac{n}{\mathrm{N}}\right)}$.
– $m = t\frac{\sigma}{\sqrt{n}}$.

11.14. Який рівень імовірності найчастіше використовують при розрахунках в аналізі аграрно-економічних явищ?
– 0,683.
* 0,954.
– 0,997.

– 0,999.

11.15. При вирішенні питання організації вибірки, яка статистична характеристика вважається критерієм?
– Середня.
– Дисперсія.
* Помилка вибірки.
– Імовірність.

11.16. За якою з наведених формул розраховують чисельність вибірки при визначенні і оцінці середньої ознаки за схемою безповторного відбору?
– $n = \frac{t^2\sigma^2}{\Delta^2}$.
* $n = \frac{t^2\sigma^2 \mathrm{N}}{\Delta^2 \mathrm{N} + t^2\sigma^2}$.
– $n = \frac{t^2 w(1-w)}{\Delta^2}$.
– $n = \frac{t^2 w(1-w)N}{\Delta^2 \mathrm{N} + t^2 w(1-w)}$.

11.17. Яким способом здійснюється розповсюдження результатів вибірки, якщо вибіркова середня помножується на відповідний показник обсягу?
* Спосіб прямого перерахунку.
– Спосіб поправочних коефіцієнтів.
– Спосіб прямого перерахунку з врахуванням поправочних коефіцієнтів.
– Шляхом розрахунку інтегрованих показників.

Тема 12. Подання статистичних даних: таблиці, графіки, карти

12.1. Назвати види статистичних таблиць.
– Прості.
– Складні.
– Групові.
* Прості, групові, комбінаційні

12.2. До якого виду відноситься статистична таблиця, побудована за трьома групувальними ознаками?
– Складна.
– Групова.
– Аналітична.

* Комбінаційна.

12.3. Яка з відповідей не відноситься до правил оформлення статистичних таблиць.

– Статистичні таблиці повинні бути замкнуті.
– Назва таблиці, її граф і рядків повинні бути стислими.
* Одиниці виміру вказуються після назви таблиці.
– Відсутність явища позначається символом «-».

12.4. Яка з відповідей дає визначення статистичного графіка?

– Зображення явищ на рисунку за допомогою символів.
– Наочне зображення статистичних даних.
* Спосіб наочного подання статистичних даних та їх співвідношень за допомогою геометричних знаків чи інших графічних засобів.
– Спосіб наочного подання статистичних даних із метою їх аналізу.

12.5. Яка з перелічених відповідей виходить за межі визначення особливостей мови статистичних графіків?

– Безперервність виразу.
– Двомірність графічних знаків.
* Своєрідність графічної системи.
– Відокремленість викладу.

12.6. У чому відмінність статистичних графіків від графіків взагалі?

– Особливість побудови.
– Відокремленість викладу.
– ДвомІрність графічних знаків.
* Предмет дослідження — масові статистичні дані.

12.7. Яка з перелічених відповідей виходить за межі визначення елементів статистичних графіків?

– Поле графіка, графічні знаки.
* Назва графіка.
– Просторові і масштабні орієнтири.
– Експлікація графіка.

12.8. Що покладено в основу наукової класифікації статистичних графіків?

– Форми і типи графіків.
– Умовні зображення та загальне призначення.
* Загальне призначення, види, форми і типи основних елементів.
– Предмет дослідження.

12.9. Як класифікуються графіки за видами їх поля?
* Діаграми, картограми, картодіаграми.
– Лінійні, стовпчикові, стрічкові.
– Прямокутні, колові.
– Фігурні.

12.10. Які існують види діаграм?
– Лінійні, стовпчикові.
* Лінійні, стовпчикові, стрічкові, прямокутні, колові, секторні, радіальні, фігурні.
– Лінійні і фігурні.
– Прямокутні та колові.

12.11. В якому виді діаграм статистичні дані зображують у вигляді прямокутників, розташованих по горизонталі?
– Стовпчикові.
* Стрічкові.
– Прямокутні.
– Секторні.

12.12. В якому виді діаграм величина явищ зображується у вигляді площ?
– Стовпчикові.
– Фігурні.
* Прямокутні.
– Квадратні.

12.13. Який вид діаграм будується для відображення структури явищ?
– Стрічкові.
* Секторні.
– Радіальні.
– Квадратні.

12.14. Який вид діаграм використовується для порівняння абсолютних величин?
– «Знак Варзара».
* Квадратні діаграми.
– Секторні діаграми.
– Радіальні діаграми.

12.15. Який вид діаграм використовується для порівняння трьох пов'язаних між собою величин?
* «Знак Варзара».
– Квадратні діаграми.

– Секторні діаграми.
– Радіальні діаграми.

12.16. Який вид діаграм будується за принципом співвідношень площ кіл як квадратів їх радіусів?
– Секторні.
– Радіальні.
* Колові.
– Кульові.

12.17. Який вид діаграм використовується для відображення явищ, котрі змінюються в замкнуті календарні строки?
– Секторні.
– Колові.
– Кульові.
* Радіальні.

12.18. Як називають метод графічного зображення явищ, який передбачає заміну геометричних фігур малюнками?
* Метод фігур-знаків.
– Метод графічних образів.
– Масштабні орієнтири.
– Просторові орієнтири.

12.19. Який вид графічних зображень застосовують для відображення явищ шляхом нанесення умовної штриховки на карту-схему?
– Діаграми.
* Картограми.
– Картодіаграми.
– Картосхеми.

12.20. Який вид радіальної діаграми використовують для зображення продовжуваного циклу діаграмованого явища?
– Замкнуті.
* Спіральні.
– Кульові.
– Секторні.

12.21. Як називаються види графіків, якщо розмір діаграмованого явища зображується у масштабі логарифмів, а дата — на осі абсцис?
– Логарифмічні.
* Напівлогарифмічні.
– Замкнуті.

– Спіральні.

12.22. Як називають статистичні графіки, якщо діаграмоване явище наноситься на карту у вигляді діаграм?

– Картосхема.
– Картограма.
* Картодіаграма.
– Фігурна діаграма.

НАУКОВО-ПІЗНАВАЛЬНІ ТЕМИ

ТЕМА 1. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

§ 1.1. Поняття про статистичні гіпотези

Гіпотеза - один з найважливіших факторів руху науки по шляху прогресу. Вона висувається як імовірний висновок у результаті спостереження, як правило, нових фактів. Виникаючи як результат спостереження за певними явищами /фактами/, гіпотеза набуває форми теоретичного припущення. Звернення до фактів зумовлює можливість перевірки такого припущення. При цьому факти, якими перевіряється гіпотеза, повинні бути науково обґрунтованими, тобто являти собою результат спостереження, що ґрунтується на наукових принципах.

Завдання перевірки статистичних гіпотез виникає у різних сферах людської діяльності, і в економіці зокрема. При порівнянні і оцінках різних явищ внаслідок наявності елементу випадковості, ця задача вирішується за допомогою методів математичної статистики. Як правило, у розпорядженні дослідників є вибіркові дані. Статистичний аналіз вибірки дозволяє зробити певний висновок щодо об'єкта дослідження шляхом обчислення статистичних оцінок (точечної, інтервальної). Але якщо задача оцінювання вирішує питання знаходження найкращої статистичної оцінки відносно параметрів або характеру розподілу вихідної сукупності, то завдання перевірки статистичних гіпотез зводиться до з'ясування, прийнятна чи ні деяка оцінка в ролі значення досліджуваної функції розподілу або параметра.

Завдання перевірки гіпотез інколи може логічно передувати завданню оцінки. Так, якщо дані спостереження містять грубі помилки, які особливо при малій вибірці істотно зміщують обчислювальні характеристики, то для одержання вірогідних оцінок необхідно переконатися спочатку в однорідності статистичної сукупності.

Гіпотеза - це наукове припущення, яке потребує перевірки, доведення. Під **статистичною гіпотезою** слід розуміти припущення про властивості випадкової величини, яке може бути перевірене за результатами статистичних спостережень. Статистичні гіпотези відносяться або до виду, або до окремих параметрів розподілу випадкової величини. Наприклад, статистичною вважають гіпотезу про те, що розподіл продуктивності однакових за фахом робітників, працюючих в рівних умовах, але на різних підприємствах, має нормальний характер розподілу.

Статистичною буде також гіпотеза про те, що середня продуктивність праці однакових за фахом робітників, працюючих в однакових умовах на різних підприємствах, не відрізняється за рівнем.

Якщо властивості випадкової величини виражаються кількісно, гіпотезу називають **параметричною,** якщо ж якісно - **непараметричною.** Наприклад, гіпотеза типу «нормальне розподіл має задану дисперсію» - параметрична; гіпотеза типу «вибіркові сукупності (дві, три и т.д.) однорідні» – непараметрична.

Висунуту гіпотезу, яку потрібно перевірити, називають **нульовою** (основною - Но). Гіпотезу протилежну нульовій, називають **конкуруючою** (альтернативною - H_I або H_A). Таким чином, перед дослідником стоїть завдання перевірити гіпотезу H_0 відносно конкуруючої гіпотези H_I за даними вибіркового спостереження з *n* незалежних змінних $x_1, x_2, x_3, ..., x_n$ над випадковою величиною x .

Суть перевірки гіпотези зводиться в цілому до умови, коли потрібно зробити висновок про вибір одного з можливих двох взаємовиключаючих рішень. Їх називають **альтернативними.** Наприклад, при випробуванні певного виду вітамінних добавок у кормових раціонах спортивних коней такими взаємовиключаючими (альтернативними) висновками будуть: а) вітамінна добавка N сприяє росту результатів спортивних змагань; б) вітамінна добавка N не сприяє росту результатів спортивних змагань.При вивченні впливу спеціалізації виробництва на його економічну ефективність можлива альтернатива: а) зростання спеціалізації виробництва сприяє підвищенню його економічної ефективності; б) зростання спеціалізації виробництва не сприяє підвищенню його економічної ефективності. При випробуванні гіпотези про роль матеріального стимулювання у справі підвищення продуктивності праці існують альтернативи; а) під-вищення рівня оплати однієї людино-години сприяє зниженню трудомісткості виробництва продукції; б) підвищення рівня оплати однієї людино-години не сприяє зниженню трудомісткості виробництва продукції.

Як видно, нульовій гіпотезі H_0 завжди протистоїть деяка альтернативна гіпотеза H_A, яка заперечує їй. При формальному підході будь-яку з конкуруючих гіпотез, здавалося б, можна розглядати як нульову. Але вибір однієї з двох гіпотез як нульової потребує обґрунтування. Аргументація спеціального обґрунтування випливає при розгляді питання про можливі помилки при перевірці статистичних гіпотез. Про це мова піде нижче.

Принципи обґрунтування та прийняття рішень в умовах випадкової варіації досліджуваних факторів розроблені в основному Е.Нейманом і К.Пірсоном, а також зустрічаються і в ряді робіт інших математиків і статистиків, присвячених питанням теорії перевірки статистичних гіпотез і розкриттю логічних основ їх оцінки.

Важливо пам'ятати, що змістовний бік математико-статистичних прийомів обробки даних має вирішальне значення в наукових дослідженнях. Незнання його або нез’ясування робить неможливим формулювання статистичної гіпотези, а також вибір відповідного прийому випробування (перевірки) даної гіпотези.

§ 1.2. Помилки при перевірці статистичних гіпотез. Статистичні критерії і критична область

Перевірка статистичних гіпотез здійснюється на основі вибіркових даних. Обмеженість обсягу вибірки зумовлює можливість прийняття неправильних рішень. Інакше кажучи, висунута гіпотеза може бути **вірного** і **невірного.** Звідси виникає необхідність перевірки правильності прийнятого рішення. Перевірку здійснюють статистичними методами, тому називають її **статистичною перевіркою.** В результаті такої перевірки існує можливість у двох випадках прийняти невірне рішення, тобто здійснити два роди помилок: або відхилити гіпотезу, коли вона вірна; або прийняти гіпотезу, коли вона невірна. Їх називають відповідно **помилками першого і другого роду.** Таким чином, помилка першого роду полягає в тому, що нульова гіпотеза H_0 приймається в той час, як є вірною гіпотеза H_I. Схематично ці випадки наведено в таблиці 96.

Таблиця 96.

Помилки при статистичній перевірці гіпотези

Нульова гіпотеза, Но	Вірна	Невірна
Відхиляється	Помилка першого роду	Вірне рішення
Приймається	Вірне рішення	Помилка другого роду

У результаті перевірки статистичної гіпотези її або приймають, або відхиляють. З цієї метою використовують випадкову величину, розподіл якої відомий. Залежно від закону розподілу цю величину в математичній статистиці прийнято позначати: а) z (абоυ)- якщо вона розподілена нормально; б*) F* – розподіл за законом Фішера-Снеденора; в) *t (або Т)*- за законом Стьюдента; г) χ^2 - за законом *"Хі*-квадрат" Пірсона і т.д.

Прийняте рішення відносно нульової гіпотези, 'таким чином, повинно опиратися на **статистичні критерії**. Під критерієм розуміють випадкову величину, на підставі якої приймається однозначне правило, що встановлює умови, при яких перевірювана гіпотеза повинна бути прийнята чи відхилена. Термін «прийняти гіпотезу» означає: «вирішити, що результати спостережень не суперечать висунутій гіпотезі». Під терміном «відхилити гіпотезу» слід розуміти протилежне – «вирішити, що результати спостережень суперечать висунутій гіпотезі».

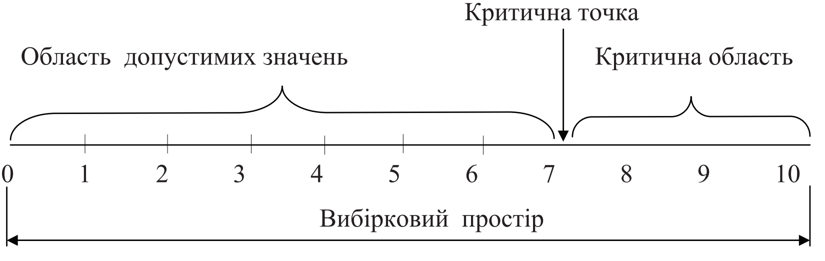


Рис. 39. Вибірковий простір випадкової змінної

Якщо уявити вибірковий простір як множину можливих вибіркових значень, розташованих у порядку зростання (рис.39), то статистичний критерій поділяє цей простір на дві непересічні області: **критичну** і **область прийняття гіпотез** (область допустимих значень, додаткова область). Точки, що відділюють критичну область від області прийняття гіпотези, називаються **критичними точками** (границями). Слід розрізняти **односторонню** (правосторонню і лівосторонню) і **двостронню критичні області.** Якщо прийняти позначення критерію через К, критичної точки – $К_{кр}$, то правосторонню критичну область можна уявити у вигляді нерівності $K \rangle K_{кр}$ *($К_{кр}$*- додатне число). Лівостороння критична область визначається нерівністю К< $K_{кр}$ ($К_{кр}$ – від'ємне число). Таким чином, **односторонньою** називають правосторонню або лівосторонню критичну область.

У випадках, якщо К < $К_1$, К > $К_2$ (при $K_2 > K_1$), критичну область називають **двосторонньою.**

На рис. 40 схематично зображено правосторонню (1), лівосторонню (2) і двосторонню критичну область (3).

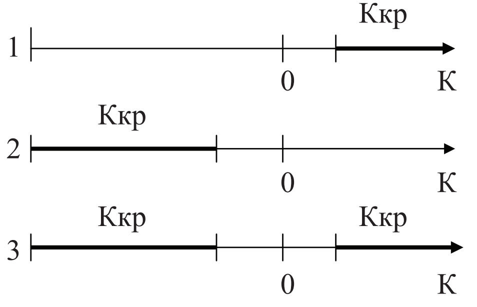


Рис. 40. Критична область:
1 - правостороння; 2 - лівостороння; 3 - двостороння

Якщо вибіркове значення статистичного критерію попадає у критичну область, то гіпотеза, яка перевіряється відхиляється, якщо у додаткову - гіпотеза приймається.

Отже, **під критичною областю** розуміється сукупність значень статистичної характеристики гіпотез, коли спостережуване значення характеристики належить до цієї сукупності, то гіпотеза $Н_0$ відхиляється на користь гіпотези $Н_I$.

Область допустимих значень (додаткова) даного критерію - це множина допустимих положень вибіркової точки у вибірковому просторі, які призводять до прийняття нульової гіпотези.

Під вибірковою точкою у математичній статистиці розуміють конкретне значення досліджуваної величини *x,* одержане у даній вибірці.

Як говорилося раніше, при перевірці статистичної гіпотези існує можливість зробити два роди помилок.

У задачах перевірки гіпотез помилка першого роду носить назву **рівня значимості** (мала ймовірність). Рівень значимості – це ймовірність зробити помилку першого роду, тобто відхилити правильну гіпотезу. Із зменшенням рівня значимості зростає ймовірність помилки другого роду при одному й тому ж обсязі вибірки.У статистичній практиці застосовують 5%-й рівень значимості. Його прийнято позначати через α- або $P_{0,05;}$ $P_{0,1}$ і т.д.

Величина, що доповнює ймовірність помилки другого роду, називається **потужністю критерію** для перевірки гіпотези проти її альтернативи. Отже, потужність критерію - це ймовірність того, що нульова гіпотеза (Но) буде відхилена, якщо вірна конкуруюча /H_I/, тобто є ймовірність не допустити помилку другого роду. Потужність критерію може бути зображена графічно (рис.41), як приклад, тут розглянемо випадок, коли в умовах дії нульової гіпотези, яка складається в визнання оцінки $\tilde{x}$ невідомого параметра $\bar{x}$,що дорівнює $\tilde{x}'$ при конкуруючій гіпотезі $\tilde{x} = \tilde{x}''$, вибіркова середня $\tilde{x}$ розподілена нормально з щільністю ймовірностей $f_0(z)$.Критерій будується так, що він є функцією $\tilde{x}$. При цьому $\bar{x}= 0$; $\sigma^2 = 1$.

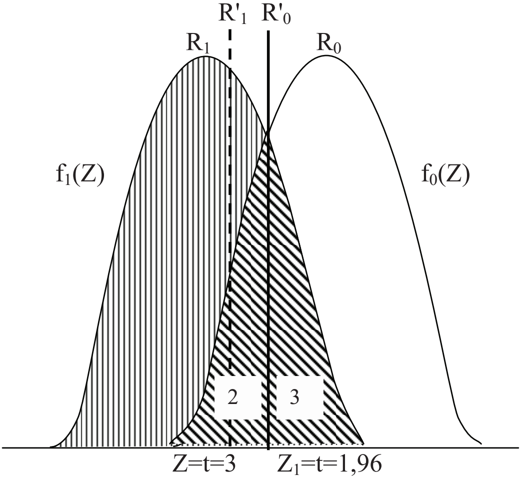


Рис. 41. Розподілу в умовах дії нульової і конкуруючої гіпотез: 1 - потужність критерію при даній альтернативі; 2 - імовірність помилки першого роду; 3 - імовірність помилки другого роду

Якщо справедлива конкуруюча гіпотеза, то критерій t розподілений з щільністю ймовірностей f(z). У цьому випадку, наприклад, значення | f | > 1.96 зустрічається з імовірністю 0,05, а |t| > 3-з імовірністю 0,003. Перпендикуляр R_0 проведений з точки z , що дорівнює 1,96, поділяє площину на дві части R_0 і R_1 . Якщо обчислене значення t попадає у точки, які належать області R_0 , то нульова гіпотеза (Но) не відхиляється.

Якщо значення *t,* попадає в область R_1 , то нульова гіпотеза відхиляється і приймається конкуруюча гіпотеза.

Перевірка нульової гіпотези ($H_{0)}$ здійснюється у такій послідовності:

1) вибирається рівень значимості;

2) визначають критичну область, яка відповідає цьому рівню значимості;

3 обчислюють вибіркове (фактичне) значення статистичної характеристики t ;

4) при попаданні значення *t* у критичну область нульова гіпотеза відхиляється; при попаданні t у область допустимих (додаткових) значень - гіпотеза не відхиляється.

§ 1.3. Перевірка статистичних гіпотез відносно середніх

Серед статистичних характеристик, відносно яких може висуватися і оцінюватися гіпотеза, найважливішим параметром є середня величина. Це зумовлює її роль як основної узагальнюючої характеристики статистичної сукупності.

Задачі, пов'язані з оцінкою гіпотез про середні величини, поділяються на дві групи: 1) дисперсія вихідної (генеральної) сукупності відома; 2) дисперсія вихідної сукупності невідома. При невідомій величині дисперсії її заміняють дисперсією вибіркових даних ($\sigma_в^2$).

Гіпотези відносно середніх перевіряються відповідно до логічних принципів у викладеній вище послідовності.

Приклад 1 (дисперсія відома). При вибірковому обстеженні 25 голів молодняка великої рогатої худоби встановлено, що середньодобовий приріст ваги однієї голови становить 888 г ($\tilde{x}$). Припустивши, що дані приросту розподіляються нормально з σ = 30*г*, перевірити на рівні значимості α = 0,05 нульову гіпотезу H_0: $\tilde{x}_{H_0}$ = 900 *г* проти конкуруючої гіпотези H_I : $\tilde{x}_{H1} \neq$ 900 *г*.

Рішення. Оскільки середньоквадратичне відхилення відомо, знаходимо спостережене значення вибіркової характеристики – розрахункове (t_p):

$$t_p = \frac{\tilde{x} - \tilde{x}H_0}{\sigma}\sqrt{n} = \frac{888 - 900}{30}\sqrt{25} = -2.$$

Оскільки конкуруюча гіпотеза H_I : $\widetilde{x}_{H1} \neq$ 900 г, вибираємо двосторонню критичну область з границями, які обчислюються із умови f (t) = $1 - \alpha = 0{,}95$.

За стандартною таблицею значень функції Лапласса (додаток 5). знаходимо, що рівню ймовірності P= 0,95 відповідає табличне |t|=1,96.

Оскільки $|t_p| > t_T$ (|2| >1,96), нульова гіпотеза (H_o) відхиляється, тобто вона протирічить вибірковим даним. Робимо висновок, що середньодобовий приріст ваги однієї голови молодняка тварин суттєво відрізняється від показника 900 г.

Приклад 2 (дисперсія невідома). При вибірковому обстеженні 10 голів молодняка тварин встановлений добовий приріст відповідно 850, 900, 910, 970, 825, 815, 827, 833, 912, 928 г.

Припустивши, що дані добових приростів ваги розподіляються нормально, необхідно перевірити на рівні значимості α = 0,05 нульову гіпотезу H_0 : $\widetilde{x}_{H0}$ = 850 г при конкуруючій гіпотезі H_I: $\widetilde{x}_{H1}$= 870 г.

Рішення. Визначаємо вибіркову середню арифметичну $\widetilde{x}$ і дисперсію σ^2.

$$\widetilde{x} = \frac{850+900+910+970+825+815+827+833+912+928}{10} = 877 \text{ г.}$$

Розраховуємо дисперсію, побудувавши допоміжну таблицю 97 .

Таблиця 97

Розрахунок дисперсії

x_i	$x_i - \widetilde{x}$	$(x_i - \widetilde{x})^2$
850	-17	289
900	33	1089
910	43	1849
970	3	9
825	-42	1764
815	-62	2704
827	-40	1600
833	-34	1156
912	45	2025
928	61	3721
8770	х	16206

$$\sigma^2 \frac{\Sigma(x_i - \widetilde{x})^2}{n} = \frac{16206}{10} = 1620.6;$$

$$\sigma = 40{,}257 \approx 40(г).$$

Оскільки відповідно умови і задачі дисперсія генеральної сукупності невідома, число ступенів вільності дорівнює $\nu = n - 1$. Розрахункове значення нормованого відхилення знаходимо за формулою:

$t_p = \frac{\tilde{x} - \tilde{x}_{H0}}{\sigma}\sqrt{n-1} = \frac{877-850}{40}\sqrt{10-1} = 2.02$. У зв'язку з тим, що ($H_1 : \tilde{x}_{H1} = 870 > \tilde{x}_{H0}$) рівно як і для ($H_1 : \tilde{x}_{H1} < \tilde{x}_{H0}$) вибираємо односторонню критичну область, границі якої обчислюємо з умови $f(t) = 1 - 2\alpha = 1 - 2 \times 0.05 = 0.90$.

За стандартною таблицею t - розподілу (див. додаток. 1) для числа ступенів вільності ν = 10 - 1 і рівня значимості α = 2 х 0,05 (0,10) знаходимо t_T =1,83. Оскільки $|t_p| > t_T$ (|2,02| > *1,83*), нульова гіпотеза відхиляється, тобто вона протирічить вибірковим даним. Робимо висновок, що середньодобовий приріст однієї голови суттєво відрізняється від 850 г.

§ 1.4. Перевірка статистичних гіпотез відносно розподілів

Розглянуті раніше питання перевірки статистичних гіпотез торкались величин параметрів досліджуваних статистичних сукупностей. При цьому міркування велись, виходячи з передбачень про однорідність сукупності і нормальності розподілу її одиниць. Тобто, сукупності що досліджуються, типові і різняться лише розмірами рівнів досліджуваних ознак. Але окремі випадки вимагають обов'язковості перевірки гіпотези відносно характеру розподілу. Тут необхідно вирішувати такі завдання:

1. Визначити відповідність емпіричного розподілу тому чи іншому теоретичному виду розподілу - нормальному, біноміальному, поліноміальному і т.п.

2. Визначити можливість належності двох і більше емпіричних розподілів до одного і того ж виду розподілів.

3. З'ясувати наявність незалежності у розподілі ознак однієї від іншої.

При перевірці гіпотез про розподіли широке застосування знаходить Xi - квадрат критерій. Детальне викладення аспектів і умов його використання буде дано нижче.

Оскільки емпіричний розподіл не завжди цілком відповідає нормальному розподілу, часто потрібно з'ясувати, сильно чи слабо розходяться емпіричні і теоретичні ряди. З цієї метою, необхідно встановити таку границю, недосягнення якої означає, що розбіжність між емпіричним і нормальним розподілом ще не така велика, щоб її враховувати, і що даний емпіричний ряд ще можна практично прийняти за нормальний. З цієї метою розраховується критерії *Xi*-квадрат (χ^2),

Величина χ^2 визначається за уже відомою формулою:

$\chi^2 = \sum_{i=1}^{n} \frac{(n_\phi - n_T)^2}{n_T}$, *де* $n_{\phi.}$ n_T – відповідно частоти емпіричного і теоретичного ряду.

Як бачимо, критерій *Xi*- квадрат являє собою суму відношень між квадратами різниць емпіричних і теоретичних частот до теоретичних частот.

Якщо при вибраному рівні ймовірності обчислені значення χ^2 перевищують табличні, то нульова гіпотеза про відповідність емпіричного розподілу теоретичному відхиляється.

Приклад . Розглянемо випадок перевірки статистичної гіпотези про відповідність емпіричного розподілу нормальному. Для прикладу використаємо ряд розподілу підприємств за показником собівартості виробництва однієї деталі (табл. 98).

Емпіричний розподіл представлено у вигляді інтервального ряду. Він заданий послідовністю рівновіддалених варіант (центр інтервалу) і відповідних частот.

Потрібно, використовуючи критерій χ^2,перевірити гіпотезу про те, що генеральна сукупність x розподілена нормально.

Для того, щоб при заданому рівні значимості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, необхідно виконати обчислювальну роботу за такими етапами:

1. Обчислити вибіркову середню $\tilde{x}$ і вибіркове середнє квадратичне відхилення σ.

2. Обчислити теоретичні частоти

$$n_T = f(t) \times \frac{ni}{\sigma},$$

де n – чисельність вибірки (сума всіх частот),

t - шаг інтервалу $t = \frac{|x_i = \tilde{x}|}{\sigma}$,

$f(t) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{t^2}{2}}$ (спеціальні таблиці – координати кривої).

Таблиця 98

Розрахунок частот нормального розподілу(вирівнювання емпіричних частот за нормальним законом)

Інтервал собівартості однієї деталі, грн. *(i=9)*	Середине значення /центр/ інтервалу, x_i	Число підприємств, n_i	Розрахункові величини				Статистичні параметри			
			$x_i\ n_i$	$x_i - \tilde{x}$	$(x_i - \tilde{x})^2$	$(x_i - \tilde{x})^2 n_i$	нормоване відхилення, $t = \frac{\lvert x_i - \tilde{x} \rvert}{\sigma}$	табличне значення функції $f(t)^*$ *	теоретична частота нормального ряду розподілу $f(t) \times \frac{ni}{\sigma}$	округлене значення теоретичної частоти, n_T
67-76	72	3	216	-28	784	2352	2,15	0,0395	1,81	2
76-85	81	5	405	-19	361	1805	1,46	0,137	6,26	6
85-94	90	15	1350	-10	100	1500	0,77	0,297	13,57	14
94-103	99	16	1584	- 1	1	16	0,08	0,398	18,19	18
103-112	108	17	1836	8	64	1088	0,62	0,329	15,04	15
112-121	117	7	819	17	289	2023	1,31	0,169	7,72	8
121-130	126	2	252	26	676	1352	2,00	0,054	2,47	2
130-139	135	1	135	35	1225	1225	2,69	0,011	0,50	1
Всього	х	66	6597	х	х	11361	х	х	65,56	66

* $f(t)^*$ знаходимо за стандартною таблицею $f(t) = \frac{1}{\sqrt{2}} e^{-\frac{t^2}{2}}$ (ординати нормальної кривої додаток 3)

Наприклад, t=2,15. Знаходимо у таблиці «2,1» на впроти «5» (0,0395).

i=9	$\tilde{x} = \frac{6597}{66} = 100$	$\sigma = \sqrt{\frac{11361}{66}} = 13$	$\frac{ni}{\sigma} = \frac{66 \times 9}{13} = 45,7$

3. Співставити емпіричні і теоретичні частоти за допомогою критерію Пірсона. Для цього необхідно :

а) скласти розрахункову таблицю (див. табл. 99), за якою знайти спостережуване значення критерію:

$$\chi_p^2 = \sum_{i=1}^{n} \frac{(n_i - n_T)^2}{n_T};$$

Таблиця 99

Розрахунок значення критерію χ^2

Серединне значення (центр) інтервалу	Частоти		Розрахункові величини		
	Емпіричне , n_i	Теоретичне, n_T	$n_i - n_T$	$(n_i - n_T)^2$	$\frac{(n_i - n_T)^2}{n_T}$
72	3	2	1	1	0,50
81	5	6	-1	1	0,17
90	15	14	1	1	0,07
99	16	18	-2	4	0,06
108	17	15	2	4	0,27
117	7	8	-1	1	0,12
126	2	2	0	0	0,00
135	1	1	0	0	0,00
Всього	66	66	х	х	1,19

$$\alpha = 0{,}05; 0{,}01; 0{,}001; \quad \gamma = 5; \quad \chi_p^2 = 1{,}19;$$

$$\chi^2_{T(0,05;5)} = 11{,}1; \quad \chi^2_{T(0,01;5)} = 15{,}1; \quad \chi^2_{T(0,001;5)} = 20{,}5;$$

б) за таблицею стандартних значень χ^2 при заданому рівні значимості α і числі ступенів вільності $\gamma = l - 3$ (*l* - кількість груп), знаходимо критичну точку $\chi^2_T(\alpha.\nu)$ правосторонньої критичної області.

Якщо $\chi_p^2 < \chi_T^2$, то гіпотеза про нормальність розподілу генеральної сукупності не відхиляється, тобто емпіричні і теоретичні частоти розрізняються незначимо (випадково). При $\chi_p^2 > \chi_T^2$ - гіпотезу відхиляють, тобто емпіричні і теоретичні частоти різняться значимо.

Нечисленні частоти *(n<5)* рекомендується об'єднувати. Відповідні їм теоретичні частоти підсумовують. За такої операції число ступенів вільності розраховують виходячи з того, що *l* дорівнює числу груп, які залишились після об'єднання частот.

Потрібно пам'ятати, що при перевірці гіпотези про нормальний розподіл генеральної сукупності за допомогою критерію χ^2 число ступенів вільності (υ) знаходять за формулою υ = *l* – 1 - р , де *p* - число параметрів, які оцінюються за вибіркою. Нормальний розподіл визначається двома параметрами: математичним очікуванням α- і середнім квадратичним відхиленням σ. Оскільки обидва ці параметри оцінювалися за вибіркою (за оцінку приймають вибіркову середню $\tilde{x}$, , а за оцінку σ - вибіркове середнє квадратичне відхилення), то р =2, тоді, υ = *(l* - 1 – 2) = l - 3.

Виконуючи розрахунки відповідно вказаній вище послідовності, необхідно побудувати розрахункові таблиці /див. табл. 38, 39/ і обчислити всі необхідні параметри для перевірки статистичної гіпотези про відповідність емпіричного розподілу нормальному. Із таблиці 39 знаходимо $\chi^2_p = 1{,}19$.

За стандартною таблицею критичних значень розподілу χ^2 (див. додаток. 7), при рівні значимості α = 0,05 і числу ступенів вільності υ = *l*-3= 8- 3=5, знаходимо критичну точку правостороньої критичної області $\chi^2_{T(0{,}05;5)} = 11{,}1$.

Оскільки $\chi^2_p < \chi^2_T$ - гіпотеза про нормальний розподіл генеральної сукупності не відхиляється, тобто емпіричні і теоретичні частоти розрізняються незначимо (випадково).

Таким чином, висунута гіпотеза приймається. Графічне зображення розподілу емпіричних і теоретичних частот показано на рисунку 42.

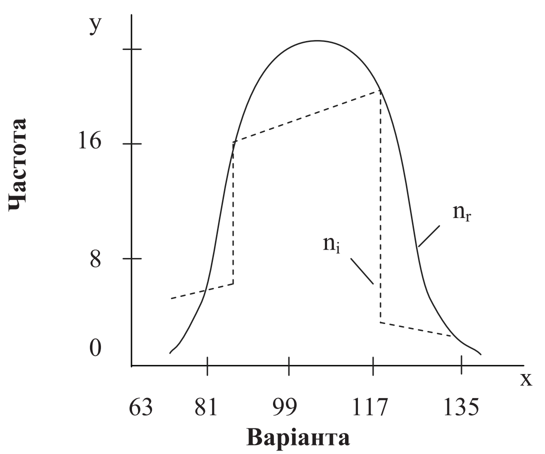


Рис. 42. Розподіл емпіричних (n_i) і теоретичних (n_T) частот

§ 1.5. Основні аспекти і умови застосування *xi* - квадрат критерію

Xi - квадрат (критерій згоди Пірсона - χ^2) є об'єктивною оцінкою близькості емпіричних розподілів до теоретичних. Використовується, як уже було сказано, у тих випадках, коли необхідно встановити відповідність двох порівнюваних рядів розподілу - емпіричного і теоретичного, або двох емпіричних. При цьому порівнюються частоти названих рядів розподілу, виявляються розбіжності між ними і визначається вірогідність цих розбіжностей.

За допомогою *Xi* - квадрат критерію можна виявити відміни в розподілі двох емпіричних рядів, порівнювати вибірки, які мають альтернативні ознаки, а також оцінювати вірогідність кореляції між альтернативними ознаками. Як і інші критерії згоди (Колмогорова λ, Романовского, Фішера F, Ястремского L), χ^2 являє собою деяку величину, яка оцінюється з певною ймовірністю. Він може приймати різні завжди додатні значення (малі й великі). При χ^2=0 слід вважати, що відміни між частотами порівнюваних рядів розподілу відсутні. Даний критерій не рекомендується використовувати для оцінки малих вибірок.

Як було показано в § 1.4, за допомогою χ^2 - критерію можна здійснити статистичну перевірку гіпотез відносно розподілів, тобто відповідність емпіричних даних розподілу деякому теоретичному закону розподілу. Таку оцінку наближення емпіричного розподілу до теоретичного дає сума співвідношень частот

$$\chi^2 = \sum_{i=1}^{n} \frac{(n_\phi - n_T)^2}{n_T},$$

де n_ϕ, n_T - відповідно частоти емпіричного і теоретичного ряду.

Збіг емпіричних і теоретичних частот зумовлює величину χ^2= 0. Це вказує на підтвердження нульової гіпотези. (Но). При наявності достовірної різниці у частотах емпіричного і теоретичного ряду величина χ^2 буде свідчити про неправильність висунутої гіпотези. Значення параметра *xi* - квадрат зростає із збільшенням різниці між частотами. Величина χ^2 також залежить від числа ступенів вільності. Чим менше значення χ^2, тим вищі його ймовірність і вірогідність. Таким чином, при зміні величини χ^2 від 0 до ∞ імовірність його змінюється від 1 до 0. У міру наближення n к ∞ розподіл χ^2 наближається до нормального.

При використанні *xi* - квадрат критерію необхідно пам'ятати про достатньо велике число одиниць вибірки *(n* > 50) і величини частот (n_s > 5). Як було сказано раніше при n_s < 5 об'єднують сусідні інтервали ряду розподілу. Якщо вибіркова сукупність досить велика, *Xi* - квадрат критерій буде обґрунтований, тобто у такому випадку він майже завжди спростовує невірну гіпотезу. Серед розроблених критеріїв згоди цей критерій забезпечує мінімальну помилку в прийнятті невірної гіпотези.

При оцінці відмінностей між емпіричним і теоретичним розподілами потрібно знати величини χ^2 , які відповідають визначеним рівням значимості. Для цієї мети К.Пірсон розробив стандартні таблиці, в яких на перетині значень χ^2 і числа ступенів вільності подані ймовірності, які оцінюють величину χ^2 (додаток 6) .

Якщо за розрахунковими даними значення ймовірності виявиться дуже малою величиною, наприклад 0,01, то відмінності між досліджуваними рядами потрібно вважати істотними, тобто нульова гіпотеза не приймається. Якщо ж імовірність виявиться не малою, розбіжності вважаються випадковими і нульова гіпотеза приймається. Р.Фішер довів, що ризик зробити помилку буде невеликим, якщо провести суміжну лінію у ймовірності Р = 0,05. Значення χ^2 , які лежать за цією лінією (0,04; 0,03; 0,02 тощо), .вказують на наявність істотних відхилень.

При розрахунку числа ступенів вільності досліджуваних частот враховують кількість обчислювальних статистичних характеристик теоретичної функції розподілу. У даному випадку використовувана кількість таких параметрів дорівнює 2 $(\tilde{x},\sigma)$, звідси υ =*l*-1- 2 .

Приклад. Розглянемо розрахунок оцінки відмінності теоретичного і емпіричного розподілів 54 підприємств за трудомісткістю виробництва одиниці продукції (табл. 100).

Для знаходження величини χ^2 використовується відома формула:

$$\chi^2 = \frac{(n_i - n_T)^2}{n_T},$$

де n_i, n_T - відповідно емпірична і теоретична частоти ряду розподілу. У нашому прикладі, підсумовуючи часткові значення величин χ^2, одержуємо розрахункове значення χ^2- критерію, що дорівнює 4,57. Знаходимо число ступенів вільності для даного випадку:υ=*l* –1-р=6-1-2=3 (*l* = 6, оскільки два останні інтервали об'єднані в один). За стандартною таблицею /додаток. б/ знаходимо $P(\chi^2 = 4{,}57)$ при числі ступенів вільності, рівному 3. Табличне значення $P(\chi^2) =$ 0,1718. Даний рівень імовірності значно відрізняється від нуля. Тому відмінність між емпиричним і теоретичним розподілами потрібно вважати випадковою і досліджуваний розподіл підприємств за трудомісткістю одиниці продукції необхідно визнати як підпорядкований закону нормального розподілу.

Таблиця 100

Розрахунок критерію χ^2 при оцінці відмінностей між емпіричним і теоретичним рядами розподілу

Серединне значення /центр/ інтервалу, x_i	Частота, n_i	Центрування, $\lvert x_i - \bar{x} \rvert$	Нормоване відхилення, $t = \frac{\lvert x_i - \bar{x} \rvert}{\sigma}$	Табличне значення функції, f(t)	Теоретична частота, n_T	Різниця частот, $(n_i - n_T)$	χ^2 - критерій, $\frac{(n_i - n_T)^2}{n_T}$
241	6	148	1,52	0,1257	5	1	0,20
307	10	82	0,84	0,2803	10	*0*	0,00
373	20	16	0,16	0,3939	15 '	5	1,60
439	7	50	0,51	0,3503	13	-6	2,77
505	7	116	1,19	0,1965	3	0	0.00
571 637	2 2	182 248	1,87 2,55	0,0695 0,0154	1	0	0,00
Всього	54	x	x	x	54	x	4,57

i =66	$\tilde{x}$ =389	σ= 97	$\frac{ni}{\sigma} = 37$

Замість значень показників ймовірності *($P\chi^2$),* розроблених К.Пірсоном, Р.Фішер розрахував стандартну таблицю значень χ^2 , які відповідають шуканим імовірностям при різному числі ступенів вільності варіації (додаток. 7).

Розрахунок, такої стандартної таблиці він аргументував тим, що в практичних обчисленнях не так важливо знати точне значення ймовірності Р, яке відповідає значенню χ^2, як визначити, в якій мірі вірогідно фактичне значення χ^2. Тому в стандартних таблицях Р.Фішера наведено значення χ^2, які відповідають певним рівням імовірності $P\chi^2$. При цьому потрібно відмітити, що названі таблиці містять значення $\chi^2<1$, які зустрічаються для малих ступенів вільності, і значення $\chi^2>30$ - для великих величин υ .

Стандартні математичні таблиці значень χ^2, Р.Фішера мають більш широкі аспекти практичного використання при розрахунках статистичних оцінок. Випадок їх використання при перевірці гіпотези про розподіли був розглянутий у § 1.4. Нижче викладені деякі інші аспекти розрахунку статистичних оцінок з застосуванням критичних значень χ^2 стандартних таблиць Р.Фішера.

За допомогою критерію *xi-* квадрат можна перевірити належність кількох вибіркових даних однієї і тієї ж генеральної сукупності при вирішенні питання про однорідність вибірки.

Приклад. Порівняти варіаційний ряд продуктивності молочного стада корів ферм великого рогатої худоби з частковою механізацією і повною механізацією виробничих процесів.

Обчислення *xi-* квадрат критерію для встановлення різниці в частотах двох емпіричних рядів полягає в тім, що за частоти невідомої генеральної сукупності приймаються величини половини суми частот по кожному інтервалу порівнюваних рядів розподілу (табл. 101).

Подальші розрахунки зводяться до знаходження різниці емпіричних і теоретичних частот і визначення суми їх співвідношень по вже відомій формулі *Xi* - квадрат критерію (табл. 102).

При цьому в ряд емпіричних частот заносять у послідовному попарної запису частоти обох рядів, у графу теоретичних частот - відповідні їм усереднені частоти теоретичного ряду. Розраховані часткові значення χ^2 підсумовують і отриману суму (χ^2_p) порівнюють із стандартним значенням (χ^2_T) при рівні значимості α = 0,05 і числі ступенів вільності варіації υ=*l*-1= 9-1=8 (додаток. 7). Як бачимо, критична точка для цих параметрів становить 15,5, тобто величина χ^2_T дорівнює 15,5 і менше χ^2_p на 4,4 при порозі імовірності Р = 0,95.

Таблиця 101

Розрахунок теоретичних частот при порівнянні двох емпіричних рядів

Середньорічний надій, кг	Емпірична частота поголів'я корів на фермах з різним рівнем механізації		Теоретичні частоти, $n_T = \frac{n_1 + n_2}{2}$
	Частково механізованих, n_1	Цілком механізованих, n_2	
3500-3700	10	4	7
3700-3900	12	8	10
3900-4100	16	10	13
4100-4300	21	11	16
4300-4500	15	21	18
4500-4700	7	17	I2
4700-4900	7	15	11
4900-5100	6	12	9
5100-6300	6	2	4
	100	100	100

Таблиця 102

Розрахунок значення критерію χ^2 при порівнянні двох емпіричних рядів

Емпіричні частоти у парнопослідовному запису, n_i	Теоретичні частоти, n_T	Розрахункові величини	
		$n_i - n_T$	$\frac{(n_i - n_T)^2}{n_T}$
10	7	3	1,29
4	7	-3	1,29
12	10	2	0,40
8	10	-2	0,40
16	13	3	0,69
10	13	-3	0,69
21	16	а	1,56
11	16	-5	1,56
15	18	-3	0,50
21	18	5	0,50
7	12	-5	2,08
17	12	5	2,08
7	11	-4	1,45
15	11	4	1,45
6	9	-3	1,00
12	9	3	1,00
6	4	2	1,00
2	4	-2	1,00
200	200	х	19,94

Оскільки $\chi^2_p > \chi^2_T$ можна зробити висновок про вірогідність різниці двох емпіричних рядів за показниками продуктивності молочного стада корів. Як видно, стан механізації виробничих процесів суттєво впливає на показники продуктивності тварин які обслуговуються . А вибіркову сукупність поголів'я корів, з обслуговуванням при повній механізації процесів, слід вважати якісно іншою- вона належить до другої генеральної сукупності.

У порівнянні із способом обчислення величини Xi -квадрата, розглянутого раніше (в §1.4), коли мова йшла про порівняння емпіричного і теоретичного ряду, в даному прикладі значення χ^2_p занижено майже в два рази. Це зумовлюється прийнятим способом обчислення, коли при обробці даних використовується не різниця частот, а напіврізниця, і кількість величин часткових значень χ^2 більша в два рази, ніж в розрахунках розбіжностей емпіричного і теоретичного рядів розподілу.

При порівнянні двох емпіричних рядів розподілу, представлених неоднаковою кількістю одиниць вибіркової сукупності $(\Sigma n_1 \neq \Sigma n_2)$, розрахунок xi - квадрат критерію має свою особливість

Приклад. Потрібно визначити, чи вірогідна розбіжність частот за показниками врожайності зернових культур в інтервальному ряді розподілу підприємств з посівами озимої і ярої пшениці (табл. 103).

Таблиця 103

Розрахунок теоретичних частот при порівнянні двох емпіричних рядів з нерівною кількістю одиниць спостережень

Урожайність зернових культур, ц з 1 га	Число підпрємств (частота) з посівами пшениці		Сума частот, n_1+n_2	Теоретичні частоти	
	озимої, n_1	ярої, n_2		$n_{T_1} = \frac{(n_1+n_2)\Sigma n_1}{\Sigma(n_1+n_{21})}$	$n_{T2} = \frac{(n_1+n_2)\Sigma n_2}{\Sigma(n_1+n_2)}$
30-35	5	8	13	9,0	4,0
35-40	20	10	30	20,8	9,2
40-45	35	15	50	34,6	15,4
45-50	20	5	25	17,3	7,7
50-55	10	2	12	8,3	3,7
Всього	90	40	130	90,0	40,0

На першому етапі розрахунків визначають теоретичні частоти для двох емпіричних рядів (табл. 104), суми за якими повинні бути рівні сумам частот відповідних рядів $(\Sigma n_i = \Sigma n_{T1} = n_{T2})$. Потім розраховують часткові значення χ^2 за раніше описаною схемою і одержані значення підсумовують (табл. 104). Розрахункова величина *xi* - квадрат критерію (χ^2_p) в розглянутому прикладі дорівнює 8,39. Число ступенів вільності $\gamma = (l_1 - 1)(l_2 - 1)$, де l_1 -число інтервалів, l_2 - число емпіричних рядів розподілу. Таким чином, υ=(5-1)(2-1)=4. За стандартною таблицею значень χ^2 (додаток 7) знаходимо теоретичне значення *xi*- квадрат критерію (χ^2_T) .

Для порогу ймовірності Р =0,95 і числа ступенів вільності 4 величина його становить 9,5. Оскільки $\chi^2_T > \chi^2_P$ можна зробити висновок про неістотність різниці в частотах двох емпіричних рядів розподілу, які являють собою неоднакову вибірку за кількістю одиниць спостереження.

Таблиця 104

Розрахунок значення критерію χ^2 при порівнянні двох емпіричних рядів з нерівною кількістю одиниць спостережень

Емпіричні частоти в парнопослідовному запису, n_i	Теоретичні частоти, n_T	Розрахункові величини	
		$n_i - n_T$	$\frac{(n_i - n_T)^2}{n_T}$
5	9,0	-4,0	1,78
8	4,0	4,0	4,00
20	20,8	-0,8	0,03
10	9,2	0,8	0,07
35	34,6	0,4	0,00
15	15,4	-0,4	0,01
20	17,3	2,7	0,42
5	7,7	-2,7	0,95
10	8,3	1,7	0,35
2	3,7	-1,7	0,78
130	130	х	8,39

Хі – квадрат критерій використовується також при порівняні двох альтернативах розподілів, представлених у вигляді кореляційної решітки. Формула, за якою обчислюється χ^2 , в цьому випадку має вигляд:

$$\chi^2 = \frac{\left[(ad - вc) - \frac{n}{2}\right]^2 n}{(a + в)(c + d)(a + c)(в + d)},$$

де *a,в,c,d* - частоти у відповідних клітинах кореляційної решітки;

n - число одиниць вибіркової сукупності. Для прикладу розглянемо таблицю розподілу робітників по двох групах альтернативних ознак (табл. 105).

Після підстановки у наведену вище формулу даних таблиці 105 одержимо:

$$\chi^2 = \frac{\left[(60-1440)-\frac{100}{2}\right]^2 100}{30\times 70\times 66\times 34} = \frac{(-1430)^2\times 100}{4712400} = \frac{204490000}{4712400} = 43{,}4\,.$$

Таблиця 105

Розрахунок значення χ^2 при порівнянні альтернативних розподілів

Кількість працівників	Кількість робітників		Сума
	які підвищили продуктивність праці	які не підвищили продуктивність праці	
Заохочуваних матеріально	6 (а)	24 (в)	а + в = 30
Не заохочуваних матеріально	60 (с)	10 (d)	с +d= 70
	а + с =66	в + d= 34	а+в+с+d= 100

Число ступенів вільності $\nu = (l_1 - 1)(l_2 - 1)$=(2-1)(2-1)=1 ($l_1.l_2$- число груп відповідно по першій і другій ознаках). Отримане розрахункове значення χi - квадрат критерію *(χ_p^2)* порівнюємо з табличним значенням (додаток 7).

Для порогу ймовірності 0,95 величина його становить 3,8. Оскільки $\chi_p^2 > \chi_T^2$ (43,4 > 3,8), можна зробити висновок, що розглядувані альтернативні розподіли різняться значимо, тобто матеріальний стимул зумовлює підвищення продуктивності праці робітників.

Одне із практичних значень χi - квадрат критерію полягає у тому, що розрахункову його величину при альтернативних ознаках *(χ_p^2)* можна використовувати для визначення ступеня тісноти зв'язку між ними. Такою статистичною характеристикою є коефіцієнт кореляції *(r_x)*.Розрахунок його проводиться за такою формулою:

$$r_x = \sqrt{\frac{\chi_p^2}{n\sqrt{(l-1)(l_2-1)}}} = \frac{\chi_p^2}{n\sqrt{\nu}}\,.$$

Для прикладу, розглянутого вище, величина коефіцієнта кореляції становитиме: $r_x = \sqrt{\frac{43,4}{100\sqrt{1}}} = \sqrt{0,434} = 0,659.$

По його рівню можна стверджувати про досить значний ступінь тісноту зв'язку між продуктивністю праці і її матеріальним стимулюванням.

Крім розглянутих випадків *Xi*- квадрат критерій може бути застосований і для других розподілів, в яких розбіжності між теоретичними (очікуваними) і фактичними частотами не відповідають якій-небудь очевидній структурі. Використовують даний критерій згоди і при з'ясуванні об'єктивності побудови ряду розподілу, тим самим виключаючи можливість підгонки одиниць спостереження з метою відповідності їх одна одній.

Проте, використання цього критерію передбачає ряд обмежень , яких необхідно дотримуватись при розрахунку χ^2 як критерію істотності. Як було відмічено раніше, при перевірці гіпотези на відповідність емпіричного розподілу теоретичному бажано мати не менше 50 одиниць спостереження, а в кожній теоретично розрахованій групі мінімально допустима границя величини частот умовно приймається рівною 5 (інколи 3). У зв'язку з цим при малочисельності груп (як правило, крайніх) їх об’єднують. При визначенні відповідності розподілів нормальному закону число ступенів вільності дорівнює числу груп (інтервалів) мінус три (l – 3). Пояснюється це тим, що обчислення теоретичних частот пов'язано тут трьома умовами (обмеженнями), які визначають нормальний розподіл: визначеним обсягом вибірки (n), середньою величиною ($\tilde{x}$), від якої знаходяться центральні відхилення, середнім квадратичним відхиленням (σ), за яким проводиться нормування центральних відхилень серед груп (інтервалів).

При об'єднанні інтервалів (у випадку наявності в ряді розподілу частот нижче мінімально допустимого рівня) число ступенів вільності коригують. Величина вторинного числа ступенів вільності дорівнюватиме різниці між числом інтервалів після їх об'єднання мінус 3 (l' -3). Тому для прикладу, розглянутого в § 1.4., останніх два інтервали потрібно було об'єднати. У зв'язку з цим число ступенів вільності становитиме υ=l - 3 = 7-3=4.

При цьому стандартне значення Xi- квадрат критерію для Р = 0,95 дорівнюватиме 9,5 (додаток 7).

Розглянуті вище випадки використання Xi - квадрат критерію свідчать про широкі можливості його застосування в практиці статистико - економічного аналізу.

§ 1.6. Оцінка розподілів з використанням критерію згоди Колмогорова

Розглянуті в попередніх параграфах способи оцінки відмінності між двома вибірковими спостереженнями ґрунтувалися на припущенні про нормальний характер розподілу генеральних сукупностей (або близькому до нормального). Але експериментатору (досліднику) не завжди відома форма розподілу даних, з яких проводиться вибірка. Тому використання критеріїв t і χ^2 може інколи привести до суб'єктивної оцінки результатів спостережень. У зв'язку з цим в математичній статистиці розроблені критерії оцінки вибірок з будь - якого виду розподілу.

Теоретичною основою їх розробки є припущення, що ряд послідовних спостережень можна розглядати як просту незалежну вибірку з незмінним розподілом. Ці критерії одержали назву **непараметричних.** Для їх розрахунків непотрібно обчислювати середню, дисперсію та інші статистичні характеристики вибіркових розподілів. У деяких випадках для розрахунку непараметричних критеріїв використовуються не безпосередні дані спостереження, а різного роду впорядковані ряди (з нагромадженими частотами, ранжировані різниці одиниць спостережень і т.п.). Критерії, які розраховуються у цьому випадку, називають порядковими.

До непараметричних критеріїв відносять: критерій λ, (лямбда) Колмогорова, критерій Уайта, критерій Уілксона.

Ступінь наближення емпіричного розподілу до обчислюваного (теоретичного), в равній мірі як і порівняльну оцінку двох однорідних варіаційних рядів, визначають за допомогою непараметричного критерію λ (лямбда). Якщо використання xi- квадрат критерію ґрунтується на використанні таких вибіркових характеристик (параметрів) як середня ($\tilde{x}$) і стандарт (σ), то при розрахунку λ - критерію їх обчислення непотрібно. Він оснований на відповідності рядів інтегральних (нагромаджених) частот досліджуваних сукупностей. Суть його полягає в розрахунку величини максимальної різниці (D) нагромаджених частот (частостей) емпіричного і теоретичного розподілів. Тобто для використання цього критерію необхідне впорядкування двох рядів розподілу у вигляді їх кумуляції. А.Н.Колмогоров довів, що при необмеженому зростанні чисельності вибірки *(n)* імовірність нерівності $D\sqrt{n} \geq \lambda$ прямує до границі

$$\lim_{n\to\infty} P(D\sqrt{n} \geq \lambda) = 1 \sum_{n\to\infty}^{+\infty} (-1)^k e^{-2\kappa^2\lambda},$$

де D - величина максимальної різниці нагромаджених частот (частостей) емпіричного і теоретичного розподілів.

Непараметричний показник λ розраховується як відношення максимальної різниці (без врахування її знака) нагромаджених частот емпіричного і теоретичного рядів розподілу до кореня квадратного із чисельності вибірки:

$$\lambda = \frac{D}{\sqrt{n}} = \frac{\max|n_i - n_T \ |}{\sqrt{n}}.$$

У випадку повного збігу порівнюваних частот в рядах розподілу λ = 0. Чим більша розбіжність в рядах, тим більша величина лямбда-критерію. Але занадто велику величину λ випадковими відхиленнями у порівнюваних рядах розподілу пояснити важко, тому робиться висновок про невідповідність вибіркового розподілу і теоретично припущеного.

Критерій згоди Колмогорова (λ), на відміну від χ*i*- квадрат критерію, дуже простий не тільки в розрахунках, але й не передбачає використання стандартних таблиць для його оцінки. Теоретично доведено, що при чисельності вибіркової сукупності приблизно більш 25 одиниць *(n* > 25) граничні значення критерію лямбда (λ_T), що відповідають трьом порогам довірчої імовірності (P = 0,95; P = 0,99; P= 0,999), дорівнюють відповідно 1,36; 1,63; 1,95. Показник числа ступенів вільності при цьому не розраховується.

Таким чином, якщо $\lambda_p \geq \lambda_T$, то з відповідною ймовірністю розбіжності між емпіричним і теоретичним розподілами визнаються значущими (істотними).

Приклад. Продемонструємо розрахунок критерію згоди Колмогорова на прикладі розподілу підприємств за урожайністю зернових культур (табл. 106).

Таблиця 106

Розрахунок критерію "ламбда" (λ) при оцінці розбіжностей між емпіричним і теоретичним рядами розподілу

Серединне значення (центр) інтервалу, x_i	Частота, n_i	Центрування, $\lvert x_i - \tilde{x} \rvert$	Нормоване відхилення, $t = \frac{\lvert x_1 - \tilde{x} \rvert}{\sigma}$	Табличне значення функції, f(t)	Теоретична частота (заокруглена), n_T	Нагромаджені частоти		$d = \lvert n_i^i - n_T^I \rvert$
						n_i^i	n_T^I	
20,5	7	6,4	1,42	0,1456	6	7	6	I
23,5	15	3,4	0,75	0,3011	12	22	18	4
26,5	16	0,4	0,10	0,3970	15	38	33	$\boxed{5}$
29,5	9	2,6	0,58	0,3372	13	47	46	1
32,5	5	5,6	1,24	0,1849	7	52	53	1
35,5	3	8,6	1,91	0,0644	3	55	56	1
38,5	2	11,6	2,58	0,0143	I	57	57	0
x	57	x	x	X	x	x	x	x

$i = 3$	$\tilde{x} = 26{,}9$	$\sigma = 4.5$	$\frac{ni}{\sigma} = 38.0$

Максимальне значення різниці нагромаджених частот емпіричного і теоретичного рядів розподілу дорівнює $n_i^i - n_T^i = 5$. Величина ламбда-критерію становитиме:

$$\lambda = \frac{5}{\sqrt{57}} = 0{,}66 .$$

Одержана величина критерію (0,66) значно менше теоретичного її рівня ($\lambda_{0,05}$ = 1,36) відповідного порога ймовірності (Р = 0,95). Тому розходження між емпіричним розподілом і нормальним визнаються невірогідними, тобто розходження між частотами знаходяться в межах випадкових коливань.

Для виявлення вірогідності двох емпіричних розподілів, одержаних в результаті вибірки з однієї і тієї ж генеральної сукупності, але які мають неоднакову кількість одиниць, критерій "ламбда" обчислюється за формулою:

$$\lambda = \left| \frac{\Sigma n_1}{n_1} - \frac{\Sigma n_2}{n_2} \right| \max \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

де $\frac{\Sigma n_1}{n_1}$ - суми нагромаджених частот по кожному інтервалу першого ряду розподілу, поділені на обсяг вибірки; $\frac{\Sigma n_2}{n_2}$ - те ж по другому ряду розподілу;

$\left| \frac{\Sigma n_1}{n_1} - \frac{\Sigma n_2}{n_2} \right| \max$ - максимальне абсолютне значення (без врахування знака) різниці часток від ділення нагромаджених частот на обсяг вибірки;

n_1, n_2 - обсяги одиниць вибіркової сукупності по першому і другому ряду розподілів.

У таблиці 107 представлені два емпіричних ряди розподілу підприємств за врожайністю зернових культур, відібраних з однієї генеральної сукупності, але які різняться обсягами вибірки.

В останній графі цієї таблиці максимальна різниця кумульованих частот представлена величиною 0,07. Підставивши її у вище наведену формулу, знайдемо значення критерію "ламбда".

$$\lambda = d \max \sqrt{\frac{n_1 \times n_2}{n_1 + n_2}} = 0.07 \sqrt{\frac{100 \times 200}{100 + 200}} = 0{,}57 .$$

Розрахована величина критерію менше граничних критичних значень λ для всіх трьох порогів імовірності ($\lambda_{0,05} = 1{,}36$; $\lambda_{0,01} = 1{,}63$; $\lambda_{0,001} = 1{,}95$). Це свідчить про неістотність розбіжностей між порівнюваними емпіричними рядами розподілу. Звідси висновок про те, що обидві вибірки репрезентують досліджувану генеральну сукупність.

У випадках, коли під знаком радикала громіздкі числа, розглянуту вище формулу приводять до вигляду:

$$\lambda^2 = d^2 \max \frac{n_1 \times n_2}{n_1 + n_2}.$$

Таблиця 107

Розрахунок критерію "ламбда" (λ) при оцінці розбіжностей між емпіричними рядами розподілу з неоднаковими обсягами вибірки

Серединне значення інтервалу, x_i	Частоти		Нагромаджені частоти		Розрахункові дані		
	n_1	n_2	$\overset{i}{n}_i$	$\overset{i}{n}_2$	$\frac{\overset{i}{n}_1}{n_1}$	$\frac{\overset{i}{n}_2}{n_2}$	$d = \left\lvert \frac{\overset{i}{n}_1}{n_1} - \frac{\overset{i}{n}_2}{n_2} \right\rvert$
20,5	3	5	3	5	0,03	0,02	0,01
23,5	10	15	13	20	0,13	0,10	0,03
26,5	12	20	25	40	0,25	0,20	0,05
29,5	20	58	45	98	0,45	0,49	0,04
32,5	19	44	64	142	0,64	0,71	$\boxed{0{,}07}$
35,5	16	20	80	162	0,80	0,81	0,01
38,5	10	15	90	177	0,90	0,88	0,02
41,5	5	10	95	187	0,95	0,93	0,02
44,5	4	8	99	195	0,99	0,97	0,02
47,5	1	5	100	200	1,00	1,00	x
x	100	200	x	X	x	x	x

Підставивши в цю формулу необхідні значення, маємо:

$$\lambda^2 = 0{,}07^2 \frac{20000}{300} = 0{,}327;$$

$$\lambda = \sqrt{0{,}327} = 0{,}57.$$

§ 1.7. Перевірка гіпотез про істотність різниць між дисперсіями по F – критерію

Використання ряду критеріїв у статистичному аналізі вибіркових сукупностей ґрунтується на припущенні про рівність дисперсій порівнювальних рядів розподілів. У деяких випадках виникає необхідність обчислення оцінок різниці між дисперсіями у вибірках. При цьому істотність різниці визначають, вивчаючи або ступінь варіювання однієї і тієї ж ознаки в кількох сукупностях, або – різних ознак в одній і тій же сукупності. На оцінці різниць дисперсій побудований метод дисперсійного аналізу – один із ефективних статистико – математичних прийомів кількісного вивчення видів причинно – наслідкових залежностей.

Як відомо, перевірка статистичних гіпотез відносно середніх при нерівності дисперсій потребує оцінки істотності різниць дисперсій. Тут необхідно пам'ятати, якщо відома природа явищ, які вивчаються, і дослідник завідомо знає, що дисперсії не рівні, то оцінка значимості різниці дисперсій не має сенсу, оскільки будь – яка найменша різниця розглядається як істотна.

Оцінку істотності різниць між дисперсіями здійснюють у тих випадках, коли припускають про рівність дисперсій, у той час як інформація вибірок свідчить про велику різницю в їх величині. Наприклад, в рядах розподілу спостерігається зміщення центра розподілу частот або відрізняється сам характер розсіювання.

Перевірка статистичних гіпотез про рівність дисперсій має велике практичне значення, особливо при вирішенні технічних завдань. Це стосується випадків, коли йдеться про точність роботи приладів, машин, обладнання, про чіткість технологічних процесів і т.д. Тут показником обчисленої дисперсії вимірюється величина розсіювання.

Перевірка гіпотези про рівність дисперсій знаходить широке застосування в біологічних дослідженнях. Здійснюються практичні кроки використання подібних розрахунків і в галузі аграрно – економічних досліджень.

Перевіряється гіпотеза про рівність дисперсій за допомогою F – критерію. Останній введений як статистична оцінка англійським вченим Р.Фішером і являє собою відношення двох вибіркових дисперсій σ_1^2 і σ_2^2 при відповідних ступенях вільності ν_1 і ν_2. Саме відношення має вигляд $F = \frac{\sigma_1^2}{\sigma_2^2}$.

Оскільки в різних вибірках (дослідах) дисперсії можуть приймати різні наперед невідомі значення, розрахункова величина F – критерію (F_p) є випадковою. Як уже відомо, її називають випадковою величиною з розподілом Фішера – Снедекора. Обчислюються F – критерій за даними вибіркових сукупностей і одержане (фактичне) значення (F_p) порівнюється з табличним (F_T) при даному для кожної дисперсії числі ступенів вільності варіації ($\nu = n-1$) і заданому порозі ймовірності (додаток 8,9). До речі, стандартні таблиці F – критерію являють собою розподіл відношень $\frac{\sigma_1^2}{\sigma_2^2}$в безкінечній множинні випадкових вибірок із генеральної сукупності. У всіх випадках чисельником відношення служить більша із порівнюваних дисперсій. Із цього слідує, що F>1. За критичну область F – критерію приймається. $F_P \geq F_T$.

Виходячи з теоретичної концепції закону нормального розподілу, потрібно вважати малоймовірним виникнення як дуже великих, так і занадто малих відношень дисперсій. Значення величин $F_P \langle F_T$ вважаються випадковими (при заданому порозі ймовірності). Якщо фактичні значення F – критерію перевищують табличні, їх признають істотними (невипадковими).

При статистичній перевірці гіпотез про рівність дисперсій ($H_0 : \sigma_1^2 = \sigma_2^2$) нульова гіпотеза відхиляється, якщо $F_P \rangle F_T$ і приймається, коли F$F_p < F_T$. В останньому випадку визначається неістотною відмінність між вибірковими дисперсіями, а це вказує на те, що генеральні сукупності, з яких взята вибірка, мають однакові статистичні характеристики.

Розглянемо послідовність перевірки статистичної гіпотези на конкретному прикладі.

Приклад. У результаті механічного 10 % – го і 20 % – го відбору підприємств з генеральної сукупності отримано дві вибірки з неоднаковою чисельністю одиниць спостереження. Для перевірки гіпотези про рівність дисперсій (тобто для оцінки істотності різниці дисперсій) показників оплати людино – дня отримані такі параметри : $n_1 = 54;\ n_2 = 27;\ \sigma_1^2 = 25{,}8;\ \sigma_2^2 = 21{,}4$.

Сформулювати нульову гіпотезу (H_0) і конкуруючу (H_1) гіпотезу, маємо – $H_0 : \sigma_1^2 = \sigma_2^2;\ H_1 : \sigma_1^2 > \sigma_2^2;\ \alpha = 0{,}05$.

Фактичне значення F – критерію становитиме : $F_p = \dfrac{\sigma_1^2}{\sigma_2^2} = \dfrac{25{,}8}{21{,}4} = 1{,}21$.

Стандартне (табличне) значення F–критерію при $\nu = n - 1 = 54 - 1 = 53;$ $\nu_2 = 27 - 1 = 26$ і Р = 0,95 дорівнює (додаток 8) 1,82. Оскільки $F_P \langle F_T$, нульова гіпотеза приймається. А це значить, що дисперсія ознаки (оплати) в меншій сукупності підприємств відрізняється незначно, тобто різниця між порівнювальними дисперсіями визнається неістотною (випадковою).

Потрібно відмітити, що у тих випадках, коли конкуруюча гіпотеза формулюється, як $H_1 : \sigma_1^2 \neq \sigma_2^2$, критерій значимості буде двостороннім. У цьому випадку користуються таблицями стандартних значень з подвійним рівнем значимості. Так, щоб знайти F_p при 5 % -ному рівні значимості, відшукують значення F – критерію за таблицями з 2,5 % – ним рівнем значимості.

При вибірках, які нараховують сукупності 50 одиниць і більше , перевірка гіпотези про рівність дисперсій може бути здійснена за допомогою t – критерію нормального розподілу. Зумовлено це тим, що при малому обсязі вибірки F–критерій, утворений як відношення незалежних X^2 – розподілів, асиметричний. При збільшенні чисельності вибіркової сукупності розподіл X^2 наближається до нормального. Цілком зрозуміло, що F- розподіл, як відношення двох нормальних в границі розподілів, також буде нормальним.

Для великих вибірок рекомендується дещо інший порядок перевірки рівності дисперсій, зокрема, через критерій t.

Розглянемо послідовність розрахунків у цьому випадку: $t = \dfrac{\sigma_1 - \sigma_2}{m}$, де m- середня із помилок вибіркових середніх квадратичних відхилень.

Тут : $m_1 = \sqrt{\dfrac{\sigma_1^2}{2(n_1 - 1)}}$; $m_2 = \sqrt{\dfrac{\sigma_2^2}{2(n_2 - 1)}}$.

Звідси : $\overline{m} = \sqrt{m_1^2 + m_2^2}$.

§ 1.8. Перевірка гіпотез про істотність різниць дисперсій за критеріями Кохрана і Бартлета

Розглянуте вище відноситься до випадків перевірки статистичної гіпотези про рівність тільки двох дисперсій . У випадку необхідності отримання оцінки істотності ряду дисперсій (більше двох) використовують інші критерії. При однаковій чисельності вибіркових сукупностей використовується критерій Кохрана, при неоднакових вибірках – критерій Бартлета.

При розрахунку критерію Кохрана (q) знаходять відношення максимальної дисперсії (із порівнюваних) до суми всіх дисперсій:

$$q = \frac{\sigma_{\max}^2}{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}.$$

Одержану величину критерію Кохрана (q_p) порівнюють з табличним значення (q_T) при числі ступенів вільності : $\nu = n - 1$(додаток 10).

Якщо $q_p > q_T$, нульова гіпотеза відхиляється. Тобто дисперсії визначаються неоднорідними, оскільки їх відмінність істотна.

Із критеріїв, що використовуються для перевірки гіпотези про однорідність дисперсій, найпотужнішим визнано критерій. Бартлета (М). Як і за допомогою критерію Кохрана, критерієм Бартлета оцінюється істотність відмінності кількох дисперсій. Теоретичною основою використання даного критерію є припущення про нормальність розподілу ознак у досліджуваних сукупностях.

Суть розрахунку критерію М полягає в порівнянні зваженої середньої арифметичної і середньої геометричної із дисперсій. Якщо порівнювані дисперсії рівні, то середня арифметична і середня геометрична із дисперсій збігатимуться.

Обчислюють середню арифметичну зважену ($\overline{\sigma_a^2}$) і середню геометричну ($\overline{\sigma_\Gamma^2}$) із дисперсій, що порівнюються :

$$\sigma_a^{-2} = \frac{\sum \sigma_i^2 ni}{\sum ni}; \quad \sigma_\Gamma^{-2} = \sqrt[\sum ni]{\pi((\sigma_i^2)^{ni})}.$$

Введений в статистику критерій Бартлета (М) для перевірки рівності дисперсій являє собою відношення:

$$M = 1n\frac{\overline{\sigma_a^2}}{\overline{\sigma_\Gamma^2}}\sum n_i.$$

Після перетворення натуральних логарифмів в десяткові формула має вигляд : М=2,3026 ($\lg\overline{\sigma_a^2}\sum n_i - \sum ni \lg\overline{\sigma_\Gamma^2}$).

Якщо прийняти відношення $\frac{\text{М}}{C}$, де

$$C = 1 + \frac{\sum \frac{1}{ni} - \frac{1}{\sum ni}}{3(m-1)},$$

то його розподіл відповідає розподілу критерію Xi – квадрат (χ^2) з числом ступенів вільності, рівним Р-1 (Р – кількість дисперсій, які порівнюються). Тому критичні значення критерію М знаходять за стандартними математичними таблицями розподілу χ^2 при обраній довірчій імовірності (Р) і числу ступенів вільності $\nu = P - 1$.

Для прикладу розглянемо вибіркові сукупності господарств трьох регіонів за оплатою людино – дня. Вибірки характеризується такими даними : $n_1 = 90$; $n_2 = 100$; $n_3 = 120; \sigma_1^2 = 25{,}16 \sigma_2^2 = 24{,}10$; $\sigma_3^2 = 23{,}00$.

Для перевірки нульової гіпотези істотності відмінностей дисперсій, отриманих із неоднакових вибірок, обчислюють такі параметри:

$$\sigma_a^{-2}(25{,}16 \times 90 + 24{,}10 \times 100 + 23{,}00 \times 120) : 310 = 23{,}98;$$

$$1g\sigma_a^2 = 1g23{,}98 = 1{,}3799;$$

$$\sum ni1g\sigma_i^2 = 901g25{,}16 + 1001g24{,}10 + 1201g23{,}00 = 427{,}667;$$

$$M = 2{,}3026(427{,}769) = 0{,}2349.$$

$$C = 1 + \frac{\sum \frac{1}{ni} - \frac{1}{\sum ni}}{3(m-1)} = 1 + \frac{\frac{1}{90} + \frac{1}{100} + \frac{1}{120} - \frac{1}{310}}{3 \times 2} = 1 + \frac{0{,}0262}{6} = 1{,}00437$$

Розрахункова величина критерію Бартлета дорівнюватиме :

$$\frac{M}{C} = \frac{0{,}2349}{1{,}00437} = 0{,}234.$$

За стандартною математичною таблицею значень χ^2 при порозі імовірності Р=0,95 і числі ступенів вільності $\nu = 3 - 1 = 2$ знаходимо критичні значення $\chi^2 = 6{,}0$ (додаток 7) .

Оскільки $\chi_p^2 < \chi_T^2$ (0,234 < 6,0), робимо висновок про неістотну відмінність в дисперсіях . Тобто, відмінності вибіркових дисперсій є випадковими, а, отже, результати спостережень не суперечать гіпотезі, що перевіряється.

Резюмуючи розгляд питань про критерії згоди, необхідно пам'ятати такі особливості використання їх в аналітичній роботі. По-перше, при порівнянні емпіричного розподілу з тим чи іншим завжди мається на увазі не вибіркова сукупність, а генеральна. Вибірка тут характеризує генеральну сукупність, а тому висновки про значимість чи невірогідність відмінностей в розподілах відносяться до генеральної сукупності. По- друге, при трактовці понять “значимість” і “невірогідність” потрібно пам’ятати, що відсутність значимих розбіжностей між емпіричним і теоретичним рядами розподілу ще не значить, що емпіричний розподіл (у генеральній сукупності) в точності слідує шуканому закону розподілу. Факт відсутності значимих розбіжностей дає можливість признати емпіричну сукупність як сукупність, розподілену за відповідним законом, що не одне і те ж.

Використовуючи досить простий за своєю конструкцією критерій ламбда, потрібно дотримуватися умови його використання – достатнє число одиниць спостереження. До оцінки нечисленних вибірок критерій згоди Колмогорова неприйнятний.

ТЕМА 2. МЕТОДИ БАГАТОМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ

§ 2.1. Загальне поняття багатомірного статистичного аналізу

Впровадження ПЕОМ в управління народним господарством зумовлює перехід від традиційних методів аналізу діяльності підприємств до більш досконалих моделей управління економікою, які дозволяють розкрити її глибинні процеси.

Широке використання в економічних дослідженнях методів математичної статистики дає можливість поглибити економічний аналіз, підвищити якість інформації в плануванні і прогнозуванні показників виробництва і аналізу його ефективності.

Складність і різноманітність зв’язків економічних показників зумовлюють багатомірність ознак і у зв’язку з цим вимагають застосування найбільш складного математичного апарату – методів багатомірного статистичного аналізу.

Поняття «багатомірний статистичний аналіз» має на увазі об’єднання ряду методів, призваних дослідити поєднання взаємопов’язаних ознак. Мова йде про розчленування (розбиття) аналізованої сукупності, яка представлена багатомірними ознаками на відносно невелику їх кількість.

При цьому перехід від великої кількості ознак до меншої переслідує мету зниження їх розмірності і підвищення інформативної ємності. Така мета досягається шляхом виявлення інформації, що повторюється, породжуваної взаємопов'язаними ознаками, встановленням можливості агрегування (об'єднання, сумування) за деякими ознаками. Останнє передбачає перетворення фактичної моделі в модель з меншою кількістю факторних ознак.

Метод багатомірного статистичного аналізу дозволяє виявляти об'єктивно існуючі, але явно не виражені закономірності, що проявляються у тих чи інших соціально - економічних явищах. З цим доводиться зустрічатися при вирішенні ряду практичних завдань у галузі економіки. Зокрема, сказане має місце, якщо необхідно накопичувати (фіксувати) одночасно значення декількох кількісних характеристик (ознак) по досліджуваному об'єкту спостереження, коли кожна характеристика схильна до неконтрольованої варіації (у розрізі об'єктів), незважаючи на однорідність об'єктів спостереження.

Наприклад, досліджуючи однорідні (за природно – економічними умовами і типом спеціалізації) підприємства по ряду показників ефективності виробництва, переконуємося, що при переході від одного об'єкту до іншого майже кожна з відібраних характеристик (ідентичних) має неоднакове числове значення, тобто знаходить так би мовити неконтрольований (випадковий) розкид. Таке «випадкове» варіювання ознак, як правило, підпорядковується деяким (закономірним) тенденціям як у плані досить визначених середніх розмірів ознак, навколо яких здійснюється варіація, так і в плані ступеня і взаємозалежності самого варіювання.

Сказане вище приводить до визначення багатомірної випадкової величини як набору кількісних ознак, значення кожної з яких піддається неконтрольованому розкиду при повтореннях даного процесу, статистичного спостереження, досліду, експерименту тощо.

Раніше було сказане, що багатомірний аналіз об'єднує ряд методів; назвемо їх: факторний аналіз, метод головних компонент, кластерний аналіз, розпізнавання образів, дискримінантний аналіз та і ін. Перші три з названих методів розглядатимуться в наступних параграфах.

Як і інші математико – статистичні методи, багатомірний аналіз може бути ефективним у своєму застосуванні при умові високої якості вихідної інформації і масовості даних спостережень, які обробляються за допомогою ПЕОМ.

§ 2.2. Основні поняття методу факторного аналізу, суть вирішуваних ним завдань

При аналізі (у рівній мірі і досліджені) соціально - економічних явищ доводиться часто зустрічатися з випадками, коли серед різноманітності (багатопараметричності) об'єктів спостереження необхідно виключати частку параметрів, або замінити їх меншою кількістю тих чи інших функцій, не заподіявши шкоди цілісності (повноті) інформації. Вирішення такого завдання має сенс у рамках певної моделі і зумовлено її структурою. Прикладом такої моделі, яка найбільш наближається до багатьох реальних ситуацій, є модель факторного аналізу, методи якого дозволяють сконцентрувати ознаки (інформацію про них) шляхом «конденсації» більшого числа в менше, інформаційне більш ємне. При цьому одержаний «конденсат» інформації повинен бути представлений найбільш істотними і визначальними кількісними характеристиками.

Поняття «факторний аналіз» не треба змішувати з широким поняттям аналізу причинно – наслідкових зв'язків, коли вивчається вплив різних факторів (їх поєднань, комбінацій) на результативну ознаку.

Суть методу факторного аналізу полягає у виключенні опису множинних характеристик що вивчаються, і заміні його меншою кількістю інформаційно більш ємних змінних, які називаються факторами і відображають найбільш суттєві властивості явищ. Такі змінні є деякими функціями вихідних ознак.

Факторний аналіз, за словами Я.Окуня[9], дозволяє мати перші наближені характеристики закономірностей, що лежать в основі явища, сформулювати перші, загальні висновки про напрями, в яких потрібно вести подальше дослідження. Далі він вказує на основне припущення факторного аналізу, яке зводитися до того, що явище, не дивлячись на свою різнорідність і мінливість можна описувати невеликою кількістю функціональних одиниць, параметрів чи факторів. Ці терміни називають по - різному: вплив, причини, параметри, функціональні одиниці, здібності, основні або незалежні показники. Використання того чи іншого терміну зумовлюється

[9] Окунь Я. Факторный анализ: Пер. с. пол. М.: Статистика, 1974.- С.16.

контекстом про фактор і знанням суті явища, що вивчається.

Етапами факторного аналізу є послідовні зіставлення різних наборів факторів і варіантів групувань з їх включенням, виключенням і оцінкою вірогідності відмінностей між групами.

В.М.Жуковська і І.Б.Мучник[10], говорячи про суть завдань факторного аналізу, стверджують, що останній не вимагає апріорного підрозділу змінних на залежні і незалежні, оскільки всі змінні в ньому розглядаються як рівноправні.

Завдання факторного аналізу зводиться до певного поняття, числа і природи найбільш суттєвих і відносно незалежних функціональних характеристик явища, його вимірників або базових параметрів – факторів. На думку авторів, важливою відмінною особливістю факторного аналізу є те, що він дозволяє одночасно досліджувати велике число взаємозалежних змінних без припущення про "незмінність всіх інших умов", так необхідного при використанні ряду інших методів аналізу. У цьому велика перевага факторного аналізу як цінного інструменту дослідження явища, зумовленого складною різноманітністю і взаємопереплетенням зв'язків.

Факторний аналіз спирається в основному на спостереження над природним варіюванням змінних.

1. При використанні факторного аналізу сукупність змінних, які вивчаються з точки зору зв'язків між ними, не вибирається довільно: цей метод дозволяє виявляти основні фактори, які здійснюють істотний вплив у даній галузі.
2. Факторний аналіз не потребує попередніх гіпотез, навпаки, він сам може служити методом висунення гіпотез, а також виступати критерієм гіпотез, що спираються на дані, одержані іншими методами.
3. Факторний аналіз не потребує апріорних здогадок відносно того, які змінні незалежні, а які залежні, він не гіпертрофує причинні зв'язки і вирішує питання про їх міру в процесі подальших досліджень.

Перелік конкретних завдань, що вирішуються з використанням методів факторного аналізу буде таким (за В.М.Жуковською). Назвемо основні з них в галузі соціально-економічних досліджень:

---

[10] Жуковская В.М., Мучник И.Б. Факторный анализ в социально-економических исследованиях. - М.- Статистика, 1976. С.4.

1. Визначення основних аспектів різниць між об'єктами спостереження (мінімізація описання).
2. Формулювання гіпотез про природу різниць між об'єктами.
3. Виявлення структури взаємозв'язків між ознаками.
4. Перевірка гіпотез про взаємозв'язки та взаємозамінності ознак.
5. Зіставлення структур наборів ознак.
6. Розчленовування об'єктів спостереження за типовими ознаками.

Викладене свідчить про великі можливості факторного аналізу в

дослідженні суспільних явищ, де, як правило, неможливо проконтролювати (експериментально) вплив окремих факторів.

Досить ефективним є використання результатів факторного аналізу в моделях множинної регресії.

Маючи попередньо сформовану кореляційно-регресійну модель досліджуваного явища у вигляді корельованих ознак, за допомогою факторного аналізу можна такий набір ознак перетворити в значно меншу їх кількість шляхом агрегування. При цьому слід відмітити, що таке перетворення ні в якій мірі не погіршує якість і повноту інформації про досліджуване явище. Створені агреговані ознаки некорельовані і являють лінійну комбінацію первинних ознак. З формальної математичної сторони постановка завдань у такому випадку може мати нескінчену множинну рішень. Але потрібно пам'ятати, що при вивченні соціально – економічних явищ одержані агреговані ознаки повинні мати економічно обґрунтоване трактування. Інакше кажучи, в будь - якому випадку використання математичного апарату в першу чергу виходять зі знань економічної суті досліджуваних явищ.

Таким чином, сказане вище дозволяє резюмувати, що факторний аналіз є специфічним методом дослідження, який здійснюється на базі арсеналу прийомів математичної статистики.

Своє практичне застосування факторний аналіз вперше знайшов в галузі психології. Можливість звести велику кількість психологічних тестів до невеликої кількості факторів дало змогу пояснити здібності людського інтелекту.

При дослідженні соціально-економічних явищ, де є труднощі в ізолюванні впливу окремих змінних, успішно може бути використаний факторний аналіз. Застосування його прийомів дозволяє шляхом певних розрахунків "профільтрувати " неістотні ознаки і продовжити дослідження в напрямку його поглиблення.

Ефективність цього методу очевидна при дослідженні таких питань (проблем): в економіці - спеціалізація і концентрація виробництва, інтенсивність ведення господарства, бюджет сімей працівників, побудова різних узагальнюючих показників. і т.ін.

§ 2.3. Математичні основи теорії факторного аналізу . Матриця даних

Математичним аспектам сучасного факторного аналізу присвячено ряд робіт (монографій) вітчизняних і зарубіжних авторів. Як вид багатомірного статистичного аналізу, він, має свої сильні і слабкі сторони. Щоб мати уяву про це, необхідно поставити завдання дослідження теоретичних аспектів факторного аналізу, як науки.

В даному параграфі будуть розглянуті лише деякі основні елементи даного статистико – математичного методу в той мірі, в якій це необхідно для розуміння його практичної реалізації в економічних розрахунках.

Перш за все тут потрібний деякий обсяг елементарних математичних знань з алгебри, аналітичної геометрії і тригонометрії, а також розділів математики, у яких висвітлюються питання матричної алгебри.

Економіко – математичні моделі, що використовуються сьогодні в різного роду економічних розрахунків, часто призначені для опису взаємозв’язку економічних структур, їх зміни в часі і просторі, залежності від ряду факторів і т.д. Вважається, що одним з найкомпактніших способів опису таких структур (інколи дуже складених) є спосіб матричного відображення.

Фактор, як розрахункова змінна у факторному аналізу, являє собою деяку нову характеристику об’єктів досліджуваної множини. Розгляд (опис) його через призму набору вихідних ознак досягається шляхом розрахунку так званої матриці. У факторному аналізі загальновживаний термін «матриця факторів» або «факторна матриця». Інколи її називають «матрицею факторних навантажень»

Інакше, матриця – це прямокутний масив чисел, розташованих за рядками і стовпчиками. Така форма подання числових даних є зручною для їх математичної обробки. Перевага матричної форми запису полягає в тому, що в невеликому наборі символів не мов спресована множина математичних операцій, що надзвичайно зручно при аналізі масових даних. Перш за все перевага виявляється на початковому етапі обробки інформаційного масиву, тобто на етапі організації вихідних даних. Наприклад, розглянемо дані про середньорічні надої залежно від рівня годівлі корів і року їх лактації (табл. 108).

Таблиця 108

Середньорічний надій корів в залежності від лактації і рівня годівлі, кг

Рік лактації	Витрати кормів на 1 голову, ц.корм.од.		
	35-40	40-45	45 і вище
1	3500	3700	3900
2	3600	3900	4300
3	4100	4300	4700
4	4500	5000	5500

Наведену табличну інформацію можна записати у такій формі:

$$\begin{matrix} 3500 & 3700 & 3900 \\ 3600 & 3900 & 4300 \\ 4100 & 4300 & 4700 \\ 4500 & 5000 & 5500 \end{matrix}$$

Змістовне значення кожного показника визначається його місцем в даному масиві. Так, число 4300 в третьому рядку і другому стовпчику являє собою надій корів на третьому році лактації при рівні годівлі 40-45 ц кормових одиниць в рік. Аналогічно визначаємо, що числа, які занесені у стовпчик, характеризують надій при однаковому рівні годівлі, але в різні лактаційні періоди, а числа, які записані в рядок, – надій корів одного і того ж року лактації, але при різному рівні годівлі. Отже місце, що займає число в масиві, характеризує період лактації і рівень годівлі, до якого відноситься надій.

Як бачимо, елементи рядка чи кожного стовпчика володіють якоюсь загальною властивістю. Елементами матриці. можуть бути числа будь - якого виду чи навіть функцій однієї чи кількох змінних. Над числовими масивами такого типу можуть бути виконані різні алгебраїчні операції. При цьому виходять з припущення, що кожний масив є єдине ціле і позначається одним символом. У подальших міркуваннях будемо виходить з того, що елементи матриці є дійсні числа – додатні, від’ємні або нуль.

Доречно буде нагадати, що під матрицю розуміють прямокутну або квадратну таблицю чисел, яку розглядають безвідносно до існуючих (чи неіснуючих) зв'язків між ними. Тобто є числа, розташовані у вигляді рядків і стовпчиків, мають формальну математичну схему, байдужну до природи заздалегідь визначених залежностей між цими числами.

Якщо позначити матрицю факторних навантажень через А, число ознак через *n,* число факторів – через *m,* то порядок або розмірність матриці складе $n \times m$. Квадратна матриця, $(n \times n)$ має розмірність *n*.

Елементи матриці позначаються двома підрядковими індексами, перший з яких означає номер рядка, другий – номер стовпчика. Наприклад, символ x_{32} вказує на елемент, що знаходиться на перетині третього рядку і другого стовпчика .

$$\begin{vmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{vmatrix} = A \,.$$

Схема квадратної матриці розмірністю 3×3

У загальному вигляді матриця записується такому вигляді

$$\begin{vmatrix} x_{11} & x_{12} & x_{13} \ldots & x_{1n} \\ x_{21} & x_{22} & x_{23 \ldots\ldots} & x_{2_n} \\ \ldots & \ldots & x_{ij} & \ldots \\ x_{m_1} & x_{m_2} & x_{m_3} & x_{mn} \end{vmatrix} .$$

Схема загального виду матриці

Загальний елемент матриці x_i означає, що індекс рядка $|i|$ може приймати послідовне значення 1,2,3,...,*m*, а індекс стовпчика $|j|$- послідовне значення 1,2,3, … *n.* Матриця А може бути виражена через загальний її елемент x_{ij}.

Якщо рядки матриці стають стовпчиками, ми одержуємо нову матрицю, яка буде транспонована по відношенню до матриці А. На схемі наведена матриця А і її транспонування (матриця A').

$$\underset{A}{\begin{vmatrix} 6 & 2 & 0 & 3 \\ 4 & 3 & 2 & 5 \\ 5 & 1 & 0 & 6 \\ & & & \end{vmatrix}} \qquad \underset{A'}{\begin{vmatrix} 6 & 4 & 5 \\ 2 & 3 & 1 \\ 0 & 2 & 0 \\ 3 & 5 & 6 \end{vmatrix}}$$

Схема транспонування матриці А

Мають місце випадки, у факторному аналізі, коли квадратна матриця збігається з транспонованою до неї матрицею, тоді матрицю А називають симетричною.

$$\begin{vmatrix} 6 & 2 & 1 \\ 2 & 6 & 4 \\ 1 & 4 & 8 \end{vmatrix} = A = A'.$$

Схема симетричної матриці.

У факторному аналізі має місце, як правило, робота з симетричними матрицями. Прикладом такого виду може бути матриця, елементи якої представлені коефіцієнтами кореляції змінних сукупності, що вивчається.

Алгебраїчні операції над числовими масивами, поданими у вигляді матриць, вивчає матрична алгебра, предметом якої є дії не з окремими елементами, а безпосередньо з масивами. Останні виступають як відокремлені і цілісні системи.

Основні арифметичні дії з матрицями – це додавання, віднімання і множення.

Розглянемо порядок здійснення названих арифметичних операцій.

Додавання матриць. Наприклад, є дані про виробництво продукції підприємством за липень (табл.109).

Таблиця 109

Обсяг виробництва продукції підприємством за липень, т

Вид продукції	Номер бригади		
	1	2	3
0_1	66	72	24
0_2	42	63	60
0_3	30	50	40

Зміст таблиці в матричній формі з розмірністю 3×3, буде таким:

$$A = \begin{vmatrix} 66 & 72 & 24 \\ 42 & 63 & 60 \\ 30 & 50 & 40 \end{vmatrix}.$$

Маючи аналогічні дані про обсяг виробництва продукції в наступному місяці (серпень), запишемо їх у такому ж порядку у вигляді матриці:

$$B = \begin{vmatrix} 34 & 42 & 30 \\ 30 & 50 & 25 \\ 26 & 40 & 42 \end{vmatrix}.$$

Загальний обсяг виробництва по виду продукції 0_1, виробленої в першій бригаді, дорівнює сумі елементів, розташованих в кожній матриці на перетині першого рядка і першого стовпчика: 66+34=100, а загальний обсяг виробництва продукції 0_2 за два місяці по третій бригаді становить 60+25=85. Матриця в такому випадку матиме вигляд :

$$\begin{vmatrix} 66+34 & 72+42 & 24+30 \\ 42+30 & 63+50 & 60+25 \\ 30+26 & 50+40 & 40+42 \end{vmatrix} = \begin{vmatrix} 100 & 114 & 54 \\ 72 & 113 & 85 \\ 50 & 90 & 82 \end{vmatrix}.$$

Числа, одержані в результаті додавання елементів двох матриць, характеризують (у даному випадку) обсяг виробництва різних видів продукції по бригадах за два місяці. Таким чином, сума двох матриць являє собою матрицю, кожний елемент якої є сума відповідних елементів матриць – доданків.

Додавання матриць можливо лише у випадку однакової їх розмірності ($m \times n$), тобто однаковій кількості рядків (m) і однаковій кількості стовпчиків (n). Такі матриці називають узгодженими для додавання.

Додавання матриць $\begin{vmatrix} 1 & 5 & 5 \\ 3 & 2 & 4 \\ & & \end{vmatrix}$ і $\begin{vmatrix} 2 & 6 \\ 8 & 5 \end{vmatrix}$ не має змісту

Якщо прийняти позначення $A = \{a_{ij}\}$ і $B = \{в_{ij}\}$, то сумою матриць А і В буде матриця $A + B = \{a_{ij} + в_{ij}\}$.

Віднімання матриць. Подамо дані про виробництво продукції за серпень за розглянутим вище прикладом з додаванням матриць у вигляді таблиці 110.

Таблиця 110

Обсяг виробництва продукції підприємством за серпень, т

Вид продукції	Номер бригади		
	1	2	3
0_1	34	42	30
0_2	30	50	25
0_3	26	40	42

У матричному відображенні зміст таблиці матиме вигляд:

$$B = \begin{vmatrix} 34 & 42 & 30 \\ 30 & 50 & 25 \\ 26 & 40 & 42 \end{vmatrix}.$$

Подамо обсяг виробленої продукції підприємством за липень також у вигляді матриці:

$$A=\begin{vmatrix}66 & 72 & 24\\ 42 & 63 & 60\\ 30 & 50 & 40\end{vmatrix}.$$

Щоб визначити зміну в обсязі виробництва продукції 0_1 в серпні порівняно з липнем, наприклад, в бригаді 3, необхідно від 30 відняти 24; по продукції 0_2 відповідно 25-60 і т.д.

Подамо різницю в обсязі виробництва за два місяці в розрізі бригад і видів продукції в матричному відображенні:

$$\begin{vmatrix}34-66 & 42-72 & 30-24\\ 30-42 & 50-63 & 25-60\\ 26-30 & 40-50 & 42-40\end{vmatrix}=\begin{vmatrix}-32 & -30 & 6\\ -12 & -13 & -35\\ -4 & -10 & 2\end{vmatrix}.$$

Одержана матриця характеризує різницю в обсягах виробництва продукції різних видів у серпні порівняно з липнем по всіх бригадах. Проведені розрахунки ілюструють приклад віднімання матриць. Схематично це виглядає на умовному прикладі так:

$$\begin{vmatrix}2 & 3\\ 6 & 8\\ 4 & 6\end{vmatrix}-\begin{vmatrix}3 & 4\\ 5 & 8\\ 2 & 3\end{vmatrix}=\begin{vmatrix}-1 & -1\\ 1 & 0\\ 2 & 3\end{vmatrix} \text{ або } B-A=\{в_{ij}-a_{ij}\}.$$

Здійснювати операцію віднімання можна тільки із матрицями, що мають однакову розмірність (порядок). Інакше кажучи, матриці узгоджені для додавання, узгоджені і для віднімання (і навпаки).

Множення матриць. Правило множення матриць випливає з правила їх додавання. Якщо подати : $A+A=\{a_{ij}\}+\{a_{ij}\}=\{2x_{ij}\}=2A$, то для випадку, коли S – ціле додатне число, можна записати у вигляді SA= A+A+A+…+A.

Згідно з правилом додавання матриць одержимо:

$$SA=\{Sx_{ij}\}.$$

Для скалярних величин (в даному випадку S) і матриць добуток виглядатиме як матриця, в якій кожний елемент помножено на S.

Наприклад, при S= 6 і матриці $\begin{vmatrix}2 & 6\\ -3 & 9\end{vmatrix}$ добуток матиме вигляд:

$$6\begin{vmatrix}2 & 6\\ -3 & 9\end{vmatrix}=\begin{vmatrix}12 & 36\\ -18 & 54\end{vmatrix}.$$

При множенні матриці на матрицю необхідно уявити многократне множення матриці на вектори. Так, при множенні матриці А на матрицю В подаємо матрицю В як набір векторів – стовпчиків. В такому випадку добуток АВ виглядатиме як послідовно записані один за одним добутки матриці А на кожний вектор – стовпчик, що утворює В.

Приклад. Маємо дві матриці:

$$A = \begin{vmatrix} 2 & 1 & 3 \\ 1 & -2 & 1 \\ 3 & 1 & 1 \\ 4 & 3 & 2 \end{vmatrix}; \quad B = \begin{vmatrix} -1 & 2 \\ 0 & 1 \\ 2 & 2 \end{vmatrix}.$$

Матрицю В подамо у вигляді двох векторів – стовпчиків:

$$x = \begin{vmatrix} -1 \\ 0 \\ 2 \end{vmatrix} \text{ і } y = \begin{vmatrix} 2 \\ 1 \\ 2 \end{vmatrix}.$$

Помноживши матрицю А на кожний з векторів – стовпчик матриці В, маємо;

$$Ax = \begin{vmatrix} 2x/-1/+ & 1\times 0 & + & 3\times 2 \\ 1x/-1/+ & /-2/\times 0 & + & 1\times 2 \\ 3x/-1/+ & 1\times 0 & + & 1\times 2 \\ 4x/-1/+ & 3\times 0 & + & 2\times 2 \end{vmatrix} = \begin{vmatrix} 4 \\ 1 \\ -1 \\ 0 \end{vmatrix};$$

$$Ay = \begin{vmatrix} 2\times 2+ & 1\times 1 & + & 3\times 2 \\ 1\times 2+ & /-2/\times 1 & + & 1\times 2 \\ 3\times 2+ & 1\times 1 & + & 1\times 2 \\ 4\times 2+ & 3\times 1 & + & 2\times 2 \end{vmatrix} = \begin{vmatrix} 11 \\ 2 \\ 9 \\ 15 \end{vmatrix}.$$

Розмістивши вектори один за одним, маємо добуток матриці АВ:

$$AB = \begin{vmatrix} 4 & 11 \\ 1 & 2 \\ -1 & 9 \\ 0 & 15 \end{vmatrix}.$$

Проведені операції в повному їх вигляді мають вид :

$$AB = \begin{vmatrix} 2 & 1 & 3 \\ 1 & -2 & 1 \\ 3 & 1 & 1 \\ 4 & 3 & 2 \end{vmatrix} \times \begin{vmatrix} -1 & 2 \\ 0 & 1 \\ 2 & 2 \end{vmatrix} = \begin{vmatrix} 4 & 11 \\ 1 & 2 \\ -1 & 9 \\ 0 & 15 \end{vmatrix}.$$

До одержаного результату можна дійти іншим шляхом, послідовно перемноживши елементи матриць А і В, рухаючись по горизонталі впродовж і – того рядка матриці А і одночасно - вниз по j – тому стовпчику матриці В, потім додавши між собою всі ці добутки.

Сума добутків відповідних елементів утворює іj – й елемент матриці – добутку АВ.

Діючи за такою схемою, послідовно множимо елементи третього рядка матриці А на елементи першого стовпчика матриці В:

$$3\times(-1)+1\times 0+1\times 2=-3+0+2=-1.$$

Таким чином, елемент, що стоїть в третьому рядку матриці А і першому стовпчику матриці В, дорівнює –1.

Добуток матриць Аі В має зміст лише у тому випадку, коли j – й стовпчик матриці В (отже, і всі інші стовпчики) налічує таку ж кількість елементів, що й і-й рядок матриці А (отже, і всі інші її рядки).

Таким чином, в матриці В повинно бути стільки ж рядків, скільки стовпчиків має матриця А. Отже, добуток матриць А і В визначено тільки у випадку, коли число стовпчиків в матриці А дорівнює числу рядків в матриці В. Такі матриці називають узгодженими для множення А на В.

Приклад. Маємо матриці:

$$A = \begin{vmatrix} 0 & 3 \\ 6 & 2 \end{vmatrix} \;;\; B = \begin{vmatrix} 1 & 2 & 5 \\ 4 & 3 & 8 \end{vmatrix}.$$

Оскільки А містить два стовпчики, а В- два рядки, існує добуток АВ.

$$AB = \begin{vmatrix} 0 & 3 \\ 6 & 2 \end{vmatrix} \times \begin{vmatrix} 1 & 2 & 5 \\ 4 & 3 & 8 \end{vmatrix} = \begin{vmatrix} 12 & 9 & 24 \\ 14 & 18 & 27 \end{vmatrix}.$$

Розмірність (порядок) добутку АВ складає 2×3. У даному випадку добуток ВА не існує, бо матриця В містить три стовпчики, а матриця А – два рядки.

Згідно з законом матричної алгебри перед виконанням операцій додавання, віднімання і множення необхідно з'ясувати питання про узгодженість матриць.

Крім розглянутих вище з матрицями, існують й інші. Це- множення векторів, множення на діагональну матрицю, лінійні перетворення, трансформування матриць і т.д. Названі операції докладно викладені в курсі матричної алгебри.

У факторному аналізі найчастіше зустрічаються такі види матриць: діагональна, скалярна, одинична, обернена.

Діагональна матриця – квадратична матриця, яка відрізняється від нуля тільки елементами, що лежать на діагоналі, яка зв'язує верхній лівий край з правим нижнім краєм матриці (головна діагональ).

Скалярна матриця – діагональна матриця, в якій всі елементи рівні між собою.

Одинична матриця – діагональна матриця, в якій всі елементи головної діагоналі дорівнюють одиниці. Цей вид матриці виконує в матричній алгебрі ту ж роль, що й одиниця в арифметиці.

Обернена матриця . Подавши найпростішу арифметичну залежність у вигляді $x \cdot \frac{1}{x} = 1,$ яка означає, що добуток будь - якого числа на обернене йому число дорівнює одиниці, можна знайти аналогічний зв’язок і в матричній алгебрі. Якщо матриця А квадратна, то існує матриця, обернена їй A^{-1}, або $A \cdot A^{-1} = 1.$

Спосіб обчислення оберненої матриці досить складний. Названі види матриць схематично представлено:

$$\begin{vmatrix} a & 0 & 0 & 0 \\ 0 & в & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{vmatrix} \begin{vmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{vmatrix} \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

а в с

Схеми видів матриць, які зустрічаються у факторному аналізі
(а- діагональна, в – скалярна, с- одинична)

§ 2.4. Формування вихідної інформації і факторних моделей

Встановивши вид матриці вихідних даних, приступають до формування інформаційного масиву. На цьому етапі визначають перелік змінних і об’єктів спостереження. Відбір ознак (змінних)і об’єктів спостереження є дуже важливим етапом роботи і виконується в суворій відповідності з метою дослідження. Якість інформації тут значною мірою залежить від знань природи явищ і ознак, які мають найбільш істотну інформацію про це явище. Важливими критеріями відбору при формуванні інформаційного масиву є точність, вірогідність і зіставність даних. При цьому ознаки не повинні різнитися за змістом і колом обстежених об’єктів (у регіональному плані і періоді часу). Недопустимим слід вважати включення в масив лінійно взаємопов’язаних за способом розрахунку ознак.

При формуванні об’єктів спостережень дається списковий їх перелік, який включає назву підприємств, районів, областей і т.ін., а також показників, які характеризують зміну явищ у часу. На етапі формулюванні мети дослідження вказується, що підлягає обстеженню – генеральна чи вибіркова сукупність. Об’єкти спостереження повинні бути однорідні за складом, тобто типові.

На наступному етапі приступають до перетворення вихідної інформації. Під перетворенням вихідних даних у факторному аналізі мають на увазі зміну характеру емпіричного розподілу відповідно до мети і завдання дослідження, зокрема, соціально – економічних явищ.Перетворення вихідної інформації повинно забезпечити: компенсацію можливих помилок у вихідних даних; зіставність значень ознак; невелювання впливу на результати розрахунків тих ознак, які мають значне відхилення від середніх їх значень.

Неоднакова розмірність ознак, яка заважає їх зіставності, вимагає зведення до єдиного масштабу. Досягається це шляхом трансформування (нормування) матриці вихідних даних. Серед існуючих різних способів трансформування матриці часто використовують таку стандартну формулу:

$$S_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j},$$

де S_{ij} – стандартні значення ознак;

x_{ij} - величина j –ї ознаки в і –му об’єкті;

$\bar{x}_j$ - середня j –ї ознаки;

σ_j - середнє квадратичне відхилення (стандарт) j – ї ознаки.

Інколи замість значення середнього квадратичного відхилення (σ_j) використовують показник розмаху варіації (R), тоді формула матиме вигляд :

$$S_{ij} = \frac{x_{ij} - \bar{x}_j}{R}.$$

Можна привести до стандартної форми ознаки, які мають різну розмірність, через таке відношення :

$$S_{ij} = \frac{x_{ij}}{x_y} \text{ і т.д.}$$

Слід пам’ятати, що перетворення матриці здійснюється завжди єдиним способом. Нормування ознак в окремих рядах може бути реалізоване шляхом підбору різних видів перетворення. Це зумовлюється постановкою завдання дослідження того чи іншого явища.

Після завершення етапів формування і перетворення вихідної інформації приступають до формування самих моделей факторного аналізу. Всі аналізовані ознаки, що входять в той чи інший об’єкт, кореляційно залежні. При цьому кореляція може проявлятися в одному випадку як вплив однієї ознаки на інші, а в другому – як вплив ознаки, що не входить в досліджувані фактори. Такі “ скриті” ознаки у факторному аналізу називають загальними факторами. Вивчення впливу цих факторів є основною метою факторного аналізу.

Таким чином, методи факторного аналізу повинні описати з усього набору факторів невелику кількість основних внутрішніх параметрів. Тобто, виходимо з припущення, що кожну з ознак досліджуваного набору можна показати, як функцію невеликої кількості загальних факторів і характерного фактору. Якщо позначити загальні фактори через $h_1.h_2.h_3...h_n$, а специфічний фактор через Z_j то в стандартній формі кожну з ознак (S_j) можна подавати у вигляді функції:

$$S_j = f(h_1, h_2, h_3, ..., h_n Z_j)\,.$$

Тут важливо з’ясувати змістовне наповнення понять “загальний фактор” і “ характерний фактор”. Перше поняття орієнтує

що кожний із факторів $(h_{1,2,3,...,n})$ має важливе значення при вивченні всіх змінних $(S_{1,2,3,...n})$. Тоді, як під другим поняттям “характерний фактор” розуміють, що зміна в факторі Z_j спричиняє зміну значення тільки однієї відповідної ознаки S_j .

Як відзначалося вище, аналізовані ознаки об’єктів, що вивчаються кореляційно залежні. Із показників кореляції у факторному аналізі найчастіше використовують коефіцієнти кореляції (r), тобто показник ступеня тісноти зв’язку для лінійних залежностей. Тут доречним буде нагадати, що показник ступеня тісноти зв’язку (η) для нелінійних (криволінійних) залежностей у факторному аналізі не використовується. Зумовлено це обмеженням основного припущення про лінійну залежність.

§ 2.5. Виділення факторів і визначення їх навантажень

Вихідним початком факторного аналізу є складання кореляційної матриці, а його метою – побудова факторної матриці. Отже, розв’язується завдання виділення факторів. Серед існуючих способів розв’язання цього завдання найпростішим і загальним методом виділення факторів є так званий центроїдний метод. При розгляді конкретного прикладу виділення факторів і визначення їх навантажень будемо користуватися названим вище методом без викладання його теоретичних аспектів. Останні розглядаються в спеціальній математичній літературі.

На початковому етапі виділення факторів складається матриця коефіцієнтів кореляції. Організувавши редуковану кореляційну матрицю, переходять до редукованої факторної матриці. Остання повинна показувати кількість загальних факторів, відображуючи кореляцію між змінними, які вивчаються. Тут число загальних факторів відповідає числу стовпчиків редукованої факторної матриці. По цій же матриці маємо навантаження кожного фактору для тієї чи іншої змінної. Це – рядки факторної матриці.

Згідно з існуючою теоремою, редукована матриця кореляції дорівнює добутку редукованої факторної матриці на транспоновану. Схематично це має такий вигляд:

$$\underbrace{\begin{array}{c|cc|} & 1 & 2 \\ 1 & C_{11} & C_{12} \\ 2 & C_{21} & C_{22} \\ 3 & C_{31} & C_{32} \end{array}}_{F} \times \underbrace{\begin{array}{c|ccc|} & 1 & 2 & 3 \\ 1 & C_{11} & C_{21} & C_{31} \\ 2 & C_{12} & C_{22} & C_{32} \end{array}}_{F'} = \underbrace{\begin{array}{c|ccc|} & 1 & 2 & 3 \\ 1 & h^2 & R_{12} & R_{13} \\ 2 & R_{21} & h_2^2 & R_{23} \\ 3 & R_{31} & R_{32} & h_3^2 \end{array}}_{R}$$

З наведеної залежності (R=F·F′) випливає рівняння, яке має важливе практичне значення, що дозволяє встановити кореляцію на підставі факторних навантажень. Наприклад, якщо маємо *n* некорольованих факторів С, загальних для змінних *а* і *в*, то кореляція межі *а* і *в* ($r_{\text{ав}}$) дорівнює сумі добутків навантажень кожного з факторів на ці змінні:

$$r_{ав} = r_a C_1 r_в C_1 + r_a C_2 r_в C_2 + ... + r_a C_n r_в C_n \text{ де}$$

$r_a C_1 r_в C_1$ - навантаження фактора C_1 при змінних *а* і *в*;

$r_a C_2 r_в C_2$ - навантаження фактора C_2 при змінних *а* і *в*;

$r_a C_n r_в C_n$- навантаження *n* – го фактора, загального для обох змінних.

Наведене вище рівняння дозволяє визначити кореляцію між двома змінними, якщо відомі навантаження загальних для цих змінних факторів. У практичних розрахунках завжди вирішується протилежне завдання: визначити факторні навантаження на підставі існуючих кореляцій.

Якщо припустити існування загального фактора C_1 при відомих кореляціях змінної α з трьома іншими змінними $в,c,d$, то кожна з змінних буде характеризуватися навантаженням загального фактора такими рівняннями:

$$r_{ав} = (r_a C_1) \times (r_в C_1);$$
$$r_{ac} = (r_a C_1) \times (r_c C_1);$$
$$r_{ad} = (r_a C_1) \times (r_d C_1).$$

Як бачимо, у правій части наведених рівнянь існує однаковий параметр $r_a C_1$. У цьому зв'язку існуюча теорема свідчить, що середня кореляція змінної з іншими змінними, розрахованаі з суми всіх кореляцій (у стовпчику), пропорційна кореляції цієї змінної з загальним фактором $r_a C_1$.

У практичних розрахунках середня кореляція розраховується шляхом ділення суми елементів одного стовпчика на корінь квадратний з суми всіх стовпчиків матриці. У цьому і полягає суть виділення факторів за матрицею парних кореляційних залежностей.

Приклад. У кореляційно-регресійну модель урожайності зернових культур (y) включено шість факторів :затрати праці на 1 га зернових (x_1); вартість основних виробничих фондів в розрахунку на 1 гектар ріллі (x_2); матеріально – грошові затрати виробництва з розрахунку на 1 гектар зернових (x_3); виробництво зерна на 1 людино - годину (x_4); вартість основних виробничих фондів з розрахунку на одного працівника рослинництва (x_5); оплата 1 людино - години в зерновому господарстві (x_6) .

Як бачимо, поставлене аналітичне завдання: одержати кількісну характеристику змін урожайності під впливом факторів інтенсифікації виробництва (x_1, x_3), фондооснащеності (x_2) і фондоозброєності (x_5), продуктивності праці (x_4) та її оплати (x_6). У вибірку включено 57 одиниць спостереження.

У результаті обробки вихідної інформації на ПЕОМ одержано кореляційна матриця: (табл. 111).

Таблиця 111

Матриця вихідних коефіцієнтів кореляції

Зміни	1	2	3	4	5	6
1	1.000	0,175	0,136	-0,659	0,073	-0,191 1
2		1,000	0,045	-0,114	0,257	0,035 2
3			1,000	-0,152	0,117	0,164 3
4				1,000	-0,059	0,383 4
5					1,000	-0,088 5
6						1,000 6

Величина одержаного множинного коефіцієнта кореляції (R) по досліджуваній моделі становить 0,761. Початку пошуку загального для всіх змінних фактора передує побудова редукованої кореляційної матриці (табл.111).

По головній діагоналі цієї матриці заносяться величини максимальних значень коефіцієнтів кореляції у стовпчику (без врахування алгебраїчних знаків). На наступному етапі розраховують навантаження першого загального фактора. З цією метою виконують такі обчислювальні операції:

а) відшукують суми параметрів по стовпчиках з врахуванням алгебраїчних знаків;

б) визначають суми сум стовпчиків. У нашому випадку ця величина (Т) становить 2,623. Потім обчислюють її корінь квадратний: $\sqrt{T} = 1{,}61957$;

в) одержані по стовпчиках суми ділять на $\sqrt{T}$, маючи, таким чином, навантаження першого фактора для шести змінних, тобто - їх кореляцію з досліджуваним фактором. У символіці навантаження першого фактора C_1 для змінної α має такий вигляд:

$$C_{1a} = \frac{\sum r_a}{\sqrt{T}}.$$

Ця характеристика записана в останньому рядку таблиці 111;

г) як критерій правильності розрахунків використовують додатково обчислену величину $\frac{1}{\sqrt{T}}$. У нашому прикладі її значення дорівнює 0,61745. Як бачимо, при умові правильності розрахунків $T \cdot \frac{1}{\sqrt{T}} = \sqrt{T}$. У прикладі : $2{,}623 \times 0{,}61745 = 1{,}61957$, що повністю збігається з розрахованою величиною $\sqrt{T}$.

Другим критерієм правильності розрахунків є сума всіх факторних навантажень . Її величина повинна також дорівнювати $\sqrt{T}$. У наших розрахунках $\sum C_1 = 1{,}620$ при $\sqrt{T}$ =1,61957. Розрахунком розглянутих критеріїв завершується аналіз редукованої кореляційної матриці з метою визначення навантажень першого, загального для всіх змінних фактора (табл.112).

Таблиця 112

Редукована кореляційна матриця досліджуваних шести змінних (Р)

Змінні	P_1	P_2	P_3	P_4	P_5	P_6	$\sum r$
P_1	**0,659**	0,175	0,136	-0,659	0,073	-0,191	0,193
P_2	0,175	**0,257**	0,045	-0,114	0,257	0,035	0,655
P_3	0,136	0,045	**0,164**	-0,152	0,117	0,164	0,474
P_4	-0,659	-0,114	-0,152	**0,659**	-0,059	0,383	0,058
P_5	0,073	0,257	0,117	-0,059	**0,257**	-0,088	0,557
P_6	-0,191	0,035	0,164	0,383	-0,088	**0,383**	0,686
$\sum r$	0,193	0,655	0,474	0,058	0,557	0,686	**2,623**
C_1	0,119	0,404	0,293	0,036	0,344	0,424	**1,620**

Т=2,623; $\sqrt{T}$ = 1,61957; $\frac{1}{\sqrt{T}} = 0{,}61745$.

Критерій $T \cdot \frac{1}{\sqrt{T}} = 1{,}61957$.

Критерій $\sum C_1 = 1{,}620$.

Для виділення навантажень решти факторів виходять з теоретичної концепції (існує теорема) про те, що кореляція двох змінних, викликана яким – небудь загальним для них фактором, дорівнює добутку навантажень цього фактора для обох змінних, тобто добутку їх кореляцій с цим фактором. Так, кореляція між першою і другою змінними P_1 і P_2, зумовлена першим фактором, являє собою добуток його навантажень по першій і другій змінних. Виходячи з наведених вище розрахунків, маємо:

$$r_{p_1p_2} = r_{P_1C_2} \times r_{P_2C_1} = 0.119 \times 0.404 = 0.048 .$$

Одержаний за розрахунками коефіцієнт кореляції між змінними P_1 і P_2 дорівнює 0,175. Щоб визначити частину дисперсії, яка може бути зумовлена іншими факторами, знаходять так званий “залишок ” шляхом віднімання з початкового коефіцієнта кореляції між змінними ($r_{p_1p_5}$ = 0,175) величини коефіцієнта кореляції, зумовленої першим фактором ($r_{p_1p_5}$ = 0,048) Тоді маємо 0,175-0,048=0,127.

У випадку одержання від’ємного залишку слід пам’ятати, що навантаження досліджуваного фактора у відповідних змінних мають від’ємний знак.

За даними нашого прикладу, для кожної пари змінних знаходимо: різницю між значенням початкових коефіцієнтів кореляції і добутком факторних навантажень:

$$r_{p_1p_1} = 0.659 - 0.119 \times 0.119 = 0.645 ;$$

$$r_{p_1p_2} = 0.175 - 0.119 \times 0.404 = 0.127 ;$$

$$r_{p_1\ p_3} = 0.136 - 0.119 \times 0.293 = 0.101 ;$$

$$r_{p_1p_4} = (-0.659) - 0.119 \times 0.036 = -0.663 ;$$

$$r_{p_1p_5} = 0.073 - 0.119 \times 0.344 = 0{,}032 ;$$

$$r_{p_1p_6} = (-0.191) - 0.119 \times 0.424 = -0.241.$$

Подібні розрахунки зручніше здійснювати у вигляді робочих таблиць. При цьому слід враховувати алгебраїчні знаки (табл. 113 і 114).

Таблиця 113

Матриця добутків факторних навантажень

Змінні	Факторні навантаження	Змінні					
		P_1	P_2	P_3	P_4	P_5	P_6
		0,119	0,404	0,293	0,036	0,344	0,424
P_1	0,119	0,014					
P_2	0,404	0,048	0,163				
P_3	0,293	0,035	0,118	0,086			
P_4	0,036	0,004	0,015	0,011	0,001		
P_5	0,344	0,041	0,139	0,101	0,012	0,118	
P_6	0,424	0,050	0,171	0,124	0,025	0,146	0,180

Таблиця 114

Матриця перших залишків кореляцій

Змінні	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0,645	0,127	0,101	-0,663	0,032	-0,241
P_2	0,127	0,094	-0,073	-0,129	0,118	-0,136
P_3	0,104	-0,073	0,078	-0,163	0,016	0,040
P_4	-0,663	-0,129	-0,163	0,658	-0,071	0,358
P_5	0,032	0,118	0,016	-0,071	0,139	-0,234
P_6	-0,241	-0,136	0,040	0,358	-0,234	0,203
Суми	0,001	0,001	0,001	-0,010	0,000	-0,010

Для розрахунку навантажень другого фактора необхідно визначити середню кореляцію кожної змінної з іншими змінними.

З цією метою розраховують суми по стовпчиках матриці перших залишків (табл.114). Слід знати, що мірилом правильності розрахунків є критична величина «0,010». Суми по стовпчиках не повинні перевищувати її рівень. У нашому прикладі розрахунки, як бачимо, вірні.

Оскільки додатні і від'ємні значення коефіцієнтів кореляції урівноважуються, сума всіх стовпчиків матриць практично буде дорівнювати нулю. Розрахунок навантажень другого фактора можна здійснювати лише при наявності додатніх сум елементів стовпчиків матриці. З цією метою необхідно виконати перетворення алгебраїчних знаків у матриці залишків кореляцій. (Ця математична процедура не змінює абсолютне значення коефіцієнта кореляції. З точки зору графічної інтерпретації конфігурація векторів змінних зберігає свій зміст, оскільки змінюється лише напрямок змін змінних).

Розрахунок навантажень другого фактору здійснюється у такій послідовності:

1. Визначають алгебраїчну суму елементів по стовпчиках, виключаючи елементи головної діагоналі (у табл.64 рядок $\sum r_0$). Знайдені суми додають по рядку ($\Sigma\Sigma r_0$). У розглядуваному прикладі ця величина дорівнює – 1,836.
2. Відшукують стовпчик з найбільшою від'ємною сумою (стовпчик P_4, - 0,668).

Ця сума з додатним знаком записується в рядок з назвою "Стовпчик 4" по вертикалі даного стовпчика.

Подальші розрахунки по рядку здійснюють у такій послідовності: до суми стовпчика додають з протилежним знаком подвоєне значення елемента цього стовпчика, який знаходиться на перетині з “перетворюваним рядком". Одержаний результат записують у рядок, з назвою “Стовпчик 4”. У нашому випадку, наприклад, величину 0,682 одержуємо: -0,644 –2 $\times$ 0,663; величину 0,165 маємо при розрахунку: - 0,093 –2 $\times$0,129 і т.д. Обчислені елементи (суми) даного рядка підсумовуємо і заносимо у графу 8 (1,262).

Таблиця 115

Розрахунки навантаження другого фактора
(перетворення знаків у матриці перших залишків кореляції)

Змінні	P_1	P_2	P_3	P_4	P_5	P_6	Σ_0
P_1	**<u>0,645</u>**	− 0,127	− 0,101	+ -0,663	− 0,032	+ -0,241	0,001
P_2	− 0,127	**<u>0,094</u>**	+ -0,073	+ -0,129	− 0,118	+ -0,136	0,001
P_3	− 0,101	+ -0,073	**<u>0,078</u>**	+ -0,163	− 0,016	− 0,040	0,001
P_4	+ -0,663	+ -0,129	+ -0,163	**<u>0,658</u>**	+ -0,071	− 0,358	-0,010
P_5	− 0,032	− 0,118	− 0,016	+ -0,071	**<u>0,139</u>**	+ -0,234	0,000
P_6	+ -0,241	+ -0,136	− 0,040	− 0,358	+ -0,234	**<u>0,203</u>**	-0,010
$\sum_0$	0,001	0,001	0,001	-0,010	0,000	-0,010	$\Sigma\Sigma r_0 =$
$\sum r_0$	-0,644	-0,093	-0,079	-0,668	-0,139	-0,213	-1,836
Стовпчик 4	0,682	0,165	0,247	**-0,668**	0,003	-0,503	1,262

Далі відшукуємо наступний стовпчик з найбільшою від’ємною сумою. Послідовність розрахунку нового рядка аналогічна описаній вище. Одержанні елементи рядків додаємо на стовпцях, одержуючи значення $\sum r$. Подальші обчислення виконують у послідовності, аналогічній описуванню розрахунків по визначенню навантажень першого фактора. Навантаження другого фактора для змінної α визначають за уже відомою формулою:

$$C_{2a} = \sum r_\alpha (\frac{1}{\sqrt{T}}), \text{ де}$$

C_{2a} - навантаження другого фактору для змінної α;

$\sum r_\alpha$ - сума по стовпчику α;

Т - загальна сума всіх коефіцієнтів матриці.

При обчисленні факторних навантажень виникають певні математичні тонкощі, зв'язані з перетворенням знаків матриці, розрахунком певних критеріїв, а також деякими методичними особливостями виділення факторів. Викладення математичних основ цієї сторони обчислювальних дій виходить за рамки нашої роботи. Тут необхідно звертатися до спеціальної літератури.

Логічним завершенням здійснюваних розрахунків у справі вичленування факторів слід назвати етап припинення виділення факторів. Серед множин існуючих методик Я. Окунь посилається на метод під назвою "Критерій Саундерса" .

Суть і послідовність обчислювальних операцій за вказаним методом така :

1. Залишки, отримані після виділення К-го фактора, підносять до квадрата і сумують, виключивши елементи головної діагоналі і позначивши число змінних n. Одержана сума множиться на $\frac{2n}{n-1}$ з метою приведення її у відповідність з повною матрицею. Одержана величина становить значення А.

2. Різниця між кількістю змінних і уже виділених факторів ділитися на число змінних. Результат підносять до квадрату. Одержують величину значення В.

3. Факторні навантаження підносять до квадрата, включивши навантаження К-го фактора, і сумують одержані величини. Число факторних навантажень тут дорівнює $K \times n$. Результат віднімають від числа змінних (*n)* і одержане значення підносять до квадрата. Результат ділять на кількість одиниць спостереження. Одержують значення С.

4. У випадку $A \prec B \times C$ виявлення факторів припиняють. При $A \succ B \times C$ вичленовують наступний фактор і здійснюється описана процедура перевірки.

Приклад. Розглянемо викладену вище методику послідовних операцій на прикладі матриці перших залишків кореляції (табл. 116.). Піднісши до квадрата перші залишки кореляцій, знаходимо їх суму, яка дорівнює 1,583939. Далі знаходимо похідні:

$$1{,}583939 \times \frac{2n}{n-1} = 3.801(A).$$

Таблиця 116

Вихідні і розрахункові дані матриці перших залишків кореляцій шести змінних

Змінні	P_1	P_2	P_3	P_4	P_5	P_6	
P_1	**0,645**	0,127	0,101	-0,663	0,032	-0,241	
P_2	0,127	**0,094**	-0,073	-0,129	0,118	-0,136	
P_3	0,101	-0,073	**0,078**	-0,163	0,016	0,040	
P_4	-0,663	-0,129	-0,163	**0,658**	-0,071	0,358	
P_5	0,032	0,118	0,016	-0,071	**0,139**	-0,234	
P_6	-0,241	-0,136	0,040	0,358	-0,234	**0,203**	
$\sum_0$	0,001	0,001	0,001	-0,010	0,000	-0,010	$\Sigma\Sigma r^2{}_0 =$ 1,583939
$\sum r^2{}_0$	0,525004	0,070519	0,036294	0,615984	0,075001	0,261137	

Різниця між числом змінних і числом уже виділених факторів становить 6-1=5 (В).

Подальші обчислювальні операції, викладені вище в пункті 3, зводяться до знаходження значення С. Обчислена сума квадратів факторних навантажень становитиме 0,562634 ($\sum C_{10}^2$). Її різниця з числом змінних дорівнює 6-0,563=5,437. Квадрат даної величини приймає значення 29,561. Знаходимо значення С: 29,561: 57=0,519 (С). Добуток $B \times C$ дорівнює 2,595. Як випливає з проведенних розрахунків $A \succ B \times C$ (3,801 $\succ$2,595).Отже за даною кореляційною моделлю врожайності необхідно продовжити дослідження, пов'язане з вичленуванням наступного фактора.

У тій же монографії Я.Окунь торкається проблеми мінімізації числа змінних (n) для визначення однозначного числа факторів (m).

Тут автор наводить формулу Терстоуна: $n = \frac{2m+1+\sqrt{8m+1}}{2}$.

Після перетворення формули для одержання числа факторів m маємо: $n = \frac{2n+1-\sqrt{8n+1}}{2}$.

Однозначне число факторів, що вичленовуються для нашого випадку становитиме: $n = \frac{2 \times 6 + 1 - \sqrt{48+1}}{2} = 3$.

Як бачимо, в розглядуваній кореляційно – регресійній моделі урожайності з шістьма змінними можна визначити не більше трьох факторів.

Стандартна таблиця співвідношень числа змінних (n) і факторів, які вичленовуються має значення:

m	1	2	3	4	5	6	7	8	9	10
n	3	5	6	8	9	10	12	13	14	15

§ 2.6. Метод головних компонент. Загальне поняття методу, його завдання

До недавнього часу метод головних компонент вважали різновидом факторного аналізу. Нині його відносять до групи самостійних статистико – математичних методів багатомірного аналізу. Вперше він був розроблений в 1901 р. англійським статистиком К.Пірсоном . Потім знайшов свій розвиток у працях Г.Хотеллінга, Г.Хармана, С. Рао, П.Андрукевича, С.Айвазяна і інших авторів. У нашій країні метод головних компонент одержує широке розповсюдження з появою ПЕОМ.

Як відомо, соціально-економічне явище можна характеризувати цілим рядом ознак. При великому наборі таких ознак в кореляційно – регресійному аналізі вплив зв’язків стає затрудненим, тому виникає необхідність стиснення інформації, тобто опис досліджуваного явища (об’єкта) більш укрупненими показниками, так званими “ головними компонентами”. Вихідним ступенем тут є кореляційна матриця, на підставі якої з використанням методу головних компонент може бути продовжено аналіз значень спостережуваних ознак.

Правильно відібрані в кореляційну модель ознаки, як правило, пов’язані між собою. Наявність таких зв’язків між ними дозволяє на основі одного фактора мати інформацію про інший. Існування тісного зв’язку між ознаками дає підставу для виключення однієї з них. Наприклад, якщо в модель урожайності включено дві змінні x_1 і x_2, які характеризують грошові витрати на гектар, перша – всі види, друга – затрати на добрива. Тут практично буде зайвим при включенні в модель ознаки x_1 досліджувати також і ознаку x_2, оскільки вона тісно пов’язана з першою . Ідея обліку однієї ознаки на підставі другої лежить в основі методу головних компонент. Слід відзначити, що мова не йде тільки про дві ознаки. У такому випадку метод головних компонент малоефективний. Його використовують, як правило, при десятках взаєпов’язаних ознак. При цьому ставиться мета “ набрати” певну частину загальної варіації результативної ознаки мінімальною кількістю змінних. Останні підбирають до тих пір, поки сума їх дисперсій не сягатиме заданої частки у дисперсії досліджуваного явища (наприклад, 60 %, 80 %, 90 % і т.д.).

Метод головних компонент розв’язує такі завдання:

1. Відшкодування скритих, об’єктивно існуючих закономірностей у зміні явищ.

2. Характеристика явища, що вивчається, числом ознак, значно меншим взятих, на початковому етапі. Число головних компонент, виділених в процесі дослідження, буде вміщувати (у компактній формі) більше інформації, ніж початково виміряні ознаки.

3. Виявлення ознак, найбільш тісно пов'язаних з головною компонентою. Інакше кажучи, вивчення стохастичного зв'язку між ними (зв'язок, при якому зі зміною однієї змінної змінюється закон розподілу другої).

4. Прогнозування рівней досліджуваних явищ на підставі рівняння регресії, яке одержане по інформації головних компонент.

Переваги такого методу прогнозування на відміну від класичного регресійного аналізу можна пояснити тим, що при останньому в модель намагаються включити максимально можливу кількість факторів, які в економічних явищах часто характеризуються істотною корельованістю (мультилінеарністю). Прогноз за такими змінними, як правило, буває не точним. Тому виникає завдання про заміну вихідних взаємопов'язаних змінних сукупністю некорельованих параметрів. Це завдання вирішується математичним апаратом - методом головних компонент, який являє собою характеристики, побудовані на підставі первинно виміряних ознак.

Реалізація практичних можливостей зазначених вище завдань, які вирішуються методом головних компонент у галузі економіки, може бути представлена різними напрямами. Назвемо їх .

1. Аналіз причинно - наслідкових взаємозв'язків показників і встановлення їх стохастичного зв'язку з головними компонентами.

2. Виділення узагальнюючих економічних показників.

3. Ранжирування результатів спостережень по головних компонентах

4. Класифікація об'єктів спостереження.

5. Список вихідної інформації.

6. Побудова рівнянь регресії за узагальнюючими економічними показниками.

Як негативну сторону методу головних компонент слід назвати складність математичного апарату, зумовленого абсолютністю знань теорії ймовірностей, математичної статистики, лінійної алгебри, а також математичного забезпечення ПЕОМ. Формальне використання стандартних програм без розуміння математичної суті обчислювальних процедур може призвести до необґрунтованих висновків. Слід також пам'ятати про професіональні знання суті досліджуваних економічних явищ. Тільки за таких умов метод головних компонент може стати могутнім математичним засобом пізнання існуючих реалей у галузі соціально - економічних явищ.

§ 2.7. Кластерний аналіз. Загальне поняття, його математичні основи та завдання

Як уже відомо, факторний аналіз найбільш яскраво відображує риси багатомірного аналізу в частині дослідження зв'язку між ознаками. Кластерний аналіз ці риси відображує з боку класифікації об'єктів. Cluster (англ.)- нагромадження груп елементів, які характеризуються якою – небудь загальною властивістю. Суть його зводиться до групування (кластеризації) сукупності з різноманітними ознаками з метою одержання однорідних груп – кластерів. При цьому межі таких груп наперед не завдані, а кількість їх може бути або завдано, або ні. Одержані в результаті розмежування групи називаються кластерами**,** а методи їх знаходження – кластер-аналізом**.** У кластерному аналізі ознаки об'єднуються в один кількісний показник схожесті (несхожесті) групуючих об'єктів.

Будь яка міра схожисті являє собою деяку функцію, яка ставить у відповідність кожній парі точок (x_i, x_j) деяке число d_{ij}, що характеризує ступінь схожості (наближеності)між об'єктами u_i, u_j. Практично використовується такі типи мір схожисті : 1) коефіцієнт подібності) так звані квантифіковані коефіцієнти зв'язку); 2) коефіцієнти зв'язку (кореляції); 3) показники відстані в метричному просторі.

Роль міри схожості відіграє функція відстані**,** введення якої веде до поняття метричного простору**.** Останній являє собою множину елементів з будь - якою природою явищ. Для будь - якої пари елементів цієї множини визначено певне уречевлене число, яке називається відстанню**.** Найбільше вживані його показники в завданнях автоматичної класифікації соціально – економічних об'єктів - це відстань по Хеммінгу та евклідова відстань.

Якщо уявити будь - яку пару елементів Е і Д, а уречевлене для них число S (Е, Д), вкажемо три властивості відстані : 1) якщо Е і Д збігаються, відстань S (Е, Д)= 0; 2) для будь - яких трьох точок Е,Д, С

S (Е, Д)≤ S (Е, С) +S (С, Д); 3) S (Е, Д)= S (Д,Е,).

Серед відомих функцій відстані найрозповсюдженіша – евклідова відстань. Емпірична формула її має вигляд :

$$d_{ij} = \sqrt{\sum_{k=1}^{n}(x_{ki} - x_{kj})^2} \ ,$$

де x_{ki}- значення к-ої ознаки в і – му об’єкті.

Враховуючи недоліки евклідової відстані, зумовлені неможливістю врахувати можливу нерівномірність осей простору, математична література рекомендує користуватися нормованою евклідовою відстанню. Розрахунок її аналогічний розрахунку величини α_{ij}, але за стандартизованими значеннями ознак. Необхідність такого способу розрахунку пояснюється тим, що можливий випадок, коли два об’єкти досить схожі за всіма ознаками і значно різняться за однією. За цією ознакою у евклідовій відстані вони будуть далекі одна від одної. Іншими словами, величина α_{ij} залежить від масштабу виміру ознак. Для забезпечення співставності ознак їх, як правило, нормують за середньоквадратичним відхиленням (σ). Евклідова відстань, як і аналогічні їй відстані Махаланобіса, відстань методу потенціальних функцій і т.п. прийнятна у розрахунках з ознаками, які мають кількісний вимір. Для якісних ознак, які приймають тільки два значення (о і 1), застосовують формулу відстані по Хеммінгу:

$$\alpha_{ij} = \sum_{k=1}^{n}\left|x_{ki} - x_{kj}\right|,$$

де x_{ki}- значення к – ої ознаки в і – му об’єкті.

Якщо ознаки класифікуються з довільним числом градацій, рекомендується використовувати формулу міри близкості двох розбивань $\alpha(E,Д)$, яка має властивості геометричної відстані:

$$\alpha(E,Д) = \frac{1}{2}\sum_{r,s=1}^{n}\left|E_{r,S} - D_{r,S}\ \right|,$$

де $E_{r,S=}1$, якщо об’єкти x_3, x_2 знаходяться в одному класі; $E_{r,S=}0$, якщо об’єкти x_S, x_r знаходяться в різних класах.

По розбиванню елемента Д розрахунки виконуються аналогічно ($Д_{r,S=}1$;$Д_{r,S=}0$,).

Суть завдання кластерного аналізу така: існуючу деяку множину об’єктів необхідно розділити за допомогою певного правила на раніше завдану або на завдану кількість класів. У символіці ці завдання можна сформулювати так: множину реалізацій, завданих у просторі x за допомогою вирішуваних функцій з γ (за критерієм схожості), потрібно розділити на таку кількість і таких елементів алфавіту А, щоб втрати інформації не перевищували завданої величини К.

Таким чином, завдання кластерного аналізу зводиться до представлення первинної інформації у стислому вигляді без її втрати. Вирішенням такого завдання (як уже зрозуміло) проходить через класифікацію ознак (вимірів), якими характеризується кожний об'єкт. Але мова йде не про класичні принципи класифікації (наприклад, комбінаційне групування), а про принципи багатомірної класифікації. Суть її зводиться до об'єднання (класифікації) об'єктів не послідовно за окремими ознаками, а одночасно за великою чисельністю ознак. Такий набір їх створює так званий "ознаковий простір ". Кожній ознаці надається змістовність координати.

Оперуючи n ознаками, ми розглядаємо будь - який об'єкт як точку в n - мірному просторі, і завдання класифікації полягає у виявлені згущення точок (об'єктів) у цьому ознаковому просторі. Загальним для згущення точок є те, що групи (кластери) формуються на підставі "схожості" (наближення) об'єктів за великою кількістю ознак, тобто класифікація здійснюється одночасно за всім комплексом ознак, які характеризують об'єкт. При цьому жодна з ознак такого комплексу не є необхідною (або достатньою) умовою належності об'єкта до даної групи.

Формування груп об'єктів, близьких за комплексом ознак, більш ефективне у порівнянні з комбінаційним групуванням. Так, для останнього об'єкт, який має відхилення від меж групувальної ознаки (норми, характерної для даної групи за однією єдиною ознакою набору), буде виключений з групи. Легко уявити ситуацію, коли дана ознака використовується при першій градації об'єктів. У цьому випадку об'єкт може виявитися у групі досить віддаленої від тієї, з якою вона (ознака) має найбільшу схожість. У комбінаційному групуванні самі групи являють собою ні що інше як сектори ознакового простору. Здійснюючи класифікацію за названим групуванням, ми інколи штучно руйнуємо ознаковий простір завданими границями інтервалів груп, тоді як реально існують відокремлено однорідні класи.

Перевага методу кластерного аналізу в тому, що його математичний апарат дозволяє знайти і виділити реально існуюче в ознаковому просторі нагромадження об'єктів (точок) на підставі одночасного групування за великою кількістю ознак.

Кластерний аналіз, як і кореляційно - регресійний, є математичним апаратом вивчення статистичних зв'язків. Це метод пошуку емпіричних закономірностей, але для більш широкого класу зв’язків. Для регресійного аналізу є цілий ряд важко виконуваних умов (вимог) його застосування. Серед них вимоги нормальності багатомірного розподілу, неможливість використання якісних ознак, обмеження, які накладаються на алгебраїчну форму зв’язку (метод найменших квадратів ефективний для лінійних рівнянь) і ін.

Для методу кластерного аналізу однорідність сукупності не є обов’язковою умовою. Більше того, сам метод дозволяє виявити і описати структурні закономірності, забезпечивши формування однорідних класів об’єктів. Дискретність кластерних моделей на відміну від неперервних регресійних моделей, зумовлена усередненням і деякими втратами інформації, забезпечує більш евристичний характер обчислювальних процедур, а також знімає обмежування, пов’язані з алгебраїчною формою зв’язку.

Нарешті, комплексне використання обох методів у вивченні статистичних зв’язків створює умови широкого використання методу кореляційно – регресійного аналізу, забезпечуючи умови для адекватного його додатка.

Викладене вище дає змогу зробити висновок про те, що застосуванню методу кластерного аналізу повинно передувати вивчення теорії і накопиченої практики цього використання. На початкових етапах використання цього методу дослідник повинен мати чітко уявлення, яке з двох завдань він вирішує. Чи це звичайне завдання типізації, при якому досліджувану сукупність спостережень слід розділити на відносно невелику кількість групувань. Тоді виконується робота, аналогічна одержанню інтервалів статистичного групування при обробці одномірних спостережень. При цьому операція здійснюється так, щоб елементи однієї області групування знаходились один від одного по можливості на невеликій відстані. Друге завдання може полягати в тому, що дослідник намагається визначити природну відстань вихідних елементів (спостережень) на чітко виражені кластери, що знаходяться один від одного на деякій відстані, але які не розбиваються на такіж віддалені одна від одної частини. Слід пам’ятати, що перше завдання (завдання типізації) завжди має рішення, друге – в своїй постановці може мати негативний результат, тобто може виявитися, що множина вихідних спостережень не виявляє природного розташування на кластери, наприклад, утворює один кластер.

Досить важливим етапом кластер – аналізу є вибір змінних (ознак). Ця стадія аналізу є основою формування однакових просторів, у яких повинно проводитися моделювання.

Вибір ознак здійснюється, як правило, у дві стадії. В основі першої з них лежить формування первинної гіпотези про набір ознак, які впливають на досліджуване явище; в основі другої – уточнення гіпотези по результатах консультацій (опитувань) спеціалістів досліджуваної галузі.

Завершеною вважається економічна постановка завдання при умові її узгодженості з вимогами використовуваного математичного апарату і можливостями обчислювальної техніки. Після цього приступають до збору вихідної інформації.

Програмований контроль знань до науково-пізнавальних тем

Тема 1. Перевірка статистичних гіпотез

1.1. Що розуміють під статистичною гіпотезою ?
– Наукове припущення, яке потребує доведення.
– Наукове припущення взагалі.
* Наукове припущення про властивості випадкової величини, яка перевіряється за результатам статистичних спостережень.
– Наукове припущення про розмір статистичних характеристик.

1.2. Як називається гіпотеза, яку необхідно перевірити ?
– Параметрична.
* Нульова.
– Конкуруюча.
– Альтернативна .

1.3. Як називається гіпотеза, протилежна нульовій ?
– Непараметрична.
– Параметрична.
* Альтеративна.
– Основна.

1.4. У скількох випадках можна прийняти неправильне рішення в результаті статистичної перевірки прийнятого рішення ?
– Один випадок .
* Два випадки.
– Три випадки.
– Три і більше випадків.

1.5. Як при перевірці статистичних гіпотез називають випадки прийняття неправильного рішення ?
– Помилки перевірки гіпотез.
– Помилки вибірки.
– Помилки першого роду.
* Помилки першого і другого родів.

1.6. Що означає термін „прийняти гіпотезу” ?
– Знайдене правильне рішення.
– Результати спостережень, суперечать науковим припущенням.
* Результати спостережень не суперечать висунутій гіпотезі.
– Знайдені правильне і неправильне рішення.

1.7. Якщо вибірковий простір уявити як множину можливих значень величин, розташованих у порядку зростання, на які частини поділяє цей простір статистичний критерій?
* Критичну область прийняття гіпотези.
– Критичну і некритичну області.
– Область прийняття гіпотези та область неприйняття гіпотези.
– Критичну область і область неприйняття гіпотези.

1.8. Як називаються точки, що відділяють критичну область від області прийняття гіпотези?
– Правосторонні .
– Лівосторонні.
– Двосторонні.
* Критичні.

1.9. Яка відповідь виходить за межі характеристики критичної області?
* Гранична.
– Правостороння.
– Лівостороння.
– Двостороння.

1.10. Якщо вибіркове значення статистичного критерію потрапляє у критичну область, який робиться висновок ?
– Гіпотеза приймається.
* Гіпотеза не приймається.
– Гіпотеза додатково перевіряється.
– Гіпотеза додатково не перевіряється.

1.11. Що являє собою область допустимих значень ?
* Множина допустимих положень вибіркової точки у вибірковому просторі, які призводять до прийняття нульової гіпотези.
– Вибірковий простір випадкової змінної.
– Критична область.
– Статистичний критерій.

1.12. Як називаються помилки першого роду в задачах перевірки статистичних гіпотез ?
* Рівень значущості (мала ймовірність).
– Рівень імовірності.
– Рівень істотності.
– Рівень вірогідності.

1.13. Що при перевірці статистичних гіпотез називається потужністю критерію?
– Помилка першого роду.
– Помилка другого роду.
* Величина, яка доповнює ймовірність помилки другого роду.
– Величина, яка доповнює ймовірність помилки першого роду.

1.14. Яка з перелічених нижче відповідей виходить за межі етапів перевірки нульової гіпотези ?
– Вибирається рівень значущості.
– Вибирається відповідно до значущості критична область.
– Обчислюється вибіркове (фактичне) значення статистичної характеристики.
* Розраховується помилка вибірки.

1.15. Яка з перелічених нижче відповідей виходить за межі вирішуваних завдань при перевірці гіпотез щодо статистичних розподілів ?
* Визначається вірогідність середньої арифметичної в емпіричному ряді розподілу.
– Визначається відповідність емпіричного розподілу теоретичному виду розподілу.
– Визначається можливість належності двох і більше емпіричних розподілів до одного й того ж виду розподілів.
– З’ясовується наявність у розподілі незалежності ознак однієї від одної.

1.16. Які відповіді виходять за межі здійснення обчислювальних етапів при перевірці гіпотези про нормальний розподіл?
– Розраховуються $\tilde{x}$ і σ.
– Обчислюються теоретичні частоти.
– Зіставляються теоретичні та емпіричні частоти за критерієм Пірсона.
* Знаходиться потужність критерію.

1.17. Яку з відповідей слід вважати правильною щодо методичних особливостей використання Хі – квадрат критерію ?
* Чисельність вибірки повинна бути достатньо великою (не менше 50), а величини частот – не менше 5.
– Чисельність вибірки може бути менша за 50.
– Величини частот можуть бути менші за 5.

– Чисельність вибірки і величини частот не обмежуються.

1.18. Який суміжний практичний аспект має розрахована величина Хі – квадрат критерію в аналітичних розрахунках?

– Використовується для визначення дисперсії.

* Використовується для визначення коефіцієнта кореляції альтернативних ознак.

– Кількість характеризує коваріацію.

– Визначає кількісний вплив окремих ознак.

1.19. Як називаються критерії згоди при оцінці сукупностей, які не підпорядковані закону нормального розподілу?

– Параметричні критерії.

– Стохастичні критерії.

* Непараметричні критерії.

– Альтернативні критерії.

1.20. Які статистичні критерії належать до непараметричних ?

– Критерії Фішера, Ст'юдента, Пірсона.

– Критерії Ст'юдента, Пірсона.

– Критерії Кохрана, Бартлета.

* Критерії Колмогорова, Уайта, Уїлкоксона.

1.21. Які статистичні критерії використовують у випадках, коли необхідна перевірка статистичних гіпотез про рівність трьох і більше дисперсій ?

– Критерій Колмогорова.

* Критерії Кохрана і Бартлета.

– Критерій Уайта.

– Критерії Колмогорова і Бартлета.

Тема 2. Методи багатомірного статистичного аналізу

2.1. Дати сутність поняття „багатомірний статистичний аналіз".

– Кількісний статистичний метод .

– Математичний метод.

* Об'єднання ряду методів, призваних дослідити поєднання взаємопов'язаних ознак.

– Розрахунково-конструктивний метод.

2.2. Дати визначення багатомірної випадкової величини.

– Ознака, яка вимірює вплив факторів у причинно - наслідкових моделях зв'язку.

– Статистична оцінка невідомих параметрів статистичної сукупності.

* Набір кількісних ознак, значення кожної з яких піддається неконтрольованому розкиду при повтореннях даного процесу, статистичного спостереження, досвіду тощо.
– Набір кількісних статистичних ознак для характеристики економічних моделей багатофакторних зв'язків.

2.3. В чому полягає і забезпечується ефективність застосування багатомірного статистичного аналізу?
* За умови якості вихідної інформації і масовості даних спостережень.
– Логічного та арифметичного контролю вихідної інформації.
– Перевірки статистичних гіпотез відносно досліджуваних процесів і явищ.
– Встановлення законів статистичних розподілів.

2.4. Сутність методу факторного аналізу як математико - статистичного методу.
– Кількісний статистичний метод дослідження неповних кореляційних зв'язків.
– Дозволяє вимірювати кількісний вплив факторних ознак на результативні ознаки.
* Дає перші наближені характеристики закономірностей розвитку явища і можливість сформулювати загальні висновки про напрямки подальших досліджень.
– Узагальнення статистичних характеристик одержаних в результаті використання методу графічних побудов і кореляційно-регресійного аналізу.

2.5. Поняття етапів факторного аналізу.
* Послідовне зіставлення різних наборів факторів і варіантів групувань з їх включенням, виключенням і оцінкою вірогідності відмінностей між групами.
– Перевірка статистичних гіпотез відносно досліджуваних причинно-наслідкових зв'язків економічних явищ і процесів.
– Встановлення законів розподілу вибіркових характеристик статистичної сукупності.
– Послідовний розрахунок статистичних характеристик відносно однорідності досліджуваної сукупності одиниць спостереження.

2.6. Яка з перелічених відповідей не відноситься до статистичної природи факторного аналізу ?

– Сукупність змінних, які вивчаються з точки зору зв'язків між ними, не вибирається довільно, а виявляються основні фактори впливу.
– Факторний аналіз спирається на спостереження над природним варіюванням ознак.
– Дослідження не потребує попередніх гіпотез і не вимагає апріорних здогадок відносно залежності чи незалежності змінних.
* В основу моделювання причинно - наслідкових зв'язків покладено побудову багатомірних дисперсійних комплексів.

2.7. Яка з перелічених позицій не відноситься до завдань факторного аналізу?

– Розчленування об'єктів спостереження за типовими ознаками.
– Виявлення структури взаємозв'язків.
* Вивчення типів явищ.
– Визначення основних аспектів відмінностей між об'єктами спостереження (мінімізація описувань).

2.8. Що лежить в основі об'єктивності практичного використання факторного аналізу при дослідженні соціально - економічних явищ?

– Математична природа методу сформована на базі теорії ймовірностей.
– Використання комплексу статистико – математичних методів незалежно від специфіки досліджуваних явищ.
– Знання математичної природи моделей множинної регресії.
* Одержані агреговані ознаки повинні мати економічно обґрунтоване трактування .

2.9. Які завдання вирішуються за допомогою факторного аналізу в галузі соціально - економічних досліджень?

– Кількісний вимір впливу факторних якісних ознак у формуванні причинно - наслідкових зв'язків.
* Встановлення відмінностей між об'єктами спостереження та формування гіпотез про ці відмінності; перевірка гіпотез про взаємозв'язок та виявлення їх структури; розчленування об'єкті спостереження за типовими

ознаками та зіставлення структур наборів ознак.
- – Вивчення причинно - наслідкових зв'язків з врахуванням типів досліджуваних об'єктів.
- – Оцінка невідомих параметрів досліджуваних статистичних сукупностей щодо розмірів і тенденцій розвитку аграрно – економічних явищ і процесів.

2.10. Критерії відбору при формуванні інформаційного масиву моделей факторного аналізу.
- – Кореляційно взаємопов'язані лінійні зв'язки.
- * Точність, вірогідність і зіставність даних статистичного спостереження, їх однорідність за змістом і колом обстежуваних об'єктів.
- – Підпорядкованість вихідної інформації закону розподілу Гаусса.
- – Лінійних характер взаємопов'язаних ознак за способом розрахунків.

2.11. Що розуміють під перетворенням вихідних даних у факторному аналізі соціально-економічних явищ?
- – Побудова статистичних рядів розподілу сукупності.
- * Зміна характеру емпіричного розподілу відповідно до мети і завдання дослідження соціально - економічних явищ.
- – Побудова матриць коефіцієнтів кореляції.
- – Описування усього набору факторів, які підлягають досліджуванню на предмет встановлення причинно-наслідкових зв'язків.

2.12. Яка з наведених формул дозволяє мінімізувати число змінних для визначення однозначного числа факторів у кореляційно – регресійній моделі факторного аналізу?
- * Формула Терстоуна:

$$n = \frac{2m + 1 + \sqrt{8m + 1}}{2}.$$

- – Формула чисельності вибірки при оцінці середньої:

$$n = \frac{t^2 \sigma^2 N}{\Delta^2 N + t^2 \sigma^2}.$$

- – Формула граничної помилки вибіркової середньої:

$$\Delta_{\bar{x}} = t\sqrt{\frac{\sigma^2}{n}(1 - \frac{n}{N})}.$$

- – Формула загальної дисперсії альтернативних ознак:

$$\sigma_w^2 = w(1 - w).$$

2.13. Сутність методу головних компонент.
- * Стиснення інформації шляхом опису досліджуваного явища (об’єкта) більш укрупненими показниками так званими „головними компонентами”.
- – Логічне продовження та поглиблення методу статистичних групувань.
- – Встановлення мультіколінеарних залежностей у кореляційно – регресійних моделях зв’язку.
- – Визначення ступеня відокремленого спільного впливу досліджуваних факторів на результативну ознаку .

2.14. Завдання методу головних компонент як методу багатомірного статистичного аналізу.
- – Теоретичне обґрунтування причинно - наслідкових зв’язків залежності досліджуваних явищ.
- – Досліджування і встановлення ступеня тісноти зв’язків у причинно- наслідкових моделях взаємозалежності явищ.
- * Встановлення об’єктивно існуючих закономірностей зміни явищ невеликою кількістю ознак, виділених у процесі дослідження та вивчення стохастичного зв’язку між ними.
- – Математичне очікування змінної величини, зумовлене зміною випадкової.

2.15. Негативна сторона методу головних компонент.
- – Громіздкість обчислювальних операції.
- – Неможливість використання простих засобів обробки інформації та обчислювальної техніки.
- * Складність математичного апарату, зумовлену абсолютністю знань теорії ймовірностей, математичної статистики, лінійної алгебри таматематичного забезпечення ЕОМ.
- – Складність існуючих стандартних електронних програм.

2.16. Сутність методу кластерного аналізу.
- – Комбінаційне статистичне групування досліджуваних ознак за факторно - результативною схемою побудови таблиць.
- * Групування (кластеризація) сукупності з різноманітними ознаками з метою одержання однорідних груп.
- – Кластеризація статистичної сукупності з метою виявлення основних чинників формування регресійних моделей зв’язку.

– Реалізація практичних можливостей математико-статистичних оцінок результатів групувань.

2.17. Як називаються групи одержані при розмежуванні статистичної сукупності в кластерному аналізі?
– Кластеризаційні групування.
– Результативні групування.
– Типологічні групування.
* Кластери.

2.18. Як називаються прийоми чи методи кластеризації сукупності в багатомірному статистичному аналізі?
– Метод головних компонент .
– Експрес-аналіз статистичних зв’язків.
* Кластер – аналіз.
– Кластер – аналіз стохастичних зв’язків.

2.19. Як називається формула для визначення міри схожості (відстані) у метричному просторі при кластер – аналізі $d_{ij} = \sqrt{\sum_{\kappa=1}^{n}(x_{ki} - x_{kj})^2}$?
– Відстань по Хеммінгу.
– Відстань Махаланобіса.
* Евклідова відстань.
– Нормативна відстань.

ДОДАТКИ

Додаток 1

Критичні точки розподілу Ст'юдента (t- розподіл)

Число ступенів вільності, ν	Рівень значимості (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,0
4	2,13	2,78	3,75	4,50	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	3,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	3,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,63	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,58	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	8,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,23
Число ступенів вільності, ν	0,05	0,025	0,01	0,00005	0,001	0,0005
	Рівень значимості (одностороння критична область)					

Додаток 2

Імовірності t- розподілу по Ст'юденту для малих вибірок (у межах ± t)

t	5	6-7	8-10	11-25	16-25	25-35	∞
0,1	538*	538	539	539	539	539	540
0,2	576	578	578	578	578	578	579
0,3	612	613	615	616	616	616	618
0,4	647	649	651	652	653	654	655
0,5	681	683	685	687	689	689	691
0,6	713	715	718	721	722	724	726
0,7	742	746	749	752	754	756	758
0,8	770	774	778	781	783	785	788
0,9	795	800	804	808	811	813	816
1,0	818	823	828	832	735	838	841
1,1	839	844	850	854	858	860	864
1,2	859	864	870	874	878	881	885
1,3	875	881	887	892	896	889	903
1,4	890	896	902	907	912	917	919
1,5	903	906	916	921	925	928	933
1,6	915	921	928	933	937	940	945
1,7	925	932	938	943	948	951	955
1,8	934	941	947	952	956	959	964
1,9	942	948	955	960	964	967	971
2,0	949	955	962	967	970	973	977
2,1	955	961	967	972	976	978	982
2,2	960	966	972	977	980	982	986
2,3	969	971	977	981	984	986	989
2,4	969	975	986	984	987	989	992
2,5	973	978	983	987	989	991	994
2,6	976	981	986	989	991	993	995
2,7	979	983	988	991	993	995	996
2,8	981	985	990	993	995	996	997
2,9	983	987	991	994	996	997	998
3,0	985	989	993	995	997	997	999

* Тут і далі значення дано після коми.

Додаток 3

Перша функція нормованого відхилення

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

(ординати нормальної кривої)

Цілі і десят-кові частки t	Соті частки t									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989*	3989	3988	3987	3984	3982	3980	3977	3973
0,1	0,3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	0,3910	3902	3894	3885	3876	3867	3857	3847	3838	3825
0,3	0,3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	0,3683	3668	3653	3637	3651	3605	3589	3572	3555	3538
0,5	0,3521	3503	3485	3467	3448	3429	3411	3391	3373	3352
0,6	0,3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	0,3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	0,2877	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	0,2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	0,2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	0,1942	1919	1895	1872	1849	1826	1784	1758	1758	1736
1,3	0,1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	0,1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	0,1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	0,1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0,0940	0925	0909	0893	0878	0863	0843	0833	0818	0804
1,8	0,0790	0775	0761	0748	0734	0721	0707	0695	0681	0669
1,9	0,0656	0655	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0,0440	0431	0422	0413	0404	03950	0387	0379	0371	0363
2,2	0,0355	0347	0339	0332	0325	0317	0310	0303	0296	0290
2,3	0,0283	0277	0270	0264	0258	0252	0256	0241	0235	0229
2,4	0,224	0219	0213	0208	0203	0198	0194	0189	0182	0180
2,5	0,0175	0171	0167	0162	0158	0154	0151	0147	0143	0139
2,6	0,0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0,0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0,0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0,0059	0058	0056	0054	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0037	0037	0036	0035	0034
3,1	0,0033	0032	0031	0030	0029	0028	0027	0026	0025	0024
3,2	0,0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0,0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0,0012	0012	0011	0011	0010	0010	0010	0010	0009	0009

3,5	0,0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0,0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0,0004	0004	0004	0004	0004	0003	0003	0003	0003	0003
3,8	0,0003	0003	0003	0003	0002	0002	0002	0002	0002	0002
3,9	0,0002	0002	0002	0002	0002	0002	0002	0001	0001	0001
4,0	0,0001	0001	0001	0000	0000	0000	0000	0000	0000	0000

* Тут і далі значення дано після коми.

Додаток 4

Друга функція нормованого відхилення $\frac{1}{2}f(t) = \int_0^t e^{-\frac{t^2}{2}}dt$ (інтеграл імовірностей)

Цілі і десяткові частки t	Соті частки t									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0040*	0080	0120	0159	0199	0239	0279	0319	0359
0,1	0,0398	0438	0478	0517	0596	0636	0675	0714	0714	0753
0,2	0,0793	0832	0871	0909	0948	0987	1026	1064	1103	1141
0,3	0,1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	0,1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	0,1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	0,2257	2291	2334	2356	2389	2421	2454	2486	2317	2549
0,7	0,2580	2611	2642	2673	2703	2734	2764	2793	2823	2852
0,8	0,2881	2910	2940	2967	2995	3023	3051	3078	3106	3133
0,9	0,3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	0,3413	3437	3461	3485	3508	3531	3554	3577	3599	3621
1,1	0,3643	3665	3686	3703	3729	3743	3770	3790	3810	3830
1,2	0,3849	3869	3888	3906	3925	3943	3962	3980	3997	4105
1,3	0,4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	0,4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	0,4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	0,4452	4463	4474	4484	4495	45005	4515	4525	4535	4545
1,7	0,4554	4564	4573	4582	4591	5599	4608	4616	4625	4633
1,8	0,4641	4648	4656	4664	4671	4678	4686	4693	4699	4706
1,9	0,4713	4719	4726	4732	4738	4744	4750	5548	4761	4767
2,0	0,4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	0,4821	5826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	0,4861	4864	4868	4871	4874	4878	4881	4884	4887	4890
2,3	0,4893	4896	4898	4901	4904	4906	4609	4911	4914	4916
2,4	0,4980	4920	4922	4924	4927	4929	4930	4932	4934	4936
2,5	0,4937	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	0,4953	4955	4956	4957	4957	4958	4960	4961	4962	4964
2,7	0,4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	0,4974	4975	4976	4977	4977	4978	4978	4979	4980	4981
2,9	0,4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3,0	0,4986	4987	4987	4987	4989	4989	4989	4989	4990	4990
3,1	0,4990	4991	4991	4991	4992	4992	4992	4992	4993	4993
3,2	0,4993	4993	4994	4994	4994	4994	4994	4995	4995	4995
3,3	0,4995	4995	4995	4996	4996	4996	4996	4996	4996	4996
3,4	0,4997	4997	4997	4997	4997	4997	4997	4997	4997	4998
3,5	0,4998	4998	4998	4998	4998	4998	4998	4998	4998	4998

3,6	0,4998	4998	4998	4999	4999	4999	4999	4999	4999	4999
3,7	0,4999	4999	4999	4999	4999	4999	4999	4999	4999	4999
3,8	0,4999	4999	4999	4999	4999	4999	4999	4999	4999	4999
3,9	0,9999	4999	5000	5000	5000	5000	5000	5000	5000	5000
4,0	0,5000	5000	5000	5000	5000	5000	-	-	-	-

* Тут і далі значення дано після коми.

Додаток 5

Функція нормованого відхилення (функція Лапласа) f(t)=P($|t| \le t_T$)

Цілі і десяткові частки t	Соті частки t									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0080*	0160	0239	0319	0399	0478	0558	0638	0717
0,1	0,0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	0,1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	0,2358	2434	2510	2536	2661	2737	2812	2866	2960	3035
0,4	0,2108	3182	3255	3228	3401	3473	3545	3616	3688	3759
0,5	0,3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	0,4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	0,5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	0,5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	0,6319	6372	6424	6476	6528	6579	6629	6779	6779	6778
1,0	0,6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	0,7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	0,7699	7737	7775	7813	7850	7887	7323	7959	7994	8029
1,3	0,8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	0,8385	8415	8444	8473	8501	8529	8557	8554	8611	8638
1,5	0,8664	8600	8715	8740	8764	8789	8812	8836	8859	8882
1,6	0,8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	0,9109	9127	9146	9164	9181	9199	9216	9233	9249	9265
1,8	0,9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	0,9426	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	0,9545	9556	9566	9576	9596	9696	9625	9516	9625	9534
2,1	0,9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	0,9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	0,9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	0,9836	9841	9845	9849	9853	9857	9861	9865	9868	9872
2,5	0,9876	9879	9883	9886	9882	9892	9895	9908	9904	9904
2,6	0,9907	9910	9912	9915	9917	9920	9922	9924	9926	9928
2,7	0,9931	9933	9935	9937	9939	9940	9942	9944	9946	9974
2,8	0,9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	0,9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	0,9973	9974	9975	9976	9976	9977	9978	9979	9979	9980
3,1	0,9981	9981	9982	9982	9982	9983	9984	9985	9985	9986
3,5	0,9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	0,9997	9997	9997	9997	9997	9997	9997	9997	9998	9998
3,7	0,9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	0,9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
…	…	…	…	…	…	…	…	…	…	…
4,0	0,9999	9999	9999	9999	9999	9999	9999	9999	9999	9999

* Тут і далі значення дано після коми.

Додаток 6

Значення ймовірностей P (χ^2) для критерію «*Xi* –квадрат»

χ^2	ν				
	1/15	2/16	3/17	4/18	5/19
1	0,3173 1,000	0,6065 1,000	0,8013 1,000	0,9098 1,000	0,9626 1,000
2	0,1574 1,000	0,3679 1,000	0,5724 1,000	0,7318 1,000	0,8491 1,000
3	0,0083 1,000	0,2231 1,000	0,3916 1,000	0,5578 1,000	0,7000 1,000
4	0455* 998*	1353* 999*	2615* 999*	4060* 000	5494* 000*
5	0254 992	0821 996	1718 998	2873 999	4159 999
6	0143 980	0498 988	1116 993	1991 996	3062 998
7	0081 958	0302 973	0719 984	1359 990	2206 994
8	0047 924	0183 949	0460 967	0916 977	1562 987
9	0027 878	0111 913	0293 940	0611 960	1091 974
10	0016 820	0067 867	0186 904	0404 932	6752 954
11	0009 753	0041 810	0117 857	0266 894	0514 824
12	0005 679	0015 744	0074 800	0174 847	0348 886
13	0003 602	0009 673	0046 736	0113 792	0234 839
14	0002 526	0006 599	0029 667	0073 729	0156 784
15	0001 451	0003 525	0018 596	0047 662	0104 723
16	0001 382	0002 453	0011 524	0030 592	0068 657
17	0000 319	0001 386	0007 454	0019 523	0045 590
18	- 263	0001 324	0004 389	0012 456	0029 522
19	- 214	0000 269	0003 329	0008 329	0019 457
20	- 172	- 220	0002 274	0005 333	0013 395
21	- 137	- 178	0001 226	0003 279	0008 337
22	- 108	- 143	0001 185	0002 232	0005 284
23	- 084	- 114	0000 150	0001 191	0003 237
24	- 065	- 090	- 119	0001 155	0002 196
25	- 050	- 070	- 095	0001 125	0001 161
26	- 038	- 054	- 074	0000 100	0001 130
27	- 029	- 042	- 053	- 079	0001 105
28	- 022	- 032	- 045	- 062	0000 083
29	- 016	- 024	- 034	- 048	- 066
30	- 012	- 018	- 026	- 037	- 052

* Тут і далі значення дано після коми.

Додаток 7

Критичні значення «*Хі* –квадрат»

ν	Рівень імовірності Р			ν	Рівень імовірності Р		
	0,95	0,99	0,999		0,95	0,99	0,999
1	3,8	6,6	10,8	21	32,7	39,9	46,7
2	6,0	9,2	13,8	22	33,9	40,3	48,3
3	7,8	11,3	16,9	23	35,2	41,6	49,7
4	9,5	13,3	18,5	24	36,4	43,0	51,3
5	11,1	15,1	20,5	25	37,7	44,3	52,6
6	12,6	16,8	22,5	26	38,9	45,6	54,1
7	14,1	18,5	24,3	27	40,1	47,0	55,5
8	15,5	20,1	26,1	28	41,3	48,3	56,9
9	16,9	21,7	27,9	29	42,6	49,6	58,3
10	18,3	23,2	29,6	30	43,8	50,9	59,7
11	19,7	24,7	31,3	32	46,2	53,5	62,4
12	21,0	26,2	32,9	34	48,6	56,0	65,20
13	22,4	27,7	34,5	36	51,0	58,6	67,9
14	23,7	29,1	36,1	38	53,4	61,6	70,7
15	25,0	30,6	37,7	40	55,8	63,7	73,4
16	26,3	32,0	39,3	50	67,5	76,2	86,4
17	27,6	33,4	40,8	60	79,1	88,4	99,6
18	28,9	34,8	42,3	70	90,5	100,4	112,3
19	30,1	36,2	43,8	80	101,9	112,3	124,8
20	31,4	37,6	45,3	90	113,1	124,1	137,1

Додаток 8

Критичні значення F- критерію при P=0,95

ν_1 ν_2	1	2	3	4	5	6	50
9	5,12	4,26	3,86	3,63	3,48	3,37	2,8
10	4,96	4,10	3,71	3,48	3,33	3,22	-
11	4,84	3,98	3,59	3,36	3,20	3,09	2,5
12	4,75	3,88	3,49	3,26	3,11	3,00	-
13	4,67	3,80	3,41	3,18	3,02	2,92	2,32
14	4,60	3,74	3,34	3,11	2,96	2,85	-
15	4,54	3,68	3,29	3,06	2,90	2,79	2,18
16	4,49	3,63	2,24	3,01	2,85	2,47	-
17	4,45	3,59	3,20	2,96	2,81	2,70	2,08
18	4,41	3,56	3,16	2,93	2,77	2,66	-
19	4,33	3,52	3,13	2,90	2,74	2,63	2,08
20	4,35	3,49	3,10	2,87	2,71	2,90	-
21	4,32	3,47	3,07	2,84	2,68	2,57	1,93
22	4,30	3,44	3,05	2,82	2,66	2,55	-
23	4,28	3,42	3,03	2,80	2,64	2,53	1,88
24	4,26	3,40	3,01	2,78	2,62	2,51	-
25	4,24	3,38	2,99	2,76	2,60	2,49	1,84
26	4,22	3,37	2,98	2,74	2,59	2,47	-
27	4,21	3,35	2,96	2,73	2,57	2,46	1,80
28	4,20	3,34	2,95	2,71	2,56	2,44	-
29	4,18	3,38	2,93	2,70	2,54	2,43	1,77
30	4,17	3,32	2,92	2,69	2,53	2,42	-
35	4,11	3,26	2,86	2,63	2,48	2,36	-
40	4,08	3,23	2,84	2,61	2,45	2,34	-
50	4,03	3,18	2,79	2,56	2,40	2,29	-
60	4,00	3,15	2,76	2,52	2,37	2,25	-
∞	3,84	2,99	2,60	2,37	2,21	2,09	-

Примітка. ν_1, ν_2 - число ступенів вільності варіації відповідно для великої і малої дисперсій.

Додаток 9

Критичні значення F- критерію при P=0,99

ν_1 ν_2	1	2	3	4	5	6	50
9	10,56	8,02	6,99	6,42	6,06	5,80	4,51
10	10,04	7,56	6,55	5,99	5,64	5,39	-
11	9,65	7,20	6,22	5,67	5,32	5,07	3,80
12	9,38	6,93	5,95	5,41	5,06	4,82	-
13	9,07	6,70	5,74	5,20	4,86	4,62	3,7
14	8,86	6,51	5,56	5,03	4,69	4,46	-
15	8,68	6,36	5,42	4,89	4,56	4,32	3,07
16	8,53	6,23	5,29	4,77	4,44	4,20	-
17	8,40	6,11	5,18	4,67	4,34	4,10	2,86
18	8,26	6,01	5,09	4,58	4,25	4,01	-
19	8,18	5,93	5,01	4,50	4,17	3,94	2,70
20	8,10	5,85	4,94	4,43	4,10	3,87	-
21	8,02	5,78	4,87	4,37	4,04	3,81	2,58
22	7,94	5,72	4,87	4,31	3,99	3,75	-
23	7,88	5,66	4,76	4,26	3,94	3,71	2,48
24	7,82	5,61	4,72	4,22	3,90	3,67	-
25	7,77	5,57	4,68	4,18	3,86	3,63	2,40
26	7,72	5,53	4,64	4,14	3,82	3,59	-
27	7,68	5,49	4,60	4,11	3,78	3,56	2,33
28	7,64	5,45	4,57	4,07	3,75	3,53	-
29	7,60	5,42	4,54	4,04	3,73	3,50	2,27
30	7,56	5,39	4,61	4,02	3,70	3,47	-
35	7,42	5,27	4,40	3,91	3,59	3,37	-
40	7,01	5,18	4,33	3,63	3,51	8,29	-
50	7,17	5,06	4,20	3,72	3,41	3,19	-
60	7,08	4,98	4,18	3,65	3,34	3,12	-
∞	6,64	4,60	3,78	3,32	3,02	2,80	-

Примітка. ν_1, ν_2- число ступенів вільності варіації відповідно для великої і малої дисперсій.

Додаток 10

Критичні значення критерію Кохрана

ν_1 ν_2	Рівень значущості 0,05										
	1	2	3	4	5	6	7	8	36	144	
2	9985*	9750	9392	9057	8772	8534	8332	8159	6602	5813	5000
3	9669	8709	7977	7457	7007	6771	6530	6333	4748	4031	3333
4	9065	7679	6841	6287	5895	5598	5365	5175	3720	3093	2500
5	8412	6838	5981	5441	5065	4783	4564	4564	4067	2513	2000
6	7803	6161	2521	4803	4447	4164	3980	3817	2612	2119	1667
7	7271	5612	4800	4307	3974	3726	3535	3384	2278	1833	1429
8	6798	5157	4377	3910	3595	3362	3185	3043	2022	1616	1250
9	6385	4775	4027	3584	3286	3067	2901	2768	1820	1446	1111
10	6020	4450	3733	3311	3029	2823	2666	2541	1655	1308	1000
12	5410	3924	3264	2880	2624	2439	2299	2187	1403	1100	0833
15	4709	3346	2758	2419	2195	2034	1911	1815	1144	0889	0667
20	3894	2705	2205	1921	1735	1602	1501	1422	0879	0675	0500
24	3434	2354	1907	1656	1496	1374	1287	1216	0743	0567	0417
30	2929	1980	1593	1377	1237	1137	1061	1002	0604	0457	0333
40	2370	1576	1259	1082	0968	0887	0827	0780	0462	0347	0250
60	1737	1131	0895	0765	0682	0625	0583	0316	0234	0167	0167
120	0998	0632	0495	0419	0571	0337	0292	0292	0165	0120	0083
∞	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

* Тут і далі значення дано після коми.

Додаток 11

Критичне значення коефіцієнта кореляції r_{yx}

Кількість одиниць спостереження	Рівень імовірності Р	
	0,95	0,99
10	0,5760	0,7079
11	0,5529	0,6835
12	0,5324	0,6614
13	0,5139	0,6411
14	0,4973	0,6226
15	0,4821	0,6055
16	0,4683	0,5897
17	0,4555	0,5751
18	0,4438	0,5614
19	0,4329	0,5487
20	0,4227	0,5368
25	0,3809	0,4869
30	0,3494	0,4487
35	0,3246	0,4182
40	0,3014	0,3932
45	0,2875	0,3721
50	0,2732	0,3541
60	0,2500	0,3248
70	0,2319	0,3017
80	0,2172	0,2830
90	0,2030	0,2673
100	0,1946	0,2540

Додаток 12

Критичне значення коефіцієнта детермінації R^2 (рівень імовірності 0,05)

Кількість одиниць спостереження	Кількість змінних							
	2	3	4	5	6	7	8	9
15	264*	393	495	582	659	729	791	896
16	247	369	466	550	624	692	754	860
18	219	329	417	494	566	628	687	792
20	197	297	378	449	514	574	630	731
22	179	270	345	411	471	527	530	677
24	164	248	317	379	435	488	538	630
26	151	221	294	351	404	454	501	588
28	140	213	273	327	377	424	468	551
30	130	199	256	306	353	397	439	518
40	097	149	192	231	268	303	337	399
50	078	120	155	186	216	244	271	323
60	065	100	130	156	181	205	228	272
80	048	075	097	117	136	154	172	205
100	038	060	078	094	109	124	138	169
200	019	030	039	047	055	062	069	083

*Тут і далі значення дано після коми.

Додаток 13

Площі кривих нормального розподілу

Цілі і десяткові частки t	Соті частки (t)									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,0040	0,0080	0,0120	0,0160	0,199	0,0239	0,0279	0,0319	0,035
0,1	0398*	0438	0478	0517	0557	0596	0636	0625	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1172	1808	1844	1879
0,5	1615	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2389	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	3413	3438	3461	3485	3508	3561	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
2,0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
3,0	4987	4987	4987	4988	4988	4989	4989	4989	4990	4990

* Тут і далі значення дано після коми.

Розділ критерію Дурбіна –Уотсона для додатної автокореляції (Р -0,95)

n	ν-1		ν-2		ν-3		ν-4		ν-5	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,96	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	0,86	1,85	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,99
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,047	1,73
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

НАВЧАЛЬНЕ ВИДАННЯ

Анатолій Трохимович ОПРЯ

СТАТИСТИКА

(модульний варіант з програмованою формою контролю знань)

НАВЧАЛЬНИЙ ПОСІБНИК

Оригінал-макет підготовлено
ТОВ «Центр учбової літератури»

Керівник видавничих проектів – Сладкевич Б. А.

Підписано до друку 24.06.2011. Формат 60х84 1/16
Друк офсетний. Папір офсетний. Гарнітура PetersburgCTT.
Умовн. друк. арк. 25,2.

Видавництво «Центр учбової літератури»
вул. Електриків, 23 м. Київ 04176
тел./факс 044-425-01-34
тел.: 044-425-20-63; 425-04-47; 451-65-95
800-501-68-00 (безкоштовно в межах України)
e-mail: office@uabook.com
сайт: www.cul.com.ua

Свідоцтво суб’єкта видавничої справи ДК № 2458 від 30.03.2006