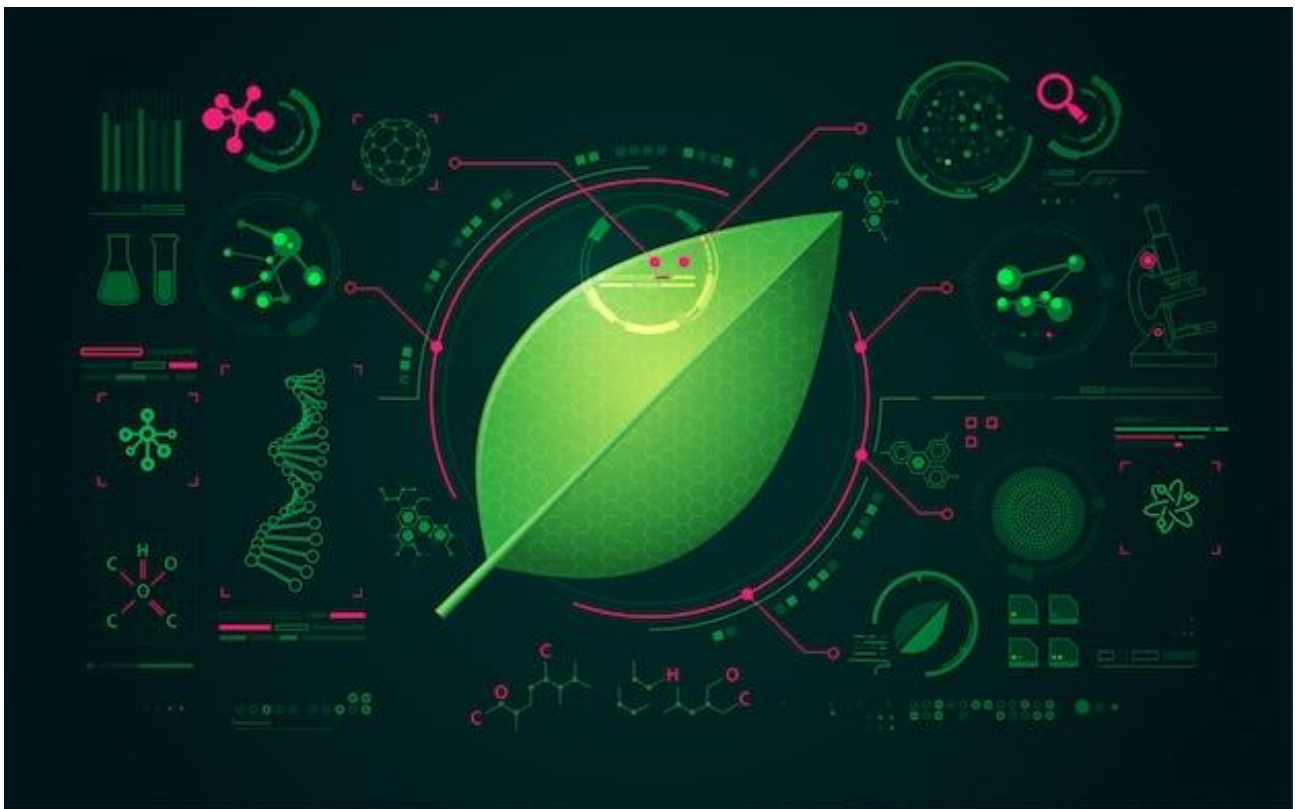


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МИКОЛАЇВСЬКИЙ НАЦІОНАЛЬНИЙ АГРАРНИЙ УНІВЕРСИТЕТ

КАФЕДРА ЕКОНОМІЧНОЇ КІБЕРНЕТИКИ, КОМП'ЮТЕКНИХ НАУК ТА
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

СТАТИСТИЧНИЙ АНАЛІЗ І МОДЕЛЮВАННЯ ПРОЦЕСІВ В АГРОНОМІЇ

Конспект лекцій для здобувачів третього освітньо-наукового рівня
вищої освіти ОНП «Агрономія» спеціальності **Н1 Агрономія** денної
форми здобуття вищої освіти



Миколаїв - 2025

Друкується за рішенням науково-методичної комісії факультету менеджменту Миколаївського національного університету від 20.11.2025 року протокол № 4.

Укладачі:

- О. В. Шебаніна – д-р екон. наук, професор, професор кафедри економічної кібернетики, комп’ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет;
- С. І. Тищенко – канд. пед. наук, доцент, доцент кафедри економічної кібернетики, комп’ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет;
- О. Ю. Пархоменко – канд. фіз.-мат. наук, доцент, доцент кафедри економічної кібернетики, комп’ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет;
- С.І. Ємельянов - доктор філософії, старший викладач кафедри економічної кібернетики, комп’ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет;
- В. О. Крайній – канд. екон. наук, старший викладач кафедри економічної кібернетики, комп’ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет;
- І. І. Хилько – старший викладач кафедри економічної кібернетики, комп’ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет.

Рецензенти:

Кравченко А.В. - заступник начальника відділу супроводження інформаційних систем, адміністрування баз даних, адміністраторів безпеки та адміністрування служби каталогів управління інформаційних технологій Головного управління ДПС у Миколаївській області

Христенко О.А. – канд. екон. наук, доцент, завідувач кафедри економіки підприємств факультету менеджменту, Миколаївський національний аграрний університет

ЗМІСТ

ВСТУП	4
ЛЕКЦІЯ 1. Особливості використання статистичних методів в агрономії.....	5
ЛЕКЦІЯ 2. Ознаки в агрономії, біології та ґрунтознавстві.....	12
ЛЕКЦІЯ 3. Ряди розподілу та їх статистичне оцінювання.....	36
ЛЕКЦІЯ 4. Аналіз тенденцій розвитку.....	50
ЛЕКЦІЯ 5. Кореляційний аналіз.....	59
ЛЕКЦІЯ 6. Регресійний аналіз.....	81
ЛЕКЦІЯ 7. Багатовимірний статистичний аналіз в агрономії.....	98
ЛЕКЦІЯ 8. Математичне формування задачі лінійного програмування	123
ЛЕКЦІЯ 9. Графічна інтерпретація задачі лінійного програмування....	134
ЛЕКЦІЯ 10. Ставлення і розв’язування транспортної задачі та її використання в агрономії.....	144
ЛЕКЦІЯ 11. Математичне моделювання виробничих процесів в агрономії.....	164
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ	186

ВСТУП

Ефективність та якість управління такою складною економічною системою, якою є сучасне виробництво сільськогосподарської продукції, в тому числі рослинництва, суттєво зростають при застосуванні статистичних методів, математичного моделювання та комп'ютерної техніки як інструментів дослідницької роботи. Стохастична природа факторів сільськогосподарського виробництва створює в сільському господарстві деяку невизначеність у динаміці розвитку та вимагає застосування дієвих інструментів і методів статистичного аналізу експериментальних даних. Використання специфічних методів статистики, математичного моделювання та новітніх інформаційних технологій дає змогу систематизації, обробки і аналізу даних спостережуваних явищ з метою встановлення притаманних для них статистичних закономірностей та застосування останніх у практичній діяльності в агрономії. Статистичні моделі використовують для діагностики стану об'єктів дослідження в агрономії, при вивченні причинно-наслідкового механізму формування варіації та динаміки агротехнологічних явищ і процесів, у моніторингу кон'юнктури ринку, при прогнозуванні та прийнятті рішень.

Для забезпечення розвитку та інтенсифікації сільського господарства необхідно широко впроваджувати сучасні досягнення науки і техніки. З метою обґрунтованого та швидкого прийняття оптимальних рішень, фахівці в галузі сільського господарства мають володіти сучасними методами статистичного аналізу і моделювання технологічних процесів в агрономії та виконувати достатньо складні розрахункові роботи, які без застосування обчислювальних засобів реалізувати важко.

Лекція 1. Особливості використання статистичних методів в агрономії

План

1. Предмет і метод статистики. Основні категорії статистичної науки.
2. Організація наукових досліджень в агрономії.
3. Математична статистика та різноманітність вирішуваних нею завдань.
4. Статистичне оцінювання. Концепція рандомізації.
5. Репрезентативність вихідного матеріалу. Поняття про помилки. Систематичні та випадкові помилки.

1. Предмет і метод статистики. Основні категорії статистичної науки

В сучасних умовах термін «статистика» має три значення:

1. **Статистика як цілеспрямована практична діяльність** по планомірному та систематичному збиранню, обробці, опрацюванню та поширенню даних про суспільні явища та процеси.

2. **Статистика як чисельні дані**, які всебічно характеризують різні аспекти розвитку суспільства.

3. **Статистика як суспільна наука**, і, відповідно, фахова навчальна дисципліна.

З точки зору вивчення у вищих навчальних закладах, курс статистики включає **теорію статистики, галузеву статистику та статистику окремих процесів**. Теорія статистики є основоположним розділом, що охоплює основні категорії, принципи та методи статистичної науки.

Суспільна наука та предмет дослідження

Статистика — це суспільна наука, яка вивчає **кількісний бік масових суспільних явищ і процесів** із врахуванням їх якісного змісту, місця і часу перебігу.

- Вона відокремилась у самостійну галузь знань завдяки специфічному об'єкту і предмету дослідження, а саме: **кількісні параметри соціально-економічного розвитку суспільства**.

- Статистика досліджує явища та процеси розвитку суспільства, що мають **масовий характер**.

- Вона досліджує якісно визначені параметри, котрі мають певний економічний або соціальний зміст, пов'язані з відповідним показником **часу та території**.

Методи статистики

Методологія науки — це її концептуальні засади, сукупність специфічних прийомів та способів дослідження об'єкту і предмету. У процесі статистичного дослідження використовуються:

- Загальнонаукові методи.
- Методи, запозичені з інших галузей знань (наприклад, математика).
- Статистичні методи, розроблені в процесі наукових досліджень і практичної діяльності.

Методи пристосовані до різних стадій дослідження:

1. **Перша стадія (збирання даних):** методи масового статистичного спостереження.

2. **Друга стадія:** методи зведення і групування, метод розрахунку узагальнений показників.

3. **Третя стадія (аналіз):** методи розрахунку відносних та середніх величин, методи вивчення варіації, взаємозв'язків, динаміки, індексний метод тощо.

Основні категорії статистичної науки

Специфічні концептуальні поняття, вироблені статистикою:

1. **Статистична сукупність:** достатньо велика кількість елементів або явищ суспільного життя, котрі мають як спільні, так й індивідуальні риси.

2. **Одиниця сукупності:** окремий елемент або явище, котрий разом з іншими утворює статистичну сукупність.

3. **Ознака:** властивість, характерна риса або особливість одиниці сукупності, яку можна спостерігати та вимірювати (оцінювати).

Якісні (атрибутивні): значення не мають чисельного виразу (наприклад, стать, галузь економіки).

○ **Кількісні:** мають чисельний вираз і відрізняються величиною. Можуть бути дискретними та інтервальними (наприклад, вік, ціна).

Альтернативні: можуть мати лише два взаємовиключних значення (наприклад, одержує стипендію / не одержує).

Варіація: коливання, мінливість або різноманітність значень ознаки у окремих одиниць статистичної сукупності. Окреме значення варіюючої ознаки називають **варіантом**.

Статистичний показник: кількісна характеристика явища або процесу в поєднанні з його якісною визначеністю.

Статистична закономірність: певна послідовність або повторюваність соціально-економічних явищ, яка проявляється в масових процесах. У складних системах закони проявляються лише у формі статистичних закономірностей.

Види закономірностей: зміни у часі (динаміки), розподілу одиниць, зміни складу та структури (структурні зрушення), взаємозв'язку.

Закон великих чисел

Закон великих чисел стверджує: в масі індивідуальних спостережень загальна статистична закономірність виявляється тим повніше і чіткіше, **чим більше одиниць сукупності охоплено спостереженням**.

2. Організація наукових досліджень в агрономії

Наукове дослідження — це вивчення конкретного об'єкта, явища або предмета з метою виявлення і розкриття закономірностей його виникнення і розвитку.

Етапи наукового дослідження

1. Аналіз існуючої наукової інформації.
2. Виявлення та обґрунтування наукової проблеми.
3. Формулювання вихідних гіпотез та їх теоретичний аналіз.

Планування, організація дослідів та його проведення.

1. Аналіз та узагальнення результатів дослідження.
2. Перевірка вихідних гіпотез, остаточне формулювання нових закономірностей і законів, наукові передбачення.

Впровадження пропозицій у виробництво.

Рівні досліджень

Розрізняють три основних рівні досліджень:

Рівень	Характеристика	Приклад в агрономії
Емпіричний (експериментальний)	Проводять експерименти, накопичують факти, аналізують їх і роблять практичні висновки. Експерименти можуть бути якісними (пошкоджені/не пошкоджені) і кількісними (врожайність, ріст рослин).	Облік показників росту рослин, урожайності.
Теоретичний	Синтезуються знання, формулюються загальні закономірності. Теорія — це система узагальнених знань, що є новим ступенем пізнання.	Формулювання загальних законів живлення рослин на основі результатів польових дослідів.
Описово-узагальнюючий	Експерименти не проводять, а описують явища, що спостерігаються безпосередньо у природі (поза експериментом).	Спостереження за ростом та розвитком рослин залежно від погодних умов.

Фундаментальні та прикладні дослідження

Поділ досліджень на фундаментальні та прикладні є умовним, оскільки вони часто переходять одне в інше.

- **Фундаментальні дослідження** спрямовані на відкриття і вивчення нових явищ і законів природи; їх результатом є закінчена система наукових знань.

Прикладні дослідження (зокрема в агрономії) спрямовані на вивчення факторів життя рослин, взаємозв'язків між рослиною і середовищем, створення нових сортів/гібридів. Мета — **розробка ефективних агрозаходів** підвищення врожайності та якості продукції. Кінцева мета — **рекомендації для впровадження одержаних результатів у виробництво**.

Основна мета наукових досліджень в агрономії: формування знань і умінь з сучасних методів агрономічних досліджень, планування, техніки закладання і проведення експериментів, **статистичного оцінювання** результатів досліджень.

3. Математична статистика та різноманітність вирішуваних нею завдань

Основні методи статистичного оцінювання результатів агрономічних досліджень здійснюються за допомогою **методів математичної статистики**.

Математична статистика — це галузь математичних знань, що розробляє раціональні прийоми систематизації, обробки і аналізу даних статистичних спостережень масових явищ з метою встановлення характерних для них статистичних закономірностей.

- Більшість методів ґрунтується на **імовірнісній природі** даних.
- Вона абстрагується від матеріального змісту явищ, озброюючи дослідника **математичним апаратом**.

Найважливіші розділи та методи

- Статистичні ряди розподілу.
- Оцінка параметрів розподілу.
- Закони розподілу вибіркового характеристик.
- Перевірка статистичних гіпотез.
- Дисперсійний, кореляційно-регресійний, коваріаційний аналіз.

Останнім часом також використовують методи **багатомірного статистичного аналізу**:

- Факторний аналіз.
- Метод головних компонент.
- Кластерний аналіз.

Основні завдання математичної статистики

Встановлення законів розподілу різних випадкових змінних.

Оскільки аналізована сукупність є вибіркою, потрібно знати, які риси досліджуваних явищ є **стійкими**, а які **випадковими**.

Перевірка статистичних гіпотез. Дослідник висуває гіпотезу про характер закономірності, притаманної явищу, і перевіряє, чи підтверджують дані спостереження цю гіпотезу (наприклад, чи дорівнює середня величина вибірки середній для всієї сукупності).

Оцінка невідомих параметрів різних розподілів. Оскільки дослідник має справу з частиною сукупності (вибіркою), рівень деяких статистичних характеристик для всієї сукупності (наприклад, середня, дисперсія) залишається невідомим, і їх потрібно оцінити.

4. Статистичне оцінювання. Концепція рандомізації

Статистична сукупність: маса однорідних елементів, які мають єдину якісну основу, але різняться певними ознаками і підлягають певному закону розподілу.

Генеральна сукупність (ГС): вся множина однорідних об'єктів чи подій, які є предметом інтересу. ГС може бути реально існуючою або гіпотетичною, скінченною чи нескінченною.

Основна мета статистичного оцінювання — визначення дійсних параметрів генеральної сукупності на основі вивчення **вибіркових показників**.

Репрезентативність та рандомізація

Щоб вибірка достатньо добре відтворювала властивості ГС (була **репрезентативною**), використовують спеціальні методи формування вибірки.

Рандомізація: розташування або вибір об'єктів у **випадковому порядку**.

- Дозволяє **нівелювати систематичні впливи факторів**, що не контролюються, та забезпечити об'єктивність.
- Застосовується для вибору порядку чергування дослідів при плануванні експериментів.
- В основу технологій рандомізації покладено процес генерації послідовності **псевдовипадкових чисел**.

Типи вибірок

Найпоширенішими вважаються:

Проста рандомізована вибірка: формується за системою відбору, що гарантує **рівну ймовірність** попадання кожного об'єкта у вибірку.

1. Систематична рандомізована вибірка: відповідає принципам рівноімовірнісного відбору; оптимальна для великих списків об'єктів.

Наприклад, вибирається перший випадковий об'єкт, а потім кожен k -й наступний.

Стратифікована вибірка: гарантує репрезентативність заздалегідь обраних властивостей, зберігаючи **пропорційне представництво** підгруп (страт) у вибірці, адекватне ГС (наприклад, 60% жінок і 40% чоловіків).

Кластерна вибірка: передбачає поетапний вибір **груп (кластерів)**, а не окремих елементів. Це здатне вирішувати проблему неповноти складу підгруп ГС. Менш надійна через наявність декількох етапів відбору, кожний з яких додає похибку.

Параметри та статистики

Генеральна сукупність (обсяг N) та **вибірка** (обсяг n) характеризуються тими ж показниками:

- Для ГС вони мають назву «**параметри**».
- Для вибіркової сукупності — «**статистики**».

Статистична оцінка — це вибіркова статистика, що використовується як наближене значення невідомого параметра генеральної сукупності.

- Оцінка є **випадковою величиною**, оскільки залежить від результатів спостережень $x_{\{1\}}, x_{\{2\}}, \dots, x_{\{n\}}$, закону розподілу та обсягу вибірки n .

5. Репрезентативність вихідного матеріалу. Поняття про помилки.

Помилка репрезентативності

Репрезентативність — властивість вибіркової сукупності представляти характеристику генеральної сукупності.

Помилка репрезентативності — ступінь відхилення статистичної структури вибірки від структури ГС. Це розбіжність між ГС (на яку спрямований теоретичний інтерес) і вибірковою сукупністю (об'єктом обстеження).

- Цей термін також називають «**помилка вибірки**».
- **Помилка вибірки** — відхилення середніх характеристик вибіркової сукупності від середніх характеристик генеральної сукупності.

Помилка вибірки обумовлюється:

1. Методом формування вибірки.
2. Розміром вибірки.

Типи помилок вибірки

Помилки вибірки діляться на **випадкові й систематичні**.

Тип помилки	Характеристика	Приклади виникнення
Випадкова помилка	Ймовірність того, що вибіркова середня вийде (або не вийде) за межі заданого інтервалу. Це статистичні погрішності, властиві самому вибіркового методу.	Пов'язані з природою будь-яких статистичних погрішностей.
Систематична помилка	Результат діяльності самого дослідника. Вони найнебезпечніші, оскільки призводять до значних зсувів результатів і не піддаються контролю й виміру. Не зменшуються при зростанні обсягу вибірки.	1) Вибірка не відповідає завданням дослідження. 2) Незнання характеру ГС. 3) Відбираються тільки «виграшні» елементи. 4) Порушення правил проведення дослідження, неадекватні способи формування вибірки.

Уникнення помилок

Систематичні помилки легше **попередити**, ніж усунути.

Прийоми уникнення помилок:

- Кожна одиниця ГС повинна мати **рівну ймовірність** потрапити у вибірку.
- Відбір бажано робити з **однорідних сукупностей**.
- Треба знати **характеристики ГС**.
- Враховувати випадкові й систематичні помилки.

Збільшувати розміри вибірки (найпростіший спосіб зменшити ймовірність помилки).

Якщо вибірка складена неправильно, помилка на кожному наступному етапі **примножується**, що може переважити цінність дослідження.

Контроль вибірки — процес наукового порівняння ГС і вибіркової сукупності, виявлення ступеня їхньої розбіжності та причин відхилення.

Ремонт вибірки — процес усунення погрішностей для підвищення якості вже зібраної інформації.

Лекція 2. Ознаки в агрономії, біології та ґрунтознавстві

План.

1. Ознаки та властивості об'єктів і їх дослідження. Кодування якісних ознак.
2. Поняття та класифікація інформаційного забезпечення наукових досліджень.
3. Сезонність в агрономії.
4. Методи раціонального планування експериментів та польових дослідів.
5. Представлення вихідного матеріалу у формі таблиць. Побудова таблиць.

1. Ознаки та властивості об'єктів і їх дослідження. Кодування якісних ознак

Кожне окремо взяте явище даної сукупності (елемент множини) називається одиницею сукупності. Одиниці сукупності характеризуються певними властивостями, які називаються ознаками. Наприклад в агрономії це будуть характеристики колосу: остистість колосу, його забарвлення та опушеності колоскових лусок, забарвлення остей і зерна, відповідно види пшениці поділяють на різновиди, число яких дуже велике. Ці ознаки, будучи якісно однорідними для всіх одиниць сукупності, різні за своїм значенням в кожній з одиниць сукупності.

За формою вираження ознаки діляться на:

- кількісні – ті, які можна виразити кількісно (числом);
- атрибутивні (якісні), які числом виразити не можна (забарвлення).

Кількісні значення ознаки при переході від однієї одиниці сукупності до іншої змінюються, варіюють. Тому кількісні зміни значень ознак в одиниць сукупності називаються варіацією, а самі ознаки – варіюючими.

Атрибутивні ознаки теж варіюють, але ця варіація в чітко обмежених межах, тому, як варіюючі, розглядаються, в основному, – кількісні. Варіація виникає в результаті впливу на індивідуальне значення ознаки різних чинників, перш за все зовнішніх і випадкових. Варіація ознаки відображається в статистичному розподілі, який має велике наукове і практичне значення. Саме наявність варіації і зумовлює необхідність статистики.

Зареєстровані у процесі масового статистичного спостереження значення ознак відбивають увесь діапазон об'єктивно існуючої в сукупності варіації. У розмаїтті поодиноких відомостей губиться загальне, у неістотному і випадковому – закономірне. Перехід від одиничного до загального відбувається завдяки зведенню.

Суть статистичного зведення полягає в тому, що матеріали спостереження класифікують та агрегують. Елементи сукупності за певними ознаками об'єднують у групи, класи, типи, а інформацію про них агрегують як у межах груп, так і в цілому по сукупності. Основне завдання зведення – виявити типові риси та закономірності масових явищ чи процесів.

Зведення є основою подальшого аналізу статистичної інформації. За зведеними даними обчислюють узагальнюючі показники, виконують порівняльний аналіз, а також аналіз причин групових відмінностей, вивчають взаємозв'язки між ознаками.

Складові статистичного зведення такі:

- 1) розробка програми систематизації та групування даних;
- 2) обґрунтування системи показників для характеристики груп і сукупності в цілому;
- 3) проектування макетів таблиць, в яких подаються результати зведення;
- 4) визначення технологічних схем обробки інформації, програмного забезпечення;

5) підготовка даних до обробки на комп'ютері, формування автоматизованих банків даних;

6) безпосереднє зведення, узагальнення, розрахунок показників.

Програма систематизації та групування даних передбачає вибір групувальних ознак і правил формування груп. Розробка програми, як і обґрунтування системи показників, залежить від мети дослідження, суті явища, яке вивчається, особливостей сукупності, ступеня варіації групувальних ознак.

Результати статистичного зведення подаються у формі статистичних таблиць, макети яких розроблюються разом з програмою обробки даних. Макет статистичної таблиці – це комбінація горизонтальних рядків і вертикальних граф, на перетині яких утворюються клітинки. Ліві бічні та верхні клітинки призначені для словесних заголовків – переліку складових сукупності та системи показників, решта – для числових даних.

На практиці використовуються різні технологічні схеми комп'ютерної обробки первинних даних. Спільними для всіх є дві операції: кодування даних і перенесення їх із документів на технічні носії інформації.

Коди – це умовні ідентифікатори ознак. Централізовано розроблені єдині класифікатори, оформлені у вигляді словників, забезпечують однозначність кодів. Сформовані автоматизовані банки даних уможливають багаторазове використання інформації, при цьому поглиблюється аналіз, зменшується кількість помилок.

За формою обробки даних зведення бувають централізованими та децентралізованими.

У статистичній практиці інформація обробляється переважно децентралізовано. Так, у разі обробки статистичної звітності зведення здійснюється від нижчої до вищої ланки управління: звіти підприємств зводяться регіональними статистичними органами, підсумки по регіонах

передаються в Державну службу статистики, де узагальнюються в цілому по країні. Характерні класифікаційні позиції, «розрізи» зведення такі: територіальна ознака (область, місто, район), підпорядкованість (міністерство, відомство), галузі господарської діяльності, форми власності.

Поділ сукупностей на групи, однорідні в тому чи іншому розумінні, пов'язаний з такими діями, як систематизація, типологія, класифікація, групування. Традиційно зазначений поділ виконують за такою схемою: із множини ознак, які описують явище, добирають розмежувальні, а потім сукупність поділяють на групи та підгрупи відповідно до значень цих ознак.

Головний принцип будь-якого поділу ґрунтується на двох положеннях:

- 1) в один клас, групу об'єднуються елементи певною мірою подібні між собою;
- 2) ступінь подібності між елементами, які належать до одного класу, значно вищий, ніж між елементами, що належать до різних класів.

У кожному конкретному дослідженні вирішуються три питання:

- 1) що взяти за основу групування;
- 2) скільки груп, позицій необхідно виокремити;
- 3) як розмежувати групи.

Основою групування може бути будь-яка атрибутивна чи кількісна ознака, що має градації. Таку ознаку називають групувальною. Залежно від складності масового явища (процесу) та мети дослідження групувальних ознак може бути одна, дві й більше.

У статистичній практиці часто вдаються до розбиття сукупностей за атрибутивними ознаками – класифікації та номенклатури. Їх розробляють міжнародні й національні статистичні органи та рекомендують як статистичний стандарт. Здебільшого йдеться про багатоступеневу класифікацію з докладною

номенклатурою груп підгруп, із чітко визначеними вимогами та умовами віднесення елементів сукупності до тієї чи іншої групи.

У міжнародній статистиці відома галузева стандартна класифікація видів економічної діяльності, класифікація занять, стандартна торговельна класифікація.

Кожній класифікаційній позиції надається код (шифр), який замінює її назву і є постійним засобом ідентифікації під час передавання інформації по каналах зв'язку, комп'ютерної обробки тощо.

Кожна класифікація є сталою, забезпечуючи порівнянність даних у просторі та часі.

Поряд з класифікацією для висвітлення певних аспектів конкретного дослідження використовують групування, на яке покладаються такі аналітичні функції:

- 1) вивчення структури та структурних зрушень;
- 2) визначення типів соціально-економічних явищ, виокремлення однорідних груп і підгруп;
- 3) виявлення взаємозв'язків між ознаками.

2. Поняття та класифікація інформаційного забезпечення наукових досліджень

Основою будь-якого наукового дослідження є інформація – сукупність відомостей (повідомлень, даних), яка визначає міру наших знань про ті чи інші явища, події та їх взаємозв'язки. Дане визначення використовується у широкому розумінні слова. У вузькому розумінні інформація – це відомості, які є об'єктом обробки, передачі і зберігання. Інформація є основним поняттям кібернетики – науки про загальні закономірності в процесі управління та передачі інформації.

Якість та ефективність інформації у науковому дослідженні визначається за такими критеріями: цілеспрямованість, цінність, своєчасність, достовірність, достатність і комплексність (повнота), швидкодіяльність, дискретність, неперервність, періодичність надходження, детермінований характер, доступність (зрозумілість), спосіб і форма подання.

Перш за все, дослідник повинен встановити цільове призначення інформації, оскільки одна і та ж інформація може використовуватися для різних цілей: створення нових концепцій, встановлення і вирішення проблем пошуку тощо. Цінність інформації визначається економічним ефектом, який дає її використання. Практичним завданням, що стоїть перед дослідником, є визначення того, яка інформація йому необхідна. Разом з тим, потрібно виключити надлишкову інформацію, яка не має прямого відношення до об'єкту дослідження.

Всі елементи дослідницької діяльності тісно пов'язані із збереженням, переробкою і зберіганням інформації (рис. 1).



Рисунок 1 – Зв'язки дослідницької інформаційної діяльності

Інформацію класифікують за різними ознаками:

За ступенем наукової новизни розрізняють:

а) нову інформацію, що відображає новизну запропонованого рішення теоретичного або практичного завдання;

б) релевантну, яка раніше містилась в аналогах (наприклад, в методичних вказівках).

За призначенням виділяють:

а) повідомлювальну інформацію, що отримана в процесі дослідження;

б) управлінську інформацію, яка необхідна для прийняття управлінських рішень.

За тривалістю періоду, протягом якого інформація зберігає свою актуальність і використовується для прийняття рішень, інформацію класифікують на:

а) теоретичну (наукову) інформацію – це результати фундаментальних чи прикладних наукових досліджень в різних областях, які широко використовуються у виробництві та управлінні;

б) стратегічну – інформація, що зберігає актуальність протягом тривалих періодів (10-15 років): довготривалі плани і прогнози, дані про повільно змінювані об'єкти, проектно-конструкторська документація;

в) тактичну (кон'юнктурну) – інформація з періодом актуальності 2-3 роки і менше;

г) оперативну – інформація, що актуальна в межах одного циклу оперативного управління.

Залежно від об'єкту, який відображає інформацію, вона буває:

а) природньонаукова – характеризує зв'язки між природними об'єктами;

б) техніко-технологічна – відображає взаємозв'язки між предметами природи, які стосуються технології та технічних засобів;

в) економічна – розкриває відносини між людьми в процесі виробництва, розподілу, обміну і споживання;

г) соціально-політична – інформація про соціальні, політичні, ідеологічні відносини між людьми.

Залежно від того, що в об'єкті відображається, інформація буває наступних видів:

а) законодавчі акти, документи уряду, положення, інструкції різних органів управління;

б) дані демографічних та соціологічних досліджень;

в) матеріали економічних теорій;

г) дані про рівень розвитку техніки, технології і тенденції їх розвитку;

д) інформація про господарські зв'язки;

є) інформація про процеси виробництва;

є) інформація про фактори виробництва; ж) інформація про макроекономічні процеси.

В сучасних умовах, із посиленням вимог до обґрунтованості наукових досліджень, зростає і роль інформації. Роль інформації важлива на всіх етапах дослідження: при виборі і конкретизації теми, вивченні історії питання, створенні гіпотези і т.д. Але найбільшу роль відіграє інформація у формуванні змісту майбутньої роботи. Залежно від складу та якості зібраної інформації може змінюватись не тільки план роботи, але і напрямок самого дослідження. В зв'язку з цим не будь-яка інформація може бути корисна для даного дослідження. Саме тому відбір найбільш значущої для даного дослідження інформації, вміння визначити її місце в ньому – необхідні умови правильного вибору змісту інформації.

Основна роль інформації в дослідженнях полягає в тому, щоб виключити суб'єктивні висновки, дати можливість отримати оптимальне рішення проблеми. Рівень наукових досліджень залежить від достовірності, ступеня

використання інформації і здатності дослідника переробити отриману інформацію. Детальніше дослідження цих зв'язків потребує вирішення питання про те, які функції повинна виконувати інформація. Такими функціями є інформативна, стимулююча та орієнтуюча.

Суть інформативної функції полягає в тому, щоб дати знання, відомості про той чи інший об'єкт і предмет дослідження. Реалізація стимулюючої функції дозволяє привести дослідників до нової постановки питання, нового його вирішення, з тим, щоб вдосконалювати практику. Орієнтуюча функція відображається у положеннях, нормах, цільових настановах, які дослідники сприймають як обов'язкову суспільну регламентацію, щоб в найкоротший термін досягти необхідних наукових результатів. Всі функції інформації взаємопов'язані і в поєднанні сприяють розвитку творчості у дослідній діяльності.

3. Сезонність в агрономії

В аграрному виробництві України рослинництво є однією з головних галузей, на яку і супутню їй галузь кормовиробництво припадає близько 93% орних земель країни, до 30% з яких відведено під кормові культури. У рослинництві 40-50 % становить побічна продукція – солома хлібів, стебла кукурудзи й сорго, жом, патока та інші, які через кормовиробництво використовуються у тваринництві. Тому гармонійне поєднання рослинництва, тваринництва і кормовиробництва – необхідна умова успішного функціонування всього аграрного комплексу країни. Отже, важливою особливістю рослинництва є сезонність цієї галузі – польові культури здатні рости і давати врожай тільки у безморозний період. Ґрунтово-кліматичні умови України досить різноманітні по зонах через неоднакові ґрунтові покриви, кількість опадів і тепла, тривалість вегетаційного періоду, умови перезимівлі, що свідчить про необхідність враховувати екологічні та біологічні особливості сільськогосподарських культур при їх розміщенні в системі землекористування.

Сезонні коливання – це більш чи менш сталі внутрішньорічні коливання в ряді динаміки, що обумовлені специфікою виробництва або споживання того чи іншого виду товару чи послуг. Для організації виробництва і реалізації продукції сезонних виробництв надзвичайно важливо вивчити тенденцію сезонних коливань, що склалися, і розробити прогноз на найближчу перспективу, головним чином, на наступний рік.

Для вивчення сезонних коливань використовуються спеціальні показники, які називаються індексами сезонності, а сукупність їх утворює сезонну хвилю. Наочне подання сезонних коливань здійснюється шляхом побудови графіка сезонної хвилі, у якому по осі Х відкладають періоди часу на протязі року, а по осі У — значення індексів сезонності.

Індексом сезонності називають процентні відношення фактичних внутрішньорічних рівнів до постійної або змінної середньої ряду:

$$i_c = \frac{\bar{y}_i}{\bar{y}}; \bar{y}_i = \frac{\sum_i y_i}{k}; \bar{y} = \frac{\sum_j \sum_i y_{ij}}{k \cdot n}.$$

де $\bar{y}_i$ – середнє значення показника за прийнятий відрізок часу;

$\bar{y}$ – середнє значення показника за весь період;

k - кількість років ($k= 1, 2, 3,4, \dots$);

n - кількість місяців ($n = 1, 2, 3, \dots$).

Розподіл сезонної хвилі спостерігають за декілька років (не менше трьох) у розрізі періодів року з метою відокремлення впливу по одному року.

Головною метою є те, щоб фактичні дані були оброблені таким чином, щоб можна було відобразити тенденцію.

При цьому.

1. Якщо ряд відображує тенденцію – аналітичне вирівнювання рядів динаміки. Формують рівняння динаміки. Після чого обчислюють фактичні дані у відсотках до вирівняних, а індекс сезонності є середнім по одноіменним внутрішньорічним періодам із цих процентних чисел за взяті роки.

2. Якщо ряд не має тенденції в розвитку індекс сезонності обчислюється по емпіричним даним.

4. Методи раціонального планування експериментів та польових дослідів

Від якості планування досліджень залежать достовірність, точність та ефективність експерименту. На сучасному етапі досліджень для планування дослідів застосовують методи математичної статистики з широким використанням комп'ютерної техніки.

1. Значення та завдання планування. Уперше математичне планування дослідів було здійснено наприкінці 20-х років минулого століття автором дисперсійного аналізу англійським математиком Р. Фішером. Таке планування підвищує надійність експерименту, дає змогу зменшити кількість дослідних варіантів і розмір дослідів, знайти оптимальні варіанти та підвищити продуктивність праці дослідника. Отже, математичне планування надзвичайно перспективним процесом у дослідній роботі.

Основне завдання планування – пошук оптимальних умов росту і розвитку рослин з метою підвищення їх продуктивності. Припустимо, що дослідник у попередніх дослідях мав підвищення врожаю від певних градацій досліджуваного фактору на 15% порівняно з контролем. Але приріст урожайності не задовольняє дослідника і спонукає його до вибору оптимальної градації того фактору, який вивчається. Вирішення таких завдань називають процесом оптимізації.

Наприклад. X – діючий фактор (удобрення, зрошення, обробіток ґрунту тощо): Y – результат цієї дії (врожай, його якість). Це є параметром оптимізації, тобто критерієм оптимальності, цільовою функцією.

Математична модель або рівняння, що пов'язує параметр оптимізації з діючими факторами, має такий вигляд:

$$Y = f(X_1, X_2, \dots, X_n),$$

де $f(X_1, X_2, \dots, X_n)$ – функція відгуку, у якому $X_1, X_2, \dots, X_n$ – діючі фактори.

Градації кожного фактору або його дози називають рівнями фактору. Набором рівнів для кожного фактору визначається кількість варіантів у досліді. Якщо кількість рівнів для всіх факторів однакова, то кількість варіантів досліду дорівнює кількості рівнів, піднесених у число факторів. При двох факторах та трьох рівнях кожного з них кількість варіантів у досліді буде $3^2 = 9$. При п'яти рівнях кожного з п'яти факторів в одному досліді буде $5^5 = 3125$ варіантів. Оскільки закласти дослід з такою кількістю варіантів практично неможливо, насамперед виключають ті з них, які є проміжними і менш ефективними. Але таке виключення не повинно бути суб'єктивним, його треба робити із застосуванням методів математичної статистики, про що йтиметься далі.

Математичне планування експерименту застосовують лише для дослідів, результати яких можна відтворити, а фактори можна регулювати. Такими факторами є сорт, гібрид, удобрення, обробіток ґрунту, глибина і строки сівби, схеми садіння тощо. До факторів, які мало регулюються, належать температура повітря і ґрунту, освітлення та ін. Однак у фітотронах ці та інші фактори повністю регулюються. Фактори, які не можна регулювати у полі – атмосферні опади, температура. Вони певною мірою змінюють процес відтворення результатів, тому у Такому разі звертаються до так званого активно-пасивного експерименту, коли зв'язки між факторами, які не регулюються, та параметрами оптимізації визначають лише за результатами спостережень.

Планування досліджу – це насамперед вибір мінімальної кількості варіантів та умов проведення досліджу з метою оптимізації. При цьому користуються двома підходами:

1) побудовою фізичної моделі процесу на основі відомих явищ (фізики, ґрунту, фізіології рослин, біології, хімії та ін.), що дає змогу мати математичну модель об'єкта досліджень у вигляді системи диференціальних рівнянь;

2) статистичним підходом, який доповнює перший. Математична модель експерименту – це рівняння, що пов'язує параметри оптимізації з факторами життя рослин.

2. Вибір параметрів досліджу. Параметр – це те, що потрібно оптимізувати, тобто це реакція на фактори, яких може бути кілька. Параметрами можуть бути врожай, його якісні показники, морозостійкість, посухостійкість рослин та їх стійкість проти шкідників, хвороб тощо. Знайти оптимальні умови для рослин легше тоді, коли правильно вибраний єдиний параметр оптимізації. При цьому всі інші параметри обмежені. Якщо єдиний параметр вибрати неможливо, вибирають узагальнений параметр оптимізації як функцію від багатьох вихідних. Слід зазначити, що правильний вибір параметра – це одна з основних умов математичного планування. Параметри оптимізації повинні відповідати таким вимогам: - параметри мають бути вимірюваними. Якщо їх не можна виміряти (наприклад, якісний параметр – стійка проти посухи рослина або нестійка), то для їх вираження користуються ранговим підходом. При цьому параметрам привласнюють ранги за шкалами: двобальний, п'ятибальний, десятибальний і т. д. Для двобальної шкали ранговий параметр має обмежену область визначення – «так» і «ні», добрий або поганий стан рослин, уражуються рослини хворобами чи ні тощо. Однак ранговий підхід більш грубий, ніж безпосереднє вимірювання кількісних параметрів (маса врожаю, висота рослин, площа листя та ін.); - параметр має бути виражений одним числом. Якщо параметр виражається як співвідношення, наприклад, вмісту азоту до фосфору 3:2, то його записують числом 1,5; -

параметр повинен бути однозначним статистично, тобто певному набору факторів має відповідати лише одне число параметра; - параметр повинен бути досить точним статистично. Якщо точність недостатня, то збільшують кількість повторностей; - параметр має бути універсальним і повним, тобто він повинен всебічно характеризувати об'єкт вивчення. Універсальним є параметр, який подається функцією кількох окремих; - кожний параметр оптимізації повинен мати фізичний зміст.

Одночасно можна оптимізувати лише одну функцію, найголовнішу з усіх. При цьому розраховують коефіцієнти парних кореляцій між основним параметром та іншими другорядними. Якщо зв'язок виявиться сильним, то другорядний параметр виключають. Також виключають параметри, які важче виміряти або зміст яких менш зрозумілий.

3. Вибір факторів досліджу. На врожай і його якість, стійкість рослин проти хвороб, шкідників їх морозо- та посухостійкість впливають різні фактори: освітлення, сорт, вологість ґрунту і повітря, температура ґрунту і повітря, повітряний і поживний режими ґрунту, обробіток ґрунту та ін. Але при математичному плануванні враховують основний з них, який в даних умовах є найбільш дієвим.

Вибраний фактор повинен задовольняти такі вимоги:

- має бути регульованим (зміна дози добрив, норми поливу, сівби, глибини оранки тощо). Температура і вологість повітря, освітлення у відкритому ґрунті – це фактори, які не можна повністю регулювати, тому їх не використовують для математичного планування польового досліджу; - щоб фактор можна було виміряти з достатньо високою точністю: - бажано, щоб фактор був однозначним; - щоб при вивченні сукупності кількох факторів їх можна безпечно поєднувати; - вибраний фактор не повинен залежати від інших, тобто між ними не повинно бути лінійної залежності (допускається криволінійний зв'язок).

4. Вибір моделі дослід. Правильно вибрана математична модель дослід дає змогу передбачити навіть ті оптимальні варіанти, які в досліді не вивчалися. Для цього користуються кроковим принципом на поверхні відгуку. Поверхня відгуку багатофакторного дослід має такі властивості як безперервність, гладкість та наявність єдиного оптимуму в певних точках даної поверхні. Якщо відомі значення параметрів у сусідніх точках поверхні відгуку, то в інших (сусідніх) можна математичними розрахунками передбачити значення іншого параметра. Так знаходять нові варіанти, яких у досліді не було і які можуть бути ефективними. Після проведення нового дослід з включенням нових варіантів знову послідовно за допомогою математичних розрахунків визначають ефективніші варіанти дослід, користуючись кроковим принципом. Якщо розраховують точки, які лежать на поверхні відгуку, то пошук називають інтерполяцією, а якщо за її межами – екстраполяцією. Чим ближче точки до області експерименту, тим точніше передбачення оптимальних варіантів.

Наприклад, для визначення залежності врожайності польових культур від діючих факторів (сорт, удобрення, зрошення тощо) будують кілька найбільш сприятливих моделей та перевіряють їх придатність. З кількох моделей використовують ту, яка за математичним виразом найпростіша, і це називають перевіркою адекватної моделі.

Якщо рух на поверхні відгуку не веде у стаціонарну область, переходять до поліномів більш високих ступенів. Те саме роблять, якщо залежність між факторами та їх параметрами криволінійна.

Суть пошуку оптимуму така:

- 1) проводять дослід з невеликими схемами;
- 2) за результатами цих дослідів будують математичні моделі, з яких вибирають найбільш придатні;
- 3) рухаючись у напрямі, який поліпшує параметр, знаходять оптимальний варіант;

4) знову закладають досліди, будують нові моделі і знаходять більш ефективні варіанти і т. д. Це і є оптимізацією планування.

Планування схем дослідів. Головним завданням планування є розробка програми досліджень. Наукове дослідження складається з таких діалектично пов'язаних складових: планування експерименту, проведення польових дослідів і виконання комплексу обліків і спостережень, аналізу й узагальнення отриманих даних.

Найбільш відповідальною частиною є планування експерименту. Воно передбачає:

- визначення завдання та об'єкта досліджень,
- розробку схеми експерименту,
- вибір земельної ділянки,
- визначення оптимальної структури експерименту.

Включаючи джерела фінансування і етапи виконання робіт, ці положення входять до програми досліджень.

При плануванні досліджень треба урахувати про такі особливості, що постійно супроводжують науковця:

- багатогранність завдань, що треба вирішити;
- наявність багатьох невідомих факторів, які треба передбачити;
- необхідність поєднання дисципліни праці та створення умов для розвитку творчої ініціативи і проявлення суб'єктивізму для кожного члена наукового колективу.

Щоб менше було похибок при плануванні, краще об'єднати зусилля усіх наукових кадрів на вирішення найважливіших теоретичних і практичних завдань, що поставлені в масштабі країни й окремих галузей.

В Україні діє така структура управління наукою: національна академія наук (НАН), галузеві академії наук (наприклад, УААН), галузеві Інститути та вищі навчальні заклади, дослідні станції. У кожній із цих установ успішно діють наукові школи, окремі організатори науки, які координують дії науковців у певних галузях.

Щоб показати масштабність, тривалість турбот і загальнодержавну необхідність вирішення глобальних проблем, досить назвати декілька напрямів досліджень, що турбують науковців як України, так і світу: - енергетика та промислова сировина, - створення і подальше удосконалення ЕОМ та систем АСУ, - глибоке пізнання природи біологічних процесів (біоніка, біотехнологія, нові сполуки і матеріали тощо).

До виконання останнього завдання причетний і Інститут цукрових буряків УААН, який, у свою чергу, є координаційним центром УААН з питань селекції і технології вирощування культури.

Програма досліджень Інституту містить завдання на довготривалі й короткі перспективи, загальноінститутські та окремі теми, теми співвиконавців інших наукових підрозділів тощо.

Етапи планування. Перед складанням схеми дослідів необхідно:

- 1) вибрати тему та визначити завдання і об'єкт досліджень;
- 2) висунути робочу гіпотезу;
- 3) вивчити сучасний стан питання;
- 4) розробити схему і методику експерименту.

Вибираючи тему та визначаючи завдання необхідно чітко з формулювати мету досліджень, побудувати логічну модель вивчаємого явища і правильно вибрати стратегію, яка визначає методи і прийоми дослідження. Об'єкт дослідження зумовлюється тим в якому вигляді мають подаватися в

експерименті його параметри, які з них будуть вхідними, а які – вихідними, змінюються в процесі експерименту, чи будуть сталими.

Робоча гіпотеза служить відправним пунктом для складання схеми або ряду схем майбутніх дослідів і розробки програми дослідження. Як правило, у більшості дослідів вони мусять бути науково обґрунтованими і базуватись на результатах попередніх досліджень. І лише іноді, як здогадка, вони можуть виникати з інтуїції дослідника.

Наступним етапом планування є вивчення сучасного стану питання. Сюди входить: вивчення літератури по вибраній проблемі, патентний пошук, інтернет-пошук, тобто переконатися, що аналогічні питання ще не розглядалися.

Складним етапом планування є розробка схеми і методики дослідів. вибір польових і лабораторних спостережень (аналізів) і обліків для оцінки і пояснювання дії досліджуваних факторів. Надійність результатів експерименту і відповідність їх поставленого завдання залежить від правильного рішення основного питання планування – розробка раціональної схеми дослідів.

Планування дослідів з повними схемами. Повні схеми – це ті, які мають всі логічно підібрані варіанти для вивчення конкретного питання це однофакторні та багатофакторні дослідів.

Однофакторні дослідів. При плануванні схем однофакторних експериментів, які кожен рік закладають на нових ділянках, слід мати на увазі два основних моменти . По-перше. варіанти в однофакторному досліді можуть розрізнятися якісно (дослідів по вивченню і порівняльній оцінці сортів та гібридів культур, форм добрив, пестицидів). По - друге, варіанти в досліді можуть мати кількісні фактори.

При розробці схем однофакторних дослідів, в яких варіанти розрізняються якісно, важливо витримати принцип єдиної логічної відміни,

правильно вибрати контрольний варіант (стандарт) і визначити супутні, які не вивчаються в досліді оптимальні умови експерименту (фон).

Для схем однофакторних дослідів з кількісними градаціями, крім перерахованих вище вимог, необхідно правильно встановити одиницю варіювання для доз досліджувального фактора і число градацій (доз). Важливо так скласти схему досліду, щоб на основі експериментальних точкокефектів варіантів можна б було побудувати криву відгуку, яка буде характеризувати залежність показників від зміни вивчаючих градацій фактора. Зазвичай зв'язок між об'єктами і збільшеними дозами одного фактора нелінійна. Тому бажано мати достатнє число доз в широкому діапазоні. Необхідно намагатись встановити або рівні інтервали між градаціями фактора, або якщо це неможливо передбачити, назначити більше градацій в місцях перегибів кривої відгуку.

Серед варіантів, які плануються, повинні бути послідовно збільшені норми чи дози факторів, від яких досліджувальний показник спочатку буде зростати досягаючи свого максимуму, а далі – знижуватись. Так, якщо при вивченні п'яти доз фосфорних добрив ($X - P_{30}, P_{60}, P_{90}, P_{120}, P_{150}$ – вони вибрані правильно, то врожайність (Y) зображується лінією, яку називають кривою відгуку.

Окремі відрізки такої кривої мають певні назви і значення: АВ – лімітуюча область. ВСД – стаціонарна. ДЕ – інгібіруюча область. В лімітуючій області починається ефект фосфорних добрив, в стаціонарній він стає найвищим, в інгібітуючій – пригнічується. Різні норми чи дози фактору називають градаціями. Різницю між наступною та попередньою дозами називають кроком експерименту – в нашому прикладі це дози фосфорних добрив.

Кожна правильно побудована схема досліду має задовольняти певні вимоги. Схема повинна мати всі градації фактору, які відповідають трьом

областям кривої відгуку – лімітуючий, стаціонарний та інгібіруючий. Це дає змогу виявити в експерименті кращі варіанти і ті у, яких ефект лише виявляється або пригнічується. Останнє необхідне, щоб запобігти застосуванню надмірних доз добрив чи норм зрошення, токсичних доз пестицидів тощо.

Необхідно правильно вибрати крок експерименту, який має бути не дуже великим, щоб не втратити проміжні ефективні варіанти. Однак він не повинен бути і дуже малим, щоб не набрати у досліді непотрібних варіантів і не ускладнювати роботу. Як правило, крок має бути таким, щоб різниця між варіантами перевищувала помилку досліду і була упевненість виявити різницю, яка існує у природі.

Багатофакторні досліді. Вивчивши в однофакторних дослідіх кращі варіанти з окремих елементів агротехніки, починають багатофакторні досліді, які мають певні переваги. У них можна виявити не тільки 100% достовірність дії факторів, а й їх взаємодії: антагонізм, тобто пригнічення дії одного фактору іншим; синергізм – посилення дії фактору іншим; адитивізм – дія факторів незалежно один від одного. Дані багатофакторного дослідіх дають змогу побудувати куполоподібну поверхню відгуку, на якій шляхом екстраполяції (якщо точки лежать поза її межами) та інтерполяції (якщо точки лежать на поверхні відгуку) можна знаходити кращі варіанти, прогнозувати і програмувати врожай та його якість.

Щоб на основі даних багато факторіального експерименту можна було б вирахувати ефекти дії і взаємодії факторів при плануванні його схеми, необхідно витримати принцип факторіальності. Суть принципу факторіальності заключається в тому, що схема повинна передбачити випробування всіх можливих поєднань, сполучень факторів та їх градацій. В факторіальних дослідіх може вивчатися дія і взаємодія, як кількісних так і якісних факторів і їх градацій. Для кількісних факторів, наприклад нульова градація (0) означає відсутність досліджуваного фактора, наприклад без добрив без пестицидів т.д. або його якийсь нижчий рівень, наприклад мінімальна норма висіву,

глибина обробітку тощо. Для якісних факторів нульова градація означає контрольний варіант (стандарт) – стандартна система обробітку, стандартний сорт чи гібрид і т.д.

Планування повних факторіальних схем полегшується використанням спеціальної символіки (кодування) варіантів. Досліджу вальні фактори зазвичай позначають великими латинськими літерами: А, В, С, Д і т.д., а їх градації – цифрами 0, 1, 2, 3, і т.д. Кодування дозволяє всі різноманітні схеми багатфакторних дослідів звести до ряду таблиць, які отримали назву матриць планування. Число стовпчиків в таблиці відповідає числу факторів, а число строк – числу варіантів.

Повна схема багатфакторного досліді (ПФД) включає всі можливі поєднання, сполучення факторів та їх градацій: 2^2 , 2^3 , 3^3 і т. д. Число, що стоїть в основі, означає кількість градацій, а число, яке зазначає ступінь – кількість факторів. Отже, число 2^2 свідчить про те, що в досліді два фактори, кожен з яких має дві градації, а ПФД має чотири варіанти.

5. Представлення вихідного матеріалу у формі таблиць. Побудова таблиць.

Результати статистичного зведення та групування, інші матеріали, одержані в результаті статистичних спостережень, оформляються в статистичні таблиці.

Результати статистичного зведення та групування, як правило, оформлюють у вигляді статистичних таблиць. Статистична таблиця – це форма раціонального та наочного представлення числових даних, які характеризують досліджувальні явища і процеси. В статистичній практиці використовують таблиці різної складності в залежності від мети дослідження, особливостей об'єкта дослідження, обсягу наявної інформації.

Статистична таблиця – форма раціонального викладення цифрового матеріалу.

За змістом: компактна, наочна, доступна.

За формою будь-яка таблиця має вертикальні графи (колонки) та горизонтальні ряди. Перетини граф та рядків утворюють клітини таблиці, в яких і наводяться відповідні числові значення. Як і любе речення, так і таблиця має підмет і присудок.

Підмет – це ті об'єкти, які описуються в таблиці (групи, окремі одиниці сукупності).

Присудок – система показників з допомогою яких характеризується підмет.

За змістом і побудовою підмета статистичні таблиці поділяються на прості, групові, комбінаційні.

Простими називаються таблиці, підмет яких містить перелік об'єктів, адміністративних і територіальних одиниць, господарських підрозділів, хронологічних дат, періодів без їх групування. Прості таблиці, у яких наводяться підсумки господарської діяльності сільськогосподарських підприємств, мають для керівників і спеціалістів велике інформаційне, а при правильно підібраному присудку – і аналітичне значення.

Якщо підмет таблиці розподілено на групи за однією ознакою, то така таблиця називається груповою. Групові таблиці мають велике пізнавальне значення для виявлення і характеристики різних типів явищ, їх структури, взаємозв'язків між явищами.

Комбінаційною називається таблиця, підмет якої розподілено на групи за двома і більше ознаками, взятими в комбінації. Комбінаційні таблиці мають важливі аналітичні властивості, оскільки детально характеризують досліджувані явища і тим самим збагачують економічний аналіз.

Правила побудови таблиці. Кожна таблиця повинна мати номер і назву. Назва таблиці повинна бути чіткою, зрозумілою і відображати її зміст.

Скорочення слів в таблиці допускається лише загальноприйнятих (га, км, м, люд.-год., %).

Значення показників в межах однієї графі округлюються за однаковим ступенем точності.

Таблиці повинні бути замкнутими, тобто мати підсумок, крім аналітичних.

Використання спеціальних позначень:

- показник відсутній (-);
- значення є, але не встановлено (...);
- величина показника є, але менша встановленої точності (0,0; 0,00);
- показник не розраховується (х);
- примітка (*);
- одиниці виміру.

Наочне представлення результатів спостереження та групування соціально-економічних явищ і процесів може бути надано на графіках.

Статистичний графік – це масштабне зображення статистичних даних за допомогою ліній, геометричних фігур та інших наочних засобів. Графіки використовуються для характеристики зміни суспільних явищ і процесів у часі, дослідження структури та порівняння явищ, контролю виконання плану, зображення явищ у просторі та в інших випадках. Вони дають змогу візуальним способом охопити всю сукупність статистичних даних і скласти в цілому картину про вивчаєме явище чи процес. Наочне зображення досліджуваних даних сприяє кращому виявленню найхарактерніших зв'язків між факторами, дозволяє виявити тенденцію у зміні та розвитку окремих явищ. Значна роль

графічної інформації у пропаганді передового досвіду, новітніх технологій, прогресивних тенденцій, в рекламних цілях тощо.

Статистичні графіки відрізняються великою різноманітністю, їх можна поділити на дві великі групи: діаграми та статистичні карти.

Лекція 3. Ряди розподілу та їх статистичне оцінювання

3.1. Значення аналізу рядів розподілу

Статистична сукупність формується під впливом різних причин та умов, з одного боку – типових, загальних для всіх елементів сукупності, а з іншого – випадкових, індивідуальних. Їх спільна взаємодія визначає як індивідуальні значення ознак, так і їх розподіл у межах сукупності. Характерні властивості складу та структури статистичної сукупності відбиваються в рядах розподілу. Тому аналіз рядів розподілу займає велике значення при проведенні статистичних досліджень.

Поглиблений аналіз рядів розподілу передбачає:

- ▶ визначення типового рівня ознаки, який є центром тяжіння;
- ▶ вимірювання варіації ознаки, ступеня згрупованості індивідуальних значень ознаки навколо центра розподілу;
- ▶ оцінювання особливостей варіації, ступеня їх відхилення від симетрії;
- ▶ оцінювання нерівномірності розподілу значень ознаки між окремими елементами сукупності, тобто ступінь їх концентрації.

3.2. Характеристики центра розподілу

Центром тяжіння будь-якої статистичної сукупності є типовий рівень ознаки, узагальнююча характеристика всього розмаїття її індивідуальних значень. Такою характеристикою є середня величина. Види середніх величин, їх значення у статистиці, порядок розрахунку середньої арифметичної та середньої гармонічної було розглянуто у главі 4.

Однак при проведенні статистичних досліджень важливе значення мають

структурні середні величини, які характеризують структуру сукупностей. До структурних середніх величин, у першу чергу, відносять моду та медіану.

Мода та медіана є конкретними характеристиками ряду розподілу, тому їх називають описовими характеристиками або характеристиками центра розподілу. На відміну від інших середніх, які залежать від усіх значень ознаки, мода та медіана залежать лише від характеру частот, тобто від структури розподілу. Порядок розрахунку моди та медіани залежить від виду ряду розподілу.

Мода – це варіанта (значення ознаки), яка найчастіше зустрічається у сукупності.

В атрибутивних та дискретних варіаційних рядах розподілу моду легко відшукати візуально без будь-яких розрахунків. У даних рядах розподілу модою є варіанта, яка має найбільшу частоту.

В інтервальних варіаційних рядах спочатку також візуально визначають модальний інтервал. *Модальний інтервал* – це інтервал, який відповідає найбільшій частоті. Потім моду обчислюють за формулою:

$$Mo = X_{Mo} + h \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)},$$

де X_{Mo} – нижня межа модального інтервалу;

h – розмір (ширина) модального інтервалу;

f_2 – частота модального інтервалу;

f_1 – частота інтервалу, попереднього модальному (передмодального);

f_3 – частота інтервалу, наступного за модальним (післямодального).

Інколи трапляються ряди розподілу, в яких не одна, а декілька варіант, які мають найбільші частоти. Це означає, що у даному випадку буде дві чи більше

моди.

Медіана – це варіанта, що поділяє впорядкований ряд розподілу на дві рівні за чисельністю частини.

При визначенні медіани використовують кумулятивні (накопичені) частоти, які полегшують пошук варіанти (значення ознаки), що знаходиться у середині упорядкованого ряду. Кумулятивні частоти отримують поступовим підсумовуванням частот. Остання кумулятивна частота завжди дорівнює сумі частот ряду розподілу.

У дискретних варіаційних рядах розподілу медіаною є варіанта, кумулятивна частота якої перевищує половину суми частот (обсягу сукупності). Тому при визначенні медіани спочатку треба визначити порядковий номер медіани, а потім встановлюється значення ознаки (варіанта), що відповідає даному порядковому номеру. Порядковий номер медіани у дискретних рядах визначається за формулою:

$$\frac{\sum_{j=1}^n f_j + 1}{2},$$

де $\sum_{j=1}^n f_j$ – сума частот.

В тому разі, коли сума частот парна, порядковий номер медіани є дробовим числом, але оскільки дробових варіантів не буває, то медіана лежить у середині сусідніх варіантів. Тому для дискретного ряду з парним числом варіант медіана розраховується як середня арифметична двох центральних варіант.

В інтервальних варіаційних рядах розподілу спочатку визначають медіанний інтервал. *Медіанний інтервал* – це інтервал, накопичена частота якого рівна або більше половини суми частот. Потім медіану визначають за формулою:

$$Me = X_{Me} + h \frac{\sum_{j=1}^n f_j}{2} - \frac{f_{нак}}{f_{ме}},$$

де X_{Me} – нижня межа медіанного інтервалу;

h – розмір (ширина) медіанного інтервалу;

$\sum_{j=1}^n f_j$ – сума частот;

$f_{нак}$ – сума накопичених частот інтервалу, попереднього медіанному (передмедіанного);

$f_{ме}$ – частота медіанного інтервалу.

Мода та медіана – це особливий вид середніх величин, вони у багатьох випадках мають переваги перед середньою арифметичною або середньою гармонічною і використовуються при вирішенні деяких практичних завдань.

Так мода має велике значення при вивченні попиту населення на товари та послуги, потоку пасажирів на транспорті тощо.

Медіана, як і мода, не залежить від крайніх значень ознаки, тому сума модулів відхилень варіант від медіани мінімальна, тобто має властивість лінійного мінімуму. Цю властивість медіани використовують при проектуванні розміщення зупинок міського транспорту.

Для характеристики структури варіаційного ряду розподілу додатково до медіани обчислюють квартилі, децилі та процентилі.

Квартилі – це варіанти, які поділяють ряд за сумою частот на чотири однакові частини, *децилі* – на десять рівних частин, *процентилі* поділяють ряд розподілу на сто рівних частин. Ці характеристики визначаються на основі кумулятивних частот за аналогією з медіаною. Порядок їх розрахунку буде розглянутий у наступному розділі.

3.3. Значення показників варіації та порядок їх визначення

В одних сукупностях індивідуальні значення ознаки щільно групуються навколо центра розподілу, в інших – значно відхиляються, варіюють.

Вимірювання ступеня коливання ознаки, тобто її варіації – це невід’ємна складова аналізу закономірностей розподілу. Термін „варіація” походить від латинського слова, що у перекладі означає зміна, коливання, різниця. У статистиці під варіацією розуміють такі кількісні зміни ознаки в межах однорідної сукупності, які зумовлені впливом різних факторів. Розрізняють варіації випадкові і систематичні. Важливе значення має аналіз систематичної варіації, так як він дає змогу оцінити залежність зміни ознаки від суттєвих для неї чинників.

Для вимірювання та оцінки варіації у статистиці існує ціла система показників варіації. *До показників варіації належать:*

- ▶ розмах варіації;
- ▶ середнє лінійне відхилення;
- ▶ дисперсія;
- ▶ середнє квадратичне відхилення;
- ▶ коефіцієнт варіації та осциляції.

У системі показників варіації найпростішим є *розмах варіації*. Даний показник характеризує діапазон варіації, це різниця між максимальним та мінімальним значенням ознаки. Безперечною перевагою розмаху варіації, як міри варіації, є простота його обчислення. Недоліком цього показника є те, що він фіксує лише крайні відхилення і зовсім не враховує відхилень решти варіантів від їхньої середньої. Тому цей показник дуже рідко використовують у практичній роботі.

Інший показник варіації – це *середнє лінійне відхилення*. Однак середнє

лінійне відхилення в статистиці також використовують мало, оскільки воно не завжди характеризує розсів значень ознаки.

Ступень варіації об'єктивніше відображує показник середнього квадрата відхилення – *дисперсія*. Дисперсія широко використовується не лише для оцінки варіації, а й для вимірювання зв'язку між явищами, перевірки статистичних гіпотез, при проведенні вибірових досліджень. Це величина не іменована.

Таблиця 3.1.

Формули для розрахунків показників варіації

Назва показників варіації		Формули розрахунків показників варіації	
		для незгрупованих даних	для згрупованих даних
<i>Розрахунок абсолютних та середніх показників варіації</i>			
Розмах варіації		$R = X_{\max} - X_{\min}$	
Середнє лінійне відхилення		$\bar{d} = \frac{\sum_{j=1}^n x_j - \bar{x} }{n}$	$\bar{d} = \frac{\sum_{j=1}^n x_j - \bar{x} f_j}{\sum_{j=1}^n f_j}$
Дисперсія		$\sigma^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n}$	$\sigma^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2 f_j}{\sum_{j=1}^n f_j}$
Середнє квадратичне відхилення		$\sigma = \sqrt{\frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n}}$	$\sigma = \sqrt{\frac{\sum_{j=1}^n (x_j - \bar{x})^2 f_j}{\sum_{j=1}^n f_j}}$
<i>Розрахунок відносних показників варіації</i>			
Коефіцієнт варіації	лінійний	$V_d = \frac{\bar{d}}{\bar{x}} \cdot 100$	
	квадратичний	$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100$	
Коефіцієнт осциляції		$V_R = \frac{R}{\bar{x}} \cdot 100$	

Корінь квадратний з дисперсії називають *середнім квадратичним відхиленням*. Середнє квадратичне відхилення є мірилом надійності середньої.

Чим менше середнє квадратичне відхилення, тим краще та повніше середня величина відображає всю сукупність.

На відміну від дисперсії, розмах варіації, середнє лінійне відхилення та середнє квадратичне відхилення є іменованими показниками, вони виражаються в тих самих одиницях вимірювання, що і значення досліджуваної ознаки.

У статистичній практиці часто виникає необхідність проводити порівняння різних, але взаємозв'язаних сукупностей або варіацію однієї ознаки у різних сукупностях. Для того, щоб забезпечити дані порівняння, потрібно розрахувати відносні показники, до яких належать лінійний та квадратичний коефіцієнт варіації, коефіцієнт осциляції. На підставі даних відносних показників можна проводити якісний аналіз явищ, робити узагальнення та відповідні рекомендації.

У статистиці для порівняння найчастіше використовують квадратичний коефіцієнт варіації. Даний показник використовується для оцінки однорідності сукупності.

Розрізняють такі значення відносних коливань:

$V_{\sigma} < 10\%$ – незначне коливання;

$10\% \leq V_{\sigma} \leq 30\%$ – середнє коливання;

$V_{\sigma} > 30\%$ – велике коливання.

На думку вчених, сукупність є однорідною, а середня величина – типовою, коли квадратичний коефіцієнт варіації не перевищує 33 %.

3.4. Види та математичні властивості дисперсії

Дисперсія посідає особливе місце у статистичному аналізі соціально-економічних явищ та процесів. Вона є невід'ємним і важливим елементом

інших статистичних методів, зокрема вибіркового спостереження, кореляційно-регресійного та дисперсійного аналізу. Тому виникає необхідність більш детального розглядання даного показника та його властивостей.

Основною метою дисперсійного аналізу є виявлення на основі величини загальної дисперсії впливу окремих чинників чи умов, які визначають варіацію ознаки. Для оцінки частки варіації, зумовленої тією чи іншою ознакою, сукупність розподіляють на групи за ознакою, вплив якої досліджується.

Це дозволяє розкласти загальну варіацію (дисперсію) ознаки на дві дисперсії, з яких одна частина варіації вивчається впливом чинника, закладеного в основу групування, а інша – варіацією, зумовленою впливом усіх інших чинників, крім того, що вивчається. Тому при розрахунках використовують декілька видів дисперсії: загальну, міжгрупову та групову (внутрішньогрупову).

Загальна дисперсія характеризує загальну варіацію ознаки під впливом усіх умов і причин, що зумовили цю варіацію. Порядок визначення загальної дисперсії вже було розглянуто раніше (у таб.3.1).

Міжгрупова дисперсія дорівнює середньому квадрату відхилень групових середніх $\bar{y}$ від загальної середньої $\bar{x}$. Вона показує результат впливу фактора, який покладено в основу групування. Міжгрупова дисперсія визначається за формулою:

$$\sigma^2 = \frac{\sum_{i=1}^m (\bar{y}_i - \bar{x})^2 k_i}{\sum_{i=1}^m k_i},$$

де $\bar{x}$ – загальна середня варіюючої ознаки;

$\bar{y}_i$ – середня величина i -ї групи;

k_i – частоти i -ї групи.

Групова (внутрішньогрупова) дисперсія дорівнює середньому квадрату відхилень окремих значень ознаки всередині групи від середньої арифметичної відповідної групи. Ця дисперсія характеризує варіацію ознаки всередині кожної групи статистичного групування. Її можна обчислити як середню просту і як зважену. На практиці найчастіше розрахунок проводять за формулою середньої арифметичної простої:

$$\sigma_i^2 = \frac{\sum_{i=1}^m (y_i - \bar{y}_i)^2}{m},$$

де y_i – значення ознаки окремо за групами;

m – кількість частот у кожній групі.

Групова дисперсія розраховується окремо для кожної i -ї групи. Для всіх груп в цілому розраховується середня з групових дисперсій.

Середня з групових дисперсій (залишкова) – це узагальнююча міра внутрішньогрупової варіації. Ця дисперсія показує результат впливу інших факторів, окрім групувального. Середня з групових дисперсій - це середня арифметична зважена з групових дисперсій:

$$\overline{\sigma^2} = \frac{\sum_{i=1}^m \sigma_i^2 k_i}{\sum_{i=1}^m k_i}.$$

Так як загальна дисперсія характеризує варіацію ознаки за рахунок впливу всіх факторів (причин), міжгрупова – за рахунок фактора, покладеного в основу групування, внутрішньогрупова – за рахунок інших факторів, не врахованих у групуванні, то між різними видами дисперсії існує певний зв'язок. Загальна дисперсія дорівнює сумі середньої з групових дисперсій та міжгрупової дисперсії:

$$\sigma^2 = \overline{\sigma^2} + \delta^2$$

Це правило додавання дисперсій має велике практичне значення. Так як дозволяє, знаючи два види дисперсій, визначити третій, а також виявити залежність результатів варіації результативної ознаки, яка пов'язана з варіацією групувальної ознаки. Дану залежність встановлюють на основі коефіцієнта детермінації, який визначають відношенням міжгрупової дисперсії до загальної:

$$\eta^2 = \frac{\delta^2}{\sigma^2}.$$

Коефіцієнт детермінації виражається у коефіцієнтах чи відсотках.

У статистиці поряд із показниками варіації кількісної ознаки використовують *показники варіації альтернативної ознаки*. Альтернативними ознаками вважаються ознаки, які властиві одним одиницям сукупності та відсутні в інших одиницях сукупності. Наприклад, робітники підприємства підрозділяються на чоловіків та жінок, за місцем проживання населення поділяється на міське та сільське, тобто у даних випадках є два взаємовиняткових варіанти.

Дисперсія альтернативної ознаки обчислюється як добуток часток, які мають цю ознаку, на частку одиниць, що її не мають:

$$\sigma^2 = d_1 \cdot d_0,$$

де d_1 – частка одиниць сукупності, яким властива ознака;

d_0 – частка решти одиниць ($d_0 = 1 - d_1$).

За відсутності первинних даних про розподіл сукупності припускають, що $d_0 = d_1 = 0,5$ (так як $d_1 + d_0 = 1$), тоді не складно допустити, що дисперсія альтернативної ознаки не може перевищувати значення 0,25.

Дисперсія альтернативної ознаки широко використовується при обробці даних соціологічних опитувань, проектуванні вибіркового дослідження тощо.

Як і будь-яка середня величина, дисперсія має певні математичні властивості:

► якщо всі значення варіант x_j збільшити (зменшити) на певну величину A , то дисперсія не зміниться.

► якщо всі значення варіант x_j збільшити (зменшити) в A разів, то дисперсія збільшиться (зменшиться) в A^2 разів, а середнє квадратичне відхилення – в A разів.

► дисперсія постійної величини дорівнює нулю.

► якщо частоти замінити частками, дисперсія не зміниться.

3.5. Форми рядів розподілу

За своєю формою ряди розподілу поділяються на:

► одновершинні;

► багатoverшинні.

Наявність двох і більше вершин свідчить про неоднорідність сукупності, про різнотиповість окремих складових та необхідність перегрупування даних з метою виявлення більш однорідних груп. У даних рядах спостерігається дві та більше моди.

Розподіли якісно однорідних сукупностей переважно одновершинні (одномодальні). Серед одновершинних розподілів є симетричні та асиметричні. Асиметричні розподіли на практиці зустрічаються найчастіше, ніж симетричні.

Якщо частоти рівновіддалені від центра значень ознаки, такий ряд розподілу називають симетричним. В симетричному розподілі спостерігається рівність трьох характеристик: $\bar{X} = Me = Mo$ (рис. 3.1, а).

Якщо вершина розподілу зміщена, тобто частоти по обидва боки від центра змінюються неоднаково, такий *ряд називають асиметричним* (скошеним). Розрізняють правосторонню та лівосторонню асиметрії. Напрямок асиметрії протилежний напрямку зміщення вершини, тобто якщо вершина зміщена ліворуч, маємо правосторонню асиметрію, і навпаки.

Асиметрія виникає внаслідок обмеженої варіації в одному напрямі або під впливом домінуючої причини розвитку, яка приводить до зміщення центру розподілу. Ступінь асиметрії різний – від помірного до значного.

В асиметричному розподілі існують певні розбіжності між характеристиками ряду розподілу:

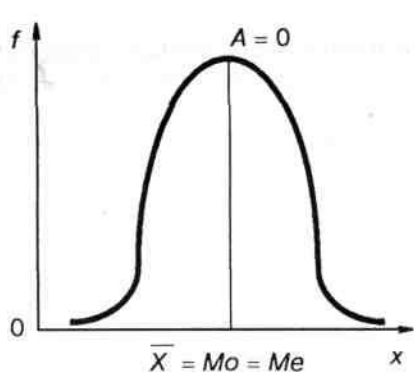
в разі правосторонньої асиметрії $\bar{X} > Me > Mo$;

в разі лівосторонньої асиметрії $\bar{X} < Me < Mo$.

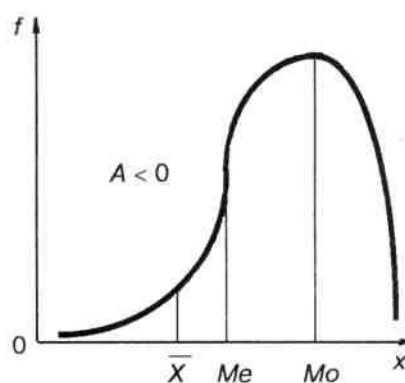
Найпростішою мірою асиметрії є *коефіцієнт асиметрії (стандартизоване відхилення)*, який характеризує напрям і міру скісності розподілу. Коефіцієнт асиметрії визначається як відношення різниці між середньою і модою (чи медіаною) до середнього квадратичного відхилення:

$$A = \frac{\bar{x} - Mo}{\sigma} \quad \text{або} \quad A = \frac{\bar{x} - Me}{\sigma}.$$

У симетричному розподілі $A=0$ (рис.5.1, а); при правосторонній асиметрії $A>0$; при лівосторонній асиметрії $A<0$ (рис. 3.1,б). Тому правосторонню асиметрію називають додатною, а лівосторонню – від’ємною. Чим більша скісність ряду, тим більше значення даного коефіцієнту.



а)



б)

Рис. 3.1. Види форм розподілу

Іншою властивістю одновершинних розподілів є ступінь зосередженості елементів сукупності навколо центра розподілу. Цю властивість називають *ексцесом розподілу*. Таким чином ексцес характеризує крутість варіаційного ряду, тобто його високовершинність (гостровершинність) або низьковершинність (плосковершинність).

Комплексне оцінювання форми розподілу проводиться з використанням моментів. За допомогою їх можна описати будь-який розподіл (для згрупованих та незгрупованих даних):

$$\mu_k = \frac{\sum_{j=1}^n (x_j - a)^k}{n} \quad \text{або} \quad \mu_k = \frac{\sum_{j=1}^n (x_j - a)^k f_j}{\sum_{j=1}^n f_j}$$

Момент розподілу – це середня арифметична k -го ступеня відхилення $x_j - a$. Залежно від величини a моменти поділяють на первинні $a = 0$, центральні $a = \bar{x}$ і умовні $a = \text{const}$. Ступінь k визначає моменти різних порядків.

Певно, що первинний момент 1-го порядку є середня арифметична $\bar{x}$; центральний момент 2-го порядку – це дисперсія, яка характеризує варіацію; центральний момент 3-го порядку характеризує асиметрію; 4-го – ексцес.

Розрахунок центральних моментів визначають за формулою:

$$\mu_k = \frac{\sum_{j=1}^n (x_j - \bar{x})^k f_j}{\sum_{j=1}^n f_j}.$$

Для інтервального варіаційного ряду розподілу центральний момент визначають за формулою:

$$\mu_k = \frac{\sum_{j=1}^n \frac{(x_j - \bar{x})^k}{h} f_j}{\sum_{j=1}^n f_j} h^k,$$

де h – ширина інтервалу.

Для порівняння ступеня асиметрії різних рядів розподілу використовують стандартизований момент за формулою:

$$A_s = \frac{\mu_3}{\sigma^3}$$

Вважають, якщо $A_s < 0,25$ – асиметрія низька, якщо A_s не перевищує 0,5 – середня, при $A_s > 0,5$ – висока.

Для вимірювання ексцесу використовують стандартизований момент четвертого порядку:

$$E = \frac{\mu_4}{\sigma^4}$$

У симетричному розподілі $E = 3$, в разі гостровершинного $E > 3$, для плосковершинного $E < 3$.

Лекція 4. Аналіз тенденцій розвитку

4.1. Поняття про ряди динаміки та їх значення. Види рядів динаміки, елементи рядів динаміки

Соціально-економічні явища та процеси безперервно змінюються у часі. Дослідження їх розвитку є одним з важливих завдань статистики. Це завдання статистика вирішує за допомогою аналізу рядів динаміки.

Ряд динаміки - це послідовність чисел, які характеризують зміну того чи іншого соціально-економічного явища та процесу у часі.

Кожний ряд динаміки складається з двох елементів. Складові елементи ряду динаміки наведено на рис. 4.1.



Рис. 4.1. Складові елементи ряду динаміки

Конкретні числові значення, які характеризують кількісну оцінку розвитку явищ у часі називають рівнями ряду. Вони можуть виражатися у вигляді абсолютних, відносних або середніх величин. Кожний рівень ряду відповідає конкретному періоду або моменту часу: $Y_1 \rightarrow t_1$, $Y_2 \rightarrow t_2$, $Y_{n-1} \rightarrow t_{n-1}$, $Y_n \rightarrow t_n$.

Таблиця 4.1.

Динаміка виробництва продукції на підприємстві „Світ” за звітний рік

Квартали	1	2	3	4	← Періоди часу (t)
Виробництво продукції, тис. грн.	190	185	192	207	← Рівні ряду (Y)

У будь-якому ряду динаміки мінімальна кількість рівнів ряду (відповідно і періодів часу) 2, максимальна – n .

Залежно від характеру ознак, згідно яких групуються ряди динаміки, їх можна поєднати у окремі групи. Класифікація рядів динаміки за різними ознаками наведена на рис. 4.2.

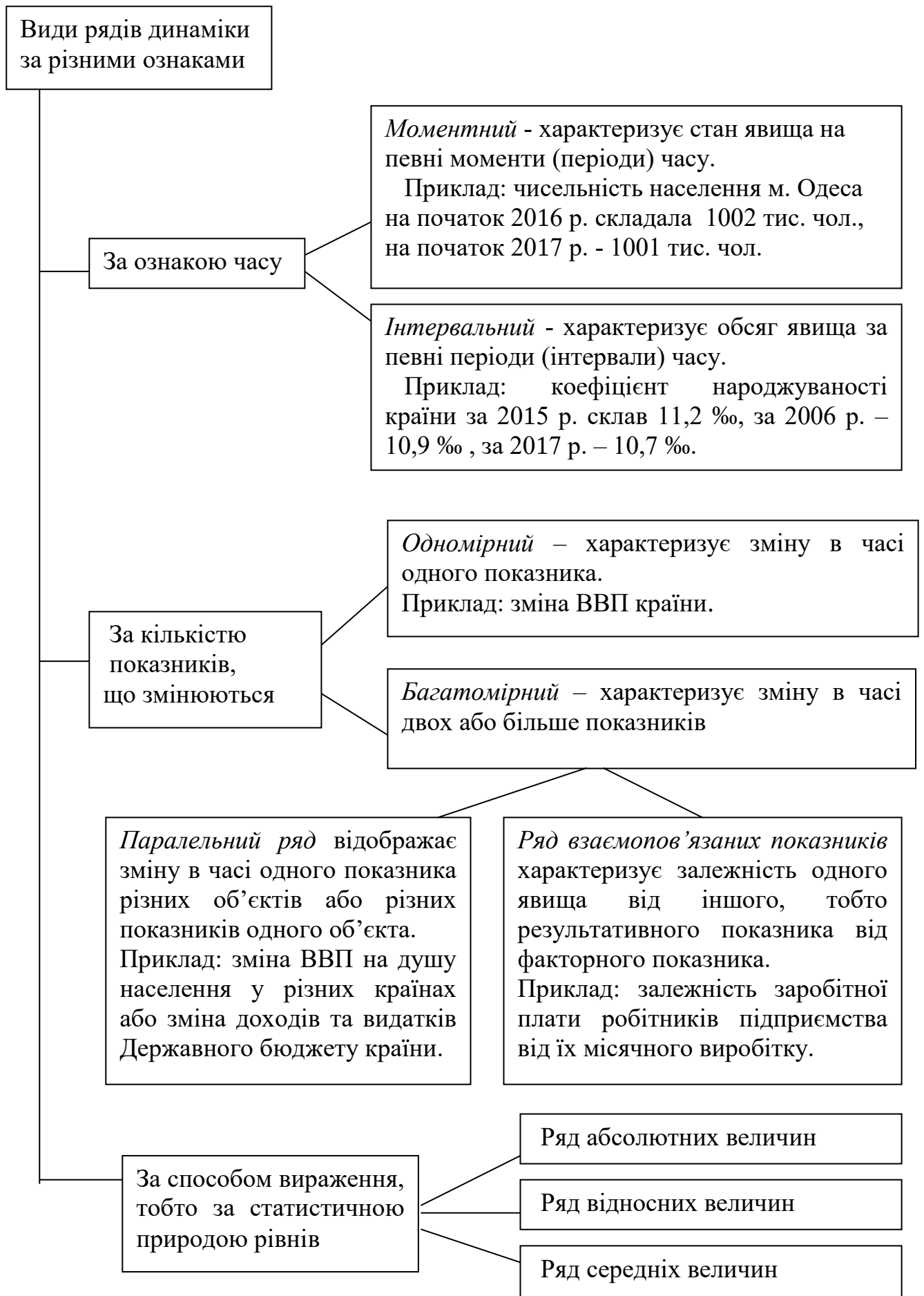


Рис. 4.2. Класифікація рядів динаміки за різними ознаками

Для отримання якісних результатів статистичних досліджень при побудові рядів динаміки необхідно дотримуватися конкретних вимог для забезпечення їх порівняння.

Важливою умовою порівнянності рядів динаміки є *вираження їх рівнів в однакових одиницях вимірювання*. Часто статистичні показники виражаються у різних одиницях. Наприклад, виробництво продукції обліковують у натуральному та грошовому вираженні. При побудові ряду слід звернути увагу на те, щоб рівні ряду динаміки завжди були виражені в однакових одиницях.

Рівні ряду динаміки мають бути порівнювальними за періодами часу, тобто не можна порівнювати виробництво продукції за місяць і за квартал, за квартал і за рік. Інтервали часу мають бути рівними.

Рівні ряду динаміки мають бути порівнювальними за методом обчислення показників. Наприклад, чисельність населення країни можна вказати на початок року, також можна розрахувати та вказати середньорічну чисельність населення. Дані показники включати до одного ряду динаміки не можна, тому що їх не порівнюють.

4.2. Показники рядів динаміки

Під час аналізу рядів динаміки використовують взаємопов'язані характеристики (показники): абсолютний приріст (зниження), темп зростання, темп приросту, абсолютне значення 1% приросту.

Показники ряду динаміки, за винятком абсолютного значення 1% приросту, розраховують двома способами: як з постійною базою порівняння (базисний спосіб), так і змінною базою (ланцюговий спосіб). Абсолютне значення 1% приросту має місце тільки при ланцюговому способі розрахунку.

Якщо кожний наступний рівень ряду зіставляють з попереднім рівнем, то отримують ланцюгові показники динаміки, а якщо кожний наступний рівень ряду зіставляють з рівнем, що узятий за базу зіставлення, то отримують базисні показники динаміки.

За базу порівняння можна прийняти не лише початковий, а й будь-який інший рівень ряду динаміки. Вибір бази порівняння залежить від мети дослідження. Схема порівняння рівнів ряду динаміки при різних способах наведена на рис. 4.3.

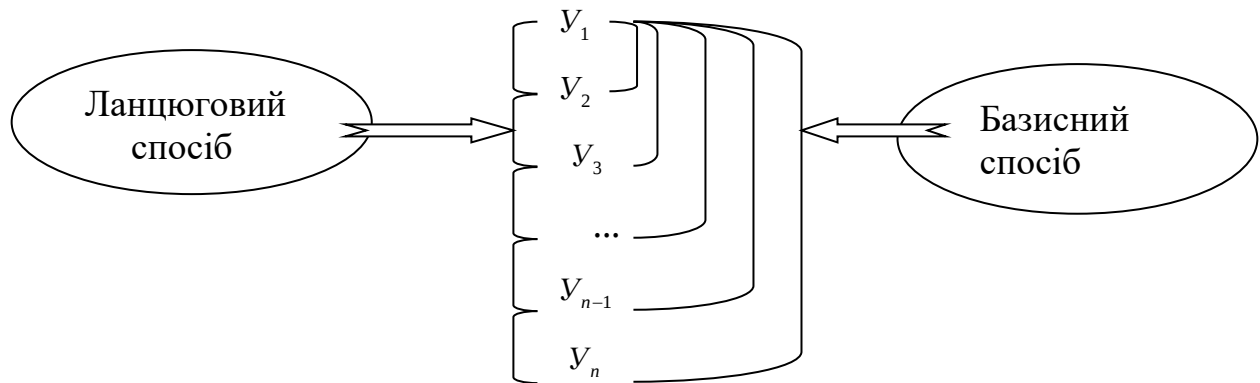


Рис. 4.3. Способи порівнянь рівнів ряду динаміки

Абсолютний приріст (Δ) відображає абсолютну зміну рівнів ряду динаміки, тобто показує, на скільки одиниць змінився рівень поточного (звітного) періоду порівняно з попереднім або базисним періодом.

Абсолютний приріст обчислюється як різниця рівнів ряду динаміки. Він виражається у тих самих одиницях вимірювання, що і рівні ряду динаміки. Знак (+ ; –) показує напрям змін (збільшення або зменшення).

Абсолютний приріст ланцюговий

$$\Delta = Y_n - Y_{n-1}$$

Абсолютний приріст базисний

$$\Delta = Y_n - Y_1,$$

де Y_n – поточний (звітний) рівень ряду;

Y_{n-1} – попередній рівень;

Y_1 – початковий (перший, базисний) рівень.

Для вичерпної та глибокої характеристики явища абсолютні величини доповнюють відносними: темпом зростання та темпом приросту.

Темп зростання характеризує інтенсивність зміни явища. Даний показник обчислюється як відношення поточного (звітного) періоду до рівня попереднього або базисного періоду. Він виражається у формі коефіцієнта чи відсотка. Темп зростання у коефіцієнтах дуже часто називається коефіцієнтом зростання. Коефіцієнт зростання показує, у скільки разів змінився рівень

поточного (звітнього) періоду порівняно з попереднім або базисним періодом.

Темп зростання ланцюговий

$$TZ = \frac{Y_n}{Y_{n-1}} \quad \text{або} \quad TZ = \frac{Y_n}{Y_{n-1}} \cdot 100$$

Темп зростання базисний

$$TZ = \frac{Y_n}{Y_1} \quad \text{або} \quad TZ = \frac{Y_n}{Y_1} \cdot 100$$

Між ланцюговими та базисними темпами зростання існує зв'язок. Добуток ланцюгових темпів (коефіцієнтів) зростання дає відповідний базисний.

Взаємозв'язок ланцюгових та базисних темпів (коефіцієнтів) зростання використовується для переходу від одних темпів зростання до інших у тих випадках, коли невідомі абсолютні рівні ряду динаміки.

Велике значення при аналізі рядів динаміки має *темп приросту*, який, на відміну від темпу зростання, завжди виражають у відсотках. Темп приросту показує, на скільки відсотків змінився рівень поточного (звітнього) періоду з рівнем попереднього або базисного періоду. Знак (+ ; –) показує напрям змін (збільшення або зменшення).

Темп приросту обчислюють як відношення абсолютного приросту до рівня ряду динаміки попереднього або базисного періоду.

Темп приросту ланцюговий

$$ТП = \frac{\Delta}{Y_{n-1}} \cdot 100$$

Темп приросту базисний

$$TZ = \frac{\Delta}{Y_1} \cdot 100$$

Однак існує також інший спосіб розрахунку даного показника, який є найбільш поширеним: $ТП = TZ - 100$.

Абсолютне значення 1 % приросту дає уяву про вагомість одного відсотка приросту. Даний показник обчислюється тільки ланцюговим способом.

Абсолютне значення 1 % приросту можна визначити шляхом ділення абсолютного приросту на темп приросту за один і той самий період часу. Однак існує і другий найбільш простий спосіб: ділення рівня попереднього періоду на 100.

$$AZ = \frac{\Delta}{ТП} \quad \text{або} \quad AZ = \frac{Y_{n-1}}{100}.$$

Порядок проведення аналізу рядів динаміки, тобто розрахунків

характеристик динаміки, розглянемо на прикладі змін доходів бюджету області.

Таблиця 4.2.

Показники доходів бюджету області за останні роки

Рік	2015	2016	2017
Доходи бюджету області, млн. грн.	467	512	524

Необхідно провести аналіз показників ряду динаміки ланцюговим та базисним способом. Зробити висновки.

Розв'язок:

Таблиця 4.3.

Динаміка доходів бюджету області за 2015 – 2017 роки

Рік	Доходи бюджету, млн. грн.	Абсолютний приріст, млн. грн.		Темп зростання, %		Темп приросту, %		Абсолютне значення 1 % приросту, млн. грн..
		ланцюговий спосіб	базисний спосіб	ланцюговий спосіб	базисний спосіб	ланцюговий спосіб	базисний спосіб	
2015	467	...	0	...	100,0	...	0	...
2016	512	45	45	109,6	109,6	9,6	9,6	4,67
2017	524	12	57	102,3	112,2	2,3	12,2	5,12

Висновок: у 2006 р. доходи бюджету області збільшилися порівняно з 2005 р. на 45 млн. грн. або у 1,096 разів, тобто на 9,6 %. У 2007 р. доходи зросли порівняно з 2006 р. на 12 млн. грн. або у 1,023 разів, або на 2,3 %, а порівняно з 2005 р. – на 57 млн. грн. або у 1,122 разів, або на 12,2 %. Кожний відсоток абсолютного приросту у 2007 р. відповідає додатковому збільшенню доходів на 5,12 млн. грн.

4.3. Середні показники рядів динаміки

При проведенні багатьох економічних розрахунків використовуються різного роду середні показники рядів динаміки:

- середні рівні рядів динаміки;
- середні показники рядів динаміки: середній абсолютний приріст; середній темп зростання; середній темп приросту.

Метод обчислення середнього рівня ряду динаміки залежить від виду ряду.

В інтервальному ряді середній рівень ряду динаміки визначають за формулою

середньої арифметичної простої:

$$\bar{Y} = \frac{\sum_{j=1}^n Y_j}{n},$$

де n – число рівнів ряду динаміки або періодів часу.

Дещо складніший порядок розрахунку середнього рівня у моментному ряді динаміки. Якщо у моментному ряді динаміки є тільки два періоди (моменту) часу, наприклад, є показники тільки на початок та кінець періоду, то середній рівень визначають за формулою:

$$\bar{Y} = \frac{Y_{\Pi} + Y_K}{2},$$

де Y_{Π} , Y_K – рівні ряду відповідно на початок та кінець періоду.

Даною формулою дуже часто користуються при розрахунках середньорічної чисельності населення. Однак цей середній рівень ряду вважається приблизною оцінкою, його застосовують нечасто.

Якщо у моментному ряді динаміки кількість періодів (моментів) часу більше двох ($n > 2$), то при розрахунку середнього рівня слід розглядати декілька варіантів. У першу чергу слід звертати увагу на те, що рівні ряду динаміки можуть знаходитися один від одного на однакових (рівних) або різних проміжках часу.

Середній рівень моментного ряду динаміки з рівними проміжками часу між датами визначають за формулою середньої хронологічної:

$$\bar{Y} = \frac{\frac{Y_1}{2} + Y_2 + Y_3 + Y_{n-1} + \frac{Y_n}{2}}{n-1}.$$

Дана формула також може мати інший вигляд:
$$\bar{Y} = \frac{\frac{Y_1 + Y_n}{2} + \sum_{j=2}^{n-1} Y_j}{n-1}.$$

Середній рівень моментного ряду динаміки з різними проміжками часу між датами визначають за формулою середньої арифметичної зваженої:

$$\bar{Y} = \frac{\sum_{j=1}^n Y_j \cdot t_j}{\sum_{j=1}^n t_j},$$

де Y – рівні ряду динаміки на певні моменти (періоди) часу;

t – тривалість часу між двома сусідніми моментами (періодами) часу.

В окремих випадках дана формула може бути модифікована, тоді вона приймає наступний вигляд:

$$\bar{Y} = \frac{\sum_{j=1}^n \bar{Y}_j \cdot t_j}{\sum_{j=1}^n t_j},$$

де $\bar{Y}_j$ – середній рівень між двома сусідніми моментами (періодами) часу.

Узагальнюючими характеристиками інтенсивності динаміки є середній абсолютний приріст, середній темп зростання, середній темп приросту.

Середній абсолютний приріст (середня абсолютна швидкість динаміки) показує, на скільки одиниць у середньому за одиницю часу за певний період змінився рівень явища, що аналізується. Даний показник можна визначити декількома методами (способами):

1 спосіб: Середній абсолютний приріст обчислюється як середня арифметична проста із ланцюгових абсолютних приростів:

$$\bar{\Delta} = \frac{\sum \Delta}{m},$$

де Δ – ланцюгові абсолютні прирости,

m – кількість ланцюгових приростів.

Кількість абсолютних приростів менша від кількості рівнів ряду динаміки на одиницю, тобто $m = n - 1$, де n – число рівнів ряду.

2 спосіб: При відсутності ланцюгових приростів середній абсолютний приріст можна обчислити за формулою:

$$\bar{\Delta} = \frac{Y_n - Y_1}{n - 1},$$

де Y_n – кінцевий (останній) рівень ряду динаміки;

Y_1 – початковий (перший) рівень ряду динаміки.

Середній темп зростання показує, у скільки разів у середньому за одиницю часу за конкретний період змінився рівень ряду явища, що аналізується. Даний показник також обчислюється декількома методами:

1 спосіб: Середній темп зростання обчислюється за формулою середньої геометричної з ланцюгових темпів зростання:

$$\overline{TЗ} = \sqrt[m]{TЗ_1 \cdot TЗ_2 \cdot TЗ_3 \dots TЗ_m},$$

де $TЗ_1, TЗ_2, \dots, TЗ_m$ – ланцюгові темпи зростання (у коефіцієнтах).

2 спосіб: При відсутності ланцюгових приростів середній темп зростання визначають за формулою: $\overline{TЗ} = \sqrt[n-1]{\frac{Y_n}{Y_1}}$

Середній темп приросту показує, на скільки відсотків у середньому за одиницю часу за певний період змінився рівень явища, що аналізується.

Середній темп приросту визначається за формулою: $\overline{ПІ} = \overline{TЗ} - 100$.

Лекція 5. Кореляційний аналіз

5.1. Види зв'язків між явищами

Усі явища суспільного життя існують неізолювано, вони органічно пов'язані між собою, залежать одні від одних, безперервно рухаються та розвиваються. Вивчення зв'язків є одним із завдань статистичного аналізу.

Зв'язки між явищами мають причинно-наслідковий характер, тому статистичний аналіз покликаний кількісно охарактеризувати відповідні зв'язки. Суспільні явища або окремі їхні ознаки, які впливають на інші, а також характеризують причини та умови зв'язку, називається *факторними* (X), суспільні явища або окремі їхні ознаки, які змінюються під впливом факторних, тобто характеризують наслідки зв'язку, називаються *результативними* (Y).

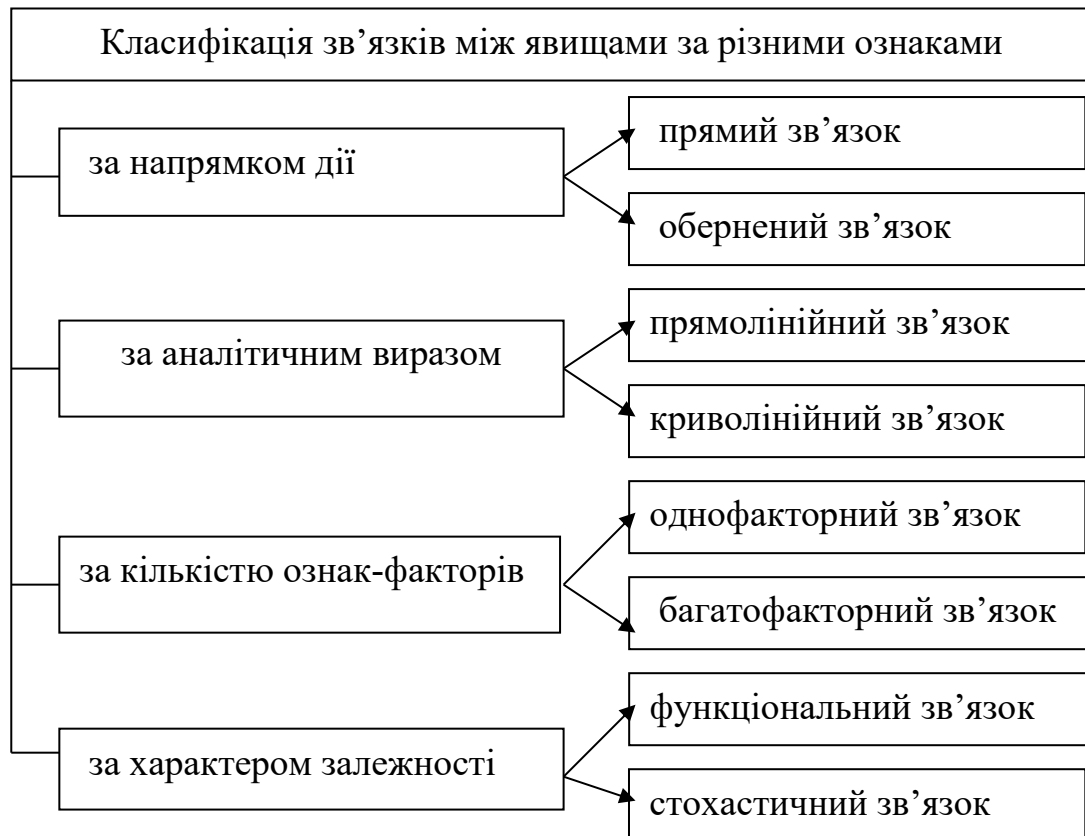


Рис.5.1. Класифікація зв'язків між явищами

Прямий зв'язок – це такий зв'язок, за яким напрямок зміни результативної ознаки співпадає з напрямком факторної ознаки, тобто зі зростанням факторної ознаки зростає і результативна або навпаки, зі зменшенням факторної ознаки результативна також зменшується.

Обернений зв'язок – це такий зв'язок, за яким напрямок зміни результативної ознаки не співпадає з напрямком факторної ознаки, тобто зі збільшенням факторної ознаки результативна зменшується або навпаки, зі зменшенням факторної ознаки результативна зростає, тобто факторна та результативна ознаки змінюються у протилежному напрямку.

Прямолінійний зв'язок – це такий зв'язок, за яким спостерігається відповідність однакових змін середніх значень факторної ознаки та приблизно однакових змін середніх значень результативної ознаки. Залежність результативної ознаки від факторної може бути вираженою рівнянням прямої лінії.

Криволінійний зв'язок – це такий зв'язок, за яким однаковим змінам середніх значень факторної ознаки відповідають нерівні зміни середніх значень результативної ознаки. Залежність результативної ознаки від факторної може бути виражена рівнянням якої-небудь кривої (гіперболи, параболи тощо).

Однофакторний зв'язок – це такий зв'язок, за яким залежність результативної ознаки спостерігається тільки від однієї факторної ознаки.

Багатофакторний зв'язок – це такий зв'язок, за яким залежність результативної ознаки спостерігається одночасно від декількох факторних ознак.

Функціональний зв'язок – це такий зв'язок, за яким певному значенню факторної ознаки відповідає одне чи кілька чітко визначених значень результативної ознаки. Функціональний – це жорстко детермінований зв'язок. Ці зв'язки вивчають точні науки: математика, фізика, хімія тощо. У статистиці вони вивчаються за допомогою індексного методу.

Стохастичний зв'язок – це такий зв'язок, за яким кожному значенню факторної ознаки відповідає певна множина значень результативної ознаки, які утворюють так званий умовний розподіл.

Кореляційний зв'язок – це підвид стохастичного зв'язку, коли зі зміною факторної ознаки змінюється середня величина результативної ознаки.

У статистиці за допомогою кореляційного зв'язку можна встановити

залежність між кваліфікацією робітників та їх денним виробітком. Між даними ознаками існує прямий зв'язок: чим вище тарифний розряд, тим більше денний виробіток. Однак, якщо ми будемо розглядати кожного певного робітника, то дана залежність не завжди спостерігається, тому що у кожному випадку, крім кваліфікації (тарифного розряду), існують різні інші фактори, які також впливають на продуктивність праці (виробітку). Кореляційний зв'язок же проявляється не в кожному окремому випадку, а в середньому для великої кількості спостережень.

5.2. Загальні методи вивчення зв'язків

Зв'язки і залежності суспільних явищ визначають різними методами. До цих методів відносять балансових метод, індексний метод, метод аналітичних групувань, кореляційно – регресійний та інші методи математичної статистики.

Найпростішим методом є *балансовий метод*. Статистичний баланс – це система показників, яка складається із сум абсолютних величин, поєднаних у рівності: $A + B = B + G$.

Балансова форма дає змогу здійснювати взаємний контроль даних, а також розраховувати невідомі показники, а балансова ув'язка - виявити неточності окремих показників та вточнити їх. За допомогою балансового методу вивчають рух робочої сили, грошових засобів, основних фондів тощо. Однак ширші можливості аналізу взаємозв'язків дає складання балансів руху матеріальних, трудових і фінансових ресурсів для сукупностей підприємств, регіонів або країні у цілому. В цьому разі баланси дають змогу виявити взаємозв'язки та пропорції в утворенні і розподілі ресурсів між підприємствами, галузями, регіонами країни.

До найважливіших методів дослідження взаємозв'язків відносять метод *аналітичних групувань*. Аби виявити залежність між ознаками за допомогою цього методу, матеріал статистичного групування групують за факторною ознакою, потім для кожної групи обчислюють середні значення факторної та результативної ознаки. Метод аналітичних групувань характеризує лише загальні риси зв'язку, однак не дає кількісної оцінки сили зв'язку. Порядок

проведення аналітичного групування був розглянутий у главі 3.

Для кількісної оцінки зв'язку між явищами на базі матеріалів аналітичного групування обчислюють коефіцієнт детермінації.

5.3. Оцінка щільності та перевірка істотності кореляційного зв'язку

Поряд із визначенням характеру зв'язку та ефектів впливу факторів x на результат y важливе значення має оцінка щільності зв'язку, тобто оцінка узгодженості варіації взаємопов'язаних ознак. Показниками щільності кореляційного зв'язку є коефіцієнт кореляції, коефіцієнт детермінації, кореляційне відношення, коефіцієнти асоціації, взаємної спряженості тощо.

Найпоширенішим є коефіцієнт кореляції. Для вимірювання щільності прямолінійних зв'язків використовується лінійний коефіцієнт кореляції. Лінійний коефіцієнт кореляції обчислюють за формулою:

$$r = \frac{\sum_{j=1}^n (x_j - \bar{x}) \cdot (y_j - \bar{y})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2 \cdot \sum_{j=1}^n (y_j - \bar{y})^2}}.$$

Однак на практиці застосовують різні модифікації наведеної формули. Значення лінійного коефіцієнта кореляції коливається в межах від (-1) до 1 . Додатне значення відповідає прямому зв'язку, а від'ємне – зворотному. Чим більше значення лінійного коефіцієнта кореляції, тим щільніший зв'язок між результативною ознакою “ y ” та факторною ознакою “ x ” і навпаки.

Таблиця 5.0.

Градація щільності зв'язку за шкалою Чеддока

Зв'язок	Лінійний коефіцієнт кореляції	
	прямий	зворотний
слабкий	0,1 – 0,3	–0,1 ... –0,3
помірний	0,3 – 0,5	–0,3 ... –0,5
помітний	0,5 – 0,7	–0,5 ... –0,7
високий	0,7 – 0,9	–0,7 ... –0,9
надто високий	0,9 – 0,99	–0,9 ... –0,99

5.4. АНАЛІЗ ПРОСТОЇ ЛІНІЙНОЇ КОРЕЛЯЦІЇ

Математичне рівняння, яке описує залежність між ознаками, називають кореляційним рівнянням або рівнянням регресії. При простій лінійній кореляції це рівняння має вигляд: $\bar{y}_x = a_0 + a_1x$, де $\bar{y}_x$ – середнє теоретичне значення результативної ознаки «у»; x значення факторної ознаки; a_0, a_1 – параметри рівняння. Кореляційне рівняння пов'язує результативну ознаку з факторною у вигляді рівняння прямої лінії, де параметр a_1 визначає середню зміну результативної ознаки (у) при зміні факторної (х) на одиницю її натурального виміру.

Невідомі параметри a_0 і a_1 визначають за способом найменших квадратів.

Для цього складають і розв'язують систему рівнянь:
$$\begin{cases} a_0n + a_1\sum x = \sum y; \\ a_0\sum x + a_1\sum x^2 = \sum xy. \end{cases}$$

Тісноту простого лінійного кореляційного зв'язку визначають з допомогою лінійного коефіцієнта кореляції (r). Абсолютна величина його коливається в межах 0 ± 1 . Чим ближче ця величина до 1, тим тіснішим вважається зв'язок між ознаками. Додатне значення величини коефіцієнта кореляції вказує на прямий зв'язок результативної і факторної ознак, від'ємне – на обернений зв'язок між ними.

Формула лінійного коефіцієнта кореляції має вигляд:

$$r_{yx} = \frac{\overline{xy} - \bar{x} \times \bar{y}}{\sigma_x \times \sigma_y},$$

де σ_x і σ_y – середні квадратичні відхилення факторної і результативної ознак.

Як приклад розглянемо кореляційну залежність між урожайністю гороху (у) і кількістю внесених органічних добрив на 1 га (Х) по 20 господарствах. За даними спостережень визначимо допоміжні величини (табл. 5.1). Підставивши у систему нормальних рівнянь обчислені сумарні значення, одержимо:

$$20a_0 + 293a_1 = 506$$

$$293a_0 + 4483a_1 = 7549$$

Таблиця 5.1. Вихідні і розрахункові дані по 20 господарствах

№	y	x	.yx	x ²	y ²
1	22	10	220	100	484
2	21	10	210	100	441
3	20	12	240	144	400
4	24	12	288	144	576
5	24	12	288	144	576
6	23	10	230	100	529
7	21	14	294	196	441
8	24	12	288	144	576
9	25	15	375	225	625
10	26	14	364	196	676
11	25	16	400	256	625
12	30	15	450	225	900
13	26	16	416	256	676
14	28	18	504	324	784
15	26	15	390	225	676
16	26	18	468	324	676
17	30	18	540	324	900
17	29	16	464	256	841
19	26	20	520	400	676
20	30	20	600	400	900
Сума	506	293	7549	4483	12978

Розв'язавши систему рівнянь отримуємо кореляційне рівняння:

$$\bar{y}_x = 14,84 + 0,71 X.$$

Економічна інтерпретація даного кореляційного рівняння така: збільшення кількості внесених органічних добрив на 1 т/га обумовлює підвищення урожайності гороху в середньому на 0,71 ц/га.

За даними тієї ж таблиці 5.1 знаходимо значення складових елементів лінійного коефіцієнта кореляції.

середнє значення фактора X: $293:20=14,65$

середнє значення результативної ознаки Y: $506:20=25,30$

середнє значення добутку ознак UX: $7549:20=377,5$

середнє квадратичне відхилення ознаки X: $(4483:20-14,65^2)^{1/2}=3,09$

середнє квадратичне відхилення ознаки Y: $(12978:20-25,3^2)^{1/2} = 2,97$.

Коефіцієнт кореляції визначаємо за формулою:

$$r_{yx} = \frac{\overline{xy} - \bar{x} \times \bar{y}}{\sigma_x \times \sigma_y}$$

У нашому прикладі $r_{yx} = (377,5-25,3*14,65)/2,97*3,09 = 0,74$. Коефіцієнт детермінації складає $d = 0,74^2*100=55\%$. Як видно, за даними розглянутого прикладу варіація урожайності гороху на 55% обумовлена впливом органічних добрив.

5.5. АНАЛІЗ ПРОСТОЇ НЕЛІНІЙНОЇ КОРЕЛЯЦІЇ

При криволінійній формі зв'язку рівномірне збільшення факторної ознаки призводить до нерівномірного збільшення (або зменшення) результативної. Інколи при цьому зростання величини результативної ознаки змінюється її зниженням, а зменшення – збільшенням.

В аналізі сільськогосподарського виробництва часто використовують рівняння параболи: $\bar{y}_x = a_0 + a_1x + a_2x^2$.

Однією з особливостей цього типу кривої є те, що вона має оптимальне значення факторної ознаки, при якому спостерігається максимальне (або мінімальне) значення результативної ознаки. Якщо в рівнянні величина a_1 виражена від'ємним числом, а a_2 – додатним, то крива змінюватиме напрям зниження на зростання тобто оптимальному значенню факторної ознаки відповідає максимальне значення результативної ознаки.

Для розрахунку параметрів рівняння параболи другого порядку використовується така система нормальних рівнянь:

$$\begin{cases} a_0n + a_1\sum x + a_2\sum x^2 = \sum y; \\ a_0\sum x + a_1\sum x^2 + a_2\sum x^3 = \sum xy; \\ a_0\sum x^2 + a_1\sum x^3 + a_2\sum x^4 = \sum x^2y. \end{cases}$$

В кореляційному аналізі економічних явищ використовуються й інші нелінійні функції.

Гіперболічна форма кореляційного зв'язку характерна для, наприклад, для залежності рівня витрат на виробництво 1 ц молока (y) від його продуктивності корів (x). Рівняння гіперболи має вигляд:

$$\bar{y}_x = a_0 + a_1 \frac{1}{x}.$$

Параметри цього кореляційного рівняння визначаємо на підставі такої системи:

$$\begin{cases} a_0n + a_1\sum \frac{1}{x} = \sum y; \\ a_1\sum \frac{1}{x} + a_2\sum \left(\frac{1}{x}\right)^2 = \sum \frac{1}{x}y. \end{cases}$$

Для аналітичного вираження явищ при вивченні темпів росту використовують рівняння експоненціальної кривої:

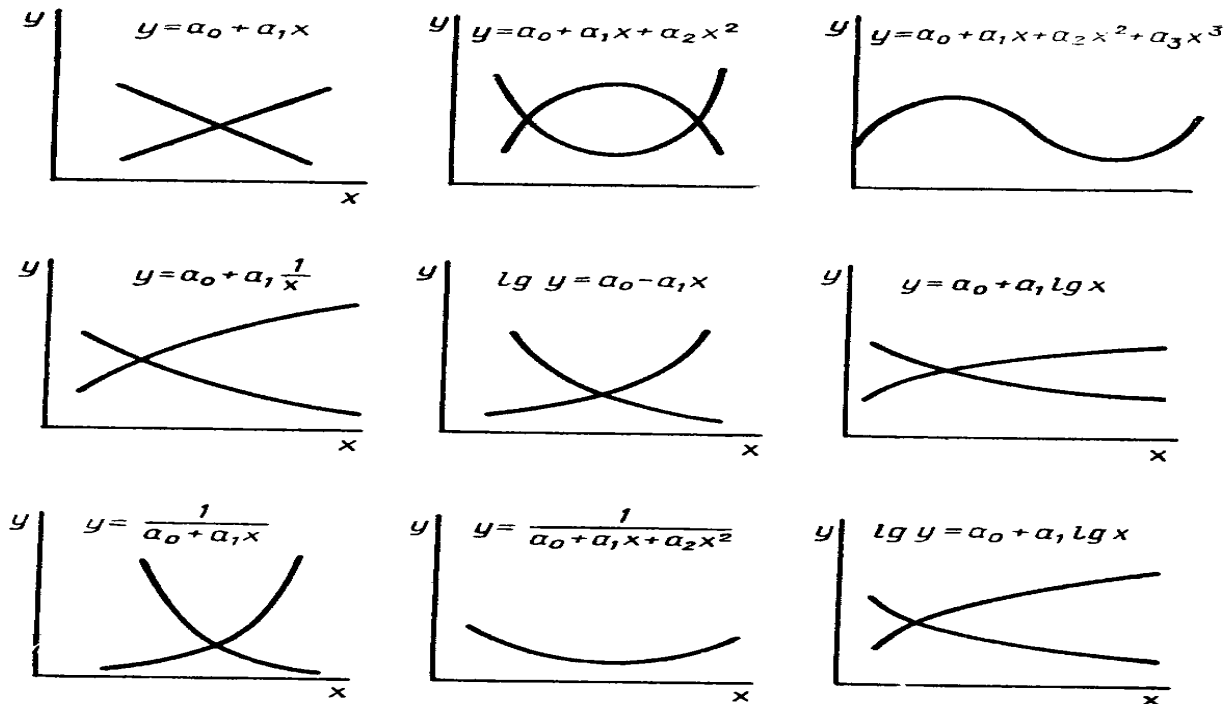
$\bar{y}_x = a_0 \times a_1^x$, де x – фактор-аргумент (порядковий номер року, який в аналітичних рівняннях динаміки одержує значення 1, 2, 3 і т. д.); a_0 – показник базисного року; a_1 – середньорічний темп приросту.

Невідомі параметри a_0 і a_1 в наведеній вище формулі визначаємо логарифмуванням, перетворивши показову функцію в пряму: $\lg y = \lg a_0 + x \lg a_1$.

Система нормальних рівнянь при цьому має вигляд:

$$\begin{cases} n \lg a_0 + \lg a_1 \sum x = \lg y; \\ \lg a_0 \sum x + \lg a_1 \sum x^2 = \lg y \sum x. \end{cases}$$

В аналізі економічних явищ часто використовують степенну функцію виду: $y = a_0 x^{a_1}$. Нелінійність щодо своїх констант зумовлюють її перетворення (шляхом логарифмування) у логарифмічно-лінійну функцію виду: $\lg y = a_0 + a_1 \lg x$. Таке перетво-



5.1. Лінії основних математичних функцій.

В аналізі економічних явищ часто використовують степенну функцію виду: $y = a_0 x^{a_1}$. Нелінійність щодо своїх констант зумовлюють її перетворення (шляхом логарифмування) у логарифмічно-лінійну функцію виду: $\lg y = a_0 + a_1 \lg x$. Таке перетворення дає змогу розв'язувати систему нормальних рівнянь методом найменших квадратів.

Логарифмічно-лінійну функцію використовують при дослідженні явищ, які характеризуються тим, що із приростом абсолютної величини факторної ознаки її вплив на результативну знижується. Цьому типу функцій притаманна пропорціональність не абсолютних (як для рівняння прямої лінії), а відносних приростів економічних показників, які вивчаються.

Якщо природа взаємозв'язку економічних явищ така, що середня арифметична результативної ознаки (y) пов'язана із середньою арифметичною факторної ознаки (x), то математичний вираз подібної залежності, тобто оцінку однієї змінної за другою, дає логарифмічна крива, гіпотетичне рівняння якої: $y = a_0 a_1 \lg x$.

Відсутність чисельних значень логарифмів для нульових і від'ємних чисел обмежує можливості використання логарифмічних функцій в окремих випадках економічного аналізу.

У випадку нелінійного зв'язку для оцінки тісноти зв'язку між ознаками обчислюють кореляційне відношення (η). Формула кореляційного відношення така:

$$\eta = \sqrt{\frac{\sigma_{\bar{y}_x}^2}{\sigma_y^2}} = \frac{\sigma_{\bar{y}_x}}{\sigma_y},$$

де $\sigma_{\bar{y}_x}^2$ – факторна дисперсія результативної ознаки; σ_y^2 – загальна дисперсія результативної ознаки.

Наведену формулу кореляційного відношення називають індексом кореляції. Всі формули кореляційного відношення (індекса кореляції) тотожні.

Розглянемо приклад параболічної залежності урожайності картоплі від внесення органічних добрив (табл.5.2).

Таблиця 5.2. Залежність урожайності картоплі (у, ц/га) від внесення органічних добрив (х, т/га)

№	X	у	X ²	X ³	X ⁴	у ²	уХ	уХ ²	у _х	(у _х -у) ²
1	11	70	121	1331	14641	4900	770	8470	73,3	238,3
2	10	70	100	1000	10000	4900	700	7000	67,9	433,7
3	12	85	144	1728	20736	7225	1020	12240	78,3	109,1
4	13	80	169	2197	28561	6400	1040	13520	82,9	34,2
5	12	75	144	1728	20736	5625	900	10800	78,3	109,1
6	11	80	121	1331	14641	6400	880	9680	73,3	238,3
7	15	90	225	3375	50625	8100	1350	20250	90,9	4,6
8	14	80	196	2744	38416	6400	1120	15680	87,1	2,7
9	14	75	196	2744	38416	5625	1050	14700	87,1	2,7
10	13	80	169	2197	28561	6400	1040	13520	82,9	34,2
11	16	100	256	4096	65536	10000	1600	25600	94,3	31,0
12	12	90	144	1728	20736	8100	1080	12960	78,3	109,1
13	22	120	484	10648	234256	14400	2640	58080	106,5	314,9
14	18	90	324	5832	104976	8100	1620	29160	100,0	125,6
15	16	85	256	4096	65536	7225	1360	21760	94,3	31,0
16	14	90	196	2744	38416	8100	1260	17640	87,1	2,7
17	16	85	256	4096	65536	7225	1360	21760	94,3	31,0
18	20	130	400	8000	160000	16900	2600	52000	104,0	233,1
19	26	100	676	17576	456976	10000	2600	67600	106,7	322,6
20	24	100	576	13824	331776	10000	2400	57600	107,4	347,6
Сума	309	1775	5153	93015	1809077	162025	28390	490020	1775	2755,9

Система рівнянь має вигляд:

$$20a_0 + 309a_1 + 5153a_2 = 1775$$

$$309a_0 + 5153a_1 + 93015a_2 = 28390$$

$$5153a_0 + 93015a_1 + 1809077a_2 = 490020$$

Розв'язавши систему рівнянь отримаємо рівняння параболі:

$$y_x = -7,66 + 9,53 \cdot X - 0,197 \cdot X^2$$

Звідси оптимальне значення фактора $X_{\text{опт.}} = (-1)9,53 : 2 \cdot (-0,197) = 24,1$ (т/га). Тобто при внесенні 24,1 т/га органічних добрив спостерігається максимальна

Для оцінки тісноти зв'язку з допомогою індекса кореляції попередньо обчислимо середню урожайність та дисперсії урожайності:

Індекс кореляції: $137,8 : 224,7 = 0,61$ або 61%.

Слід підкреслити, що з допомогою індекса кореляції можна визначати тісноту також лінійного зв'язку. Він в цьому випадку завжди буде дорівнювати значенню лінійного коефіцієнта кореляції.

У моделях множинної кореляції залежна змінна у розглядається як функція кількох (в загальному випадку) незалежних змінних x .

Припущення про існування лінійного рівняння множинної регресії може бути представлене в такому вигляді:

$$\bar{y}_{x_1, \dots, x_n} = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \dots + a_n x_n.$$

Із погляду геометрії це рівняння визначає в просторі площини відповідних змінних $X_1, X_2, X_3, \dots, X_n$ та y .

Параметри $a_0, a_1, a_2, \dots, a_n$ наведеного рівняння знаходять, розв'язавши систему нормальних рівнянь:

[illegible]

Множинне кореляційне рівняння встановлює зв'язок між досліджуваними ознаками і дає змогу вирахувати очікувані значення результативної ознаки під дією включених в аналіз ознак -факторів, пов'язаних із даним рівнянням. У вузькому розумінні рішення кореляційних моделей охоплює операції по знаходженню числових значень параметрів досліджуваних залежностей.

При вивченні множинного кореляційного зв'язку результативної ознаки, наприклад, із двома факторними аналітичне рівняння регресії має вигляд:

$$\bar{y}_{x_1x_2} = a_0 + a_1x_1 + a_2x_2.$$

Параметри a_0 , a_1 і a_2 розраховують розв'язуючи систему нормальних рівнянь:

$$\begin{cases} a_0n + a_1\sum x_1 + a_2\sum x_2 = \sum y; \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 = \sum x_1y; \\ a_0\sum x_2 + a_1\sum x_1x_2 + a_2\sum x_2^2 = \sum x_2y. \end{cases}$$

Для оцінки тісноти зв'язків у множинних лінійних кореляційних моделях застосовують три види коефіцієнтів кореляції: 1) прості коефіцієнти кореляції 2) частинні коефіцієнти кореляції 3) множинні коефіцієнти кореляції.

У множинних кореляційних моделях коефіцієнти простої кореляції між результативною ознакою і факторними, а також між самими факторними ознаками розраховують за вже відомими формулами (для двофакторної моделі буде три простих коефіцієнти кореляції):

$$r_{yx_1} = \frac{yx_1 - \bar{y} \times \bar{x}_1}{\sigma_y \times \sigma_{x_1}}; r_{yx_2} = \frac{yx_2 - \bar{y} \times \bar{x}_2}{\sigma_y \times \sigma_{x_2}}; r_{x_1x_2} = \frac{x_1x_2 - \bar{x}_1 \times \bar{x}_2}{\sigma_{x_1} \times \sigma_{x_2}}.$$

Оскільки існує залежність між факторами, які включені до множинної моделі, то простий коефіцієнт кореляції не може характеризувати «чистого» впливу даного фактора на результативну ознаку. В простому коефіцієнті кореляції крім «чистого» впливу даного фактора є також певна частина опосередкованого впливу інших факторів.

Для оцінки частинної (чистої) кореляції між двома досліджуваними ознаками, обчислюють частинні коефіцієнти кореляції. Отже, при дослідженні впливу двох факторних ознак частковий коефіцієнт кореляції характеризує тісноту лінійного зв'язку результативної ознаки (y) з однією із факторних ознак (x_1) при виключенні дії іншої факторної ознаки (x_2) і навпаки, вплив ознаки x_2 при виключенні впливу ознаки x_1 .

Формули для визначення значень величин частинних коефіцієнтів кореляції мають вигляд:

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} \times r_{x_1x_2}}{\sqrt{1 - r_{yx_2}^2} \sqrt{1 - r_{x_1x_2}^2}}; r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} \times r_{x_1x_2}}{\sqrt{1 - r_{yx_1}^2} \sqrt{1 - r_{x_1x_2}^2}}.$$

Перша з наведених формул характеризує ступінь тісноти зв'язку результативної ознаки із факторною x_1 при виключенні впливу фактора x_2 , друга – аналогічну залежність при виключенні впливу фактора x_1 .

Величина частинного коефіцієнта кореляції подібно до простого коефіцієнта може мати додатне і від'ємне значення, тобто вказувати на прямий чи обернений зв'язок між досліджуваними ознаками, а за модулем знаходиться між "0" і "1".

Використання в аналізі частинної кореляції дає змогу поглибити кількісне вивчення досліджуваних зв'язків між соціально-економічними процесами і явищами, виявляючи вплив конкретних причин варіації результативної ознаки.

Для оцінки ступеня впливу тісноти зв'язку між результативною і сукупністю

кількох факторних ознак обчислюють коефіцієнт множинної (сукупної) кореляції.

Коефіцієнт множинної кореляції характеризує інтенсивність кореляції (або ступінь тісноти зв'язку) між залежною змінною та кількома незалежними змінними. Його величина не може бути меншою, ніж абсолютна величина будь-якого коефіцієнта простої чи частинної кореляції. Множинний коефіцієнт кореляції завжди має додатний знак і змінюється від 0 до 1.

Загальний математичний вираз коефіцієнта множинної кореляції має вигляд:

$$R_{y x_1, \dots, x_n} = \sqrt{\frac{\sigma_r^2}{\sigma_y^2}},$$

де σ_r^2 – факторна (відтворена) дисперсія; σ_y^2 – загальна дисперсія.

Існують й інші формули множинного коефіцієнта кореляції.

Коефіцієнт множинної кореляції можна обчислити за значеннями величин коефіцієнтів простої кореляції. Так, для двофакторної моделі величину його розраховують за формулою:

$$R_{y x_1 x_2} = \sqrt{\frac{r_{y x_1}^2 + r_{y x_2}^2 - 2r_{y x_1} \times r_{x_1 x_2} \times r_{y x_2}}{1 - r_{x_1 x_2}^2}}.$$

В кореляційному аналізі можна визначити не тільки ступінь тісноти у вигляді коефіцієнтів кореляції чи індексів кореляції, а й встановити, на скільки процентів варіація результативної ознаки обумовлена впливом кожного фактора зокреми або усіх факторів разом взятих. Для цього обчислюють коефіцієнти детермінації.

Для простої кореляції коефіцієнт детермінації визначається як квадрат простого коефіцієнта кореляції (індекса кореляції): $d = r^2 \cdot 100$.

Аналогічно визначається множинний коефіцієнт детермінації: $D = R^2 \cdot 100$.

Частинний коефіцієнт детермінації (наприклад, для фактора x_1) визначають за формулою: $d_1 = r_{y x_1} \cdot a_1 \cdot (\sigma_{x_1} : \sigma_y) \cdot 100$.

Коефіцієнти детермінації мають додатні значення. Сума частинних коефіцієнтів детермінації для даної кореляційної моделі завжди дорівнює множинному коефіцієнту детермінації.

Розглянемо приклад аналізу множинної кореляції за даними табл.5.3.

Система рівнянь має вигляд:

$$20a_0 + 293a_1 + 1865a_2 = 506$$

$$293a_0 + 4483a_1 + 28005a_2 = 7549$$

$$1865a_0 + 28005a_1 + 178825a_2 = 47710$$

Розв'язавши систему рівнянь отримуємо кореляційне рівняння залежності урожайності гороху від двох факторів:

Таблиця 5.3.Залежність урожайності гороху (У,ц/га) від внесення органічних (X1, т/га) та мінеральних (X2,кг д.р./га)

№	У	X ₁	X ₂	У ²	X ₁ ²	X ₂ ²	УX ₁	УX ₂	X ₁ X ₂	УX ₁ X ₂	(УX ₁ X ₂ -Ус) ²
1	22	10	70	484	100	4900	220	1540	700	21,9	11,7
2	21	10	80	441	100	6400	210	1680	800	22,0	10,7
3	20	12	75	400	144	5625	240	1500	900	23,3	4,1
4	24	12	90	576	144	8100	288	2160	1080	23,5	3,2
5	24	12	80	576	144	6400	288	1920	960	23,3	3,8
6	23	10	90	529	100	8100	230	2070	900	22,2	9,7
7	21	14	90	441	196	8100	294	1890	1260	24,8	0,2
8	24	12	100	576	144	10000	288	2400	1200	23,7	2,7
9	25	15	80	625	225	6400	375	2000	1200	25,3	0,0
10	26	14	85	676	196	7225	364	2210	1190	24,7	0,3
11	25	16	90	625	256	8100	400	2250	1440	26,1	0,7
12	30	15	85	900	225	7225	450	2550	1275	25,4	0,0
13	26	16	100	676	256	10000	416	2600	1600	26,3	1,0
14	28	18	95	784	324	9025	504	2600	1710	27,5	5,0
15	26	15	100	676	225	10000	390	2600	1500	25,6	0,1
16	26	18	95	676	324	9025	468	2470	1710	27,5	5,0
17	30	18	90	900	324	8100	540	2700	1620	27,5	4,7
18	29	16	110	841	256	12100	464	3190	1760	26,4	1,3
19	26	20	120	676	400	14400	520	3120	2400	29,2	15,5
20	30	20	140	900	400	19600	600	4200	2800	29,5	18,0
Σ	506	293	1865	12978	4483	178825	7549	47710	28005	506	97,8

$$У_{x_1x_2} = 14,21 + 0,66X_1 + 0,015X_2$$

З рівняння робимо висновок: збільшення органічних добрив на 1 т/га обумовлює підвищення урожайності гороху в середньому на 0,66 ц/га незалежно від внесення мінеральних добрив; збільшення мінеральних добрив на 1 кг/га обумовлює підвищення урожайності гороху в середньому на 0,015 ц/га незалежно від внесення органічних добрив.

Обчислимо параметри, які необхідні для визначення показників тісноти зв'язку.

Середні значення:

$$Ус = 506:20=25,3$$

$$X_{1c} = 293:20=14,65$$

$$X_{2c} = 1865:20 =93,25$$

Середні квадрати:

$$У^2=12978:20=648,9$$

$$X_1^2=4483:20=224,2$$

$$X_2^2=178825:20=8941,25$$

Середні добутки:

$$УX_1=7549:20=377,5$$

$$УX_2=47710:20=2385,5$$

$$X_1X_2=28005:20=1400,3$$

Дисперсії:

$$\sigma_y^2 = 648,9 - 25,3^2 = 8,81$$

$$\sigma_{x_1}^2 = 224,2 - 14,65^2 = 9,53$$

$$\sigma_{x_2}^2 = 8941,25 - 93,25^2 = 245,69$$

Середні квадратичні відхилення:

$$\sigma_y = 2,97$$

$$\sigma_{x_1} = 3,09$$

$$\sigma_{x_2} = 15,7$$

Далі обчислюємо коефіцієнти кореляції (формули наведені вище).

Прості:

$$r_{yx_1} = 0,74$$

$$r_{yx_2} = 0,56$$

$$r_{x_1x_2} = 0,71$$

Частинні:

$$r_{yx_1 \cdot x_2} = 0,59$$

$$r_{yx_2 \cdot x_1} = 0,09$$

Множинний:

$$R = 0,74$$

Коефіцієнти детермінації:

частинні

$$d_1 = 50,9\%$$

$$d_2 = 4,6\%$$

множинний

$$D = 55,5\%$$

Тепер висновки про тісноту зв'язку. Варіація урожайності гороху на 55,5% обумовлена впливом двох факторів. З них на 50,9% - впливом фактора органічних добрив і на 4,6% - впливом мінеральних добрив.

5.7. КОРЕЛЯЦІЙНИЙ АНАЛІЗ В ЗАСТОСУНКУ EXCEL

Як видно з наведених вище розрахунків, кореляційний аналіз пов'язаний з досить громіздкими обчисленнями. Використовуючи сучасну комп'ютерну техніку можна суттєво скоротити затрати часу на обробку даних. Зокрема в середовищі Excel є пакети програм з аналізу статистичних даних, в тому числі й з кореляційного аналізу.

Розглянемо методику кореляційного аналізу в середовищі Excel (російська версія). Пакети кореляційного аналізу знаходяться в Меню "Сервис". Використовується послідовно три програми: "Описательная статистика", "Регрессионный анализ", «Корреляция».

Оцінка передумов кореляційного аналізу з допомогою програми

"Описательная статистика"

Застосування методу кореляційного аналізу передбачає наявність деяких передумов. Нехтування цими передумовами може призвести до знецінення результатів аналізу. Серед передумов кореляційного аналізу, як показує практичний досвід, особливої уваги заслуговують дві: 1) чітке уявлення про природу та структуру взаємозв'язків між досліджуваними ознаками з точки

зору причинно-наслідкових залежностей; 2) однорідність сукупності спостережень відносно досліджуваних ознак.

Що стосується першої передумови, то вона визначається на основі логічного якісного аналізу природи досліджуваних ознак та зв'язків між ними, а не кількісних обчислень. Слід підкреслити, що не завжди кореляція між ознаками свідчить про реальну безпосередню залежність. Наприклад, легко виявити кореляцію між рівнем витрат сімей на якісні продукти харчування та витратами на придбання модного одягу. Але ця кореляція не відображає реального безпосереднього зв'язку між цими ознаками. Адже витрати сім'ї на модний одяг зростають не тому, що більше витрачається коштів на якісні продукти. Очевидно що обидві ці ознаки залежать від спільного "прихованого" фактора – рівня матеріальної забезпеченості сім'ї.

Так само в сільському господарстві кореляція між показниками продуктивності праці та фондівдачі не відображає реальної безпосередньої залежності. Обидва ці показники залежать від спільних факторів (забезпечення виробничими ресурсами та ін.) тому і корелюють між собою. В економічному аналізі бажано вивчати лише ті кореляції, які відображають реальні безпосередні причинно-наслідкові залежності. Наприклад, кореляцію між добривами та урожайністю с.г.культур, продуктивністю праці та фондоозброєністю праці і т.п.

В багатофакторних кореляційних моделях необхідно звертати увагу на те, що одні фактори впливають на результативну ознаку безпосередньо, а інші опосередковано. Наприклад, розглянемо залежність рівня рентабельності зерна від факторів застосування добрив та урожайності. Фактор добрив має подвійний вплив на рентабельність: з одного боку він збільшує витрати на виробництво, а отже, знижує рентабельність, а з другого боку за рахунок добрив підвищується урожайність і (опосередковано через урожайність) зростає рентабельність. В кінцевому рахунку вплив добрив на рентабельність може бути позитивним або негативним в залежності від того чи окуповуються витрати на добрива вартістю додаткової урожайності. Вплив фактора урожайності на рентабельність також не є простим. Він охоплює як безпосередній вплив (незалежний від добрив), так і опосередкований вплив добрив через урожайність. Бажано в кореляційних моделях віддавати перевагу факторам одного рівня (особливо первинним факторам) щоб уникнути по можливості опосередкованих впливів, а отже і уникнути дублювання впливів факторів.

Не менш важливе значення має і друга передумова кореляційного аналізу – однорідність досліджуваної сукупності. Під однорідністю розуміють однотипність, однакісність. Наприклад не можна вважати однорідною сукупність підприємств, які знаходяться в різних природно-економічних умовах виробництва. Питання про однорідність визначається також на основі логічного аналізу. Але поряд з цим в статистиці розроблені методи кількісної оцінки однорідності сукупності. Найпростішим з них є обчислення критерія належності сумнівної варіанти до однорідної сукупності (позначимо його буквою K). Сумнівними варіантами завжди є мінімальне та максимальне

значення кожної досліджуваної ознаки. Відповідно формули для обчислення критерія такі : для мінімального значення

$$Kx(\min) = \frac{X_{\min} - \bar{X}}{S}$$

для максимального значення

$$Kx(\max) = \frac{X_{\max} - \bar{X}}{S}$$

Якщо значення критерія за модулем менші 3, то вважається, що сумнівне значення належить до однорідної сукупності. Якщо ж критерій перевищує 3, то сумнівне значення ознаки не належить до однорідної сукупності і дана одиниця спостереження виключається з аналізу.

Розглянемо приклад залежності рівня продуктивності праці на виробництві молока (затрати праці на 1 ц молока, люд.-год. –ознака Y) від факторів : щільність поголів'я корів на 100 га с.г.угідь (гол.-ознака X1), річні затрати кормів на 1 корову (грн.- ознака X2), річні затрати праці на 1 корову (люд.-год.- ознака X3). В середовищі EXCEL вкажемо вихідні дані у вигляді таблиці 5.4.

Таблиця 5.4. Вихідні дані до кореляційного аналізу

№	y	x1	x2	x3
1	10	20	600	250
2	8	30	300	200
3	12	10	310	300
4	10	22	550	260
5	8	29	810	240
6	11	10	300	300
7	12	12	300	290
8	10	29	530	290
9	11	12	350	270
10	11	10	610	250
11	8	30	600	200
12	9	29	590	230
13	12	9	300	210
14	7	14	780	220
15	13	25	320	280
16	9	28	590	240
17	7	28	790	220
18	11	9	340	280
19	9	21	540	250
20	7	28	800	230
21	10	29	500	290
22	8	28	730	210
23	12	28	320	280
24	13	11	340	280

25	10	22	600	240
26	10	9	790	200
27	8	27	350	270
28	11	17	570	260
29	9	13	770	210
30	12	24	390	270

В головному меню **Сервис** опція **Анализ данных** вибираємо програму **Описательная статистика**. В робочому вікні вказуємо: **Входной интервал** (блок вихідних даних), **Группирование по столбцам**, **Метки в первой строке**, **Выходной интервал**, **Итоговая статистика**, **Уровень надежности (95%)**, **К-ый наименьший (1)**, **К-ый наибольший (1)** і отримуємо такі розрахункові показники (табл.5.5).

Таблица 5.5.Розрахункові показники програми “Описательная статистика”

	У	X1	X2	X3
Среднее	9,93	20,43	522,33	250,67
Стандартная ошибка	0,33	1,48	33,58	5,87
Медиана	10,00	22,00	545,00	250,00
Мода	10,00	28,00	300,00	280,00
Стандартное отклонение (S)	1,80	8,10	183,92	32,16
Дисперсия выборки	3,24	65,63	33825,40	1034,02
Эксцесс	-1,03	-1,69	-1,38	-1,25
Асимметричность	-0,05	-0,26	0,20	-0,14
Интервал	6,00	21,00	510,00	100,00
Минимум	7,00	9,00	300,00	200,00
Максимум	13,00	30,00	810,00	300,00
Сумма	298,00	613,00	15670,00	7520,00
Счет	30,00	30,00	30,00	30,00

На основі наведених в табл.5.5 показників можна обчислити значення критерія належності сумнівних варіант до однорідної сукупності господарств.Спочатку обчислимо критерій для мінімальних значень ознак:

$$K_u(\text{мін}) = /7,00-9,93/ : 1,80 = 1,63;$$

$$K_{x1}(\text{мін}) = /9,00-20,43/ : 8,10 = 1,41;$$

$$K_{x2}(\text{мін}) = /300-522,33/ : 183,92 = 1,20;$$

$$K_{x3}(\text{мін}) = /200-250,67/ : 32,16 = 1,57.$$

Оскільки значення критерія менші 3, можна зробити висновок про те, що мінімальні значення усіх 4-ох ознак належать до однорідної сукупності господарств. Аналогічно обчислимо критерій для максимальних значень:

$$K_u(\text{мах}) = (13-9,93) : 1,80 = 1,70;$$

$$K_{x1}(\text{мах}) = (30-20,43) : 8,10 = 1,18;$$

$$K_{x2}(\text{мах}) = (810-522,33) : 183,92 = 1,56;$$

$$K_{x3}(\text{мах}) = (300-250,67) : 32,16 = 1,53.$$

Як видно з розрахунків, максимальні значення усіх ознак також належать до однорідної сукупності господарств. Отже, можна переходити до власне кореляційного аналізу.

Оцінка параметрів кореляційного рівняння з допомогою програми “Регрессия”

Загальний вигляд рівняння залежності рівня продуктивності праці від факторів має такий вигляд : $Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_3$. Обчислимо його з допомогою програми “Регрессия”). Для цього в головному меню **Сервис** опція **Анализ данных** виберемо програму **Регрессия** і вкажемо: **Входной интервал** для Y та для блоку факторів X (вихідні дані – табл.4), **Метки**, **Уровень надежности (95%)**, **Выходной интервал**. Результати обчислень (**Вывод итогов**) наведені в табл.5.6.

Таблица 5.6. Розрахункові показники програми “Регрессия”

ВЫВОД ИТОГОВ					
<i>Регрессионная статистика</i>					
Множественный R	0,8211				
R-квадрат	0,6743				
Нормированный R-квадрат	0,6367				
Стандартная ошибка	1,0844				
Наблюдения	30,0000				
<i>Дисперсионный анализ</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значим. F</i>
Регрессия	3,0000	63,29244	21,09748	17,9411	1,623E-06
Остаток	26,0000	30,57423	1,17593		
Итого	29,0000	93,86667			
	<i>Коэфф.</i>	<i>Ст. ошибка</i>	<i>t-стат.</i>	<i>P-Знач.</i>	
Y-пересечение	9,5309	2,56212	3,71992	0,0010	
x1	-0,0779	0,02564	-3,03966	0,0053	
x2	-0,0042	0,00141	-3,01356	0,0057	
x3	0,0168	0,00795	2,11184	0,0445	

З табл.5.6 видно, що рівняння залежності рівня продуктивності праці від факторів має такий конкретний вигляд: $Y = 9,53 - 0,0779 \cdot X_1 - 0,0042 \cdot X_2 + 0,0168 \cdot X_3$.

Далі обчислимо коефіцієнти кореляції з допомогою програми “Корреляция”. В головному меню **Сервис** опція **Анализ данных** вибираємо програму **Корреляция** і вкажемо: **Входной интервал** (вихідні дані – табл.5.4), **Метки**, **Входной интервал**. Результати обчислень наведені в табл.5.7.

Таблиця 5.7. Коефіцієнти кореляції (програма “Корреляция”)

	у	x1	x2	x3
у	1			
x1	-0,511	1,000		
x2	-0,703	0,241	1,000	
x3	0,633	-0,186	-0,615	1

Тут обов’язково треба звернути увагу на знаки при коефіцієнтах кореляції (табл.5.7 стовпчик у) та знаки при відповідних коефіцієнтах регресії (табл.5.6 стовпчик *коэфф.*). Ці знаки повинні бути однаковими. В такому випадку отримані кореляційні характеристики вважаються коректними тобто мають реальний економічний зміст. Якщо ж знаки відносно певного фактора не співпадають, то отримані результати обчислень є некоректними і це не дозволяє пояснити економічний зміст параметрів кореляційної моделі. Як правило, некоректність моделі виникає внаслідок тісних міжфакторних зв’язків (мультиколінеарності). В такому випадку необхідно виключити з моделі “некоректний” фактор і заново виконати кореляційні обчислення з факторами, що залишилися.

У нашому прикладі модель коректна. Адже знаки при відповідних коефіцієнтах кореляції та коефіцієнтах регресії співпадають. Для фактора X1 маємо $-0,0779$ (табл.5.6) та $-0,511$ (табл.5.7). Для фактора X2 відповідно $-0,0042$ та $-0,703$. Для фактора X3 відповідно $0,0168$ та $0,633$.

Після цього обчислюють коефіцієнти детермінації (d_i) за окремими факторами на основі даних табл.5.5-5.7.

Формула коефіцієнта детермінації така: $d_i = a_i \cdot r_{ux_i} \cdot S_{x_i} / S_y$, де i - номер фактора, a_i – коефіцієнт регресії i -го фактора в табл.5.8, r_{ux_i} – коефіцієнт кореляції ознаки у з i -им фактором в табл.5.9, S_{x_i} – стандартне відхилення i -го фактора в табл.5.7, S_y – стандартне відхилення ознаки у в табл.5.6.

Наведемо обчислення:

$$d_1 = (-0,0779) \cdot (-0,511) \cdot 8,10 / 1,80 = 0,179 \text{ або } 18\%;$$

$$d_2 = (-0,0042) \cdot (-0,703) \cdot 183,92 / 1,80 = 0,302 \text{ або } 30\%;$$

$$d_3 = (0,0168 \cdot 0,633 \cdot 32,16) / 1,80 = 0,190 \text{ або } 19\%.$$

Сума коефіцієнтів за факторами повинна дорівнювати множинному коефіцієнту детермінації (R-квадрат в табл.5.6). В нашому прикладі: $0,179 + 0,302 + 0,190 = 0,671$. В табл.5.3. R-квадрат = $0,674$. Розходження в тисячних долях можна пояснити округленнями чисел в процесі обчислень.

Тепер можна приступати до пояснення економічного змісту отриманих в процесі кореляційного аналізу характеристик тобто висновків. Це найважливіший етап аналізу. Адже саме для цього і здійснюється аналіз. Висновки мають три складових: 1) оцінка тісноти зв’язку (за коефіцієнтами детермінації); 2) оцінка пропорції впливу факторів на результативну ознаку (за коефіцієнтами регресії); 3) оцінка суттєвості (достовірності) впливу факторів (за критерієм Фішера – F та критерієм Стюдента – t-статистика).

Як свідчать коефіцієнти детермінації, рівень продуктивності праці на виробництві молока (y) на 67% залежить від сумарного впливу трьох досліджуваних факторів, в т.ч.: на 18 % від фактора щільності поголів'я корів (x_1), на 30% від затрат кормів на 1 корову (x_2) і на 19% від затрат праці на 1 корову (x_3).

Коефіцієнти регресії (табл.5.6) показують, що збільшення щільності поголів'я корів на 1 голову з розрахунку на 100 га с.г.угідь обумовлює зниження затрат праці на 1 ц молока в середньому на 0,0779 люд.-год.; збільшення затрат кормів на 1 грн. з розрахунку на 1 корову обумовлює зниження затрат праці на 1 ц молока в середньому на 0,0042 люд.-год.; збільшення затрат праці на 1 люд.-год з розрахунку на 1 корову обумовлює підвищення затрат праці на 1 ц молока в середньому на 0,0168 люд.-год.

Достовірність впливу кожного фактора на результативну ознаку можна визначити з допомогою критерія Стюдента (t – статистика в табл.5.6). Для цього необхідно спочатку знайти в таблиці критичних значень t -критерія (табл. 7,8) відповідне нашому прикладові критичне значення. Номер рядка в таблиці додатка 1 (число ступенів вільності) визначається як різниця між кількістю одиниць сукупності (в нашому прикладі 30 господарств) та кількістю досліджуваних ознак (в нашому прикладі 4). Отже, маємо 26-ий рядок (30-4). В рядку 26 табл. 7,8 вказане критичне значення t -критерія: $t_k = 2,06$. Оскільки розрахункове значення t для кожного фактора (табл.5.6) більше ніж критичне значення, є підстави стверджувати, що вплив кожного з факторів на результативну ознаку достовірний (суттєвий). Якби розрахункове значення t було меншим критичного або дорівнювало йому, то не було б підстав вважати достовірним вплив відповідного фактора.

Достовірність кореляційної моделі в цілому можна визначити з допомогою F –критерія (табл.5.6). Спочатку в табл.5.9 знаходимо критичне значення F -критерія для нашого прикладу. Номер рядка таблиці визначається як різниця між кількістю одиниць досліджуваної сукупності (в нашому прикладі 30 господарств) та кількістю параметрів в рівнянні регресії (в нашому прикладі 4), Отже, маємо 26-ий рядок (30-4). Номер стовпчика таблиці визначається як кількість параметрів рівняння регресії мінус 1. Тобто маємо 3-й стовпчик (4-1). На перетині 26-го рядка та 3-го стовпчика в табл. 5.9 знаходимо критичне значення F -критерія, яке дорівнює 2,98. Оскільки розрахункове значення F -критерія (17,94 в табл.5.6) більше ніж критичне значення (2,98), то є підстави вважати достовірним множинний коефіцієнт детермінації або кореляційну модель в цілому. Якби розрахункове значення F -критерія було меншим критичного або дорівнювало йому, то не було б підстав вважати достовірною кореляційну модель і в такому разі її не слід пояснювати економічний зміст отриманих кореляційних характеристик.

**Таблиця 5.8. Критичні значення t-критерія Стюдента
при ймовірності $P=0,95$**

Число ступенів вільності	Критичне значення t-критерія
10	3,23
11	2,20
12	2,18
13	2,16
14	2,14
15	2,13
16	2,12
17	2,11
18	2,10
19	2,09
20	2,09
21	2,08
22	2,07
23	2,07
24	2,06
25	2,06
26	2,06
27	2,05
28	2,05
29	2,05
30	2,04
40	2,02
50	2,00
60	1,98
~	1,96

Таблиця 5.9. Критичні значення F-критерія при ймовірності P=0,95

	1	2	3	4	5
10	4,96	4,10	3,71	3,48	3,33
12	4,75	3,88	3,49	3,26	3,11
14	4,60	3,74	3,34	3,11	2,96
16	4,49	3,63	3,24	3,01	2,85
18	4,41	3,56	3,16	2,93	2,77
20	4,35	3,49	3,10	2,87	2,71
22	4,30	3,44	3,05	2,82	2,66
24	4,26	3,40	3,01	2,78	2,62
26	4,22	3,37	2,98	2,74	2,59
28	4,20	3,34	2,95	2,71	2,56
30	4,17	3,32	2,92	2,69	2,53
35	4,11	3,26	2,86	2,63	2,48
40	4,08	3,23	2,84	2,61	2,45
50	4,03	3,18	2,79	2,56	2,40
60	4,00	3,15	2,76	2,52	2,37
~	3,84	2,99	2,60	2,37	2,21

Лекція 6. Регресійний аналіз

План

- 6.1. Види взаємозв'язків
- 6.2. Регресійний аналіз
- 6.3. Оцінка щільності та перевірка істотності кореляційного зв'язку
- 6.4. Рангова кореляція
- 6.5. Оцінка узгодженості варіації атрибутивних ознак

6.1. Види взаємозв'язків

Усі явища навколишнього світу, соціально-економічні зокрема, взаємозв'язані й взаємозумовлені. У складному переплетенні всеохоплюючого взаємозв'язку будь-яке явище є наслідком дії певної множини причин і водночас — причиною інших явищ. Причини та наслідки пов'язані неперервними ланцюгами прямо або опосередковано, що схематично ілюструє рис. 6.1. Так, незалежне в межах зображеного графа зв'язку явище x_1 є причиною явищ x_2 , x_3 , x_5 . Із них явище x_3 , у свою чергу, впливає на x_4 , а x_4 — на x_5 .

Поряд із причинними існують зв'язки паралельних явищ, на які впливає спільна причина. На рис. 6.1 це зв'язок між x_2 і x_3 , які мають спільну причину x_1 .

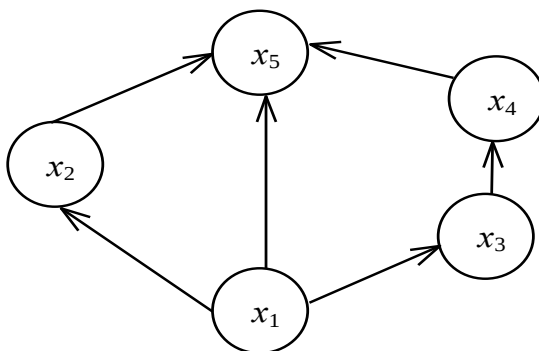


Рис. 6.1. Ланцюжок взаємозв'язків

Визначальна мета вимірювання взаємозв'язків — виявити і дати кількісну характеристику причинних зв'язків. Суть причинного зв'язку полягає в тому, що за певних умов одне явище спричинює інше. Причина сама по собі не визначає наслідку, останній залежить також від умов, в яких діє причина. Вивчаючи закономірності зв'язку, причини та умови об'єднують в одне поняття «фактор». Відповідно **ознаки**, які характеризують фактори, називаються **факторними**, а ті, що характеризують наслідки, — **результативними**.

Аналіз характеру взаємозв'язків та оцінювання сили впливу факторів на результат є передумовою розробки науково обґрунтованих управлінських рішень, прогнозування й регулювання складних соціально-економічних явищ і процесів.

Розрізняють два типи зв'язків — *функціональні* та *стохастичні*. У разі функціонального зв'язку кожному значенню фактора x відповідає одне або кілька чітко визначених значень y . Такою, наприклад, є залежність довжини ртутного стовпчика від температури навколишнього середовища. Знаючи x , можна в кожному окремому випадку точно визначити результат y . Скажімо, при проведенні валютних операцій для переведення суми в національній валюті C в еквівалентну їй суму в іноземній валюті S використовують валютний курс L :

$$S = C : L \text{ і навпаки } C = S \cdot L.$$

У соціально-економічних науках до функціонального типу належать зв'язки між показниками — адитивні ($a + b + c$) або мультиплікативні ($a = bc$, $c = a/b$), а також залежність середніх величин від структури сукупності (див. підрозд. 9.5—9.6).

На відміну від функціональних, стохастичні зв'язки неоднозначні. Наприклад, залежність захворюваності населення від екологічного стану довкілля. На забруднених радіонуклідами територіях, як і на інших, стан здоров'я мешканців коливається від «тяжко хворого» до «практично здорового». Проте в середньому в таких регіонах порівняно з екологічно чистими захворюваність значно вища.

Стохастичні зв'язки виявляються як узгодженість варіації двох чи більше ознак. У ланці зв'язку « $x \rightarrow y$ » кожному значенню ознаки x відповідає певна множина значень ознаки y , які утворюють так званий **умовний розподіл**. Стохастичний зв'язок, відбиваючи множинність причин і наслідків, виявляється в зміні умовних розподілів, що схематично ілюструє табл. 6.1.

Якщо умовні розподіли замінюються одним параметром — середньою $\bar{y}_i$, то такий зв'язок називають **кореляційним**. Отже, кореляційний зв'язок є різновидом стохастичного і виявляється зміною середніх умовних розподілів.

Таблиця 6.1 - ВИДИ ВЗАЄМОЗВ'ЯЗКІВ І ЇХ ОСОБЛИВОСТІ

Факторна ознака x_i	Результативна ознака y за наявності зв'язку		
	функціонального	стохастичного	кореляційного
x_1	y_1	$y_1 y_2$	$\bar{y}_1$
x_2	y_2	$y_1 y_2 y_3$	$\bar{y}_2$
x_3	y_3	$y_2 y_3 y_4$	$\bar{y}_3$
...	...	...	...
x_m	y_m	$y_{m-1} y_m$	$\bar{y}_m$

Наявність стохастичного зв'язку можна виявити, скориставшись комбінаційним розподілом елементів сукупності. Такий розподіл наведено в табл. 6.2. Сукупність шахт регіону поділено на групи за двома ознаками: x — глибиною розробки вугільних пластів і y — фондомісткістю видобутку вугілля. Кожна група за глибиною розробки пласта характеризується своїм особливим розподілом шахт за фондомісткістю видобутку вугілля. Це умовні розподіли. Порівняння умовних розподілів указує на тенденцію підвищення

фондомісткості зі зростанням глибини розробки пластів. Звичайно, для кожної окремої шахти така залежність може не виявитись через вплив інших факторів. Певні межі варіації фондомісткості характерні для кожної групи. Так, на шахтах, де глибина розробки пластів 500 ... 700 м, фондомісткість коливається в межах від 18 до 26 грн. за тону. Проте середній рівень фондомісткості в цій групі вищий порівняно з попередньою групою (300 ... 500 м) і нижчий порівняно з наступною (700 і більше):

$$\bar{y}_1 = \frac{19 \cdot 9 + 21 \cdot 7 + 23 \cdot 1}{17} = 20,0 ;$$

$$\bar{y}_2 = \frac{21 \cdot 8 + 23 \cdot 27 + 25 \cdot 5}{40} = 22,9 ;$$

$$\bar{y}_3 = \frac{23 \cdot 6 + 25 \cdot 15 + 27 \cdot 4}{25} = 24,8 ;$$

$$\bar{y}_4 = \frac{25 \cdot 8 + 27 \cdot 10}{18} = 26,1 .$$

Середні рівні фондомісткості видобутку вугілля наведено в останній графі таблиці. Зростання групових середніх від групи до групи свідчить про наявність кореляційного зв'язку між глибиною розробки пласта і фондомісткістю вугілля. Отже, кореляційний зв'язок, як і стохастичний, — це властивість сукупності в цілому, а не окремих її елементів.

Таблиця 6.2- КОМБІНАЦІЙНИЙ РОЗПОДІЛ ШАХТ ЗА ГЛИБИНОЮ РОЗРОБКИ ПЛАСТІВ ТА ФОНДОМІСТКІСТЮ ВУГІЛЛЯ

Глибина розробки пласта, м	Кількість шахт з рівнем фондомісткості, грн. / т						Середній рівень фондомісткості, грн. / т
	До 20	20—22	22—24	24—26	26 і більше	Разом	
До 300	9	7	1			17	20,0
300—500		8	27	5		40	22,9
500—700			6	15	4	25	24,8
700 і більше				8	10	18	26,1
По сукупності в цілому	9	15	34	28	14	100	23,5

Отже, можна не лише стверджувати, що існує кореляційний зв'язок між факторною x і результативною y ознаками, а й визначати, як у середньому змінюється y зі зміною x на одиницю. **Ефекти впливу** x на y визначаються відношенням приростів середніх групових цих величин $\Delta y : \Delta x$. Наприклад, у другій групі порівняно з першою глибина розробки вугільного пласта більша на 200 м, а фондомісткість видобутку вугілля на $22,9 - 20,0 = 2,9$ грн. / т. Звідси

$$\frac{\Delta y}{\Delta x} = \frac{2,9}{200} = 0,0145 .$$

Тобто, зі зростанням глибини розробки пласта на 100 м фондомісткість зростає в середньому на 1,45 грн. / т.

Аналогічно розраховані ефекти впливу глибини розробки пласта на фондомісткість вугілля у третій групі становлять 0,95, у четвертій — 0,65 грн. на тону вугілля.

6.2. Регресійний аналіз

Важливою характеристикою кореляційного зв'язку є *лінія регресії* — емпірична в моделі аналітичного групування і теоретична в моделі регресійного аналізу. *Емпірична лінія регресії* представлена груповими середніми результативною ознаки $\bar{y}_j$, кожна з яких належить до відповідного інтервалу значень групувального фактора x_j . *Теоретична лінія регресії* описується певною функцією $Y = f(x)$, яку називають *рівнянням регресії*, а Y — *теоретичним рівнем результативної ознаки*.

На відміну від емпіричної, теоретична лінія регресії неперервна. Так, вважають, що маса дорослої людини в кілограмах має бути на 100 одиниць менша за її зріст у сантиметрах. Співвідношення між масою і зростом можна записати у вигляді рівняння: $Y = -100 + x$, де y — маса; x — зріст.

Безперечно, така форма зв'язку між масою та зростом людини надто спрощена. Насправді збільшення маси не жорстко пропорційне до збільшення зросту. Люди одного зросту мають різну масу, проте в середньому зі збільшенням зросту маса зростає. Для точнішого відображення зв'язку між цими ознаками в рівняння слід увести другий параметр, який був би коефіцієнтом пропорційності при x , тобто $Y = -100 + bx$.

Рівняння регресії в такому вигляді описує числове співвідношення варіації ознак x і y в середньому. Коефіцієнт пропорційності при цьому відіграє визначальну роль. Він показує, на скільки одиниць у середньому змінюється y зі зміною x на одиницю. У разі прямого зв'язку b — величина додатна, у разі оберненого — від'ємна.

Подаючи y як функцію x , тим самим абстрагуються від множинності причин, штучно спрощуючи механізм формування варіації y . Аналіз причинних комплексів здійснюється за допомогою множинної регресії.

Різні явища по-різному реагують на зміну факторів. Для того щоб відобразити характерні особливості зв'язку конкретних явищ, статистика використовує різні за функціональним видом регресійні рівняння. Якщо зі зміною фактора x результат y змінюється більш-менш рівномірно, такий зв'язок описується лінійною функцією $Y = a + bx$. Коли йдеться про нерівномірне співвідношення варіацій взаємозв'язаних ознак (наприклад, коли природи значень y зі зміною x прискорені чи сповільнені або напрям зв'язку змінюється), застосовують нелінійні регресії, зокрема:

$$\text{степеневу} \quad Y = ax^b;$$

$$\text{гіперболічну} \quad Y = a + \frac{b}{x};$$

$$\text{параболічну} \quad Y = a + bx + cx^2 \text{ тощо.}$$

Вибір та обґрунтування функціонального виду регресії ґрунтується на теоретичному аналізі суті зв'язку. Нехай вивчається зв'язок між урожайністю та кількістю опадів. Надто мала і надто велика кількість опадів спричиняють зниження врожайності, максимальний її рівень можливий за умови оптимальної кількості опадів, тобто зі збільшенням факторної ознаки (опадів) урожайність спершу зростає, а потім зменшується. Залежність такого роду описується параболою $Y = a + bx + cx^2$.

Вивчаючи зв'язок між собівартістю y та обсягом продукції x , використовують рівняння гіперболи $Y = a + \frac{b}{x}$, де a — пропорційні витрати на одиницю продукції, b — постійні витрати на весь випуск.

Зауважимо, що теоретичний аналіз суті зв'язку, хоча й дуже важливий, лише окреслює особливості форми регресії і не може точно визначити її функціонального виду. До того ж у конкретних умовах простору і часу межі варіації взаємозв'язаних ознак x і y значно вужчі за теоретично можливі. І якщо кривина регресії невелика, то в межах фактичної варіації ознак зв'язок між ними досить точно описується лінійною функцією. Цим значною мірою пояснюється широке застосування лінійних рівнянь регресії:

$$Y = a + bx.$$

Параметр b (*коефіцієнт регресії*) — величина іменована, має розмірність результативної ознаки і розглядається як *ефект впливу* x на y . Параметр a — вільний член рівняння регресії, це значення y при $x = 0$. Якщо межі варіації x не містять нуля, то цей параметр має лише розрахункове значення.

Параметри рівняння регресії визначаються методом найменших квадратів, основна умова якого — мінімізація суми квадратів відхилень емпіричних значень y від *теоретичних* Y :

$$\sum (y - Y)^2 = \min.$$

Математично доведено, що значення параметрів a та b , при яких мінімізується сума квадратів відхилень, визначаються із системи нормальних рівнянь:

$$\begin{aligned}\sum y &= na + b \sum x, \\ \sum xy &= a \sum x + b \sum x^2.\end{aligned}$$

Розв'язавши цю систему, знаходимо такі значення параметрів:

$$\begin{aligned}b &= \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - \sum x \sum x}, \\ a &= \bar{y} - b \bar{x}.\end{aligned}$$

Розглянемо порядок обчислення параметрів лінійної регресії на прикладі зв'язку між урожайністю зернових і кількістю внесених добрив (у центнерах

діючої поживної речовини — д. р.). Значення взаємозв'язаних ознак та необхідні для розрахунку параметрів величини наведено в табл. 6.3.

$$\sum x = 12; \sum y = 224; \sum xy = 342,8; \sum x^2 = 18,68;$$

$$\bar{x} = 12 : 8 = 1,5; \bar{y} = 224 : 8 = 28.$$

Таблиця 6.3 - ДО РОЗРАХУНКУ ПАРАМЕТРІВ ЛІНІЙНОЇ РЕГРЕСІЇ,
ТЕОРЕТИЧНИХ РІВНІВ І ЗАЛИШКОВИХ ВЕЛИЧИН

Номер господарства	Кількість внесених добрив x , д. р	Урожайність зернових y , ц/га	xy	x^2	Y	$y - Y$	$(y - Y)^2$
1	1,1	23	25,3	1,21	24	-1	1
2	1,4	25	35,0	1,96	27	-2	4
3	1,2	26	31,2	1,44	25	1	1
4	2,0	33	66,0	4,00	33	0	0
5	1,5	27	40,5	2,25	28	-1	1
6	1,3	28	36,4	1,69	26	2	4
7	1,8	30	54,0	3,24	31	-1	1
8	1,7	32	54,4	2,89	30	2	4
Разом	12,0	224	342,8	18,68	224	$\times$	16

Користуючись цими величинами, визначаємо:

$$b = \frac{8 \cdot 342,8 - 12 \cdot 224}{8 \cdot 18,68 - 12 \cdot 12} = \frac{54,4}{5,44} = 10,0 \text{ (ц/га);}$$

$$x = 28 - 10,0 \cdot 1,5 = 13,0.$$

Отже, рівняння регресії має вигляд

$$Y = 13,0 + 10,0x,$$

тобто кожний центнер внесених добрив (у перерахунку на діючу поживну речовину) дає приріст урожайності в середньому 10 ц/га. Якщо добрива зовсім не вносити ($x = 0$), то урожайність зернових не перевищить 13,0 ц/га.

Рівняння регресії відбиває закон зв'язку між x і y не для окремих елементів сукупності, а для сукупності в цілому; закон, який абстрагує вплив інших факторів, виходить з принципу «за інших однакових умов». За цих умов очікувана врожайність зернових при внесенні добрив у обсязі 1,1 ц д. р. на 1 га становить $Y = 13 + 10 \cdot 1,1 = 24$ (ц/га). Для інших значень факторної ознаки x теоретичні рівні врожайності наведено в табл. 6.3. Вплив інших окрім x факторів зумовлює відхилення емпіричних значень y від теоретичних у той чи інший бік. Відхилення ($y - Y$) називають **залишками** і позначають символом e . Залишки, як правило, менші за відхилення від середньої, тобто $(y - Y) \leq (y - \bar{y})$.

У нашому прикладі

$$\sum_1^n (y - \bar{y})^2 = 84, \quad \sum_1^n (y - Y)^2 = 16.$$

Відповідно **загальна дисперсія** врожайності

$$\sigma_y^2 = \frac{1}{n} \sum_1^n (y - \bar{y})^2 = \frac{84}{8} = 10,5,$$

залишкова дисперсія

$$\sigma_e^2 = \frac{1}{n} \sum_{i=1}^n (y - Y)^2 = \frac{16}{8} = 2.$$

У невеликих за обсягом сукупностях коефіцієнт регресії схильний до випадкових коливань. Тому слід перевірити його істотність. Коли зв'язок лінійний, істотність коефіцієнта регресії перевіряють за допомогою t -критерію (Ст'юдента), статистична характеристика якого для гіпотези $H_0: b = 0$ визначається відношенням коефіцієнта регресії b до власної стандартної похибки μ_b , тобто $t = b/\mu_b$.

Стандартна похибка коефіцієнта регресії залежить від варіації факторної ознаки σ_x^2 , залишкової дисперсії σ_e^2 і числа ступенів свободи $k = n - m$, де m — кількість параметрів рівняння регресії:

$$\mu_b = \sqrt{\frac{\sigma_e^2}{\sigma_x^2(n-m)}}.$$

Для лінійної функції $m = 2$. За даними табл. 6.3 маємо:

$$\sigma_x^2 = \frac{18,68}{8} - 1,5^2 = 0,085, \quad \sigma_e^2 = 2.$$

Звідси $\mu_b = \sqrt{\frac{2}{0,085(8-2)}} \approx 2,0$ (ц/га), а $t = \frac{b}{\mu_b} = \frac{10}{2} = 5$, що перевищує критичне

значення двостороннього t -критерію $t_{0,95(6)} = 2,45$ (табл. 6.3). Гіпотеза про випадковий характер коефіцієнта регресії відхиляється, а отже, з імовірністю 0,95 вплив кількості внесених добрив на врожайність зернових визнається істотним.

Для коефіцієнта регресії, як і для будь-якої іншої випадкової величини, визначаються довірчі межі $b \pm t\mu_b$. У нашому прикладі довірчі межі коефіцієнта регресії з імовірністю 0,95 ($t = 2,45$) становлять $10,0 \pm 2,45 \cdot 2,0$.

Важливою характеристикою регресійної моделі є відносний ефект впливу фактора x на результат y — **коефіцієнт еластичності**:

$$\gamma = b \frac{\bar{x}}{\bar{y}}.$$

Він показує, на скільки процентів у середньому змінюється результат y зі зміною фактора x на 1%. У нашому прикладі $\gamma = 10,0 \cdot \frac{1,5}{28} = 0,8035$ тобто збільшення кількості внесених добрив на 1% спричинює приріст урожайності зернових у середньому на 0,8%.

Оцінити відносний ефект впливу фактора x на результат y можна безпосередньо на основі степеневі функції $Y = ax^b$, параметр b якої є коефіцієнтом еластичності. Степенева функція зводиться до лінійного виду логарифмуванням $\lg Y = \lg a + b \lg x$. До класу степеневих належать функції споживання, виробничі функції тощо.

6.3. Оцінка щільності та перевірка істотності кореляційного зв'язку

Поряд із визначенням характеру зв'язку та ефектів впливу факторів x на результат y важливе значення має оцінка щільності зв'язку, тобто оцінка

узгодженості варіації взаємозв'язаних ознак. Якщо вплив факторної ознаки x на результативну y значний, це виявиться в закономірній зміні значень y зі зміною значень x , тобто фактор x своїм впливом формує варіацію y . За відсутності зв'язку варіація y не залежить від варіації x .

Для оцінювання щільності зв'язку статистика використовує низку коефіцієнтів з такими спільними властивостями:

- за відсутності будь-якого зв'язку значення коефіцієнта наближається до нуля; при функціональному зв'язку — до одиниці;
- за наявності кореляційного зв'язку коефіцієнт виражається дробом, який за абсолютною величиною тим більший, чим щільніший зв'язок.

Серед мір щільності зв'язку найпоширенішим є **коефіцієнт кореляції** Пірсона. Позначається цей коефіцієнт символом r . Оскільки сфера його використання обмежується лінійною залежністю, то і в назві фігурує слово «лінійний». Обчислення лінійного коефіцієнта кореляції r ґрунтується на відхиленнях значень взаємозв'язаних ознак x і y від середніх.

За наявності прямого кореляційного зв'язку будь-якому значенню $x_i > \bar{x}$ відповідає значення $y_i > \bar{y}$, а $x_k < \bar{x}$ відповідає $y_k < \bar{y}$. Узгодженість варіації x і y схематично показано на рис. 6.2 у вигляді кореляційного поля зі зміщеною системою координат.

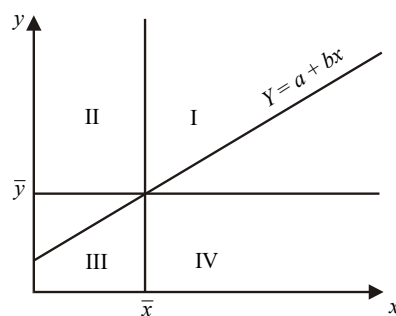


Рис. 6.2. Узгодженість варіації взаємозв'язаних ознак

Точка, координатами якої є середні $\bar{x}$ і $\bar{y}$, поділяє кореляційне поле на чотири квадранти, в яких по-різному поєднуються знаки відхилень від середніх:

Квадрант	$(x - \bar{x})$	$(y - \bar{y})$
I	+	+
II	—	+
III	—	—
IV	+	—

Для точок, розміщених у I та III квадрантах, добуток $(x - \bar{x})(y - \bar{y})$ додатний, а для точок з квадрантів II і IV — від'ємний. Чим щільніший зв'язок між ознаками x і y , тим більша алгебраїчна сума добутків відхилень $\sum_1^n (x - \bar{x})(y - \bar{y})$.

Гранична сума цих добутків дорівнює $\sqrt{\sum_1^n (x - \bar{x})^2 \sum_1^n (y - \bar{y})^2}$.

Коефіцієнт кореляції визначається відношенням зазначених сум:

$$r = \frac{\sum_1^n (x - \bar{x})(y - \bar{y})}{\sqrt{\sum_1^n (x - \bar{x})^2 \sum_1^n (y - \bar{y})^2}}.$$

Очевидно, що в разі функціонального зв'язку фактична сума відхилень дорівнює граничній, а коефіцієнт кореляції $r = \pm 1$; при кореляційному зв'язку абсолютне його значення буде тим більшим, чим щільніший зв'язок.

На практиці застосовують різні модифікації наведеної формули коефіцієнта кореляції. Для оцінювання щільності зв'язку між кількістю внесених добрив та врожайністю зернових скористаємося однією з модифікацій зазначеної формули:

$$r = \frac{\sum_1^n xy - n \bar{x} \bar{y}}{\sqrt{\sigma_x^2 \sigma_y^2}}.$$

За даними табл. 6.3 $\sum_1^n xy = 342,8$; $\bar{x} = 1,5$; $\sigma_x^2 = 0,085$; $\bar{y} = 28$; $\sigma_y^2 = 10,5$.

Згідно з цими значеннями коефіцієнт кореляції становить 0,900, що свідчить про вагомий вплив кількості внесених добрив на врожайність зернових:

$$r = \frac{342,8 - 8 \cdot 1,5 \cdot 28}{\sqrt{0,085 \cdot 10,5}} = 0,900.$$

Коефіцієнт кореляції, оцінюючи щільність зв'язку, указує також на його напрям: коли зв'язок прямий, r — величина додатна, а коли він зворотний — від'ємна. Знаки коефіцієнтів кореляції і регресії однакові, величини їх взаємозв'язані функціонально:

$$r = b \frac{\sigma_x}{\sigma_y}; \quad b = r \frac{\sigma_y}{\sigma_x}.$$

Завдяки цьому один коефіцієнт можна обчислити, знаючи інший. Наприклад:

$$r = 10,0 \sqrt{\frac{0,085}{10,50}} = 0,900.$$

Вимірювання щільності нелінійного зв'язку ґрунтується на співвідношенні варіацій теоретичних та емпіричних (фактичних) значень результативної ознаки y . Як зазначалося в підрозд. 5.6, відхилення індивідуального значення ознаки y від середньої $(y - \bar{y})$ можна розкласти на дві складові. У регресійному аналізі це відхилення від лінії регресії $(y - Y)$ та відхилення лінії регресії від середньої $(Y - \bar{y})$.

Відхилення $(Y - \bar{y})$ є наслідком дії фактора x , відхилення $(y - Y)$ — наслідком дії інших факторів. Взаємозв'язок факторної та залишкової варіацій описується правилом декомпозиції варіації:

$$\sigma_y^2 = \delta_Y^2 + \sigma_e^2,$$

де $\sigma_y^2 = \frac{1}{n} \sum_1^n (y - \bar{y})^2$ — загальна дисперсія ознаки y ; $\delta_Y^2 = \frac{1}{n} \sum_1^n (Y - \bar{y})^2 = \frac{1}{n} (a \sum x + b \sum xy) - \bar{y}^2$ — факторна дисперсія; $\sigma_e^2 = \frac{1}{n} \sum_1^n (y - Y)^2$ — залишкова дисперсія.

Очевидно, значення факторної дисперсії δ_Y^2 буде тим більшим, чим сильніший вплив фактора x на y . Відношення факторної дисперсії до загальної розглядається як міра щільності кореляційного зв'язку і називається **коефіцієнтом детермінації**:

$$R^2 = \frac{\delta_Y^2}{\sigma_y^2}.$$

Якщо за даними табл. 6.3 $\sigma_y^2 = 10,5$, $\sigma_e^2 = 2,0$, то $\delta_Y^2 = 10,5 - 2,0 = 8,5$.

Аналогічний результат дають такі обчислення:

$$\sigma_Y^2 = \frac{1}{8} (13 \cdot 224 + 10 \cdot 342,8) - 28^2 = 8,5.$$

Коефіцієнт детермінації становить $R^2 = \frac{8,5}{10,5} = 0,81$,

тобто 81% варіації врожайності зернових залежить від варіації кількості внесених добрив, а 19% припадає на інші фактори.

Корінь квадратний з коефіцієнта детермінації називають **індексом кореляції** R . Коли зв'язок лінійний, $R = |r|$, що підтверджують обчислення: $R = \sqrt{R^2} = \sqrt{0,81} = 0,90$. Тому за відомим лінійним коефіцієнтом кореляції r можна визначати внесок ознаки x у варіацію ознаки y . Так, при $r = 0,6$ можна сказати, що 36% варіації y залежить від варіації x .

На таких самих засадах ґрунтується оцінювання щільності зв'язку за даними аналітичного групування. Мірою щільності зв'язку є **кореляційне відношення**

$$\eta^2 = \frac{\delta^2}{\sigma^2},$$

де δ^2 — міжгрупова дисперсія, яка вимірює варіацію ознаки y під впливом фактора x , а σ^2 — загальна дисперсія.

Застосуємо кореляційне відношення для оцінювання щільності зв'язку між глибиною розробки вугільних пластів і фондомісткістю видобутку вугілля (див. табл. 6.2). Розрахунки загальної та факторної дисперсій подано в табл. 6.4 та 6.5. Згідно з розрахунками загальна дисперсія становить 5,19, факторна — 3,86:

$$\sigma^2 = \frac{\sum_1^{m_y} (y_i - \bar{y})^2 f_i}{\sum_1^{m_y} f_i} = \frac{519}{100} = 5,19;$$

$$\delta^2 = \frac{\sum_1^{m_x} (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^{m_x} f_j} = \frac{386,6}{100} = 3,866.$$

Кореляційне відношення

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{3,866}{5,19} = 0,745,$$

тобто 74,5% варіації фондомісткості вугілля на шахтах регіону пояснюється варіацією глибини розробки пластів.

Таблиця 6.4 - ДО РОЗРАХУНКУ ЗАГАЛЬНОЇ ДИСПЕРСІЇ
ФОНДОМІСТКОСТІ ВУГІЛЛЯ ($\bar{y} = 23,5$)

Фондомісткість, грн. / т	18—20	0—22	22—24	24—26	26—28	Разом
Кількість шахт f_i	9	15	34	28	14	100
y_i	19	21	23	25	27	×
$y_i - \bar{y}$	− 4,5	− 2,5	− 0,5	1,5	3,5	×
$(y_i - \bar{y})^2 f_i$	182,25	93,75	8,5	63,0	171,5	519

Таблиця 6.5- ДО РОЗРАХУНКУ ФАКТОРНОЇ ДИСПЕРСІЇ
ФОНДОМІСТКОСТІ ВУГІЛЛЯ ($\bar{y} = 23,5$)

Глибина розробки пластів, м	f_j	$\bar{y}_j$	$\bar{y}_j - \bar{y}$	$(\bar{y}_j - \bar{y})^2 f_j$
До 300	17	20,0	− 3,5	208,25
300 — 500	40	22,9	− 0,6	14,40
500 — 700	25	24,8	1,3	42,25
700 і більше	18	26,1	2,6	121,68
У цілому	100	23,5	×	386,58

Обчислення та інтерпретація коефіцієнта детермінації R^2 і кореляційного відношення η^2 показують: ці характеристики щільності зв'язку за змістом ідентичні, вони характеризують внесок фактора x у загальну варіацію результату y .

Перевірка істотності кореляційного зв'язку ґрунтується на порівнянні фактичних значень R^2 і η^2 з критичними, які могли б виникнути за відсутності зв'язку. Якщо фактичне значення R^2 чи η^2 перевищує критичне, то зв'язок між ознаками не випадковий. Гіпотеза, що перевіряється, формулюється як нульова:

$$H_0: R^2 = 0 \text{ або } H_0: \eta^2 = 0.$$

Критичні значення характеристик щільності зв'язку для рівня істотності $\alpha = 0,05$ і відповідного числа ступенів свободи для факторної дисперсії k_1 і залишкової k_2 наведено в табл. 6.6. Ступені свободи залежать від обсягу сукупності n та числа груп або параметрів функції m , тобто $k_1 = m - 1$, $k_2 = n - m$.

Таблиця 6.6- КРИТИЧНІ ЗНАЧЕННЯ КОЕФІЦІЄНТА ДЕТЕРМІНАЦІЇ R^2
І КОРЕЛЯЦІЙНОГО ВІДНОШЕННЯ η^2 ДЛЯ РІВНЯ ІСТОТНОСТІ $\alpha = 0,05$

k_2/k_1	1	2	3	4	5
5	0,569	699	764	806	835
6	500	632	704	751	785
7	444	575	651	702	739
8	399	527	604	657	697
9	362	488	563	618	659
10	332	451	527	582	624
12	283	394	466	521	564
14	247	348	417	471	514
16	219	312	378	429	477
18	197	283	345	394	435
20	179	259	318	364	404
24	151	221	273	316	353
28	130	193	240	279	314
32	115	171	214	250	282
36	102	153	192	226	256
40	093	139	176	207	234
50	075	113	143	170	194
60	063	095	121	144	165
80	047	072	093	110	127
100	038	058	075	090	103
120	032	049	063	075	087
200	019	030	038	046	053

Так, критичне значення коефіцієнта детермінації для $k_1 = 2 - 1 = 1$ і $k_2 = 8 - 2 = 6$ становить $R_{0,95}^2(1,6) = 0,500$. Обчислений за даними табл. 6.3 коефіцієнт детермінації $R^2 = 0,81$ перевищує критичне значення, що з імовірністю 0,95 підтверджує істотність зв'язку між кількістю внесених добрив і врожайністю зернових.

Аналогічно визначимо критичне значення кореляційного відношення для $k_1 = 4 - 1 = 3$ та $k_2 = 100 - 4 = 96$. Оскільки значення $k_2 = 96$ у табл. 6.6 відсутнє, можна використати найближче до нього число $k_2 = 100$. Критичне значення $\eta_{0,95}^2(3,100) = 0,075$.

Розраховане за даними табл. 6.2 кореляційне відношення $\eta^2 = 0,745$ значно перевищує критичне, а отже, гіпотеза про випадковий характер відхилень групових середніх відхиляється. Зв'язок між глибиною розробки вугільних пластів і фондомісткістю видобутку вугілля з імовірністю 0,95 визнається істотним.

Розглянута процедура перевірки істотності зв'язку є складовою дисперсійного аналізу, розробленого Р. Фішером. Характеристика критерію Фішера — дисперсійне відношення F — функціонально пов'язана з

кореляційним відношенням $F = \frac{\eta^2}{1-\eta^2} \frac{k_2}{k_1}$, а тому результати перевірки будуть ідентичні.

6.4. Рангова кореляція

Взаємозв'язок між ознаками, які можна зранжувати, передусім на основі бальних оцінок, вимірюється методами рангової кореляції. **Рангами** називають числа натурального ряду, які згідно зі значеннями ознаки надаються елементам сукупності і певним чином упорядковують її. Ранжування проводиться за кожною ознакою окремо: перший ранг надається найменшому значенню ознаки, останній — найбільшому або навпаки. Кількість рангів дорівнює обсягу сукупності. Очевидно, зі збільшенням обсягу сукупності ступінь «розпізнаваності» елементів зменшується. З огляду на те, що рангова кореляція не потребує додержання будь-яких математичних передумов щодо розподілу ознак, зокрема вимоги нормальності розподілу, рангові оцінки щільності зв'язку доцільно використовувати для сукупностей невеликого обсягу.

Ранги, надані елементам сукупності за ознаками x і y , позначають відповідно Rx_j та Ry_j . Залежно від ступеня зв'язку між ознаками певним чином співвідносяться й ранги. При прямому функціональному зв'язку $Rx_j = Ry_j$, тобто відхилення між рангами $d_j = Rx_j - Ry_j = 0$, отже, і сума квадратів відхилень $\sum_1^m d_j^2 = 0$. При зворотному функціональному зв'язку $\sum d_j^2 = \frac{1}{3}n(n^2 - 1)$, де n — число рангів. Якщо зв'язок між ознаками відсутній, $\sum_1^n d_j^2$ являє собою середню арифметичну цих крайніх значень:

$$\sum_1^n d_j^2 = \frac{1}{2} \left[0 + \frac{1}{3}n(n^2 - 1) \right] = \frac{1}{6}n(n^2 - 1), \quad \text{а отже,} \quad \frac{6 \sum_1^n d_j^2}{n(n^2 - 1)} = 1.$$

Спираючись на зазначену математичну тотожність, К. Спірмен запропонував формулу для коефіцієнта рангової кореляції:

$$\rho = 1 - \frac{6 \sum_1^n d_j^2}{n(n^2 - 1)}.$$

Цей коефіцієнт має такі самі властивості, як і лінійний коефіцієнт кореляції: змінюється в межах від -1 до $+1$, водночас оцінює щільність зв'язку та вказує на його напрям.

Визначимо коефіцієнт рангової кореляції за даними експертних оцінок ефективності економіки та ступеня політичного ризику для семи країн з перехідною економікою (табл. 6.7). Оскільки експертні оцінки представлені балами, необхідно провести ранжування країн. За оцінками ефективності економіки країні з найбільшим балом надається ранг 1, з найменшим — ранг

$n = 7$. За оцінками ступеня політичного ризику, навпаки, ранг 1 надається країні з найменшим ризиком, а ранг 7 — країні з найбільшим ризиком.

Таблиця 6.7- ДО РОЗРАХУНКУ КОЕФІЦІЄНТА РАНГОВОЇ КОРЕЛЯЦІЇ

№ з/п	Експертні оцінки, балів		Ранги		$d_j = R_{x_j} - R_{y_j}$	d_j^2
	Ефективність економіки (max = 10)	Ступінь політичного ризику (max = 100)	R_{x_j}	R_{y_j}		
1	6,6	64,5	1	7	-6	36
2	5,8	57,8	2	6	-4	16
3	2,9	23,6	6	1	5	25
4	3,4	36,2	5	4	1	1
5	4,5	45,3	3	5	-2	4
6	2,7	28,4	7	2	5	25
7	4,2	32,7	4	3	1	1
Разом	×	×	×	×	×	108

Сума квадратів відхилень рангів $\sum_1^n d_j^2 = 108$, а коефіцієнт рангової кореляції

$$\rho = 1 - \frac{6 \cdot 108}{7(49 - 1)} = 1 - \frac{648}{336} = -0,928.$$

Значення коефіцієнта рангової кореляції свідчить про наявність зворотного і досить високого рівня зв'язку між ефективністю економіки і ступенем політичного ризику. Критичне значення коефіцієнта рангової кореляції (табл. 6.8) для рівня істотності $\alpha = 0,05$ і $n = 7$ $\rho_{0,95}(7) = 0,71$. Отже, з імовірністю 0,95 істотність зв'язку доведено.

Таблиця 6.8- КРИТИЧНІ ЗНАЧЕННЯ КОЕФІЦІЄНТА РАНГОВОЇ КОРЕЛЯЦІЇ СПІРМЕНА ПРИ $\alpha = 0,05$

Обсяг вибірки n	5	6	7	8	9	10	11	12
$\rho_{0,95}(n)$	0,90	0,83	0,71	0,64	0,60	0,56	0,53	0,50

Якщо два і більше елементів сукупності мають однакові значення ознаки, їм надається середній ранг. Нехай, наприклад, друге за розміром значення ознаки мають три елементи сукупності (№ 2, 3, 4), тоді всім їм надається ранг $\frac{1}{3}(2 + 3 + 4) = 3$, а щільність зв'язку можна оцінити за формулою лінійного коефіцієнта кореляції.

6.5. Оцінка узгодженості варіації атрибутивних ознак

Взаємозв'язки між атрибутивними ознаками аналізуються на підставі таблиць взаємної спряженості (співзалежності). Як приклад розглянемо табл. 6.9, в якій

наведено результати соціологічного опитування населення щодо намірів прилучитися до ринку цінних паперів. Тих, хто не боїться ризикувати, класифікували як ризикованих інвесторів, тих, хто не уявляє ризику без гарантій, — обережними, а хто ризику уникає взагалі, — неризикованими.

Частоти комбінаційного розподілу респондентів за віком і схильністю до ризику концентруються навколо діагоналі з верхнього лівого кута в нижній правий. Серед молодих більшість готова ризикувати на ринку цінних паперів, у середній віковій групі готовий ризикувати один з п'яти, а половина не уявляє ризику без гарантій, у третій віковій групі на одного обережного припадають два неризиковані.

Таблиця 6.9 - РОЗПОДІЛ РЕСПОНДЕНТІВ ЗА ВІКОМ І СХИЛЬНІСТЮ ДО РИЗИКУ

Вік x , років	Тип інвестора y			Разом f_{i0}
	Ризикований	Обережний	Неризикований	
16—30	24	12	4	40
31—50	20	50	30	100
51 і більше	6	18	36	60
Разом f_{0j}	50	80	70	200

Характер розподілу частот, концентрація їх уздовж головної діагоналі свідчать про наявність стохастичного зв'язку між віком і схильністю до ризику.

Оцінка щільності стохастичного зв'язку ґрунтується на відхиленнях частот (часток) умовного та безумовного розподілів, тобто на відхиленнях фактичних частот f_{ij} від теоретичних F_{ij} , пропорційних до підсумкових:

$$F_{ij} = \frac{f_{i0} f_{0j}}{n},$$

де f_{i0} — підсумкові частоти за ознакою x ; f_{0j} — підсумкові частоти за ознакою y ; n — обсяг сукупності $\left(n = \sum_i^{m_x} f_{i0} = \sum_j^{m_y} f_{0j} \right)$.

Якби схильність до ризику не залежала від віку, то кількість ризикованих серед молоді становила б

$$F_{11} = \frac{40 \cdot 50}{200} = 10,$$

обережних у другій віковій групі

$$F_{22} = \frac{100 \cdot 80}{200} = 40,$$

неризикованих у третій віковій групі

$$F_{33} = \frac{60 \cdot 70}{200} = 21.$$

Абсолютну величину відхилень фактичних частот f_{ij} від пропорційних F_{ij} характеризує квадратична спряженість χ^2 Пірсона:

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - F_{ij})^2}{F_{ij}} = n \left[\sum_i \sum_j \frac{f_{ij}^2}{f_{i0} f_{0j}} - 1 \right].$$

За відсутності стохастичного зв'язку $\chi^2 = 0$. На основі розподілу ймовірностей χ^2 перевіряється істотність зв'язку. Критичні значення χ^2 для $\alpha = 0,05$ і числа ступенів свободи $k = (m_x - 1)(m_y - 1)$ наведено в табл. 6.10. Так, для $k = (3 - 1)(3 - 1) = 4$ критичне значення $\chi_{0,95}^2(4) = 9,49$. Фактичне значення

$$\chi^2 = 200 \left[\frac{24^2}{40 \cdot 50} + \frac{12^2}{40 \cdot 80} + \frac{4^2}{40 \cdot 70} + \frac{20^2}{100 \cdot 50} + \frac{50^2}{100 \cdot 80} + \frac{30^2}{100 \cdot 70} + \frac{6^2}{60 \cdot 50} + \frac{18^2}{60 \cdot 80} + \frac{36^2}{60 \cdot 70} - 1 \right] = 49,5,$$

що значно перевищує критичне, а отже, з імовірністю 0,95 істотність зв'язку між віком і схильністю до ризику доведено.

Відносною мірою щільності стохастичного зв'язку слугує **коефіцієнт взаємної спряженості** (співзалежності). За умови, що $m_x = m_y$ використовують формулу Чупрова:

$$C = \sqrt{\frac{\chi^2}{n(m_x - 1)(m_y - 1)}},$$

де m_x — число груп за ознакою x ; m_y — число груп за ознакою y . Оскільки за відсутності зв'язку між ознаками $\chi^2 = 0$, то і $C = 0$. При функціональному зв'язку $C \rightarrow 1$. У разі, коли $m_x \neq m_y$, віддають перевагу коефіцієнту спряженості Крамера:

$$C = \sqrt{\frac{\chi^2}{n(m_{\min} - 1)}},$$

де $m_{\min}$ — мінімальне число груп (m_x або m_y).

У нашому прикладі $m_x = m_y = 3$, а тому наведені формули коефіцієнта взаємної спряженості тотожні:

$$C = \sqrt{\frac{49,5}{200(3-1)}} = \sqrt{0,124} = 0,352,$$

що свідчить про наявність зв'язку.

Таблиця 6.10 - КРИТИЧНІ ЗНАЧЕННЯ $\chi_{0,95}^2(k)$

k	1	2	3	4	5	6	7	8
χ^2	3,84	5,99	7,81	9,49	11,07	12,59	14,07	15,51

Якщо обидві взаємозв'язані ознаки альтернативні, тобто кількість груп $m_x = m_y = 2$, то за відсутності зв'язку добутки діагональних частот однакові: $f_{11}f_{22} = f_{12}f_{21}$. Саме на відхиленнях добутків частот ґрунтуються характеристики зв'язку:

$$\chi^2 = n \frac{(f_{11}f_{22} - f_{12}f_{21})^2}{f_{01}f_{02}f_{10}f_{20}},$$

$$C = \sqrt{\frac{\chi^2}{n}} = \frac{f_{11}f_{22} - f_{12}f_{21}}{\sqrt{f_{01}f_{02}f_{10}f_{20}}}.$$

У літературі зі статистики коефіцієнт c для 4-клітинкової таблиці називається **коефіцієнтом контингенції** або **асоціації**. Очевидно, що за змістом він ідентичний коефіцієнту взаємної спряженості, а з χ^2 пов'язаний функціонально: $\chi^2 = nC^2$.

За допомогою коефіцієнта контингенції оцінимо щільність зв'язку між шкідливою звичкою палити і хворобами легенів (табл. 6.11).

Таблиця 6.11 - РОЗПОДІЛ ПАЦІЄНТІВ КЛІНІКИ ЗА РЕЗУЛЬТАТАМИ ЛЕГЕНЕВИХ ПРОБ

Наявність звички палити	Результати легеневих проб		Разом
	Аномальні	Нормальні	
Палить	20	5	25
Не палить	10	15	25
Разом	30	20	50

$$C = \frac{20 \cdot 15 - 10 \cdot 5}{\sqrt{30 \cdot 20 \cdot 25 \cdot 25}} = \frac{250}{612,3} = 0,408.$$

Значення $\chi^2 = nC^2 = 50 \cdot 0,408^2 = 8,32$ перевищує критичне $\chi_{0,95}^2(1) = 3,89$. Істотність зв'язку доведено з імовірністю 0,95.

Корисною мірою при аналізі 4-клітинкових таблиць взаємної спряженості є відношення перехресних добутків або **відношення шансів**

$$W = \frac{f_{11} f_{22}}{f_{12} f_{21}}.$$

Відношення шансів характеризує міру відносного ризику. У нашому прикладі

$$W = \frac{15 \cdot 20}{5 \cdot 10} = 6.$$

Отже, імовірність легеневих хвороб у тих, хто палить, у 6 разів вища порівняно з тими, хто не палить.

Зауважимо, що методи аналізу таблиць взаємної спряженості можна використати і для кількісних ознак. Будь-які технічні перешкоди відсутні. Проте слід пам'ятати, що коефіцієнт спряженості оцінює лише узгодженість фактичного розподілу з пропорційним. При переставлянні рядків чи стовпців значення коефіцієнта C не зміниться. Міри щільності кореляційного зв'язку — коефіцієнт детермінації R^2 і кореляційне відношення η^2 — оцінюють не лише узгодженість частот, а й порядок, послідовність, в якій поєднуються різні значення ознак. Отже, ці характеристики зв'язку більш потужні. А загалом вибір методу вимірювання зв'язку і характеристик його щільності має ґрунтуватись на попередньому теоретичному аналізі суті явищ, характеру взаємозв'язків, наявній інформації.

Тема 7. Багатовимірний статистичний аналіз в агрономії

7.1. Поняття багатовимірних статистичних методів та їх роль у агрономії.

Навколишній світ складається з різноманітних об'єктів, підприємств, регіонів, зайнятих, безробітних, кожен з яких характеризується набором якісних і кількісних ознак. Наприклад, промислові підприємства розрізняються за обсягом виробленої продукції, за розміром основних виробничих засобів, за кількістю працівників, за рівнем рентабельності, за собівартістю продукції та іншими якісними та кількісними показниками. Всього можна назвати приблизно 100 техніко-економічних показників бізнесу.

Інакше кажучи, об'єкти навколо нас відрізняються багатогранністю, тобто дозволяють описувати їх з різних боків. З цієї точки зору одновимірні об'єкти (число описових ознак дорівнює 1), які характеризує традиційна статистика є простими, частковими. Розрахунок таких показників, як середня, дисперсія, коефіцієнтів варіації та інших виконується для окремих ознак. На відміну від класичної статистики, багатовимірні статистичні методи дозволяють вивчати велику кількість показників одночасно, тобто оцінювати саме багатовимірні об'єкти. Хоча потрібно зауважити, що окремі елементи багатомірності зустрічаються у статистиці. Наприклад, комбінаційне групування. Воно використовується під час вивчення складних явищ. Як відомо, його схема передбачає послідовний розподіл вихідної сукупності об'єктів спочатку за першою ознакою, потім у середині кожної групи за другою ознакою і таке інше.

У випадку трьох і більше групувальних ознак утворюються складна громіздка таблиця, серед багаточисленних підгруп якої є незначні чи пусті. Головним недоліком такого підходу є необхідність розташовувати значні за обсягом спостереження, ймовірність отримання пустих груп, складна процедура групування.

Така проблема традиційними одномірними методами вирішується утруднено, але добре розв'язується за допомогою алгоритмів багатовимірного аналізу, насамперед, кластерного аналізу. Об'єднання об'єктів у однорідні групи відбувається за принципами їх близькості у багатовимірному просторі ознак. Розрахунок відстаней між точками простору дозволяє знайти скупчення й вирішувати завдання багатовимірного групування об'єктів незалежно від їхньої кількості.

Поряд з багатовимірним групуванням постає питання віднесення нових об'єктів до відомих, вже існуючих груп (кластерів). Наприклад, віднесення будьякої країни до групи економічно безпечних або групи економічно

небезпечних країн з метою інвестування. Це завдання також вирішується в рамках багатовимірних статистичних методів і виконується за допомогою дискримінантного аналізу.

Доволі часто на практиці необхідно оцінити деякі невимірювані внутрішні латентні якості об'єктів, які на поверхні виявляються у вигляді зовнішніх факторів-симптомів. Це так звані латентні показники. Наприклад, продуктивність праці «білих комірців» виявляється через винахідництво, раціоналізаторство, прибутковість підприємств, ритмічність праці та її оплату, впровадження нових технологій та інше.

Означені проблеми можна розв'язати за допомогою методів багатовимірного статистичного аналізу.

Багатовимірні методи мають велику кількість напрямків таких, як кластерний аналіз, дискримінанти аналіз, метод головних компонент, факторний аналіз та інші.

Основними можливостями використання багатовимірних статистичних методів є три ключові позиції: по-перше, одночасне використання інформаційних даних з різнорідних джерел; по-друге, просіювання даних з метою виявлення прихованих закономірностей, латентних характеристик, що сприяє підвищенню конкурентоспроможності; по-третє, новий, більш розвинутий рівень порівняльного аналізу багатовимірних об'єктів.

Отже, багатовимірний статистичний аналіз – це сукупність формалізованих статистичних методів, які базуються на поданні результативної інформації в багатовимірному геометричному просторі та дозволяють визначати приховані (латентні), але об'єктивно існуючі закономірності в організаційній структурі та тенденціях розвитку соціально-економічних явищ і процесів. За змістом багатовимірний статистичний аналіз можна умовно розбити на три складові:

1. Багатовимірний статистичний аналіз спільних розподілів та їх основних характеристик.
2. Багатовимірний статистичний аналіз характеру та структури взаємозв'язків між компонентами досліджуваної багатовимірної ознаки.
3. Багатовимірний статистичний аналіз геометричної структури досліджуваної сукупності багатовимірних спостережень.

7.2. Аналітичні завдання, які вирішуються багатовимірними методами

Поряд з проблемами багатовимірного групування та класифікацій у агрономії виникає задача виявити та виміряти деякі внутрішні властивості об'єктів, які проявляються на поверхні у вигляді зовнішніх факторів-симптомів.

Ці внутрішні (чи приховані) ознаки, які проявляються у вигляді факторів – симптомів називаються латентними.

Наприклад, рівень життя проявляється в середніх показниках доходу на душу населення, динаміки індексів цін споживачів товарів, тривалості життя, народжуваності, смертності, тощо. Динаміка захворювань відбувається на підставі його зовнішніх проявів – температури хворого, різноманітних аналізів, кашля. У такому випадку спочатку потрібно виявити захворювання (класифікувати), тобто провести дискримінацію, а потім оцінити латентний показник. Аналогічно про технічний рівень виробництва роблять висновки на основі таких показників, як фондоозброєність праці, енергоозброєність, механоозброєність та інші.

Отже, за допомогою багатовимірних статистичних методів можна виявити та подати у компактній формі означені приховані ознаки, тобто відтворити досліджувані латентні показники.

Під час багатовимірного аналізу виконується групування вихідної інформації за такими напрямками:

1. Виділення груп об'єктів зі схожими поєднаннями значень ознак.
2. Виділення груп ознак, які найбільше відображують образ досліджуваного латентного показника.
3. Скорочення вихідного простору ознак без значної втрати інформації.
4. Вимірювання тісноти кореляційного зв'язку між групами ознак.
5. Багатовимірне шкалювання.

Під час проведення дослідження в реальних економічних умовах означені напрямки взаємно переплітаються.

Розв'язання перших двох напрямків передбачає використовувати методи розпізнання образів. Для цього доцільно використовувати кластерний аналіз (розпізнання без вчителя) і дискримінантний аналіз (розпізнання образу з вчителем).

Завдання другого та третього напрямків, тобто скорочення вихідного простору, вирішуються за допомогою вимірювання відстані між об'єктами та еталонами чи антіеталонами. Даний підхід відомий як таксономічний аналіз. Він дозволяє у 4-5 разів скоротити кількість вихідних ознак.

Однак головним методом оцінки латентних показників, вимірювання кореляційних зв'язків між ознаками є факторний аналіз і метод канонічних кореляцій. Спочатку будуються нові змінні у вигляді лінійних комбінацій вихідних ознак, а потім між ними вимірюється тіснота кореляційних зв'язків. Такий підхід дозволяє виявити залежності між окремими групами показників. Це по суті є найвищий розвиток багатфакторної статистики.

Багатовимірне шкалювання є змішаним підходом, який поєднує проблематику розпізнання образів, скорочення простору ознак, обробки даних, які отримані на підставі експертних оцінок. У межах цього підходу розв'язуються задачі метричного та неметричного шкалювання, вибору індивідуальних думок експертів.

Оскільки статистика вивчає кількісний бік масових явищ та процесів, а багатовимірний аналіз є подальшим розвитком статистики на більш високому рівні, то будемо припускати, що всі ознаки об'єктів є кількісними, а не атрибутивними. Якщо для опису явища необхідне використання атрибутивної ознаки, то потрібно перевести їх за допомогою змінних 0 і 1 до кількісної форми.

До впровадження комп'ютерів з потужною базою для обробки великих масивів, означені питання розв'язувались теоретично, оскільки без відповідної техніки розв'язувати складні задачі неможливо. Тільки сучасні комп'ютерні технології дозволяють з успіхом здійснювати обробку багатовимірних даних з метою їх класифікації, ранжування, оцінки латентних показників.

Отже, методи багатовимірного статистичного аналізу дозволяють розв'язувати різноманітні задачі дослідження багатовимірних процесів: групування, стиснення даних, моделювання складних систем, оцінювання параметрів багатовимірних сукупностей та ін. Це дозволяє використовувати отримані результати для прийняття ефективних управлінських рішень.

Зауважимо, що методи багатовимірного статистичного аналізу тісно пов'язані зі статистикою, матричною алгеброю, теорією ймовірностей, аналітичною геометрією.

7.3. Етапи багатовимірного статистичного дослідження

Обробка багатовимірних статистичних даних має певні особливості:

- 1) застосування методів багатовимірного статистичного аналізу вимагає творчого підходу до розв'язання аналітичних задач, оскільки оброблюються багатовимірні сукупності даних;
- 2) для багатовимірних статистичних методів характерна глибока формалізація та складна логіко-математична конструкція;
- 3) практичне застосування потребує використання обчислювальної техніки.

Розглянемо етапи дослідження за допомогою багатовимірного статистичного аналізу:

- 1) постановка задачі включає опис предметної області об'єктів, визначення обсягів виділених ресурсів (час, трудовитрати і т. д.);

2) визначення набору методів багатовимірної статистичної аналізи та порядку їх використання;

3) збирання вихідних даних, визначення способів збирання інформації, форм її подання;

4) аналіз даних (перевірка однорідності вибірки, відповідність законам розподілу, виявлення грубих помилок і т. д.);

5) уточнення математичної постановки задачі та визначення можливості застосування раніше відібраних методів (у разі необхідності набір методів змінюється);

6) реалізація, тобто проведення обчислень, за допомогою програмного забезпечення математичного інструментарію;

7) оцінювання адекватності моделі, визначення несуперечності математичних результатів і економічних висновків.

На практиці всі перелічені етапи необов'язково присутні і не розмежовані. Деякі з них можуть об'єднуватися або виключатися. Знання всіх етапів дозволяє раціонально планувати реалізацію багатовимірних методів та враховувати майбутні обсяги дослідження.

Етапи кластерного аналізу (1-2)

Розглянемо етапи кластерного аналізу з позицій математичного змісту і методики.

Етап. Подання вихідних даних у вигляді матриці

На початковому етапі вихідні дані подаються у вигляді матриці. Очевидно, що матриця може мати як прямокутну, так і квадратну форму.

Прямокутна матриця вихідних даних в багатовимірному статистичному аналізі частіше згадується як матриця "об'єкт - властивість" і є найбільш поширеною формою представлення вихідних даних.

У табл. 7.2 наочно зображена в загальному вигляді прямокутна матриця вихідних даних з n елементами - об'єктами дослідження. На цій посаді можуть виступати підприємства машинобудування, комерційні банки, студенти університету, які описуються за допомогою k ознак.

У матриці значення X_{ij} відповідає i -го об'єкту, який характеризується j -м показником, де $i = 1, 2, \dots, n, j = 1, 2, \dots, k$.

У разі якщо ми виробляємо попарні порівняння об'єктів по деякому властивості, матриця вихідних даних являє собою квадратну матрицю. Наприклад, відстань між містами Росії, результати футбольного матчу між командами і т.п. У табл. 7.3 представлена квадратна матриця вихідних даних, у якій число рядків і стовпців однаково.

Таблиця 7.2

Прямокутна матриця вихідних даних

об'єкт (Елемент)	Мінлива			
	x_1	x_2	...	x_n
1	x_{11}	x_{12}	...	x_{1n}
2	x_{21}	x_{22}	...	x_{2n}
...	...	...		...
n	x_{n1}	x_{n2}	...	x_{nn}

Таблиця 7.3

Квадратна матриця вихідних даних

		Об'єкти (елементи)			
		1	2		n
об'єкти (Елементи)	1	x_{11}	x_{12}	...	x_{1n}
	2	x_{21}	x_{22}	...	x_{2n}
		...	...		...
	n	x_{n1}	x_{n2}	...	x_{nn}

У багатовимірному статистичному аналізі така матриця називається "матрицею парних порівнянь". Уявімо зміст таблиці у вигляді матриці

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{pmatrix} \quad (7.1)$$

де $i = 1, 2, \dots, n$; $m = 1, 2, \dots, n$; x_{ij} - результат порівняння i -го об'єкта з m -м об'єктом.

На результати обчислень при кластерному аналізі впливає те, які використовуються одиниці виміру і яку величину (масштаб) мають змінні. Усунути залежність змінних від одиниць виміру і масштабу дозволяє процедура стандартизації вихідних даних. Таким чином, змінні набувають однакову вагу при кластеризації.

В ході стандартизації здійснюється перехід від матриці вихідних даних x_{ij} до матриці стандартизованих значень z_{ij} , у. Найчастіше використовується формула переходу:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

де $\bar{x}_j$ - середнє значення у-ї ознаки; σ_j - середньоквадратичне відхилення у-ї ознаки.

Середнє значення у-ї ознаки розраховується за такою формулою:

де n - число елементів (об'єктів).

Середнє квадратичне відхилення у-ї ознаки розраховується як:

При процедурі стандартизації все числові пропорції зберігаються, але всі змінні не мають одиниць виміру. Це дозволяє проводити з ними обчислювальні операції, як для змінних, що мають однакові одиниці виміру.

У тому випадку, якщо вихідні змінні мають однакові одиниці виміру, проводити стандартизацію немає необхідності.

Етап. Визначити подібність об'єктів

Визначити подібність об'єктів пов'язано з поняттям однорідності об'єктів. Визначення критерію однорідності об'єктів є найбільш важким етапом кластерного аналізу і в зв'язку з цим найменш формалізованим в завданню автоматичної класифікації. Однорідність об'єктів визначається мірою ступеня близькості (подібності) об'єктів або відмінності між об'єктами.

В якості запобіжного відмінності частіше використовується відстань між об'єктами, що позначається як du (від слова "distance"), що забезпечує більш простий алгоритм дій. До того ж цей алго

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (7.3)$$

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n}} \quad (7.4)$$

ритм добре реалізований в комп'ютерних програмах, таких як Statistica, SPSS та ін.

Залежно від вихідних даних може бути використаний той чи інший вид відстаней. Важливо відзначити, що критерієм вибору виду відстані є правильна (точніше сказати адекватна) інтерпретація профілю (особливостей) кластерів.

Основні заходи відстаней

Основні заходи відстаней для змінних, вимірюваних на метричних шкалах.

1. Евклідова відстань.

Багатовимірне евклідова відстань ($1u$ між двома об'єктами i та j) визначається за формулою

$$d_{ij} = \sqrt{\sum_{k=1}^k (z_{ik} - z_{jk})^2}, \quad (7.5)$$

де $i = 1, 2, \dots, pp$, k - число змінних.

Всі змінні стандартизовані, не мають розмірності. Це забезпечує можливість підсумовування відстаней за різними змінним.

Іншими словами, d_{ij} - це геометричне відстань між двома об'єктами в багатовимірному просторі. Багатовимірним простором називають простір, що має число вимірів більше трьох. У нашому випадку у формулі (7.5) маємо n -мірним простір.

Обчислення багатовимірного евклідового відстані за формулою (7.5) частіше називають простим евклідовим відстанню.

Виважена евклідова відстань застосовується в тому випадку, якщо змінні ранжовані між собою за ступенем важливості, тобто їм присвоєні ваги. Вага показника показує, наскільки важливо врахувати при класифікації дана ознака, тобто при розрахунку міри подібності враховується важливість показника, оцінена шляхом додаткових досліджень, наприклад, експертним шляхом.

Виважена евклідова відстань розраховується наступним чином:

$$d_{ij} = \sqrt{\sum_{j=1}^k (z_{ij} - z_{ij'})^2 w_j}, \quad (7.6)$$

$w_j > 0, \sum_{j=1}^k w_j = 1$

де w_j - вага j -го показника,

Сума всіх ваг повинна дорівнювати 1.

Якщо важко визначитися з важливістю показника і все ваги рівнозначні, то рекомендується використовувати просте евклідова відстань.

2. Квадрат евклидова відстані:

$$d_{ij}^2 = \sum_{j=1}^k (z_{ij} - z_{ij'})^2, \quad (7.7)$$

де i - номер об'єкта, $i = 1, 2, \dots, n$; n - обсяг вибірки.

Квадрат евклидова відстані знаходиться як відстань між двома елементами z_i / через суму квадратів різниці значень всіх змінних. Квадрат евклидова відстані використовується для додання великих ваг найбільш віддаленим один від одного об'єктів. Особливо це важливо використовувати для стандартизованих змінних.

3. Відстань Чебишева:

$$d_{ij} = \max_j |z_{ij} - z_{ij'}|, \quad (7.8)$$

де $i = 1, 2, \dots, n$.

Відстань Чебишева дорівнює максимальному відстані між відповідними координатами об'єктів. Відстань Чебишева використовують тоді, коли потрібно визначити відмінність двох об'єктів z_i / з якої-небудь однієї координаті. Відстань Чебишева є грубою мірою відмінності, так як значна частина наявної інформації ігнорується.

4. **Відстань Хеммінга** (відстань міських кварталів або Манхеттенський відстань):

$$d_{ij} = \sum_{j=1}^k |z_{ij} - z_{ij'}|, \quad (7.9)$$

де $i = 1, 2, \dots, n$.

Відстань Хеммінга обчислюється як сума абсолютних значень координатних відстаней. У більшості випадків ця міра відстаней призводить до таких же результатів, як і просте евклідова відстань.

5. Відстань Маньківського:

$$d_{ij} = \left(\sum_{j=1}^k |z_{ij} - z_{ij}|^p \right)^{\frac{1}{p}}, \quad (7.10)$$

де $i = 1, 2, \dots, n$.

При $p = 2$ формула відстані Маньківського набирає вигляду евклідова відстані; при $p = 1$ отримуємо відстань Хеммінга.

Відстані між об'єктами, розраховані з якої-небудь з перерахованих вище формул, представляють у вигляді матриці відстаней:

$$D = \begin{pmatrix} 0 & d_{12} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & 0 \end{pmatrix}, \quad (7.11)$$

де $d_{ij} \geq 0$.

Як бачимо, матриця відстаней являє собою квадратну матрицю типу "об'єкт - об'єкт" (близько n), де в якості елементів виступають відстані між об'єктами в метричному просторі. Діагональні елементи такої матриці дорівнюють нулю.

Матриця відстаней має такі властивості (аксіомами):

- 1) $d_{ij} = d_{ji}$ (аксіома симетрії). Відстань між об'єктами i та j дорівнює відстані між об'єктами j та i ,
- 2) $d_{ij} = 0$ тоді і тільки тоді, коли $i = j$ (аксіома тотожності). Відстань між об'єктом і їм самим дорівнює 0, тому на головній діагоналі матриці ми бачимо значення рівні 0.
- 3) $d_{ij} \leq d_{ik} + d_{kj}$ (аксіома трикутника або нерівність трикутника). Відстань між двома будь-якими точками завжди менше або дорівнює (в разі, якщо три точки лежать на одній прямій) суми відстаней від даних точок до якоїсь третьої точки.

Як зазначено вище, поряд з різними видами відстаней однорідність об'єктів може бути визначена за допомогою заходів ступеня близькості (подібності). В якості запобіжного близькості (подібності) може бути використаний **лінійний коефіцієнт кореляції**:

$$r_{ij} = \frac{\sum_{j=1}^k z_{ij} \cdot z_{ij}}{\sqrt{\sum_{j=1}^k z_{ij}^2 \cdot \sum_{j=1}^k z_{ij}^2}}, \quad (7.12)$$

де $i = 1, 2, \dots, n$.

Використання міри схожості в кластерному аналізі забезпечує комбінацію двох методів багатовимірною статистичного аналізу: кореляційного і кластерного аналізів. У цьому випадку в результаті отримуємо кореляційну матрицю або матрицю кореляцій (13):

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & 1 \end{pmatrix}, \quad (7.13)$$

де $-1 \leq r_{ik} \leq 1$; $k = 1, 2, \dots, n$.

Кореляційна матриця має такі властивості (аксіомами):

- 1) $d_{ij} = d_{ji}$ (аксіома симетрії). Кореляція (тіснота зв'язку) між об'єктами i та j дорівнює кореляції між об'єктами j та i ;
- 2) $d_{ii} = 1$ тоді і тільки тоді, коли $i = j$ (аксіома тотожності). Об'єкт схожий сам на себе, тому на головній діагоналі значення кореляцій рівні 1.

Етапи кластерного аналізу (3-5)

Етап. Вибір методу об'єднання об'єктів в кластери

Наступним етапом кластерного аналізу є об'єднання об'єктів в кластери при використанні матриці відстаней.

Методи об'єднання об'єктів в кластери, або методи кластерного аналізу, підрозділяються на ієрархічні і ітеративні.

Ієрархічні методи поділяються на дивізімні і агломеративні.

У **дивізімних** процедурах на першому етапі всі об'єкти об'єднані в один кластер, а на останньому етапі кожен об'єкт представляє окремий кластер. Іншими словами, агломеративного процедура - це об'єднання об'єктів в кластери, а дивізімная процедура - його розбиття сукупності об'єктів на кластери.

У **агломеративного** процедурах на першому етапі кожен об'єкт являє собою окремий кластер, а на останньому етапі - один об'єднаний кластер з усіх об'єктів.

Розглянемо ієрархічні агломеративні методи кластеризації даних, як найбільш бажані в комп'ютерних програмах. Метою цих методів є класифікація об'єктів в кластери на основі міри схожості або відстані.

Перевага ієрархічних процедур в тому, що вони дозволяють проаналізувати структуру досліджуваного безлічі спостережень і наочно представити результати кластеризації.

Результатом такої кластеризації є ієрархічне дерево, графік якого називається дендрограма.

До недоліків ієрархічних процедур слід віднести громіздкість обчислень і не завжди оптимальні розбиття на кластери.

Перерахуємо п'ять методів об'єднання (зв'язку), які найчастіше використовуються в ієрархічних агломеративного методах кластеризації даних:

- 1) **міжгрупове зв'язування (between groups linkage)**. Інша назва цього методу кластеризації - зв'язування середніх всередині груп. Обчислюється найменше середнє значення відстані між усіма нарами груп. На кожному кроці об'єднуються кластери або об'єкти, відстань між якими мінімально;
- 2) **одиначне зв'язування (nearest neighbor)**. Цей метод у вітчизняній літературі відомий під назвою "метод найближчого сусіда". Принцип дії цього

методу полягає в тому, що за відстань між кластерами приймають відстань між їхніми найближчими об'єктами. Іншими словами, спочатку об'єднуються два найближчих спостереження, що мають мінімальну відстань, після чого матриця відстаней перераховується наново. На останньому кроці все спостереження будуть об'єднані в один кластер;

- 3) **повне зв'язування (*furthest neighbor*)**. Найчастіше цей метод згадується, як "метод далекого сусіда", в якому за відстань між кластерами приймають відстань між найбільш віддаленими один від одного їх об'єктами;

- 4) **центроїдного кластеризація (*centroid clustering*)**. При центроїдного кластеризації відстань між кластерами знаходиться як відстань між їх центрами тяжіння;

- 5) **метод Варда (*Ward's method*)**. Для оцінки відстаней між кластерами використовується метод дисперсійного аналізу. Характеристикою якості кластеризації даних є досягнення мінімального значення внутрикласової дисперсії.

Ітеративний метод кластерного аналізу відрізняється від ієрархічного тим, що число кластерів визначається відразу, на самому початку аналізу. Найчастіше в ітеративному кластерному аналізі застосовується **метод К-середніх**, який є найбільш популярним методом кластеризації. При ієрархічному кластерному аналізі оптимальне число кластерів визначається на заключному етапі.

Метод **До-Середній** був запропонований в 1950-х рр. математиком Г. Штейнгаузом і майже одночасно С. Ллойдом. алгоритм

методу **До-Середній** полягає в мінімізації суми квадратів відхилень точок кластерів від центрів цих кластерів. На кожній ітерації центр має перераховуватися для кожного кластера, отриманого на попередньому кроці. Алгоритм виконується до тих пір, поки відбувається зміни кластерів.

Метод **До-Середній** істотно відрізняється від ієрархічного агломеративного методу і є ітеративною процедурою, в результаті якої на кожній ітерації об'єкти переміщуються в різні кластери. Метод / С-середніх передбачає, що задається число кластерів, на які треба розбити вихідні об'єкти. Алгоритм цього методу така: випадковим чином відбираються k об'єктів, які стають центрами груп. Потім склад кластерів змінюється з тим, щоб мінімізувати мінливість всередині кластерів і максимізувати мінливість між кластерами. Кожен наступний об'єкт відноситься до тієї групи, міра подібності з центром тяжкості якої мінімальна. Далі обчислюється новий центр тяжіння для кластера. Алгоритм буде повторюватися до тих пір, поки склад кластерів не перестане змінюватися.

Етап. Визначення оптимального числа кластерів

Як було зазначено вище, при ітеративному алгоритмі кластерного аналізу число кластерів має бути задано.

Відзначимо, що проблема визначення оптимального числа кластерів при ієрархічних процедурах кластерного аналізу є невирішеною. Однак є деякі підходи до її вирішення.

Одним з результатів ієрархічного кластерного аналізу є таблиця злиття, яка містить коефіцієнти. Ці коефіцієнти дорівнюють значенню відстані між кластерами, які об'єднуються на даному етапі. Використовуючи таблицю злиття, можна оцінити оптимальне число кластерів. Необхідно проаналізувати динаміку різниці між сусідніми коефіцієнтами по кроках кластеризації, іншими словами - абсолютний приріст коефіцієнта. Потрібно знайти той крок, при якому абсолютний приріст коефіцієнта максимальний, і визначити оптимальне число кластерів. Для цього знаходимо різницю між числом об'єктів n і порядковими номерами кроків (окремо), між якими виявлено найбільший абсолютний приріст коефіцієнта.

Так, наприклад, в розрахунках по країнам "великої двадцятки" ($n = 20$) і в таблиці послідовного злиття найбільший абсолютний приріст коефіцієнта виявлено при переході з 15-го на 16-й крок: він дорівнює 4,585. Отже, оптимальне число кластерів дорівнює п'яти кластерів ($20 - 15 = 5$) або чотирьох кластерів ($20 - 16 = 4$). Вибір числа кластерів (5 або 4) залежить від більш адекватної інтерпретації результатів. У нашому прикладі ми зупинилися на виборі п'яти кластерів.

Слід зазначити, що ієрархічні процедури кластерного аналізу розміщують об'єкти по групах, які можуть істотно відрізнятися за складом при використанні різних методів кластеризації. Кластерний метод привносить структуру в дані, і ця структура може не збігатися з реальною. Тому важливо відрізнити реальні угруповання від кластерів, отриманих методами кластерного аналізу.

Етап. Інтерпретація кластерів і якості розбиття

Для інтерпретації результатів необхідно мати "профіль" кластера, який добре описується середніми значеннями показників об'єктів, які увійшли в кластер, і середніми квадратичними відхиленнями. Розрахунок цих значень і їх аналіз дозволяє дати умовну назву кластерам і їх характеристику.

Важливим питанням на цьому етапі є оцінка якості розбиття на кластери. Найчастіше мірою якості виступає сума внутрікласових дисперсій відстаней або сума попарних внутрікласових відстаней між внутрікластерними елементами.

Приклад кластерного аналізу

Виконаємо кластерний аналіз в комп'ютерній програмі IBM SPSS 17 для даних, що характеризують рівень енергоефективності країн "Групи двадцяти". Країни цієї групи представлені 19 основними промисловими і формують ринки країнами: Аргентина, Австралія, Бразилія, Великобританія, Німеччина, Індія, Індонезія, Італія, Канада, Китай, Мексика, Туреччина, Росія, Саудівська Аравія, США, Франція, ПАР, Південна Корея, Японія, і окремо в групу входить Європейський Союз. У країнах "великої двадцятки" зосереджено дві третини населення світу, близько 90% світового ВВП. Вони представляють 80% світової

торгівлі; на частку цих країн припадає 84% викидів в атмосферу від згоряння викопного палива ^[1].

Мета кластерного аналізу - виділити однорідні групи країн за рівнем енергоефективності в 2012 р. Слід зазначити, що в "Групу двадцяти" входять країни, різні по території, чисельності та щільності населення, рівнем економічного потенціалу, структурі енергетики та енергетичної політики.

Етап. Подання вихідних даних у вигляді матриці

Для створення інформаційної бази нами були використані офіційно опубліковані статистичні дані огляду світової енергетики, представлені **ВР** ^[2]: виробництво електроенергії, ТВт • год і викиди CO₂ в атмосферу, млн т. Макроекономічні дані надані на офіційному сайті Федерального статистичного відомства Німеччини ^[3]: ВВП, млрд дол. США і чисельність населення, млн осіб.

Це найбільш важливі, на наш погляд, показники, що характеризують економічну, енергетичну та екологічну спрямованість розвитку кожної країни. На основі даних нами розраховані індикатори енергоефективності для кожної країни "Групи двадцяти" за 2012 р.:

X₁ - електроємність ВВП (витрати електроенергії на одиницю ВВП), кВт•год / дол. ;

X₂ - виробництво електроенергії на душу населення, МВт • год / чел. ;

x₃ - ВВП на одиницю викидів CO₂, дол. / кг;

X₄ - ВВП на душу населення, тис. Дол. / чел. ;

X_Г, - викиди CO₂ на душу населення, т / чел. ;

X_б - викиди CO₂ на одиницю виробництва електроенергії, кг / Вт • год.

Вихідні дані для кластерного аналізу, представлені в табл. 7.4, мають вигляд прямокутної матриці "об'єкт - ознака".

Таблиця 7.4

Матриця вихідних даних для кластерного аналізу

№	країни "Група двадцяти"	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
1	Австралія	0,16	10,41	4,04	65,55	17,28	1,66
2	Аргентина	0,29	3,38	2,48	11,51	4,64	1,37
3	Бразилія	0,23	2,79	4,90	12,33	2,52	0,90
4	Великобританія	0,15	5,75	4,62	38,81	8,39	1,46
5	Німеччина	0,18	7,54	4,27	42,48	9,95	1,32

6	Європейський Союз	0,19	6,41	4,40	34,40	7,81	1,22
7	Індія	0,59	0,85	0,98	1,44	1,47	1,73
8	Індонезія	0,22	0,81	1,87	3,76	2,01	2,47
9	Італія	0,14	4,86	5,09	33,94	6,67	1,37
10	Канада	0,34	17,48	2,91	51,72	17,75	1,02
11	Китай	0,62	3,66	0,87	5,92	6,82	1,86
12	Мексика	0,24	2,41	2,44	10,00	4,11	1,70
13	Росія	0,53	7,43	1,19	14,09	11,88	1,60
14	Саудівська Аравія	0,39	8,89	1,06	23,04	21,74	2,44
15	США	0,27	13,56	2,70	49,73	18,43	1,36
16	Туреччина	0,29	3,23	2,57	11,04	4,30	1,33
17	Франція	0,21	8,53	7,07	41,28	5,84	0,68
18	ПАР	0,61	5,04	0,94	8,20	8,71	1,73
19	Південна Корея	0,45	10,45	1,52	23,28	15,27	1,46
20	Японія	0,18	8,63	4,24	46,87	11,04	1,28
Середнє значення		0,31	6,61	3,01	26,47	9,33	1,50

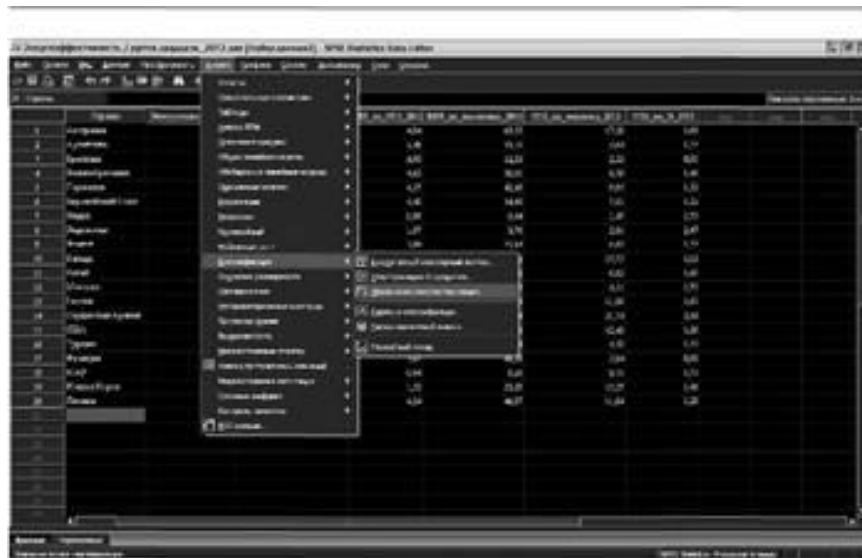
В отриманій матриці значення X_{ij} відповідає i -й країні "Групи двадцяти", яка характеризується j -м показником енергоефективності, де $i = 1, 2, \dots, 20$ і $j = 1, 2, \dots, 6$.

Всі показники енергоефективності розраховані як відношення двох абсолютних показників, тому є відносними показниками. Проводимо стандартизацію вихідних змінних з тим, щоб усунути залежність змінних від одиниць виміру.

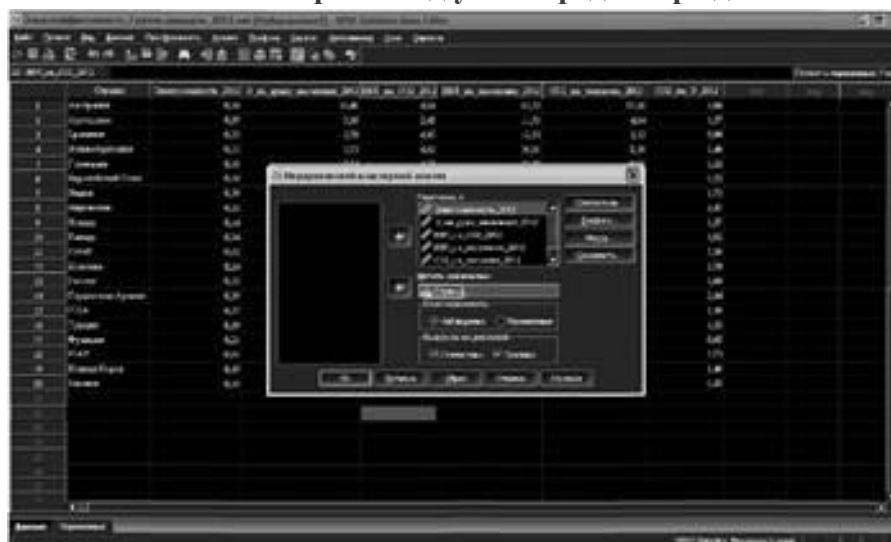
Етап. Визначити подібність об'єктів

Проводиться вибір команд у вікні редактора даних: *Аналіз - Класифікація - ієрархічна кластеризація* (рис. 7.1).

Вибір змінних для кластеризації, представлений на рис. 7.2, передбачає вказівку всіх змінних енергоефективності. В поле *мітити значень* поміщаємо якісну змінну *Країни*, яка виступає в якості ідентифікатора.



Мал. 7.1. Вибір команд у вікні редактора даних



Мал. 7.2. Вибір змінних для кластеризації

Графічно кластеризацію представляють у вигляді дендрограмми. На рис. 7.3 показано діалогове вікно *Ієрархічний кластерний аналіз : Графіки*, в якому в якості діаграми пропонується побудувати Дендрограма.



Мал. 7.3. Діалогове вікно Ієрархічний кластерний аналіз: Графіки

В якості методу кластеризації використовуємо *міжгрупові зв'язку*, при якому в кластери об'єднуються об'єкти, відстань між якими мінімально. Виберемо *Квадрат відстані Евкліда* (формула (7.7)) в якості міри близькості (рис. 7.4).



Мал. 7.4. Вибір способу вимірювання відстані і методу формування кластерів

Для прояви стандартизації змінних в поле *Стандартизація* задається вид формули. Вибираємо *пана значення* (див. Формулу (7.2)).

У табл. 7.5 представлений фрагмент матриці відстаней, яка представляє собою квадратну матрицю типу "об'єкт - об'єкт" (див. Формулу (7.11)).

Таблица 7.5

Фрагмент матриці відстаней

	Австр алія	Аргент ина	Брази лія	Великобри танія	Німечч ина	Європейс ький Союз	Інд ія
Австралія	0,0	17,1	20,5	5,7	4,1	7,1	33, 5

Аргентина	17,1	0,0	3,4	5,1	5,9	3,9	5,8
Бразилія	20,5	3,4	0,0	5,3	6,4	3,5	14,2
Великобританія	5,7	5,1	5,3	0,0	0,5	0,5	18,7
Німеччина	4,1	5,9	6,4	0,5	0,0	0,4	20,0
Європейський Союз	7,1	3,9	3,5	0,5	0,4	0,0	17,2
Індія	33,5	5,8	14,2	18,7	20,0	17,2	0,0

Етап. Вибір методу об'єднання об'єктів в кластери

Для визначення оптимального числа кластерів застосуємо таблицю послідовності злиття, яка дозволяє переглянути черговість злиття кластерів і вибрати їх оптимальне число (табл. 7.6).

Таблиця 7.6

Таблиця злиття

етап	Кластер об'єднаний з		коефіцієнти	Етап першого появи кластера		Наступний етап	Абсолютний приріст коефіцієнта *
	кластер 1	кластер 2		кластер 2	кластер 1		
1	2	16	0,016	0	0	6	-
2	5	20	0,160	0	0	7	0,144
3	4	9	0,310	0	0	5	0,150
4	11	18	0,312	0	0	10	0,002
5	4	6	0,502	3	0	7	0,190
е	2	12	0,793	1	0	11	0,291
7	4	5	1,052	5	2	13	0,259

8	13	19	1,438	0	0	14	0,386
9	10	15	1,672	0	0	12	0,235
10	7	11	1,991	0	4	14	0,318
11	2	3	3,932	6	0	15	1,942
12	1	10	4,937	0	9	18	1,005
13	4	17	5,168	7	0	15	0,231
14	7	13	5,793	10	8	16	0,625
15	2	4	6,897	11	13	18	1,105
16	7	14	11,482	14	0	17	4,585
17	7	8	12,926	16	0	19	1,444
18	1	2	13,416	12	15	19	0,490
19	1	7	16,241	18	17	0	2,825

* Абсолютні прирости коефіцієнтів розраховані нами самостійно, так як в програмі цього розрахунку не здійснюється.

Стовпець таблиці "Коефіцієнти" містить відстані для відповідних кластерів. Кожен рядок показує склад кластера на даному етапі класифікації. На першому етапі об'єднані об'єкти (країни) № 2 і № 16, на другому - № 5 і № 20, на третьому - № 4 та № 9 і т.д.

Етап. Визначення оптимального числа кластерів

У стовпці "Коефіцієнти" розташовані відстані між двома кластерами, розраховані на основі обраної міри відстаней (в нашому прикладі це квадрат евклидова відстані). Кластеризації потрібно зупинити на тому етапі, коли зміна коефіцієнта відбувається стрибкоподібно, тому що в протилежному випадку в одному кластері виявляться занадто далекі один від одного об'єкти.

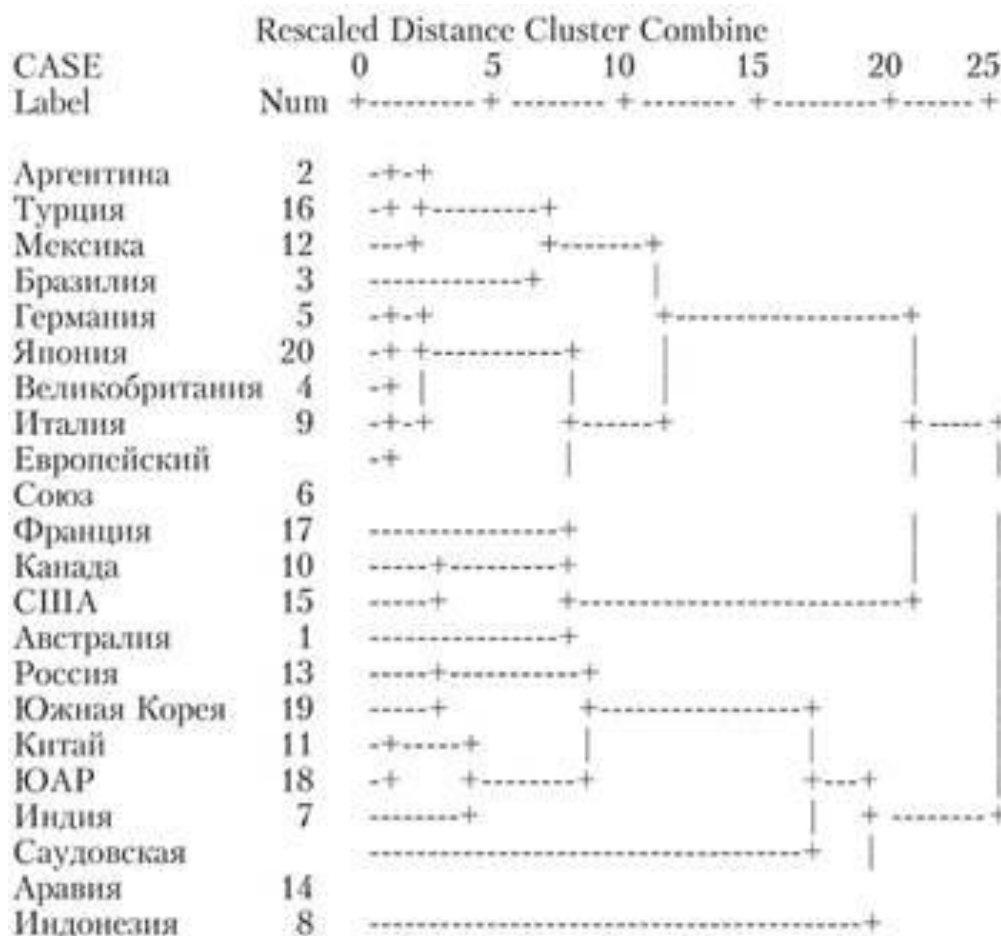
Оптимальним вважається число кластерів, рівну різниці числа спостережень і номери етапу, після якого відбувається стрибок. Останній стовпець табл. 7.6 демонструє приріст коефіцієнта. Як було зазначено вище, цей розрахунок ми зробили самостійно для того, щоб виявити стрибок. Ми бачимо, що коефіцієнт приростає стрибкоподібно, нерівномірно, і найбільший стрибок спостерігається на етапі з 15-го на 16-й. Виходить, що в першому

випадку оптимальне число кластерів дорівнює п'яти ($20 - 15 = 5$), у другому випадку - чотирьом ($20 - 16 = 4$). Провівши якісний аналіз кластерів в обох випадках, зупинимося на виборі п'яти кластерів (рис. 7.5). Цей вибір дозволяє дати більш чітку характеристику однорідних груп.



Мал. 7.5. Ієрархічний кластерний аналіз: вибір числа кластерів

На рис. 7.6 представлена горизонтальна деревоподібна діаграма, звана дендрограмма. На горизонтальній осі представлені відстані, на вертикальній осі перераховані країни "Групи двадцяти".



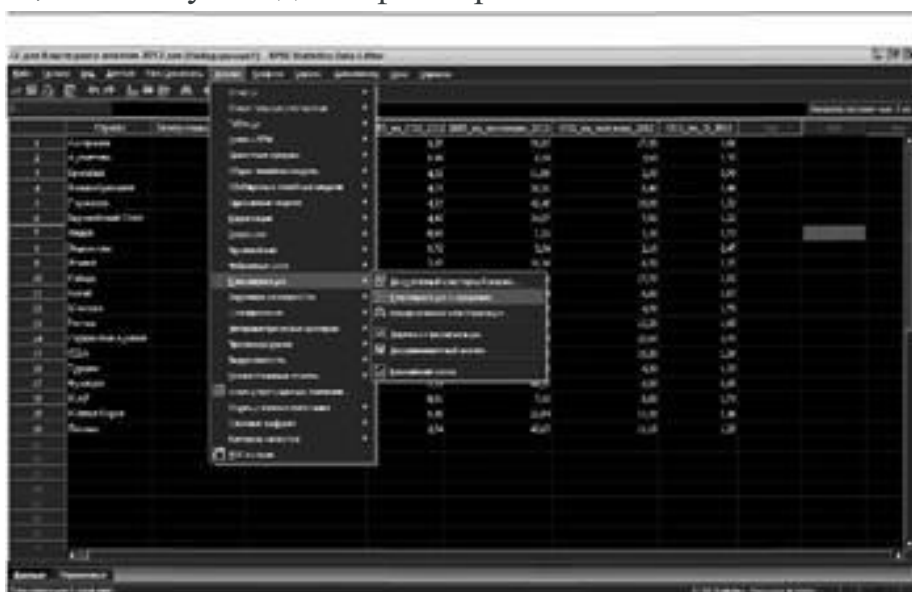
Мал. 7.6. Дендрограмма

Метод К-середніх

Як було зазначено вище, найбільш популярним методом кластеризації є метод *До-Середній*, який відноситься до ітеративному увазі кластерного аналізу. На відміну від ієрархічного виду кластерного аналізу, в методі *До-Середній* число кластерів визначається на початку аналізу.

Проведений вище ієрархічний аналіз дозволив нам визначити оптимальну кількість кластерів: 5. Вибираємо команди у вікні редактора даних: *Аналіз - Класифікація - Кластеризація К-середніми* (рис. 7.7).

Вибір змінних для кластеризації методом *К-середній* представлений на рис. 7.8. В поле *Змінні* вказуємо всі змінні для аналізу енергоефективності країн "Групи двадцяти". В поле *мітити спостереження* поміщаємо якісну змінну *Країни*, яка виступає ідентифікатором.



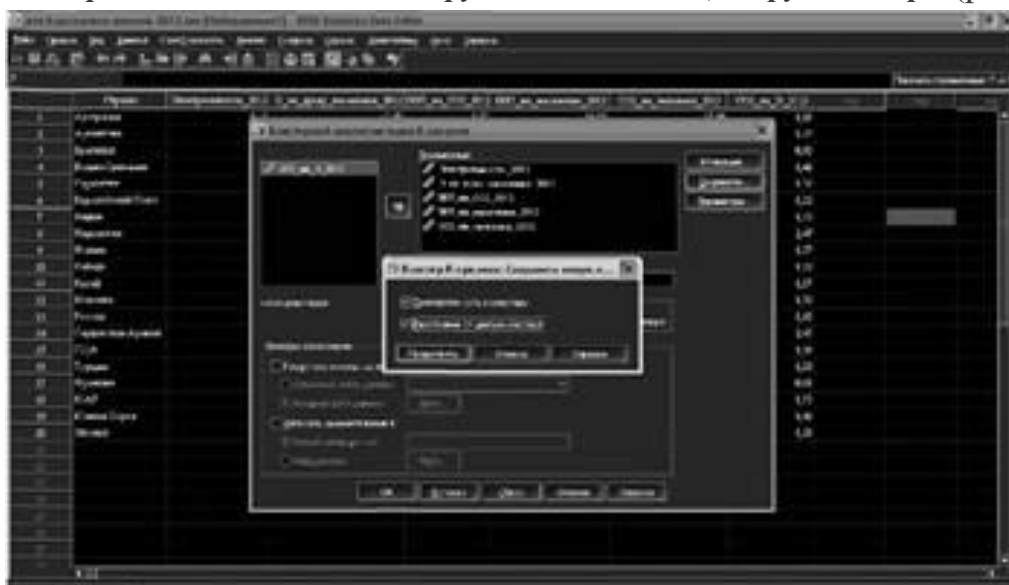
Мал. 7.7. Вибір команд у вікні редактора даних

Важливо вказати число кластерів, на яке будемо розбивати досліджувану сукупність, в поле *Число кластерів*. Для нашого прикладу в цьому полі вказуємо число 5.



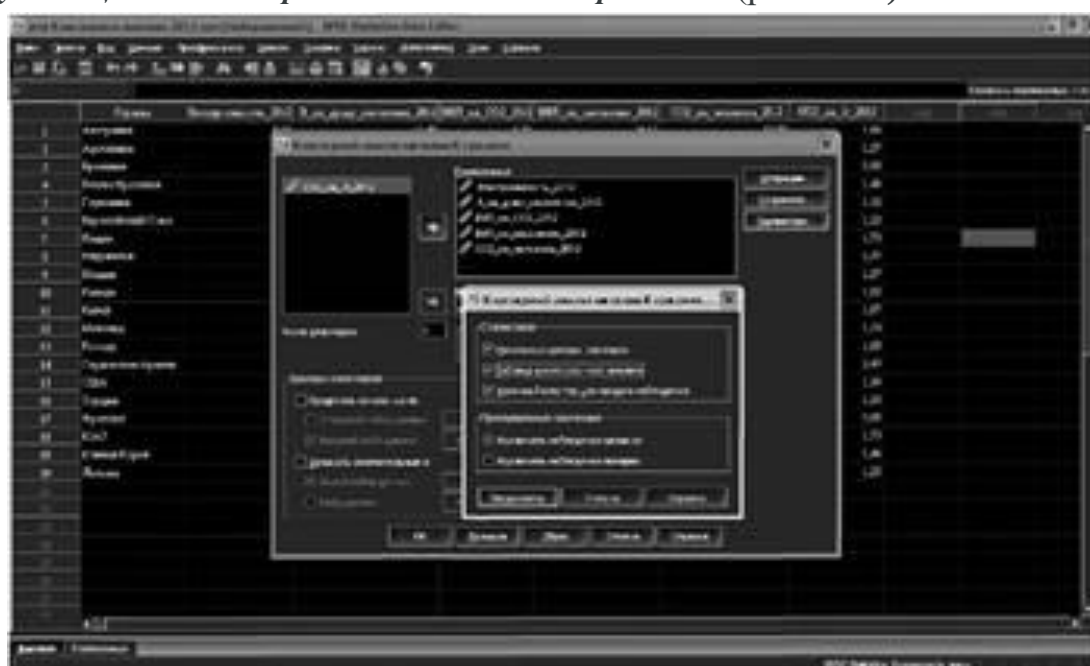
Мал. 7.8. Вибір змінних для кластеризації методом K -середній

Далі вибираємо опцію *Зберегти ...* для того, щоб отримати результати: *Приналежність до кластеру* і *Відстань від центру кластера* (рис. 7.9).



Мал. 7.9. Кластеризація методом K -середній: Зберегти нову приналежність до кластеру

Дати якісний аналіз нових труни (кластерів) дозволяє описова статистика енергоефективності. Для цього вибираємо опцію *Параметри ...* і робимо вибір на нулі *Статистики: Початкові центри кластерів, Таблиця дисперсійного аналізу* і *Кінцевий кластер для кожного спостереження* (рис. 7.10).



Мал. 7.10. Вибір параметрів для кластерного аналізу методом K -середніх

Виконавши вищеперелічені команди, отримаємо таблицю, в якій показана належність країн "Групи двадцяти" до кластерів (табл. 7.7).

Таблиця 7.7

Належність країн "Групи двадцяти" до кластерів

№	країни	кластер	відстань
1	Австралія	1	0,000
2	Аргентина	4	3,580
3	Бразилія	4	5,479
4	Великобританія	3	1,366
5	Німеччина	3	4,993
6	Європейський Союз	3	3,851
7	Індія	4	7,521
8	Індонезія	4	5,308
9	Італія	3	4,716
10	Канада	5	5,248
11	Китай	4	3,627
12	Мексика	4	2,049
13	Росія	2	7,640
14	Саудівська Аравія	2	6,203
15	США	5	2,792
16	Туреччина	4	3,098
17	Франція	3	4,586
18	ПАР	4	5,090
19	Південна Корея	2	3,671
20	Японія	5	7,122

Для якісного аналізу груп розбиття в програмі розраховується середнє значення для кожного показника енергоефективності по країнам "великої двадцятки". Кінцеві центри кластерів і середні значення по країнах "великої двадцятки" представлені в табл. 7.8.

Таблиця 7.8

Кінцеві центри кластерів

показники енергоефективності	Середні значення по країнах "великої двадцятки"	кластер				
		1	2	3	4	5
Електроємність ВПП, кВт • год / дол.	0,31	0,16	0,46	0,17	0,39	0,26
Виробництво електроенергії на душу населення, МВт • год / чол.	6,61	10,41	8,92	6,62	2,77	13,22
ВВП на одиницю викидів CO ₂ , дол / кг	3,01	4,04	1,26	5,09	2,13	3,28
ВВП на душу населення, тис. Дол / чол.	26,47	65,55	20,14	38,18	8,03	49,44
Викиди CO ₂ на душу населення, т / чол.	9,33	17,28	16,30	7,73	4,32	15,74
Викиди CO ₂ на одиницю виробництва електроенергії, кг / Втч	1,50	1,66	1,83	1,21	1,64	1,22

Етап. інтерпретація кластерів

З огляду на результати, представлені в табл. 7.7 і 7.8, проводимо порівняння середніх значень по кожному кластеру.

Перший кластер складається з однієї країни: Австралія, яка характеризується хорошим рівнем наступних показників: електроємність ВВП, ВВП на одиницю викидів CO₂ і ВВП на душу населення. Відзначимо, що три

показники мають значення гіршими за середні: виробництво електроенергії на душу населення, викиди CO₂ на душу населення, викиди CO₂ на одиницю виробництва електроенергії.

У другій кластер увійшли три країни: Росія, Саудівська Аравія та Південна Корея, у яких всі показники енергоефективності гірше, ніж середні значення по всіх країнах.

У третій кластер в результаті кластерного аналізу визначено п'ять країн: Великобританія, Німеччина, Європейський Союз, Італія і Франція. У цю групу увійшли країни, у яких всі показники енергоефективності вище середнього значення для всіх країн "Групи двадцяти".

Четвертий кластер представляють вісім країн: Аргентина, Бразилія, Індія, Індонезія, Китай, Мексика, Туреччина і ПАР. Ця група країн характеризується високою електроємна і низьким ВВП на душу населення. У той же час спостерігаються нижче середнього виробництво електроенергії на душу населення і викиди CO₂ на душу населення.

У п'ятий кластер потрапили три країни: Канада, США і Японія. Цей кластер складають країни з високою енергоефективністю, так як майже всі показники мають значення краще середніх значень, крім двох показників, які мають значення гірше середнього значення за сукупністю "Групи двадцяти": виробництво електроенергії па душу населення і викиди CO₂ на душу населення.

Дамо назви отриманим кластерам і розподілимо ці кластери в порядку зростання: перший кластер - це кластер, що включає країни з низьким рівнем енергоефективності, п'ятий - кластер, що включає країни з високим рівнем енергоефективності (табл. 7.9). В результаті первісна нумерація зміниться, але для логічного узагальнення результатів це зручніше.

Попит на енергію в усьому світі зростає і буде продовжувати рости. Підвищення енергоефективності використання всіх видів енергії є одним із пріоритетних завдань в світі. Як правило, це обумовлено високим рівнем енергоемності ВВП, необхідністю модернізації економіки, підвищення конкурентоспроможності промисловості та поліпшення стану навколишнього середовища.

Таблиця 7.9

Результати кластерного аналізу

кластер	Назва кластера за рівнем енергоефективності в країні	Число країн в кластері	Країни, певні в кластер
1	низький	3	Росія, Саудівська Аравія, Південна Корея
2	Нижче середнього	8	Аргентина, Бразилія, Індія, Індонезія, Китай, Мексика, Туреччина, ПАР
3	середній	1	Австралія
4	Вище середнього	3	Канада, США, Японія
5	високий	5	Великобританія, Німеччина, Європейський Союз, Італія, Франція
всього		20	

Лекція 8. Математичне формування задачі лінійного програмування

Велика кількість планово-виробничих і економічних задач пов'язана з розподілом яких-небудь, як правило, обмежених ресурсів (сировини, робочої сили, енергії, палива і т.ін.). Часто розподіл ресурсів можна здійснити не єдиним чином. Наприклад, дану продукцію можна отримати різними способами, по-різному вибираючи технологію, сировину, застосовуючи обладнання, організацію процесу. При цьому кожний спосіб розподілу ресурсів, що оцінюється з позиції деякого критерію (прибуток, об'єм випуску продукції і т.ін.), характеризується певним значенням показника цього критерію. Природним тому є намір знайти такий варіант розподілу (програму, план), який би гарантував найбільший економічний ефект. Такий план називають оптимальним.

Реальні економічні процеси досить складні. При їх математичному описанні доводиться враховувати багато різних факторів. Тому математична модель містить велике число умов обмежень із багатьма невідомими. Якщо невідомі входять в модель тільки в першій степені, то задача належить до розділу лінійного програмування, в протилежному випадку – до розділу нелінійного програмування. Оптимізаційні задачі, в яких потрібно враховувати послідовність дій або фактор часу, розглядаються в розділі динамічного програмування. Якщо в задачі фігурують параметри, що є випадковими величинами, то вона відноситься до задач стохастичної оптимізації.

Предметом дослідження математичного програмування є математичні моделі, пов'язані у більшості випадків із визначеними економічними процесами, що описують економіку підприємства, промислового об'єднання, народного господарства або окремих економічних процесів у них.

Математичне програмування – це розділ математики, який розробляє теорію і чисельні методи розв'язання багатовимірних задач з обмеженнями, тобто задач на екстремум функції багатьох змінних з обмеженнями на область зміни цих змінних. На відміну від класичної теорії екстремальних задач, яка є частиною математичного програмування, основна увага приділяється тим задачам, в яких активно беруть участь обмеження на область зміни змінних.

8.1. Приклади задач лінійного програмування

1. Задача оптимального виробничого планування

Однією із найбільш розповсюджених задач даної групи є задача про максимальний випуск продукції із наявних обмежених запасів сировини при різних технологіях виробництва. Сюди ж можна віднести задачу досягнення

максимальної рентабельності підприємства при виробництві із наявних запасів ресурсів різних видів продукції.

Приклад 1. Для виготовлення трьох видів виробів А, В і С використовується токарне, фрезерне, зварювальне і шліфувальне обладнання. Затрати часу на обробку одного виробу для кожного з типів обладнання вказані в таблиці 3.1. У ній же вказаний загальний фонд робочого часу кожного із типів використовуваного обладнання, а також прибуток від реалізації одного виробу кожного виду.

Таблиця 3.1

Тип обладнання	Затрати часу (верстато-год.) на обробку одного виробу виду			Загальний фонд робочого часу обладнання (год.)
	А	В	С	
Фрезерне	2	4	5	120
Токарне	1	8	6	280
Зварювальне	7	4	5	240
Шліфувальне	4	6	7	360
Прибуток (грн.)	10	14	12	

Вимагається визначити, скільки виробів і якого виду потрібно виготовити підприємству, щоб прибуток від їх реалізації був максимальним. Скласти математичну модель задачі.

Розв’язання. Припустимо, що буде виготовлено x_1 одиниць виробів виду А, x_2 одиниці – виду В і x_3 одиниці – виду С. Тоді для виробництва такої кількості виробів потрібно затратити $2x_1 + 4x_2 + 5x_3$ верстато-годин фрезерного обладнання.

Оскільки загальний фонд робочого часу верстатів даного типу не може перевищувати 120, то повинна виконуватися нерівність

$$2x_1 + 4x_2 + 5x_3 \leq 120.$$

Аналогічні міркування відносно можливого використання токарного, зварювального і шліфувального обладнання призведуть до таких нерівностей:

$$x_1 + 8x_2 + 6x_3 \leq 280,$$

$$7x_1 + 4x_2 + 5x_3 \leq 240,$$

$$4x_1 + 6x_2 + 7x_3 \leq 360.$$

При цьому через те, що кількість виробів, які виготовляються, не може бути від’ємною, то

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \quad (1)$$

Далі, якщо буде виготовлено x_1 одиниць виробів виду А, x_2 одиниць виробів виду В і x_3 одиниць виробів виду С, то прибуток від їх реалізації складе $F = 10x_1 + 14x_2 + 12x_3$.

Таким чином, приходимо до наступної математичної задачі: дана система

$$\begin{cases} 2x_1 + 4x_2 + 5x_3 \leq 120, \\ x_1 + 8x_2 + 6x_3 \leq 280, \\ 7x_1 + 4x_2 + 5x_3 \leq 240, \\ 4x_1 + 6x_2 + 7x_3 \leq 360 \end{cases} \quad (2)$$

чотирьох лінійних нерівностей з трьома невідомими x_j ($j = \overline{1,3}$) і лінійна функція відносно цих змінних

$$F = 10x_1 + 14x_2 + 12x_3; \quad (3)$$

потрібно серед усіх невід'ємних розв'язків системи нерівностей (2) знайти такий, при якому функція (3) набуває максимального значення. Як це зробити, буде показано пізніше.

Лінійна функція (3), максимум якої потрібно визначити, разом із системою нерівностей (2) і умовою невід'ємності змінних (1) утворюють математичну модель початкової задачі.

Оскільки функція (3) лінійна, а система (2) містить тільки лінійні нерівності, то задача (1) – (3) є задачею лінійного програмування.

Приклад 2. Міський молочний завод виготовляє молоко, кефір і сметану, розфасовані в пляшки. На виробництво 1т молока, кефіру і сметани потрібно відповідно 1010, 1010 і 9450 кг молока. При цьому затрати робочого часу при розливі 1т молока і кефіру складають 0,18 і 0,19 машино-годин. На розфасовці 1т сметани зайняті спеціальні автомати протягом 3,25 год. Усього для виробництва молочних виробів завод може використати 136 000 кг молока. Основне обладнання може бути зайняте протягом 21,4 машино-годин, а автомати з розфасовки сметани – впродовж 16,25 год. Прибуток від реалізації 1 т молока, кефіру і сметани відповідно становить 30, 22 і 136 грн. Завод повинен щодня виробляти не менше 100 т молока, розфасованого в пляшки. На виробництво іншої продукції немає ніяких обмежень.

Потрібно визначити, яку продукцію і в якій кількості потрібно щодня виробляти заводу, щоб прибуток від її реалізації був максимальний. Скласти математичну модель задачі.

Розв'язання. Припустимо, що молочний завод буде щодня виготовляти x_1 тонну молока, x_2 тонни кефіру і x_3 тонни сметани. Тоді йому для виробництва цієї продукції необхідно $1010x_1 + 1010x_2 + 9450x_3$ тонни молока.

Оскільки завод може використовувати щодня не більше 136 000 т молока, то повинна виконуватися нерівність

$$1010x_1 + 1010x_2 + 9450x_3 \leq 136000.$$

Аналогічні міркування, проведені відносно можливого використання ліній розливу молочної продукції і автоматів з розфасовки сметани, дозволяють записати такі нерівності:

$$\begin{aligned} 0,18x_1 + 0,19x_2 &\leq 21,4, \\ 3,25x_2 &\leq 16,25. \end{aligned}$$

Оскільки щодня повинно вироблятися не менше 100 т молока, то $x_1 \geq 100$. Далі, за своїм економічним значенням змінні x_2 і x_3 можуть набувати тільки невід'ємних значень: $x_2 \geq 0$, $x_3 \geq 0$. Загальний прибуток від реалізації x_1 тонни молока, x_2 тонни кефіру і x_3 тонни сметани дорівнює $30x_1 + 22x_2 + 136x_3$ грн. Таким чином, переходимо до наступної математичної задачі: дана система

$$\begin{cases} 1010x_1 + 1010x_2 + 9450x_3 \leq 136000 \\ 0,18x_1 + 0,19x_2 \leq 21,4, \\ 3,25x_2 \leq 16,25, \\ x_1 \geq 100 \end{cases} \quad (4)$$

чотирьох лінійних нерівностей із трьома невідомими x_1 , x_2 , x_3 і лінійна функція відносно цих же змінних

$$F = 30x_1 + 22x_2 + 136x_3. \quad (5)$$

Потрібно серед усіх невід'ємних розв'язків системи нерівностей (4) знайти такий, при якому функція (5) матиме максимальне значення. Оскільки система (4) являє собою співвідношення лінійних нерівностей і функція (5) лінійна, то вихідна задача є задачею лінійного програмування.

2. Задача про оптимальний склад суміші

Досить широкий клас задач програмування становлять так звані задачі на складання сумішей або задачі на використання замінників.

Приклад 3. При відгодівлі тварин кожна тварина щодня повинна одержувати не менше 60 одиниць поживної речовини A , не менше 50 одиниць речовини B і не менше 12 одиниць речовини C . Вказані поживні речовини містяться в трьох видах корму. Склад одиниць поживних речовин в 1кг кожного з видів корму наведений у таблиці 3.2:

Таблиця 3.2

Поживні речовини	Кількість одиниць поживних речовин в 1кг корму виду		
	I	II	III
A	1	3	4
B	2	4	2
C	1	4	3

Скласти денний раціон, що забезпечує отримання необхідної кількості поживних речовин при мінімальних грошових витратах, якщо ціна 1 кг корму I виду складає 9 коп., корму II виду — 12 коп. і корму III виду — 10 коп.

Приклад 4. Для виробництва чавунного литва використовується n різних початкових шихтових матеріалів (чавун різних марок, сталевий брухт, ферофосфор та ін.). Хімічний склад чавунного литва визначається вмістом в ньому хімічних елементів (кремнію, марганцю, фосфору та ін.). Готовий чавун повинен мати строго визначений хімічний склад, який задається величинами H_j що є частками (у відсотках) j -ого хімічного елементу в готовому продукті. При цьому відомі величини: h_{ij} — вміст (у відсотках) j -го хімічного елементу в i -му початковому шихтовому матеріалі; c_i — ціна одиниці кожного i -го шихтового матеріалу. Визначити склад шихти, що забезпечує отримання литва заданої якості при мінімальній загальній вартості використовуваних шихтових матеріалів.

3. Задача про оптимальний план перевезень

Цю задачу часто називають транспортною. У простішому варіанті транспортна задача виникає при необхідності найбільш раціонального перевезення деякого однорідного вантажу. При цьому споживачам байдуже, з яких пунктів він надходить, важливо, щоб був задоволений попит, а кожен постачальник має можливість постачати вантаж будь-якому споживачу. Зворотні перевезення не передбачаються.

Приклад 5. У трьох пунктах відправлення зосереджений однорідний вантаж у кількостях, відповідно рівних 420, 380 і 400 т. Цей вантаж необхідно перевезти в три пункти призначення в кількостях, відповідно рівних 260, 520 і 420 т. Вартості перевезень 1 т вантажу з кожного пункту відправлення в кожен пункт призначення є відомими величинами і задаються матрицею

$$(c_{ij}) = \begin{pmatrix} 2 & 4 & 3 \\ 7 & 5 & 8 \\ 6 & 9 & 7 \end{pmatrix}.$$

Знайти план перевезень, що забезпечує вивіз наявного в пунктах відправлення і завезення необхідного в пункти призначення вантажу при мінімальній загальній вартості перевезень.

4. Задача про оптимальне розміщення виробництва

Це одна з важливих модифікацій транспортної задачі.

Приклад 6. У m пунктах можуть бути розміщені підприємства, що виробляють деяку однорідну продукцію. Ця продукція надходить у n пунктів її споживання, причому в j -му пункті потреби в продукції рівні a_j одиницям. Витрати, пов'язані з доставкою одиниці продукції з i -го пункту відправлення в

j -й пункт споживання, складають c_{ij} грн. Відомо, що в i -му пункті виготовлення продукції максимальний об'єм її виробництва не може перевищувати b_i одиниць, а витрати, пов'язані з виготовленням одиниці продукції, складають d_i грн. Визначити таке розміщення підприємств, при якому забезпечуються потреби в продукції у кожному з пунктів її споживання при якнайменших загальних затратах, пов'язаних із виробництвом і доставкою продукції.

5. Задача про раціональний розкрій матеріалів

Модель цієї задачі має важливе значення для економії матеріалів та сировини.

Приклад 7. На швейній фабриці тканина може бути розрізана кількома способами для виготовлення потрібних деталей швейних виробів. Нехай при j -му варіанті розкрою ($j = \overline{1, n}$) 100 м^2 тканини виготовляється b_{ij} деталей i -го виду ($i = \overline{1, m}$), а розміри відходів при даному варіанті розкрою дорівнюють $c_j \text{ м}^2$. Знаючи, що деталей i -го виду потрібно виготовляти B_i штук, потрібно розкроїти тканину так, щоб було одержано необхідну кількість деталей кожного виду при мінімальних загальних відходах. Скласти математичну модель задачі.

6. Задача динаміки виробництва і створення запасів

Ця задача полягає в оптимальному розподілі деякої продукції та її запасів. В умовах діючого підприємства всяка зміна об'єму випуску продукції пов'язана з додатковими витратами. Зберігання готової продукції також вимагає певних витрат. У випадках, коли попит на готову продукцію в окремі відрізки часу не постійний, виникає потреба пошуку такого компромісного плану випуску, при якому сумарні затрати на розширення і згортання виробництва, а також на зберігання залишків продукції були б мінімальними за умови своєчасного і повного задоволення потреб.

7. Стохастична задача комплектування станочного парку

Приклад 8. На авторемонтний завод протягом деякого відрізка часу надходять замовлення на виконання ремонтних робіт. Наперед невідомі час надходження замовлень і їх кількість. Однак зрозуміло, що якщо верстатів для виконання різноманітних замовлень недостатньо, то це призведе до затримки у здійсненні ремонту, а замовники звернуться за послугами до інших, більш укомплектованих підприємств, і даний завод зазнає збитків у зв'язку з недоотриманням прибутку.

З другого боку, якщо набір різних верстатів занадто розширити, то більшу частину часу вони будуть простоювати, а завод, витративши гроші на їх

закупівлю, змушений буде і далі терпіти збитки у зв'язку з утримуванням надлишкової кількості оснащення.

У даному випадку прибуток, що отримує завод, є випадковою величиною, а тому говорити про його максимізацію немає сенсу. Тому на практиці як цільова функція вибирається або математичне сподівання прибутку, обчислене на основі відомих ймовірностей надходження замовлень (цей випадок зводиться до звичайної задачі лінійного програмування), або ймовірності того, що розмір доходу буде не меншим заданої величини (тут потрібні спеціальні методи досліджень).

8.2. Загальна і основна задачі лінійного програмування

У попередньому параграфі були розглянуті приклади задач лінійного програмування. У всіх цих задачах потрібно знайти максимум або мінімум лінійної функції за умови, що її змінні набували невід'ємних значень і задовольняли деяку систему лінійних рівнянь або лінійних нерівностей або систему, що містить як лінійні нерівності, так і лінійні рівняння. Кожна з цих задач є окремим випадком загальної задачі лінійного програмування.

Загальною задачею лінійного програмування називається задача, яка полягає у визначенні максимального (мінімального) значення функції

$$F = \sum_{j=1}^n c_j x_j \quad (1)$$

за умов

$$\sum_{j=1}^n a_{ij} x_j \leq b_i \quad (i = \overline{1, k}), \quad (2)$$

$$\sum_{j=1}^n a_{ij} x_j = b_i \quad (i = \overline{k+1, m}), \quad (3)$$

$$x_j \geq 0 \quad (j = \overline{1, l}, \quad l \leq n), \quad (4)$$

де a_{ij}, b_i, c_j — задані постійні величини і $k \leq m$.

Функція (1) називається *цільовою функцією* (або *лінійною формою*) задачі (1) – (4), а умови (2) – (4) — обмеженнями даної задачі.

Стандартною (або *симетричною*) *задачею лінійного програмування* називається задача, яка полягає у визначенні максимального значення функції (1) при виконанні умов (2) і (4), де $k = m$ і $l = n$.

Канонічною (або *основною*) *задачею лінійного програмування* називається задача, яка полягає у визначенні максимального значення функції (1) при виконанні умов (3) і (4), де $k = 0$ і $l = n$.

Сукупність чисел $X = (x_1, x_2, \dots, x_n)$, що задовольняють обмеженням задачі (2) – (4), називається *допустимим розв'язком* (або *планом*).

План $X^* = (x_1^*, x_2^*, \dots, x_n^*)$, при якому цільова функція задачі (1) набуває свого максимального (мінімального) значення, називається *оптимальним*.

Значення цільової функції (1) при плані X будемо позначати через $F(X)$. Отже, X^* — *оптимальний* план задачі, якщо для будь-якого X виконується нерівність $F(X) \leq F(X^*)$ [відповідно $F(X) \geq F(X^*)$].

Вказані вище три форми задачі лінійного програмування еквівалентні в тому значенні, що кожна з них за допомогою нескладних перетворень може бути переписана у формі іншої задачі. Це означає, що якщо є спосіб знаходження розв'язку однієї з вказаних задач, то тим самим може бути визначений оптимальний план будь-якої з трьох задач.

Щоб перейти від однієї форми запису задачі лінійного програмування до іншої, потрібно в загальному випадку вміти, по-перше, зводити задачу мінімізації функції до задачі максимізації, по-друге, переходити від обмежень-нерівностей до обмежень-рівностей і навпаки, по-третє, замінювати змінні, які не підлягають умові невід'ємності.

У тому випадку, коли вимагається знайти мінімум функції $F = c_1x_1 + c_2x_2 + \dots + c_nx_n$, можна перейти до знаходження максимуму функції $F_1 = -F = -c_1x_1 - c_2x_2 - \dots - c_nx_n$, оскільки $\min F = -\max(-F)$.

Обмеження-нерівність початкової задачі лінійного програмування, що має вигляд « $\leq$ », можна перетворити в обмеження-рівність додаванням до його лівої частини додаткової невід'ємної змінної, а обмеження-нерівність виду « $\geq$ » — в обмеження-рівність відніманням із його лівої частини додаткової невід'ємної змінної. Таким чином, обмеження-нерівність

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \leq b_i$$

перетвориться в обмеження-рівність

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n + x_{n+1} = b_i \quad (x_{n+1} \geq 0), \quad (5)$$

а обмеження-нерівність

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \geq b_i$$

— в обмеження-рівність

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n - x_{n+1} = b_i \quad (x_{n+1} \geq 0). \quad (6)$$

У той же час кожне рівняння системи обмежень

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i$$

можна записати у вигляді нерівностей:

$$\begin{cases} a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \leq b_i, \\ -a_{i1}x_1 - a_{i2}x_2 - \dots - a_{in}x_n \leq -b_i. \end{cases} \quad (7)$$

Число додаткових невід'ємних змінних, що вводяться, при перетворенні обмежень-нерівностей в обмеження-рівність дорівнює числу перетворюваних нерівностей.

Додаткові змінні, що вводяться, мають цілком певний економічний зміст. Так, якщо в обмеженнях початкової задачі лінійного програмування відображається витрата і наявність виробничих ресурсів, то числове значення додаткової змінної в плані задачі, записаної у формі основної, дорівнює об'єму невживаного відповідного ресурсу.

Відзначимо, нарешті, що якщо змінна x_k не підлягає умові невід'ємності, то її слід замінити двома невід'ємними змінними u_k і v_k , прийнявши $x_k = u_k - v_k$.

Приклад. Записати у формі основної задачі лінійного програмування наступну задачу: знайти максимум функції $F = 3x_1 - 2x_2 - 5x_4 + x_5$ за умов

$$\begin{cases} 2x_1 + x_3 - x_4 + x_5 \leq 2, \\ x_1 - x_3 + 2x_4 + x_5 \leq 3, \\ 2x_2 + x_3 - x_4 + 2x_5 \leq 6, \\ x_1 + x_4 - 5x_5 \geq 8, \\ x_1, x_2, x_3, x_4, x_5 \geq 0. \end{cases}$$

Розв'язання. У даній задачі вимагається знайти максимум функції, а система обмежень містить чотири нерівності. Отже, щоб записати її у формі основної задачі, потрібно перейти від обмежень-нерівностей до обмежень-рівностей. Оскільки число нерівностей, що входять у систему обмежень задачі, дорівнює чотирьом, то цей перехід може бути здійснений введенням чотирьох додаткових невід'ємних змінних. При цьому до лівих частин кожної з нерівностей типу « $\leq$ » відповідна додаткова змінна додається, а від лівих частин кожної з нерівностей типу « $\geq$ » віднімається. У результаті обмеження набувають виду рівнянь:

$$\begin{cases} 2x_1 + x_3 - x_4 + x_5 + x_6 = 2, \\ x_1 - x_3 + 2x_4 + x_5 + x_7 = 3, \\ 2x_2 + x_3 - x_4 + 2x_5 + x_8 = 6, \\ x_1 + x_4 - 5x_5 - x_9 = 8, \\ x_1, x_2, \dots, x_9 \geq 0. \end{cases}$$

Отже, дана задача може бути записана у формі основної задачі таким чином: максимізувати функцію $F = 3x_1 - 2x_2 - 5x_4 + x_5$ за умов

$$\begin{cases} 2x_1 + x_3 - x_4 + x_5 + x_6 = 2, \\ x_1 - x_3 + 2x_4 + x_5 + x_7 = 3, \\ 2x_2 + x_3 - x_4 + 2x_5 + x_8 = 6, \\ x_1 + x_4 - 5x_5 - x_9 = 8, \end{cases}$$

$$x_1, x_2, \dots, x_9 \geq 0.$$

Приклад. Записати задачу, що полягає в мінімізації функції

$$F = -x_1 + 2x_2 - x_3 + x_4$$

за умов

$$\begin{cases} 2x_1 - x_2 - x_3 + x_4 \leq 6, \\ x_1 + 2x_2 + x_3 - x_4 \geq 8, \\ 3x_1 - x_2 + 2x_3 + 2x_4 \leq 10, \\ -x_1 + 3x_2 + 5x_3 - 3x_4 = 15, \\ x_1, x_2, x_3, x_4 \geq 0 \end{cases}$$

у формі основної задачі лінійного програмування.

Розв'язання. У даній задачі вимагається знайти мінімум цільової функції, а система обмежень містить три нерівності. Отже, щоб записати її у формі основної задачі, замість знаходження мінімуму функції F потрібно знайти максимум функції $F_1 = -F$ при обмеженнях, що виходять з обмежень початкової задачі, додаванням до лівих частин кожного від обмежень-нерівностей типу « $\leq$ » додаткової невід'ємної змінної і відніманням додаткових змінних від лівих частин кожного з обмежень-нерівностей типу « $\geq$ ».

Отже, початкова задача може бути записана у формі основної задачі лінійного програмування так: знайти максимум функції $F_1 = x_1 - 2x_2 + x_3 - x_4$ за умов

$$\begin{cases} 2x_1 - x_2 - x_3 + x_4 + x_5 = 6, \\ x_1 + 2x_2 + x_3 - x_4 - x_6 = 8, \\ 3x_1 - x_2 + 2x_3 + 2x_4 + x_7 = 10, \\ -x_1 + 3x_2 + 5x_3 - 3x_4 = 15, \\ x_1, x_2, \dots, x_7 \geq 0. \end{cases}$$

Приклад. Записати у формі стандартної задачі лінійного програмування наступну задачу: знайти максимум функції $F = 6,5x_1 - 7,5x_3 + 23,5x_4 - 5x_5$ за умов

$$\begin{cases} x_1 + 3x_2 + x_3 + 4x_4 - x_5 = 12, \\ 2x_1 - x_3 + 12x_4 - x_5 = 14, \\ x_1 + 2x_2 + 3x_4 - x_5 = 6, \\ x_1, x_2, \dots, x_5 \geq 0 \end{cases}$$

Розв'язання. Методом послідовного виключення невідомих зведемо дану задачу до наступної: знайти максимум функції $F = 6,5x_1 - 7,5x_3 + 23,5x_4 - 5x_5$ за умов

$$\begin{cases} x_2 + x_3 + x_4 = 6, \\ 3x_3 + 10x_4 + x_5 = 26, \\ x_1 + x_3 + 11x_4 = 20, \end{cases}$$

$$x_1, x_2, \dots, x_5 \geq 0.$$

Остання задача записана у формі основної для задачі, що полягає у знаходженні максимального значення функції $F = x_3 + 2x_4$ за умов

$$\begin{cases} x_3 + x_4 \leq 6, \\ 3x_3 + 10x_4 \leq 26, \\ x_3 + 11x_4 \leq 20, \end{cases}$$

$$x_3, x_4 \geq 0.$$

Цільова функція задачі перетворена за допомогою підстановки замість x_1 і x_5 їх значень відповідно до рівнянь системи обмежень задачі.

Лекція 9. Графічна інтерпретація задачі лінійного програмування

Розглянемо основну задачу лінійного програмування. Вона полягає у визначенні максимального значення функції $F = \sum_{j=1}^n c_j x_j$ за умов $\sum_{j=1}^n a_{ij} x_j = b_i$ ($i = 1, m$), $x_j \geq 0$ ($j = 1, n$).

Перепишемо цю задачу у векторній формі: знайти максимум функції

$$F = CX \quad (1)$$

за умов

$$x_1 P_1 + x_2 P_2 + \dots + x_n P_n = P_0, \quad (2)$$

$$X \geq 0, \quad (3)$$

де $C = (c_1; c_2; \dots; c_n)$, $X = (x_1; x_2; \dots; x_n)$; CX – скалярний добуток; $P_1, \dots, P_n$, і P_0 – m -вимірні вектор-стовпці, складені з коефіцієнтів при невідомих і вільних членах системи рівнянь задачі:

$$P_0 = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}; \quad P_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}; \quad P_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}; \quad \dots; \quad P_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

План $X = (x_1; x_2; \dots; x_n)$ називається *опорним планом* основної задачі лінійного програмування, якщо система векторів P_j , що входять у розклад (2) з додатними коефіцієнтами x_j , лінійно незалежна.

Оскільки вектори P_j є m -вимірними, то з визначення опорного плану випливає, що число його позитивних компонент не може бути більше, ніж m .

Опорний план називається *невиродженим*, якщо він містить рівно m позитивних компонент, у протилежному випадку він називається *виродженим*.

Властивості основної задачі лінійного програмування (1) – (3) тісно пов'язані з властивостями опуклих множин.

Нехай $X_1, X_2, \dots, X_n$ – довільні точки евклідового простору E_n . *Опуклою лінійною комбінацією* цих точок називається сума $\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n$, де α_i — довільні невід'ємні числа, сума яких дорівнює 1:

$$\sum_{i=1}^n \alpha_i = 1, \quad \alpha_i \geq 0 \quad (i = \overline{1, n}).$$

Множина називається *опуклою*, якщо разом із будь-якими двома своїми точками вона містить і їх довільну опуклу лінійну комбінацію.

Точка X опуклої множини називається *кутовою*, якщо вона не може бути представлена у вигляді опуклої лінійної комбінації яких-небудь двох інших різних точок даної множини.

Теорема 1. Множина планів основної задачі лінійного програмування є опуклою (якщо вона не порожня).

Непорожня множина планів основної задачі лінійного програмування називається *багатогранником розв'язків*, а всяка кутова точка багатогранника розв'язків — *вершиною*.

Теорема 2. Якщо основна задача лінійного програмування має оптимальний план, то максимального значення цільова функція задачі набуває в одній з вершин багатогранника розв'язків. Якщо максимального значення цільова функція задачі набуває більш ніж в одній вершині, то вона набуває його у будь-якій точці, що є опуклою лінійною комбінацією цих вершин.

Теорема 3. Якщо система векторів $P_1, P_2, \dots, P_k$ ($k \leq n$) в розкладі (2) лінійно незалежна і така, що

$$x_1 P_1 + x_2 P_2 + \dots + x_k P_k = P_0, \quad (4)$$

де всі $x_j \geq 0$, то точка $X = (x_1; x_2; \dots; x_k, 0, \dots, 0)$ є вершиною багатогранника розв'язків.

Теорема 4. Якщо $X = (x_1; x_2; \dots; x_n)$ — вершина багатогранника розв'язків, то вектори P_j , що відповідають додатним x_j в розкладі (2), лінійно незалежні.

Сформульовані теореми дозволяють зробити такі висновки.

Непорожня множина планів основної задачі лінійного програмування утворює опуклий багатогранник. Кожна вершина цього багатогранника визначає опорний план. В одній із вершин багатогранника розв'язків (тобто для одного з опорних планів) значення цільової функції є максимальним (за умови, що функція обмежена зверху на множині планів). Якщо максимального значення функція набуває більш ніж в одній вершині, то цього ж значення вона набуває в будь-якій точці, яка є опуклою лінійною комбінацією даних вершин.

Вершину багатогранника розв'язків, у якій цільова функція набуває максимального значення, знайти порівняно просто, якщо задача, записана у формі стандартної, містить не більше двох змінних, або якщо задача, записана у формі основної, містить не більше двох вільних змінних.

Знайдемо розв'язок задачі, що полягає у визначенні максимального значення функції

$$F = c_1 x_1 + c_2 x_2 \quad (5)$$

за умов

$$a_{i1} x_1 + a_{i2} x_2 \leq b_i \quad (i = \overline{1, \kappa}), \quad (6)$$

$$x_j \geq 0 \quad (j = 1, 2). \quad (7)$$

Кожна з нерівностей (6), (7) системи обмежень задачі геометрично визначає напівплощину відповідно з граничними прямими $a_{i1} x_1 + a_{i2} x_2 \leq b_i$

($i = \overline{1, k}$) $x_1 = 0$ і $x_2 = 0$. У тому випадку, якщо система нерівностей (6), (7) сумісна, областю її розв'язків є множина точок, що належать усім вказаним напівплощинам. Оскільки множина точок перетину даних напівплощин опукла, то областю допустимих розв'язків задачі (5) – (7) є опукла множина, яка називається багатокутником розв'язків (введений раніше термін „багатогранник розв'язків” звичайно використовується, якщо $n \geq 3$). Сторони цього багатокутника лежать на прямих, рівняння яких отримуються з початкової системи обмежень заміною знаків нерівностей на знаки точної рівності.

Таким чином, початкова задача лінійного програмування полягає в знаходженні такої точки багатокутника розв'язків, в якій цільова функція F набуває максимального значення. Ця точка існує тоді, коли багатокутник розв'язків не порожній і на ньому цільова функція обмежена зверху. За вказаних умов в одній із вершин багатокутника розв'язків цільова функція набуває максимального значення. Для визначення даної вершини побудуємо лінію рівня $c_1x_1 + c_2x_2 = h$ (де h — деяка стала), яка проходить через багатокутник розв'язків, і пересуватимемо її у напрямі вектора $\vec{C} = (c_1; c_2)$ доти, поки вона не пройде через останню її спільну точку з багатокутником розв'язків. Координати вказаної точки і визначають оптимальний план даної задачі.

Закінчуючи розгляд геометричної інтерпретації задачі (5) – (7), відзначимо, що при знаходженні її розв'язку можуть зустрітися випадки, зображені на рис. 9.1–9.4. Рис. 9.1 характеризує такий випадок, коли цільова функція набуває максимального значення в єдиній точці A . Із рис. 9.2 видно, що максимального значення цільова функція набуває в будь-якій точці відрізка AB . На рис. 9.3 зображений випадок, коли цільова функція не обмежена зверху на множині допустимих розв'язків, а на рис. 9.4 — випадок, коли система обмежень задачі несумісна.

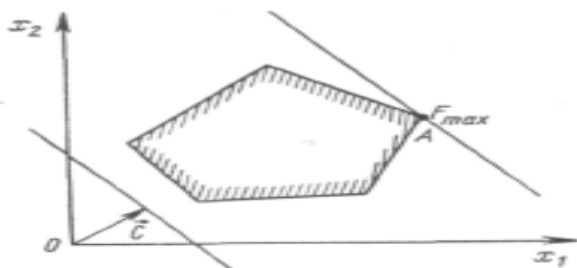


Рис. 1.1

Рис.9.1

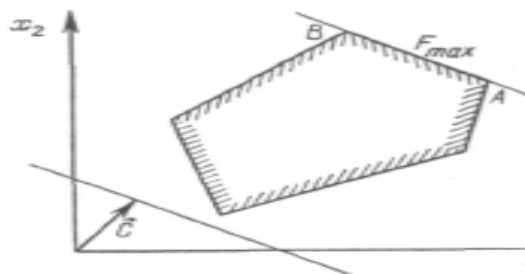


Рис. 1.2

Рис.9.2

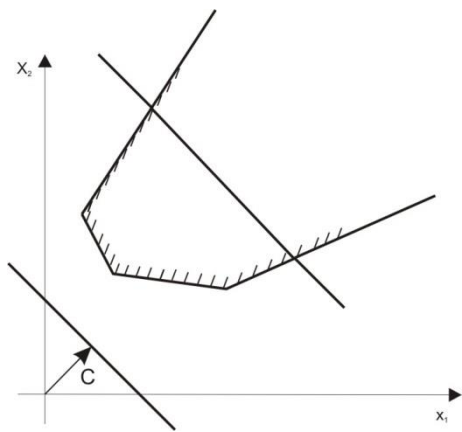


Рис. 9.3

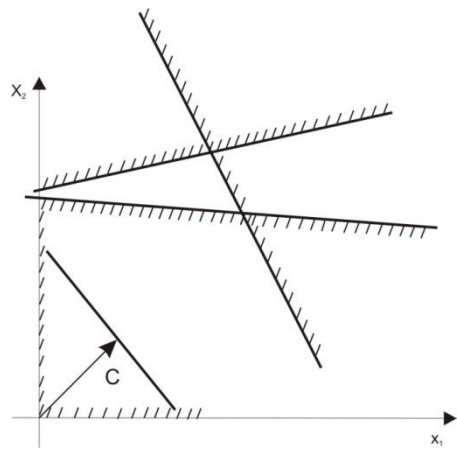


Рис. 9.4

Відзначимо, що знаходження мінімального значення лінійної функції при даній системі обмежень відрізняється від знаходження її максимального значення за тих же обмежень лише тим, що лінія рівня $c_1x_1 + c_2x_2 = h$ пересувається не у напрямі вектора $\vec{C} = (c_1; c_2)$, а в протилежному напрямі. Таким чином, відзначені вище випадки, що зустрічаються при знаходженні максимального значення цільової функції, мають місце і при визначенні її мінімального значення.

Отже, знаходження розв'язку задачі лінійного програмування (5) – (7) на основі її геометричної інтерпретації включає наступні етапи:

1. Будують прямі, рівняння яких отримуємо в результаті заміни в обмеженнях (6) і (7) знаків нерівностей на знаки точної рівності.
2. Знаходять напівплощини, що визначаються кожним з обмежень задачі.
3. Знаходять багатокутник розв'язків.
4. Будують вектор $\vec{C} = (c_1; c_2)$.
5. Будують пряму $c_1x_1 + c_2x_2 = h$, яка проходить через багатокутник розв'язків.

6. Рухають лінію рівня $c_1x_1 + c_2x_2 = h$ у напрямі вектора $\vec{C}$. Остання спільна точка (точки) лінії рівня і багатокутника розв'язків і є точкою, в якій цільова функція набуває максимального значення. Або встановлюють необмеженість зверху функції на множині планів.

7. Визначають координати точки максимуму функції і обчислюють значення цільової функції в цій точці.

Приклад. Для виробництва двох видів виробів А і В підприємство використовує три види сировини. Норми витрат сировини кожного виду на виготовлення одиниці продукції даного виду наведені в таблиці 9.3. У ній же

вказаний прибуток від реалізації одного виробу кожного виду і загальна кількість сировини даного виду, яка може бути використана підприємством.

Таблиця 9.3

Вид сировини	Норми витрат сировини (кг) на один виріб		Загальна кількість сировини (кг)
	<i>A</i>	<i>B</i>	
I	12	4	300
II	4	4	120
III	3	12	252
Прибуток від реалізації одного виробу (грн.)	30	40	

Враховуючи, що вироби *A* і *B* можуть виготовлятися в будь-яких співвідношеннях (збут забезпечений), вимагається встановити такий план їх випуску, при якому прибуток підприємства від реалізації всіх виробів буде максимальним.

Розв'язання. Припустимо, що підприємство виготовить x_1 виробів виду *A* і x_2 виробів виду *B*. Оскільки виробництво продукції обмежене кількістю сировини кожного виду і кількість виробів, що виготовляється, не може бути від'ємною, то повинні виконуватися нерівності:

$$\begin{cases} 12x_1 + 4x_2 \leq 300, \\ 4x_1 + 4x_2 \leq 120, \\ 3x_1 + 12x_2 \leq 252, \\ x_1, x_2 \geq 0. \end{cases}$$

Загальний прибуток від реалізації x_1 виробів виду *A* і x_2 виробів виду *B* складе $F = 30x_1 + 40x_2$.

Таким чином, ми приходимо до наступної математичної задачі: серед усіх невід'ємних розв'язків даної системи лінійних нерівностей вимагається знайти такий, при якому функція F набуває максимального значення.

Знайдемо розв'язок сформульованої задачі, використовуючи її геометричну інтерпретацію. Спочатку визначимо багатокутник розв'язків. Для цього в нерівностях системи обмежень і в умовах невід'ємності змінних знаки нерівностей замінимо на знаки точної рівності і побудуємо відповідні прямі:

$$\begin{cases} 12x_1 + 4x_2 = 300, & (I) \\ 4x_1 + 4x_2 = 120, & (II) \\ 3x_1 + 12x_2 = 252, & (III) \\ x_1 = 0, & (IV) \\ x_2 = 0. & (V) \end{cases}$$

Ці прямі зображені на рис. 9.5. Кожна з побудованих прямих ділить площину на дві півплощини. Координати точок однієї півплощини задовольняють початковій нерівності, а іншої — ні. Щоб визначити шукану півплощину, потрібно взяти яку-небудь точку, що належить одній із півплощин, і перевірити, чи задовольняють її координати даній нерівності. Якщо координати взятої точки задовольняють даній нерівності, то шуканою є та напівплощина, якій належить ця точка, в протилежному випадку — інша півплощина.

Знайдемо, наприклад, півплощину, що визначається нерівністю $12x_1 + 4x_2 < 300$. Для цього, побудувавши пряму $12x_1 + 4x_2 = 300$ (на рис. 9.5 ця пряма I), візьмемо яку-небудь точку, що належить одній із двох отриманих півплощин, наприклад, точку $O = (0; 0)$. Координати цієї точки задовольняють нерівність $12 \cdot 0 + 4 \cdot 0 < 300$. Отже, півплощина, якій належить точка $O = (0; 0)$, визначається нерівністю $12x_1 + 4x_2 \leq 300$. Це і показано стрілками на рис. 9.5.

Перетин отриманих півплощин і визначає багатокутник розв'язків даної задачі.

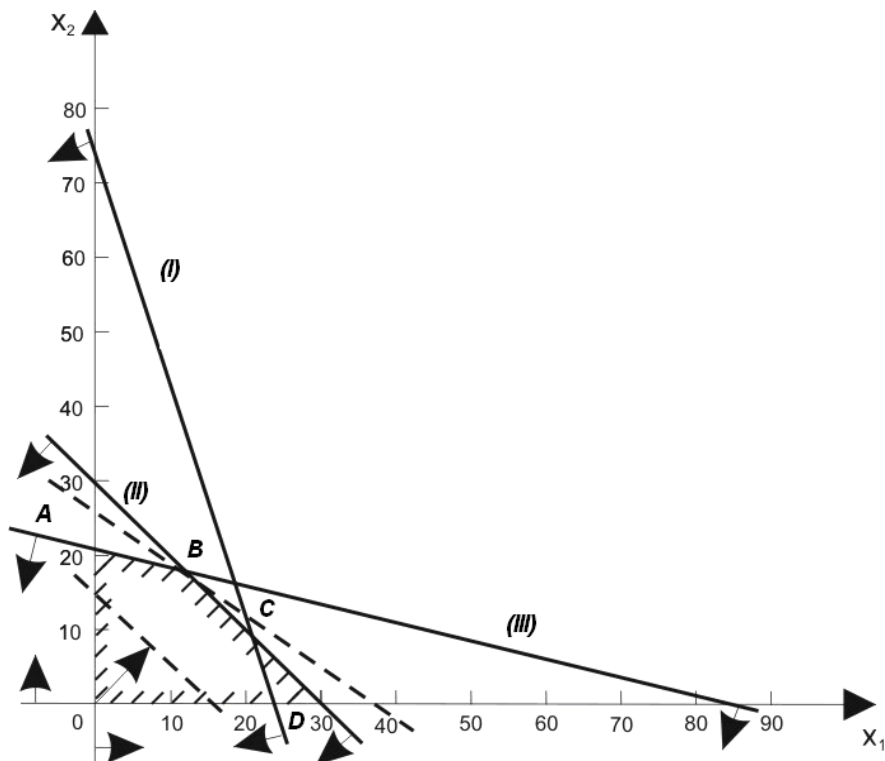


Рис. 9.5

Як видно з рис. 9.5, багатокутником розв'язків є п'ятикутник $OABCD$. Координати будь-якої точки, що належить цьому п'ятикутнику, задовольняють даній системі нерівностей і умові невід'ємності змінних. Тому сформульована задача буде розв'язана, якщо ми зможемо знайти точку, що належить п'ятикутнику $OABCD$, в якій функція F набуває максимального значення. Щоб знайти вказану точку, побудуємо вектор $\vec{C} = (30; 40)$ і пряму $30 \cdot 1 + 40 \cdot 2 = h$, де h – деяка стала така, що пряма $30 \cdot 1 + 40 \cdot 2 = h$ має спільні точки з багатокутником розв'язків. Покладемо, наприклад, $h = 480$ і побудуємо пряму $30x_1 + 40x_2 = 480$ (рис. 9.5).

Якщо тепер взяти яку-небудь точку, що належить побудованій прямій і багатокутнику розв'язків, то її координати визначають такий план виробництва виробів A і B , при якому прибуток від їх реалізації дорівнює 480 грн. Далі, вважаючи h рівним деякому числу, більшому, ніж 480, ми одержуватимемо різні паралельні прямі. Якщо вони мають спільні точки з багатокутником розв'язків, то ці точки визначають плани виробництва виробів A і B , при яких прибуток від їх реалізації перевершить 480 грн.

Рухаючи побудовану пряму $30x_1 + 40x_2 = 480$ у напрямі вектора $\vec{C}$, бачимо, що останньою спільною точкою її з багатокутником розв'язків задачі слугує точка B . Координати цієї точки і визначають план випуску виробів A і B , при якому прибуток від їх реалізації є максимальним.

Знайдемо координати точки B як точки перетину прямих (II) і (III). Отже, її координати задовольняють рівнянням цих прямих

$$\begin{cases} 4x_1 + 4x_2 = 120, \\ 3x_1 + 12x_2 = 252. \end{cases}$$

Розв'язавши цю систему рівнянь, отримаємо $x_1^* = 12$, $x_2^* = 18$. Отже, якщо підприємство виготовить 12 виробів виду A і 18 виробів виду B , то воно отримає максимальний прибуток, рівний $F_{\max} = 30 \cdot 12 + 40 \cdot 18 = 1080$ (грн.).

Приклад. Знайти максимум і мінімум функції $F = x_1 + x_2$ за умов

$$\begin{cases} 2x_1 + 4x_2 \leq 16, \\ -4x_1 + 2x_2 \leq 8, \\ x_1 + 3x_2 \geq 9, \\ x_1, x_2 \geq 0. \end{cases}$$

Розв'язання. Побудуємо багатокутник розв'язків (рис. 9.6).

Як видно з рис. 9.6, багатокутником розв'язків є трикутник ABC . Координати точок цього трикутника задовольняють умові невід'ємності і нерівностям системи обмежень задачі. Отже, задача буде розв'язана, якщо серед точок трикутника ABC знайти такі, в яких функція $F = x_1 + x_2$ набуває

максимального і мінімального значення. Для знаходження цієї точки побудуємо пряму $x_1 + x_2 = 4$ і вектор $\vec{C} = (1;1)$.

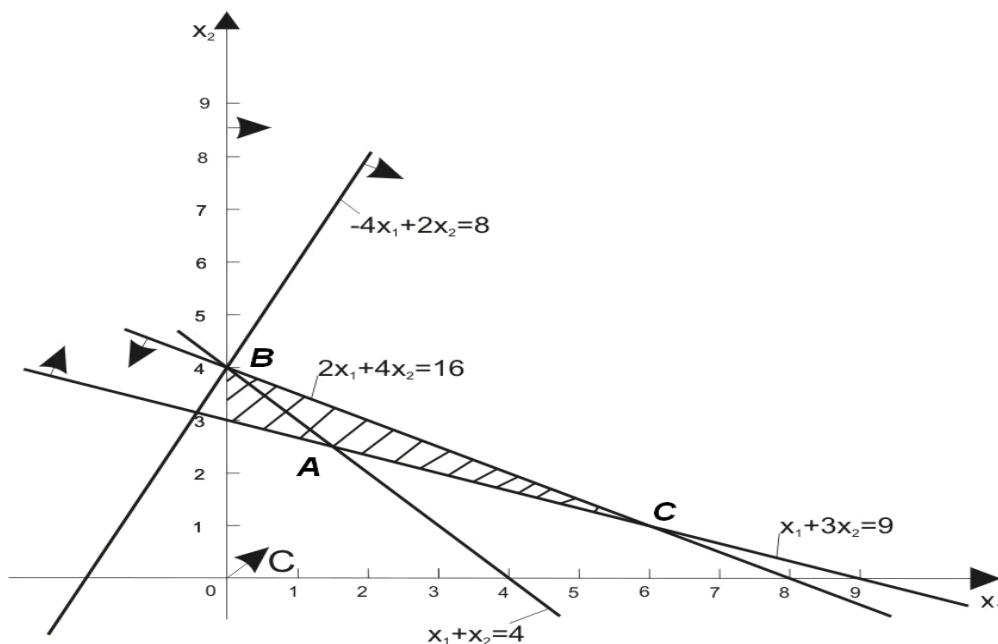


Рис. 9.6

Рухаючи дану пряму паралельно самій собі у напрямі вектора $\vec{C}$, бачимо, що її останньою спільною точкою з багатокутником розв'язків задачі є точка C . Отже, в цій точці функція F матиме максимальне значення. Оскільки C — точка перетину прямих (I) і (II), то її координати задовольняють рівнянням цих прямих:

$$\begin{cases} 2x_1 + 4x_2 = 16, \\ x_1 + 3x_2 = 9. \end{cases}$$

Розв'язавши цю систему рівнянь, отримаємо $x_1^* = 6$, $x_2^* = 1$. Таким чином, максимальне значення функції $F_{\max} = 7$.

Для знаходження мінімального значення цільової функції задачі рухаємо пряму $x_1 + x_2 = 4$ в напрямку, протилежному напрямку вектора $\vec{C} = (1;1)$. Для визначення координат точки A розв'язуємо систему рівнянь

$$\begin{cases} x_1 + 3x_2 = 9, \\ x_1 = 0, \end{cases}$$

звідки $x_1^* = 0$, $x_2^* = 3$. Підставляючи знайдені значення змінних у цільову функцію, отримаємо $F_{\min} = 3$.

Приклад. Знайти максимальне значення функції

$$F = -16x_1 - x_2 + x_3 + 5x_4 + 5x_5$$

за умов

$$\begin{cases} 2x_1 + x_2 + x_3 = 10, \\ -2x_1 + 3x_2 + x_4 = 6, \\ 2x_1 + 4x_2 - x_5 = 8, \\ x_1, x_2, x_3, x_4, x_5 \geq 0. \end{cases}$$

Розв'язання. На відміну від розглянутих вище задач, у даній задачі обмеження задані у вигляді рівнянь. При цьому число невідомих дорівнює п'яти. Тому дану задачу потрібно звести до задачі, в якій число невідомих дорівнювало б двом. Для цього перейдемо від запису заданої задачі у формі основної до задачі, записаної у стандартній формі:

$$F = 2x_1 + 3x_2 \rightarrow \max$$

за умов

$$\begin{cases} 2x_1 + x_2 \leq 10, \\ -2x_1 + 3x_2 \leq 6, \\ 2x_1 + 4x_2 \geq 8, \\ x_j \geq 0, j = \overline{1,2}. \end{cases}$$

Із цільової функції початкової задачі змінні x_3, x_4, x_5 виключені за допомогою підстановки їх значень із відповідних рівнянь системи обмежень.

Побудуємо багатокутник розв'язків отриманої задачі (рис. 9.7).

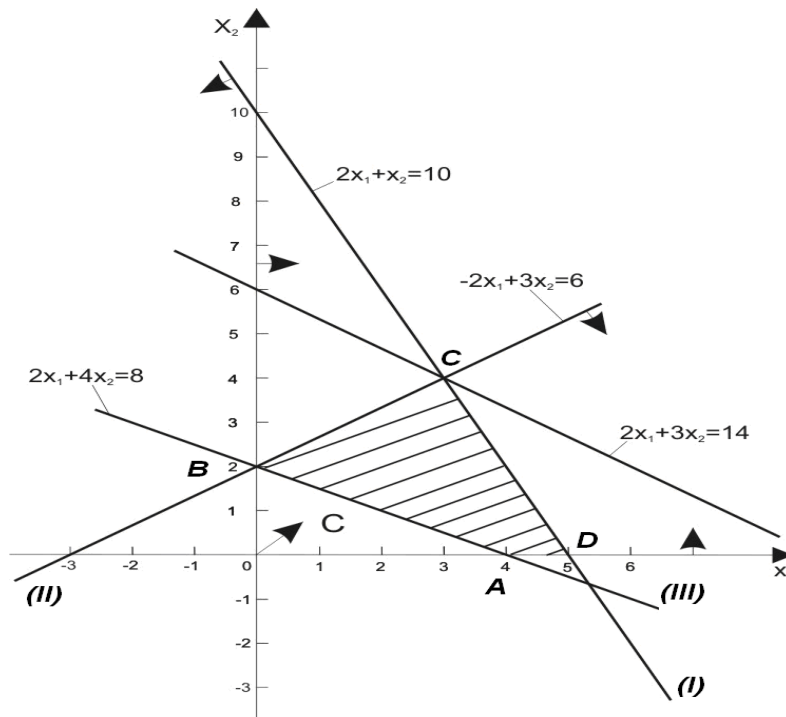


Рис. 9.7

Як видно з рис. 9.7, максимального значення цільова функція задачі набуває в точці C перетину прямих I і II . Тому в кожній із вершин отриманого багатокутника розв'язків останньої задачі принаймні дві змінні початкової задачі набувають нульових значень. Так, у точці C маємо $x_3 = 0$, $x_4 = 0$. Підставляючи ці значення в перше і друге рівняння системи обмежень початкової задачі, отримаємо $x_1^* = 3$, $x_2^* = 4$. Підставляючи знайдені значення x_1 і x_2 у третє рівняння системи обмежень початкової задачі, знайдемо $x_5 = 14$.

Отже, оптимальним планом заданої задачі буде $X^* = (3; 4; 0; 0; 14)$. При цьому плані $F_{\max} = 18$.

Лекція 10. Ставлення і розв'язування транспортної задачі та її використання в агрономії

Постановка транспортної задачі, умова існування її розв'язку

Загальна постановка транспортної задачі полягає в знаходженні оптимального плану перевезень певного однорідного вантажу із m пунктів відправлення $A_1, A_2, \dots, A_m$ в n пунктів призначення $B_1, B_2, \dots, B_n$. При цьому критерієм оптимальності є або мінімальна вартість перевезень усього вантажу, або мінімальний час його доставки.

Позначимо через c_{ij} тарифи перевезення одиниці вантажу із i -го пункту відправлення до j -го пункту призначення, через a_i - запаси вантажу в i -му пункті відправлення, через b_j - потреби в вантажі у j -му пункті призначення, а через x_{ij} - кількість одиниць вантажу, що перевозиться із i -го пункту відправлення до j -го пункту призначення. Тоді математична постановка задачі полягає у визначенні мінімального значення функції

$$F = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \quad (10.1)$$

за умов

$$\sum_{i=1}^m x_{ij} = b_j \quad (j=1, \dots, n), \quad (10.2)$$

$$\sum_{j=1}^n x_{ij} = a_i \quad (i=1, \dots, m), \quad (10.3)$$

$$x_{ij} \geq 0 \quad (i=1, \dots, m; \quad j=1, \dots, n) \quad (10.4)$$

Кожне не невід'ємне рішення системи лінійних рівнянь (10.2) та (10.3), що визначається матрицею $X = (x_{ij}) \quad (i=1, \dots, m; \quad j=1, \dots, n)$, називається **планом** транспортної задачі.

План $X^* = (x_{ij}^*) \quad (i=1, \dots, m; \quad j=1, \dots, n)$, за якого функція (10.1) набуває свого мінімального значення, називається **оптимальним планом** транспортної задачі.

Звичайно вихідні дані транспортної задачі записують у вигляді таблиці (див. табл. 10.1).

Таблиця 10.1

Пункти відправлення	Пункти призначення					Запаси
	B_1	...	B_j	...	B_n	
A_1	c_{11} x_{11}	...	c_{1j} x_{1j}	...	c_{1n} x_{1n}	a_1
...	...	...	...	...	...	...
A_i	c_{i1} x_{i1}	...	c_{ij} x_{ij}	...	c_{in} x_{in}	a_i
...	...	...	...	...	...	...
A_m	c_{m1} x_{m1}	...	c_{mj} x_{mj}	...	c_{mn} x_{mn}	a_m
Потреби	b_1	...	b_j	...	b_n	

Якщо загальні потреби у вантажі в пунктах призначення дорівнюють запасам вантажу в пунктах відправлення

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j, \quad (10.5)$$

то модель транспортної задачі називається **закритою**. Якщо означена умова не виконується, то модель називається **відкритою**.

Теорема 10.1. Для розв'язності транспортної задачі необхідно й досить, щоб запаси вантажу в пунктах відправлення дорівнювали потребам у вантажі в пунктах призначення (що теж саме – виконувалось рівняння (10.5)).

У разі перевищення запасів над потребами $\sum_{i=1}^m a_i > \sum_{j=1}^n b_j$ вводиться фіктивний $(n+1)$ -й пункт призначення з потребами $b_{n+1} = \sum_{i=1}^m a_i - \sum_{j=1}^n b_j$, та відповідні тарифи вважаються рівними нулю: $c_{in+1} = 0$ ($i = 1, \dots, m$). Тепер отримана задача є транспортною, для якої виконується рівність (10.5).

Аналогічно при $\sum_{i=1}^m a_i < \sum_{j=1}^n b_j$ вводиться фіктивний $(m+1)$ -й пункт

відправлення з запасом вантажу $a_{m+1} = \sum_{j=1}^n b_j - \sum_{i=1}^m a_i$, і тарифи вважаються рівними нулю: $c_{m+1j} = 0$ ($j = 1, \dots, n$). Цим задача зводиться до стандартної транспортної задачі.

Число змінних x_{ij} в транспортній задачі дорівнює $m \cdot n$, а число рівнянь у системах (10.2) і (10.3) дорівнює $n + m$. Так як ми вважаємо, що виконується умова (10.5), то число лінійно незалежних рівнянь дорівнює $n + m - 1$. Отже, опорний план транспортної задачі може мати не більш $n + m - 1$ ненульових невідомих.

Якщо в опорному плані число відмінних від нуля компонент дорівнює в точності $n + m - 1$, то план є **не виродженим**, а якщо менше – то **виродженим**.

10.1. Побудувати економіко математичну модель наступної транспортної задачі. Є три постачальники і чотири споживачі. Потужність постачальників і потреби споживачів, а також витрати на перевезення одиниці вантажу для кожної пари “постачальник-споживач” зведені в табл. 10.2.

Таблиця 10.2

Постачальники	Запаси	Споживачі та їх потреби			
		1	2	3	4
		20	110	40	110
1	60	1 x_{11}	2 x_{12}	5 x_{13}	3 x_{14}
2	120	1 x_{21}	6 x_{22}	5 x_{23}	2 x_{24}
3	100	6 x_{31}	3 x_{32}	7 x_{33}	4 x_{34}

У лівому верхньому кутку клітини записані коефіцієнти витрат, x_{ij} – обсяг перевезень від i -го постачальника до j -го споживача.

Треба знайти обсяги перевезень для кожної пари “постачальник-споживач” таким чином, щоб:

1. потужності всіх постачальників були реалізовані;
2. попит усіх споживачів - задоволений;
3. сумарні затрати на всі перевезення - мінімальними.

Це означає, що система обмежень має вигляд:

- у кожному рядку

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} = 60, \\ x_{21} + x_{22} + x_{23} + x_{24} = 120, \\ x_{31} + x_{32} + x_{33} + x_{34} = 100, \end{cases} \quad (10.6)$$

- у кожному стовпчику

$$\begin{cases} x_{11} + x_{21} + x_{31} = 20, \\ x_{12} + x_{22} + x_{32} = 110, \\ x_{13} + x_{23} + x_{33} = 40, \\ x_{14} + x_{24} + x_{34} = 110, \end{cases} \quad (10.7)$$

а цільова функція

$$\begin{aligned} F = & 1 \cdot x_{11} + 2 \cdot x_{12} + 5 \cdot x_{13} + 3 \cdot x_{14} + \\ & + 1 \cdot x_{21} + 6 \cdot x_{22} + 5 \cdot x_{23} + 2 \cdot x_{24} + \\ & + 6 \cdot x_{31} + 3 \cdot x_{32} + 7 \cdot x_{33} + 4 \cdot x_{34} \rightarrow \min. \end{aligned} \quad (10.8)$$

Пошук оптимального опорного плану перевезень за методом потенціалів

Знаходження **оптимального плану** транспортної задачі починають з знаходження якого-небудь її **опорного плану**. Опорний план можна знайти:

- методом північно-західного кута;
- методом мінімального елемента;
- методом апроксимації Фогеля.

Сутність цих методів полягає в тому, що опорний план знаходять послідовно за $n + m - 1$ кроків, на кожному з яких в таблиці умов задачі заповнюють одну клітину, яку називають **зайнятою**. Заповнення однієї з клітинок забезпечує повністю або задоволення потреб у вантажі одного з пунктів призначення, або вивіз вантажу із одного з пунктів відправлення.

Метод північно-західного кута. Заповнення клітинок таблиці умов задачі починається з лівої верхньої клітинки для невідомого x_{11} („північно-західний кут”). Далі заповнюють перший із пунктів відправлення, що залишився, та перший із пунктів призначення, що залишився. Закінчується клітинкою для невідомого x_{mn} - заповнення йде ніби по діагоналі таблиці.

10.2. Задача. Умови задачі зведені в табл. 10.3.

Таблиця 10.3.

Пункти відправлення	Пункти призначення					Запаси
	B_1	B_2	B_3	B_4	B_5	
A_1	2	3	4	2	4	140
A_2	8	4	1	4	1	180
A_3	9	7	3	7	2	160
Потреби	60	70	120	130	100	480

Знайти опорний план перевезень даної транспортної задачі методом північно-західного кута.

Розв'язок. Кількість пунктів відправлення $m = 3$, а кількість пунктів призначення $n = 5$. Таким чином, опорний план задачі визначається числами, що стоять у $5 + 3 - 1 = 7$ заповнених клітинках.

Заповнення таблиці починаємо з клітинки (A_1, B_1) для невідомого x_{11} . Дано змінній x_{11} максимально можливе значення, тобто максимально можливу поставку в клітинку (A_1, B_1) – “північно-західний” кут таблиці поставок: $x_{11} = \min(60, 140) = 60$. Після цього попит першого споживача буде задоволений, та перший стовпчик таблиці випаде із подальшого розгляду. У таблиці поставок знайдемо новий “північно-західний” кут – клітинку (A_1, B_2) і дамо їй максимально можливу поставку. Враховуючи, що перший постачальник вже віддав 60 одиниць товару і в нього залишилося тільки $140 - 60 = 80$ одиниць товару, що перевищує потреби пункту B_2 , то знаходимо $x_{12} = 70$. Тимчасово виключимо із розгляду стовпець B_2 . У пункті A_1 запаси вважаємо рівними 10 одиницям. Далі розглядаємо клітинку (A_1, B_3) – потреби пункту B_3 більше залишків запасів пункту A_1 . Тому $x_{13} = 10$. Виключимо із розгляду рядок A_1 . Потреби пункту B_3 вважаємо рівними $120 - 10 = 110$ одиницям.

Далі перейдемо до заповнення клітинки для невідомого x_{23} і т.д. Отримані результати заносимо до табл. 10.10.

Таблиця 10.4

Пункти відправлення	Пункти призначення					Запаси
	B_1	B_2	B_3	B_4	B_5	
A_1	2 60	3 70	4 10	2	4	140
A_2	8	4	1 110	4 70	1	180
A_3	9	7	3	7 60	2 100	160
Потреби	60	70	120	130	100	480

Таким чином, отримуємо опорний план:

$$X = \begin{pmatrix} 60 & 70 & 10 & 0 & 0 \\ 0 & 0 & 110 & 70 & 0 \\ 0 & 0 & 0 & 60 & 100 \end{pmatrix}.$$

Згідно з цим опорним планом перевезень, загальна вартість перевезень вантажу складає

$$S = 2 \cdot 60 + 3 \cdot 70 + 4 \cdot 10 + 1 \cdot 110 + 4 \cdot 70 + 7 \cdot 60 + 2 \cdot 100 = 1380 \text{ гр.од.}$$

Метод мінімального елемента. Сутність цього методу полягає у виборі на кожному кроці клітинки з мінімальним тарифом та задоволенні попиту відповідного споживача.

10.3. Задача. Знайти опорний план транспортної задачі 10.2 методом мінімального елемента.

Розв'язок. Вихідні дані задачі записані в табл. 10.2. Мінімальний тариф, рівний 1, знаходиться в клітинках для змінних x_{23} та x_{25} . Порівнюємо максимально можливі поставки для цих клітинок: потреби клітинки (A_2, B_3) - 120 одиниць, потреби (A_2, B_5) - 60 одиниць. Задовольняємо максимально можливий попит - заповнюємо клітинку (A_2, B_3) , тоді $x_{23} = 120$ (табл. 10.5). Виключимо тимчасово із розгляду третій стовпчик. Потужності пункту відправлення A_2 вважаємо рівними $180 - 120 = 60$ одиниць. У рядку A_2 в клітинках, що залишилися, мінімальний тариф, рівний 1, відповідає стовпчику B_5 . Вносимо $x_{25} = 60$ у відповідну клітинку таблиці. Потужності другого постачальника використані повністю. Далі в незаповнених клітинках

мінімальний тариф дорівнює 2 (клітинки (A_1, B_1) , (A_1, B_4) та (A_3, B_5)). Заповнюємо клітинку (A_1, B_1) $x_{11} = 60$. Аналогічно заповнюємо всі клітинки табл. 10.5.

Таблиця 10.5

Пункти відправлення	Пункти призначення					Запаси
	B_1	B_2	B_3	B_4	B_5	
A_1	2 60	3 0	4 0	2 80	4 0	140
A_2	8 0	4 0	1 120	4 0	1 60	180
A_3	9 0	7 70	3 0	7 50	2 40	160
Потреби	60	70	120	130	100	480

Опорний план дорівнює

$$X = \begin{pmatrix} 60 & 0 & 0 & 80 & 0 \\ 0 & 0 & 120 & 0 & 60 \\ 0 & 70 & 0 & 50 & 40 \end{pmatrix}.$$

Загальна вартість перевезень вантажу складає

$$S = 2 \cdot 60 + 2 \cdot 80 + 1 \cdot 120 + 1 \cdot 60 + 7 \cdot 70 + 7 \cdot 50 + 2 \cdot 40 = 1380 \text{ гр.од.}$$

Пошук оптимального плану перевезень за методом потенціалів.

Теорема 10.2. Якщо для будь-якого опорного плану $X^* = (x_{ij}^*)$ ($i = 1, \dots, m$; $j = 1, \dots, n$) транспортної задачі існують такі числа $\alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_n$, що

$$\beta_j + \alpha_i = c_{ij} \quad \text{при} \quad x_{ij} > 0 \quad (10.9)$$

та

$$\beta_j + \alpha_i \leq c_{ij} \quad \text{при} \quad x_{ij} = 0 \quad (10.10)$$

для всіх $i = 1, \dots, m$ та $j = 1, \dots, n$, то $X^* = (x_{ij}^*)$ - оптимальний опорний план транспортної задачі.

Числа α_i та β_j ($i = 1, \dots, m$; $j = 1, \dots, n$) називаються **потенціалами** відповідно пунктів призначення та пунктів відправлення.

Сформульована теорема дозволяє побудувати **алгоритм знаходження рішення транспортної задачі**.

1. Знаходимо опорний план одним із розглянутих вище методів. При цьому число заповнених клітинок повинно дорівнювати $n + m - 1$.

2. Для кожного із пунктів відправлення та призначення знаходимо потенціали α_i та β_j ($i = 1, \dots, m$; $j = 1, \dots, n$) із системи рівнянь

$$\beta_j + \alpha_i = c_{ij} \quad (10.11)$$

де c_{ij} - тарифи, що стоять у заповнених клітинках таблиці умов транспортної задачі.

Оскільки число невідомих перевищує на одиницю число рівнянь, одне із невідомих можна вважати рівним довільному числу, наприклад $\alpha_i = 0$. Далі послідовно із системи (10.11) знаходимо значення всіх невідомих.

3. Для кожної із незаповнених клітинок знаходимо числа

$$\alpha_{ij} = \alpha_i + \beta_j - c_{ij} \quad (10.12)$$

Якщо всі числа $\alpha_{ij} \leq 0$, то опорний план є оптимальним.

Якщо для деякої клітинки $\alpha_{ij} > 0$, то вихідний опорний план не є оптимальним. Треба перейти до нового опорного плану. Для цього серед усіх $\alpha_{ij} > 0$ вибираємо максимальне. Для цієї клітинки будемо цикл перерахунку та проводимо **перехід по циклу**.

Циклом в таблиці умов транспортної задачі називається ламана лінія, вершини якої розташовані в зайнятих клітинках таблиці, а ланки – вздовж рядків та стовпчиків, причому в кожній вершині зустрічаються рівно дві ланки, одна з яких знаходиться в рядку, а друга – в стовпчику.

Приклади деяких циклів наведені на рис. 10.1.

Для кожної вільної клітинки можна побудувати лише один цикл. Після його побудови переходимо до нового опорного плану. Для цього треба перемістити вантажі по клітинках, зв'язаних з даною вільною клітинкою. Це переміщення проводимо згідно з наступними правилами:

- кожній клітинці, зв'язаній циклом, приписуємо певний знак, причому вільній клітинці – знак плюс, а всім іншим – по черзі знаки плюс та мінус (плюсові та мінусові клітинки);

- у дану вільну клітинку переносимо найменше із чисел x_{ij} , що стоять в мінусових клітинках. Одночасно це число додають до відповідних чисел, що знаходяться в мінусових клітинках. Клітинка, що була вільною, стає зайнятою, а мінусова клітинка, в якій стояло мінімальне із чисел x_{ij} , вважається вільною.

Після переміщень вантажу в межах клітинок, зв'язаних циклом з даною вільною клітинкою, знаходимо новий опорний план транспортної задачі.

Описаний вище перехід від одного опорного плану до іншого опорного плану називається **переходом по циклу перерахунку**.

4. Отриманий новий опорний план транспортної задачі перевіряємо на оптимальність. Для цього переходимо до пункту 2.

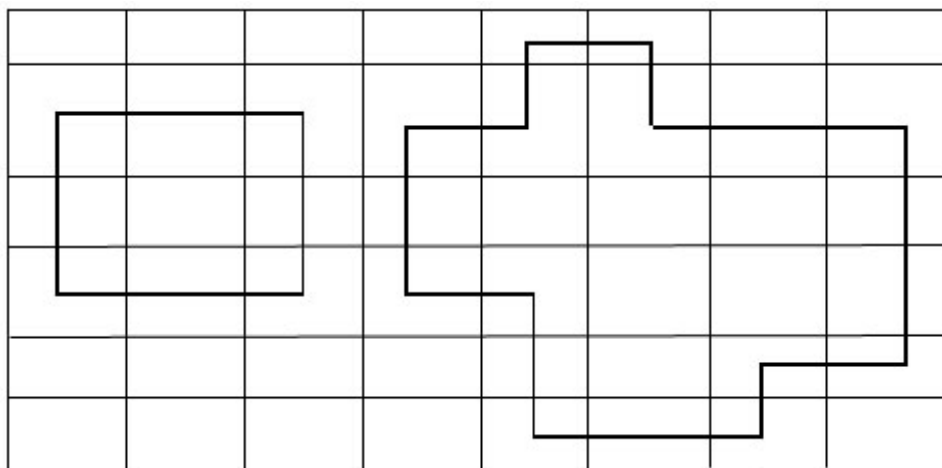


Рис. 10.1. Приклади циклів перерахунку.

10.4. Задача. Знайти оптимальний план транспортної задачі, вихідні дані якої наведені в табл. 10.6.

Таблиця 10.6

Пункти відправлення	Пункти призначення				Запаси
	B_1	B_2	B_3	B_4	
A_1	1 30	2 20	4	1	50
A_2	2	3 10	1 10	5 10	30
A_3	3	2	4	4 10	10
Потреби	30	30	10	20	90

Розв'язок. Методом північно-західного кута знаходимо опорний план (наведений в табл. 10.6).

Перевіряємо його на оптимальність. Знаходимо потенціали пунктів відправлення та призначення із системи

$$\begin{aligned} \beta_1 + \alpha_1 &= 1, \quad \beta_2 + \alpha_1 = 2, \quad \beta_2 + \alpha_2 = 3, \\ \beta_3 + \alpha_2 &= 1, \quad \beta_4 + \alpha_2 = 5, \quad \beta_4 + \alpha_3 = 4. \end{aligned}$$

Вважаємо $\alpha_1 = 0$. Далі знаходимо $\beta_1 = 1$, $\beta_2 = 2$, $\alpha_2 = 1$, $\beta_3 = 0$, $\beta_4 = 4$, $\alpha_3 = 0$. Для кожної вільної клітинки знаходимо числа $\alpha_{ij} = \alpha_i + \beta_j - c_{ij}$:
 $\alpha_{13} = -4$, $\alpha_{14} = 3$, $\alpha_{21} = \alpha_{32} = 0$, $\alpha_{31} = -2$, $\alpha_{33} = -4$.

Заносимо знайдені числа в табл. 10.7.

Таблиця 10.7

Пункти відправлення	Пункти призначення				Запаси
	B_1	B_2	B_3	B_4	
A_1	1 30	2 - 20	4 -4	1 + 3	50
A_2	2 0	3 + 10	1 10	5 - 10	30
A_3	3 -2	2 0	4 -4	4 10	10
Потреби	30	30	10	20	90

Так як серед чисел α_{ij} є додатні, то план не є оптимальним і треба перейти до нового опорного плану. Найбільше серед додатних чисел є $\alpha_{14} = 3$, тому для цієї вільної клітинки будемо цикл перерахунку та робимо перехід по цьому циклу. Найменше із чисел в мінусових клітинках дорівнює 10 (клітинка (A_2, B_4)). Ця клітинка стає вільною в новій табл. 10.8. До плюсових клітинок табл. 10.7 додаємо 10, а з мінусових – віднімаємо 10. Клітинка (A_2, B_4) стає вільною.

Після цих перерахунків отримуємо новий опорний план (табл. 10.8).

Знову перевіряємо план на оптимальність. Складаємо систему рівнянь:

$$\beta_1 - \alpha_1 = 1, \quad \beta_2 - \alpha_1 = 2, \quad \beta_4 - \alpha_1 = 1,$$

$$\beta_2 - \alpha_2 = 3, \quad \beta_3 - \alpha_2 = 1, \quad \beta_4 - \alpha_3 = 4.$$

Вважаємо $\alpha_1 = 0$. Знаходимо $\beta_1 = \beta_4 = 1$, $\beta_2 = 2$, $\beta_3 = 0$, $\alpha_2 = -1$, $\alpha_3 = -3$. Для кожної вільної клітинки знаходимо числа α_{ij} :
 $\alpha_{13} = -2$, $\alpha_{21} = 0$, $\alpha_{24} = -3$, $\alpha_{31} = 1$, $\alpha_{32} = 3$, $\alpha_{33} = -1$.

Заносимо знайдені числа в табл. 10.9.

Таблиця 10.8

Пункти відправлення	Пункти призначення				Запаси
	B_1	B_2	B_3	B_4	
A_1	1 30	2 - 10	4 -4	1 + 10	50
A_2	2 0	3 20	1 10	5 -3	30
A_3	3 1	2 + 3	4 -1	4 - 10	10
Потреби	30	30	10	20	90

Таблиця 10.9.

Пункти відправлення	Пункти призначення				Запаси
	B_1	B_2	B_3	B_4	
A_1	1 30	2 0	4 -4	1 20	50
A_2	2 0	3 20	1 10	5 -3	30
A_3	3 -2	2 10	4 -4	4 -1	10
Потреби	30	30	10	20	90

Так як всі оцінки $\alpha_{ij} < 0$, то отриманий план є оптимальним:

$$X^* = \begin{pmatrix} 30 & 0 & 0 & 20 \\ 0 & 20 & 10 & 0 \\ 0 & 10 & 0 & 0 \end{pmatrix}.$$

Вартість перевезень вантажу складає:

$$S = 1 \cdot 30 + 1 \cdot 20 + 3 \cdot 20 + 1 \cdot 10 + 2 \cdot 10 = 140 \text{ гр.од.}$$

Відкрита модель транспортної задачі.

Відкритою називається така транспортна задача, в якій потужність виробників не дорівнює попиту споживачів.

Відкрита транспортна задача розв'язується зведенням її до закритої. З цією метою у відкритій транспортній задачі вводять фіктивного постачальника, потужність якого дорівнює різниці сумарного попиту і сумарних потужностей, або фіктивного споживача, якщо потужність постачальників більша попиту споживачів. У цьому випадку фіктивний попит дорівнює різниці сумарних потужностей постачальників і сумарного попиту споживачів. Коефіцієнти затрат приймаються рівними витратам від недозавантаження споживачів або витратам на зберігання матеріалів на складах у протилежному випадку. Якщо така інформація відсутня, то вони приймаються рівними нулю.

Після таких перетворень задача стає закритою і розв'язується за звичайними правилами.

10.5. Задача. У господарстві заготовили 1740т сіна. Сіно розташоване на п'яти різних ділянках в наступній кількості: на першій ділянці—750т, другому—430т, третій — 120т, четвертому—170т і п'ятому—270т. Сіно потрібно доставити до чотирьох тваринницьких ферм: на першу ферму—150т, другу—420т, третю—270т і четвертую — 800т.

Відстань перевезення сіна від кожної ділянки до ферм така (табл. 10.10.).

Таблиця 10.10.

Відстань від ділянок до ферм (км.)

Ділянки	Ферми			
	1-а	2-а	3-а	4-а
1-й	3	11	3	6
2-й	2	9	10	7
3-й	9	5	6	12
4-й	7	4	2	3
5-й	3	5	10	8

Потрібно скласти такий план перевезень сіна, щоб транспортні витрати були мінімальними.

Вказівки до виконання завдання. Транспортне завдання є одному із завдань лінійного програмування, оскільки всі її умови можна виразити у вигляді системи лінійних обмежень і записати цільову функцію. Тому дане завдання може бути вирішена за допомогою симплексного методу. Проте існують простіші в порівнянні з симплексним методи вирішення завдань такого роду. Серед них найбільш простим і ефективним є метод потенціалів.

При вирішенні транспортних завдань перш за все перевіряється умова рівності ресурсів постачальників потребам споживачів. Якщо ця умова не виконується, то вводиться фіктивний постачальник або споживач. При цьому фіктивні об'єми ресурсів або потреб включаються в завдання з нульовими оцінками.

Потім заповнюється розрахункова таблиця і складається перший опорний план, який можна отримати декількома способами. Більш близький до оптимальному опорний план може бути отриманий з використанням методу «якнайкращого» елементу в таблиці. При цьому способі складання плану починається з клітки з мінімальною оцінкою при рішенні задачі на мінімум або з максимальної — при вирішенні на максимум. Якщо в таблиці є декілька кліток з однаковими «кращими» оцінками, то заповнюється перш за все клітка, в яку можна записати найбільше постачання.

Загальне число заповнених кліток повинне бути рівне $m+n-1$ (m — число рядків, n - число стовпців). Якщо число заповнених кліток опиниться менше цієї величини, то проводиться перерозподіл постачань або ж в одну з вільних кліток ставиться нульове постачання і ця клітка вважається заповненою.

Після складання першого опорного плану за допомогою алгоритму методу потенціалів проводиться перевірка його на оптимальність і, якщо план не оптимальний, то здійснюється його поліпшення.

Алгоритм методу потенціалів (рішення задачі на мінімум).

1. Для всіх заповнених кліток розраховуються потенціали за формулою:

$$u_i + v_j = c_{ij}$$

де u_i - потенціали рядків;

v_j - потенціали стовпців;

c_{ij} - оцінки.

Для розрахунку потенціалів одному з них спочатку надають будь-яке значення. Зазвичай $u_1=0$

2. Для всіх вільних кліток розраховуються характеристики за формулою:

$$d_{ij} = c_{ij} - (u_i + v_j)$$

Якщо в таблиці немає жодної вільної клітки з негативною (при рішенні на максимум—позитивною) характеристикою, то план вважається оптимальним.

3. Серед від'ємних (при рішенні на максимум — додатніх) характеристик вибирається максимальна по абсолютній величині, і для клітки з цією характеристикою будується ланцюг. Для цього з вибраної вільної клітки по рядку або стовпцю проводиться пряма лінія до зайнятої клітки, потім під кутом 90° лінія проводиться до наступної зайнятої клітки і так до тих пір, поки ланцюг не замкнеться в початковій клітці.

4. У вершинах ланцюга, починаючи з вибраної вільної клітки, по черзі проставляються знаки плюс і мінус. У клітках із знаком мінус вибирається мінімальне постачання, яке перерозподіляється по ланцюгу: там, де стоїть знак «плюс», вона додається, а де «мінус» — віднімається. Початкова вільна клітка стає зайнятою, а клітка, в якій вибрано мінімальне постачання, — вільною. У випадку якщо є декілька кліток з мінімальним постачанням, то вільною стає

клітка з максимальною (при рішенні на максимум—мінімальною) оцінкою, а в останніх проставляється постачання «0».

Складається новий план і розраховується значення цільової функції.

5. Прейти до 1.

Приклад розв'язання транспортної задачі лінійного програмування.

Умова задачі.

В пунктах відправлення $A_1, A_2, A_3, \dots$ знаходиться однорідний вантаж в кількості $a_1, a_2, a_3, \dots$ відповідно, який необхідно перевезти в пункти призначення $B_1, B_2, B_3, \dots$, потреба кожного з яких становить $b_1, b_2, b_3, \dots$. Відома відстань між пунктами перевезень (оцінки).

Маємо три постачальника з такими запасами вантажу:

$A_1 = 26$ т, $A_2 = 33$ т, $A_3 = 45$ т;

і чотири споживача з потребою на цей вантаж:

$B_1 = 14$ т, $B_2 = 19$ т, $B_3 = 20$ т, $B_4 = 5$ т.

Відома відстань між пунктами перевезень.

Відстань між пунктами перевезень, км (C_{ij}).

Постачальники	Споживачі			
	B_1	B_2	B_3	B_4
A_1	6	4	15	19
A_2	13	17	28	3
A_3	5	20	6	10

Необхідно знайти кількість вантажу, який повинен отримати кожен споживач від того або іншого постачальника, щоб загальна кількість тонно-кілометрів була мінімальною.

Розв'язок.

Рішення транспортної задачі лінійного програмування здійснюється в таблицях. Перша таблиця (опорний план) може бути заповнена декількома методами:

1. метод північно-західного кута;
2. метод найменшої вартості;
3. метод подвійної переваги.

Задача, в якій загальні запаси дорівнюють загальним потребам називається закритою транспортною задачею:

- загальні запаси = $26+33+45 = 104$ т;

- загальні потреби = $14+19+20+5 = 58$ т.

Отже задача відкрита, тому необхідно ввести фіктивного споживача, якому слід перевезти $V_{\text{фікт}} = 104 - 58 = 46$ т.

Клітини з фіктивним споживачем (постачальником) заповнюються в останню чергу, порушення умови оптимальності в цих клітинах розглядається лише після того, як в основних клітинах порушення умови оптимальності не спостерігається.

Першу транспортну таблицю заповнимо методом північно-західного кута: першою заповнюється верхня ліва клітина X_{11} , в яку заноситься або всі запаси A_1 , або вся потреба B_1 в залежності від того, що менше. На кожному наступному кроці розглядається перший з тих, що залишилися, пункт призначення або пункт відправлення:

- якщо це пункт призначення, то потреба його визначається за правилом: $X_{ij} = a_i$ - вивезений вантаж з A_i ;

- якщо це пункт відправлення, то запаси його визначається за правилом: $X_{ij} = b_j$ - завезений вантаж в B_j .

Так заповнюються всі клітини таблиці з X_{11} по X_{mh} (це заповнення має вигляд східців по діагоналі таблиці).

В клітину X_{11} записуємо 14 і відповідно покривається вся потреба першого споживача за рахунок першого постачальника, в якого залишається вантаж в кількості: $26 - 14 = 12$ тон, який можна відвезти в другий пункт призначення. Тимчасово не розглядається стовпець B_1 . Потреба другого пункту споживання B_2 становить 19 тон, а за рахунок першого постачальника A_1 можна завезти лише 12 тон, 7 тон завозиться з другого пункту відправлення. Тимчасово не розглядається рядок A_1 . В постачальника A_2 було 33 тони, з яких 7 тон перевозиться в пункт споживання B_2 , 20 тон в пункт споживання B_3 та 1 тонна в фіктивний пункт споживання $B_{\text{фікт}}$. Потреба фіктивного пункту споживання покривається за рахунок другого та третього постачальника.

Постачальники	Споживачі										Запаси	Потенціали, U_i
	B_1		B_2		B_3		B_4		$B_{\text{фікт}}$			
A_1	14	6	12	4		15		19		0	26	0
A_2		13		17		-		28		3		
A_3		5		7	20		5	1		33	13	
						6		10		0		
					+	**			45	-	45	13
Потреба	14		19		20		5		46		104	
Потенціали V_i	6		4		15		-10		-13			

При такому плані перевезень загальна кількість тонно-кілометрів буде становити:

$$F_{\min} = 14 \cdot 6 + 12 \cdot 4 + 7 \cdot 17 + 20 \cdot 28 + 5 \cdot 3 + 1 \cdot 0 + 45 \cdot 0 = 826 \text{ т-км.}$$

За допомогою методу потенціалів можна визначити оптимальний план транспортної задачі. Задача має оптимальне значення, коли сума потенціалів в усіх клітинах таблиці не перевищує оцінок, тобто при рішенні задачі на мінімум план вважається оптимальним в тому випадку, якщо для всіх вільних клітинок дотримується вимога: $U_i + V_j \leq C_{ij}$, а для зайнятих $U_i + V_j = C_{ij}$.

Для перевірки плану на оптимальність використовують систему оцінок (потенціалів) рядків і стовпчиків, що розраховуються за правилом: для будь-якої заповненої клітинки сума потенціалів відповідного рядка і стовпчика повинна дорівнюватися оцінці (вартість перевезення одиниці вантажу або відстань) цієї клітини.

$$U_i + V_j = C_{ij}$$

де: C_{ij} - оцінка клітинки;

i - рядки ($i = 1, 2, \dots, m$),

j - стовпчики ($j = 1, 2, \dots, h$)

U_i - оцінка i -го рядка ($i = 1, 2, \dots, m$)

V_j - оцінка j -го стовпця ($j = 1, 2, \dots, h$)

Тобто, якщо відомий потенціал U_i , то $V_j = C_{ij} - U_i$, а якщо відомий V_j , то $U_i = C_{ij} - V_j$. При цьому один з рядків або один з стовпчиків отримує довільну оцінку, а оцінки останніх рядків і стовпчиків треба розрахувати за вказаними формулами.

Нехай $U_1 = 0$, тоді потенціали першого та другого стовпчиків будуть становити:

$$V_1 = C_{11} - U_1 = 6 - 0 = 6 \quad \text{для } B_1$$

$$V_2 = C_{12} - U_1 = 4 - 0 = 4 \quad \text{для } B_2$$

Тепер можна визначити потенціал другого рядка:

$$U_2 = C_{22} - V_2 = 17 - 4 = 13$$

Визначимо потенціали для B_3 , B_4 і $B_{\text{фікт}}$:

$$V_3 = C_{23} - U_2 = 28 - 13 = 15 \quad \text{для } B_3$$

$$V_4 = C_{24} - U_2 = 3 - 13 = -10 \quad \text{для } B_4$$

$$V_5 = C_{25} - U_2 = 0 - 13 = -13 \quad \text{для } B_{\text{фікт}}$$

Тепер можна визначити потенціал A_3 :

$$U_3 = C_{35} - V_5 = 0 - (-13) = 13$$

В усіх зайнятих клітинах умова оптимальності виконується, а для вільних клітин виконуються розрахунки по перевірці плану на оптимальність :

$$\text{Для } X_{13} \text{ сума потенціалів } U_1 + V_3 = 0 + 15 = 15, \text{ що } \leq 15$$

$$\text{Для } X_{14} \text{ сума потенціалів } U_1 + V_4 = 0 + (-10) = -10, \text{ що } \leq 19$$

$$\text{Для } X_{15} \text{ сума потенціалів } U_1 + V_5 = 0 + (-13) = -13, \text{ що } \leq 0$$

$$\text{Для } X_{21} \text{ сума потенціалів } U_2 + V_1 = 13 + 6 = 19, \text{ що } > 13$$

$$\text{Для } X_{31} \text{ сума потенціалів } U_3 + V_1 = 13 + 6 = 19, \text{ що } > 5$$

$$\text{Для } X_{32} \text{ сума потенціалів } U_3 + V_2 = 13 + 4 = 17, \text{ що } \leq 20$$

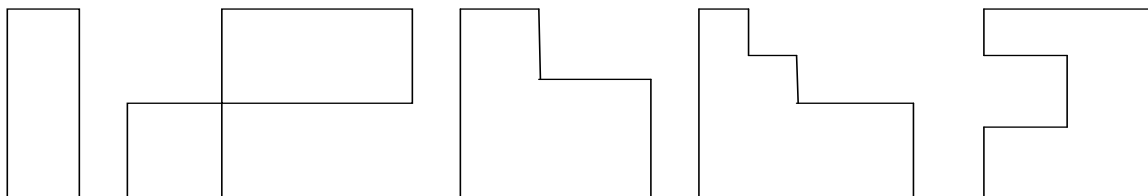
$$\text{Для } X_{33} \text{ сума потенціалів } U_3 + V_3 = 13 + 15 = 28, \text{ що } > 6$$

$$\text{Для } X_{34} \text{ сума потенціалів } U_3 + V_4 = 13 + (-10) = 3, \text{ що } \leq 10$$

Отже, даний план перевезень не є оптимальним, тому що в клітинках X_{21} , X_{31} , X_{33} не виконується вимога $U_i + V_j \leq C_{ij}$. Необхідно перейти до наступного плану.

Для цього вибирається клітина, в якій сума потенціалів найбільш перевищує оцінку, це клітина X_{33} і вона називається клітиною перерахунку.

Починаючи з клітини перерахунку будується по зайнятим клітинам фігура перерахунку (замкнутий контур, вершини якого знаходяться в зайнятих клітинах), яка може мати вигляд (деякі варіанти фігури):



Клітини, які увійшли до фігури перерахунку позначаються по черзі знаками “+”, “-”, “+”, “-”, ... , починаючи з клітини перерахунку.

Заповнюється нова транспортна таблиця:

- кількість вантажу в клітинах, які не увійшли до фігури перерахунку переноситься в нову таблицю без змін;

- для клітин фігури перерахунку: кількість вантажу в клітинках зі знаком “+” збільшується, а в клітинках зі знаком “-” зменшується на одну і ту ж величину, а саме, на найменшу кількість вантажу у клітинках, позначених знаком “-”.

Виправлений таким чином план перевезень записується у відповідну таблицю і досліджується на оптимальність, як і попередній.

Постачальники	Споживачі					Запаси	Потенціали, U_i
	B_1	B_1	B_1	B_1	$B_{\text{фікт}}$		
A_1	6 - 14	4 + 12	15	19	0	26	0
A_2	13	17 - 7	28	3 5	0 + 21	33	13
A_3	5 ** +	20	6 20	10	0 - 25	45	13
Потреба	14	19	20	5	46	104	
Потенціали U_j	6	4	-7	-10	-13		

При такому плані перевезень загальна кількість тонно-кілометрів буде становити:

$$F_{\min} = 14 \cdot 6 + 12 \cdot 4 + 7 \cdot 17 + 5 \cdot 3 + 21 \cdot 0 + 20 \cdot 6 + 25 \cdot 0 = 386 \text{ т-км.}$$

Обчислимо потенціали за правилом, описаним вище, і перевіримо складений план на оптимальність.

Для вільних клітин зробимо розрахунки:

$$\text{для } X_{13} \text{ сума потенціалів } U_1 + V_3 = 0 + (-7) = -7, \text{ що } \leq 15$$

$$\text{для } X_{14} \text{ сума потенціалів } U_1 + V_4 = 0 + (-10) = -10, \text{ що } \leq 19$$

$$\text{для } X_{15} \text{ сума потенціалів } U_1 + V_5 = 0 + (-13) = -13, \text{ що } \leq 0$$

$$\text{для } X_{21} \text{ сума потенціалів } U_2 + V_1 = 13 + 6 = 19, \text{ що } > 13$$

$$\text{для } X_{23} \text{ сума потенціалів } U_2 + V_3 = 13 + (-7) = 6, \text{ що } \leq 28$$

$$\text{для } X_{31} \text{ сума потенціалів } U_3 + V_1 = 13 + 6 = 19, \text{ що } > 5$$

$$\text{для } X_{32} \text{ сума потенціалів } U_3 + V_2 = 13 + 4 = 17, \text{ що } \leq 20$$

$$\text{для } X_{34} \text{ сума потенціалів } U_3 + V_4 = 13 + (-10) = 3, \text{ що } \leq 10$$

Даний план перевезень не є оптимальним, тому що в клітинках X_{21} , X_{31} , не виконується умова оптимальності $U_i + V_j \leq C_{ij}$.

Вибирається клітина X_{31} , починаючи з неї будується фігура перерахунку.

Наступна транспортна таблиця має вигляд:

Постачальники	Споживачі					Запаси	Потенціал $U(i)$
	B_1	B_2	B_3	B_4	$B_{\text{фікт}}$		
A_1	6	4	15	19	0	26	0
	-7	19			28		
A_2	13	17	28	3	0	33	-1
				5			
A_3	5	20	6	10	0	45	-1
	+7		20		-18		
Потреба	14	19	20	5	46	104	
Потенціали, V_j	6	4	7	4	1		

При такому плані перевезень загальна кількість тонно-кілометрів буде становити:

$$F_{\min} = 7 \cdot 6 + 19 \cdot 4 + 5 \cdot 3 + 28 \cdot 0 + 7 \cdot 5 + 20 \cdot 6 + 18 \cdot 0 = 288 \text{ т-км.}$$

Обчислимо потенціали і для вільних клітин зробимо розрахунки по перевірці плану на оптимальність :

для X_{13} сума потенціалів $U_1 + V_3 = 0 + 7 = 7$, що ≤ 15

для X_{14} сума потенціалів $U_1 + V_4 = 0 + 4 = 4$, що ≤ 19

для X_{15} сума потенціалів $U_1 + V_5 = 0 + 1 = -1$, що > 0

для X_{21} сума потенціалів $U_2 + V_1 = -1 + 6 = 5$, що ≤ 13

для X_{22} сума потенціалів $U_2 + V_2 = -1 + 4 = 3$, що ≤ 17

для X_{23} сума потенціалів $U_2 + V_3 = -1 + 7 = 6$, що ≤ 28

для X_{32} сума потенціалів $U_3 + V_2 = -1 + 4 = 3$, що ≤ 20

для X_{34} сума потенціалів $U_3 + V_4 = -1 + 4 = 3$, що ≤ 10

Даний план перевезень не є оптимальним, тому що в клітинці X_{15} не дотримується вимога оптимальності $U_i + V_j \leq C_{ij}$. Побудуємо фігуру перерахунку та розрахуємо нову таблицю.

Малюємо фігуру перерахунку та розрахуємо нову таблицю.

Поста- чальник и	Споживачі								Запаси	Потенці али, U_i		
	B_1		B_2		B_3		B_4				Вфіктивн ий	
A_1		6		4		15		19		0	26	0
			19						7			
A_2		13		17		28		3		0	33	0
							5		28			
A_3		5		20		6		10		0	45	0
	14				20				11			
Потреба	14		19		20		5		46		104	
Потенці али, V_j	5		4		6		3		0			

При такому плані перевезень загальна кількість тонно-кілометрів буде становити:

$$F_{\min} = 7 \cdot 6 + 19 \cdot 4 + 5 \cdot 3 + 28 \cdot 0 + 7 \cdot 5 + 20 \cdot 6 + 18 \cdot 0 = 281 \text{ т.-км.}$$

Обчислимо потенціали та для вільних клітин зробимо розрахунки по перевірці плану на оптимальність :

для X_{11} сума потенціалів $U_1 + V_1 = 0 + 5 = 5$, що ≤ 6

для X_{13} сума потенціалів $U_1 + V_3 = 0 + 6 = 6$, що ≤ 15

для X_{14} сума потенціалів $U_1 + V_4 = 0 + 3 = 3$, що ≤ 19

для X_{21} сума потенціалів $U_2 + V_1 = 0 + 5 = 5$, що ≤ 13

для X_{22} сума потенціалів $U_2 + V_2 = 0 + 4 = 4$, що ≤ 17

для X_{23} сума потенціалів $U_2 + V_3 = 0 + 6 = 6$, що ≤ 28

для X_{32} сума потенціалів $U_3 + V_2 = 0 + 4 = 4$, що ≤ 20

для X_{34} сума потенціалів $U_3 + V_4 = 0 + 3 = 3$, що ≤ 10

Даний план перевезень є оптимальним, тому що в усіх клітинах таблиці виконується умова оптимальності.

При оптимальному плані перевезень загальна кількість тонно-кілометрів буде становити 281 т.-км.

Так, перевезення вантажу слід організувати таким чином:

- з пункту A_1 необхідно перевезти 19тон в пункт B_2 та 7тон в пункт $B_{\text{фікт}}$;
- з пункту A_2 необхідно перевезти 5тон в пункт B_4 та 28тон в пункт $B_{\text{фікт}}$;
- з пункту A_3 необхідно перевезти 14тон в пункт B_1 , 20тон в пункт B_3 та 11тон в пункт $B_{\text{фікт}}$.

Лекція 11. Математичне моделювання виробничих процесів в агрономії

Економіко-математична модель оптимізації структури посівних площ.

Загальна постановка задачі.

В сільському господарстві земля - основний засіб виробництва. Її необхідно використовувати з максимальним економічним ефектом, постійно дбаючи про родючість. Успішне рішення задачі раціонального використання землі багато в чому залежить від обґрунтованої структури посівних площ.

Оптимальна структура посівних площ являє собою відображення спеціалізації рослинництва у вигляді такого співвідношення посівних площ по культурах, яке забезпечує отримання максимальної кількості сільськогосподарської продукції.

Постановка задачі: виходячи з виробничих ресурсів (земельних, трудових, матеріальних, тощо) визначити оптимальну структуру посівних площ, яка забезпечить виконання плану продажу продукції по видах, внутрішні потреби господарства в продукції, агрономічні та сівозмінні вимоги при максимальному економічному ефекті.

В якості критерію оптимальності може бути:

- максимум отримання прибутку від виробництва;
- максимум виробництва валової та товарної продукції в грошовому виразі;
- максимум виробництва конкретної продукції (зерна, картоплі, тощо).

Для розробки економіко-математичної моделі необхідна така інформація:

- розмір площі ріллі, сінокосів, пасовищ;
- перелік сільськогосподарських культур, що вирощуються в даній кліматичній зоні, їх можлива врожайність;
- затрати праці та коштів на 1 га посіву;
- виручка від реалізації продукції (з 1 га або одиниці продукції);
- наявність виробничих ресурсів в господарстві, норми витрат їх на 1 га;
- потреба тварин в кормах по видах;
- план продажу та внутрішня потреба господарства по видах продукції;
- агротехнічні вимог та можливі межі входження окремої культури або групи культур в сівозміну.

Економіко-математичну модель можна представити в загальному виді.

Знайти максимум функції:

$$Z_{\max} = \sum_{j=1}^n C_j X_j$$

при обмеженнях:

- виробничі ресурси господарства обмежені:

$$\sum_{j=1}^n a_{ij} X_j \leq b_i;$$

- окремі культури та групи культур в сівозміну вводяться в обґрунтованих агротехнічними вимогами межах:

$$\overline{X1_j} \leq \sum_{j=1}^n X_j \leq X2_j ;$$

- дотримується співвідношення окремих культур (наприклад: для отримання збалансованих концентрованих кормів, страхові культури, ...):

$$\sum_{j=1}^l W_{ij} X_j - \sum_{j=l+1}^n W'_{ij} X_j \leq 0;$$

- обмеження, на отримання продукції даного виду не менше заданої кількості:

$$\sum_{j=1}^n U_{ij} X_j \geq P_i;$$

- при умові невід'ємності змінних:

$$X_j \geq 0; \overline{X_j} \geq 0;$$

де:

C_j - прибуток з 1га j -ї культури;

X_j - площа посіву j -тої культури;

a_{ij} - затрати ресурсів i -го виду на 1га j -ї культури;

b_i - наявність виробничих ресурсів i -го виду;

$\overline{X1_j}$ - нижня межа входження окремої культури або групи культур в сівозміну;

$\overline{X2_j}$ - верхня межа входження окремої культури або групи культур в сівозміну;

U_{ij} - урожайність j -тої культури;

P_i - необхідна кількість i -го виду продукції;

W_{ij}, W'_{ij} - коефіцієнти пропорційності.

Приклад економіко-математичної моделі оптимізації структури

посівних площ.

Умова задачі.

Господарство має 1200 га оранки, на яких може вирощувати зернові та просапні культури.

Площа зернових у структурі посівних площ повинна бути від 40 до 60 %, площа цукрових буряків не повинна перевищувати 15%, а площа соняшника або картоплі – не більше 10%. Площа озимих повинна бути не більше 50% від максимальної площі зернових. Площа ярої пшениці співвідноситься до площі озимої пшениці як 1:10.

Підприємство повинно виробити не менше:

- 15600ц зерна (в тому числі 7050ц пшениці);
- 10500ц цукрового буряку;
- 1050ц соняшника або 8720ц картоплі;
- 530ц проса або гречки.

Крім цього для потреб тваринництва необхідно:

- соломи – 17400ц;
- кукурудзи на силос (зелений корм) – 15570ц

В якості критерію оптимальності виступає прибуток від товарної продукції рослинництва.

Розробити економіко-математичну модель, змінними в якій будуть площі посіву кожної культури ($X_1, X_2, \dots, X_8$), валові збори культур ($X_9, X_{10}, \dots, X_{16}$), загальні витрати на виробництво продукції (X_{17}), витрати на виробництво товарної продукції (X_{18}), виручка від реалізації товарної продукції рослинництва (X_{19}).

Культури, які будуть вирощуватися в господарстві

Зернові					Просапні		
озима пшениця	яра пшениця	горох	гречка	жито	цукровий буряк	картопля	кукурудза на силос

Числові дані для розробки економіко-математичної моделі.

Культура	Урожайність, ц/га		Затрати на 1 га, грн.	Ціна реалізації 1ц, грн.
	основної продукції	Побічної продукції		
Озима пшениця	32	40	540	40
Яра пшениця	30	30	320	40
Гречка	9		380	44
Жито	25	25	356	28
Горох	16	16	397	45
Цукровий буряк	226		1865	16
Картопля	190		1921	50
Кукурудза на силос	178		302	

Розв'язування.

На 19-ть змінних задачі накладено 25 обмежень:

1-ше обмеження стосується використання наявної площі:

$$1X_1+1X_2+1X_3+1X_4+1X_5+1X_6+1X_7+1X_8 \leq 1200, \text{ де:}$$

$X_1, X_2, \dots, X_8$ – площі сільськогосподарських культур.

2-ге та 3-тє обмеження задають верхню та нижню межу насиченості сівозміни зерновими культурами, питома вага яких повинна бути від 40% до 60%.

$$1X_1+1X_2+1X_3+1X_4+1X_5 \leq 20 \quad (60\%)$$

$$1X_1+1X_2+1X_3+1X_4+1X_5 \geq 480 \quad (40\%), \text{ де}$$

X_1, X_2, X_3, X_4, X_5 – площі зернових культур.

4-те, 5-те та 6-те обмеження описують максимальну площу цукрових буряків, озимих та картоплі відповідно, так 4-те стосується площі цукрових буряків:

$$1X_6 \leq 180 \quad (15\%), \text{ де:}$$

X_6 – площа цукрових буряків.

7-ме обмеження характеризує співвідношення між площею посіву озимої пшениці та ярої пшениці:

$$1X_1 - 10X_2 = 0, \text{ де:}$$

X_1 – площа озимої пшениці;

X_2 – площа ярої пшениці.

Група обмежень з 8-го по 12-те задають нижню межу виробництва певної продукції, так 8-ме обмеження характеризується валового збору зерна, який повинен бути на менше 15600 ц:

$$1X_9+1X_{10}+1X_{11}+1X_{12}+1X_{13} \geq 15600; \text{ де:}$$

$X_9, X_{10}, X_{11}, X_{12}, X_{13}$ – валові збори озимої пшениці, ярої пшениці, гречки, жита, гороху відповідно.

13-те та 14-те обмеження забезпечать задоволення потреб тваринництва в соломі та силосі відповідно, так 13-те має вигляд:

$$40X_1+30X_2+25X_4+16X_5 \geq 17400; \text{ де:}$$

X_1, X_2, X_4, X_5 – площі зернових культур: озимої пшениці, ярої пшениці, жита та гороху відповідно (коефіцієнти при цих змінних - урожайність соломи відповідних культур).

Обмеження з 15-го по 22-ге забезпечують зв'язок між площею посіву та валовим збором по кожній сільськогосподарській культурі, 15-те обмеження стосується озимої пшениці:

$$32X_1 - X_9 = 0; \text{ де:}$$

X_1 – площа озимої пшениці (32 – урожайність озимої пшениці);

X_9 – валовий збір озимої пшениці.

23-те обмеження акумулює затрати на виробництво продукції:

$$540X_1 + 320X_2 + 380X_3 + 356X_4 + 397X_5 + 1865X_6 + 1921X_7 + 302X_8 - X_{17} = 0; \text{ де:}$$

$X_1, X_2, \dots, X_8$ – площі сільськогосподарських культур (коефіцієнти при змінних – затрати на 1 га, грн.);

X_{17} – загальні затрати на виробництво продукції, грн.

24-те обмеження характеризує затрати на виробництво товарної продукції:

$$540X_1 + 320X_2 + 380X_3 + 356X_4 + 397X_5 + 1865X_6 + 1921X_7 - X_{18} = 0; \text{ де:}$$

$X_1, X_2, X_3, X_4, X_5, X_6, X_7$ – площі товарних сільськогосподарських культур (коефіцієнти при змінних – затрати на 1 га, грн.);

X_{18} – загальні затрати на виробництво товарної продукції, грн.

25-те обмеження стосується виручки від реалізації товарної продукції:

$$40X_9 + 40X_{10} + 44X_{11} + 28X_{12} + 45X_{13} + 16X_{14} + 50X_{15} - X_{19} = 0; \text{ де:}$$

$X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}$ – валові збори с/г культур (коефіцієнти при змінних – ціна реалізації 1 ц продукції, грн.);

X_{19} – виручка від реалізації товарної продукції рослинництва, грн.

В якості критерію оптимальності виступає прибуток від виробництва товарної продукції рослинництва:

$$1X_{19} - 1X_{18} \rightarrow \max.$$

Економіко-математичну модель оптимізації посівних площ можна представити в табличному вигляді.

Економіко-математична модель оптимізації структури посівних площ

[illegible]

14.	Потреба тваринництва в силосі							178												>= 15570	14
15.	Виробництво пшениці озимої	32							-1											= 0	15
16.	Виробництво пшениці ярої		30							-1										= 0	16
17.	Виробництво гречки			9							-1									= 0	17
18.	Виробництво жита				25							-1								= 0	18
19.	Виробництво гороху					16							-1							= 0	19
20.	Виробництво цукрового буряку						226							-1						= 0	20
21.	Виробництво картоплі							190							-1					= 0	21
22.	Виробництво кукурудзи							178								-1				= 0	22
23.	Витрати на виробництво продукції	540	320	380	356	397	1865	1921	302								-1			= 0	23
24.	Затрати на виробництво товарної продукції	540	320	380	356	397	1865	1921										-1		= 0	24
25.	Виручка від реалізації товарної продукції									40	40	44	28	45	16	50			-1	= 0	25
Цільова функція (прибуток), у.о.																		-1	1	max	

Економіко-математична модель оптимізації виробничої структури садового агропромислового підприємства

1. Загальна постановка задачі

Необхідно знайти оптимальний план виробництва, зберігання, переробки і реалізації продукції в агропромисловому підприємстві, в якому визначається:

- розмір і структура насаджень в розрізі сортових, видових і технологічних особливостей вирощування культур;
- завантаження потужності сховищ по їх типам і помісячна програма зберігання продукції;
- щомісячне завантаження потужностей переробки за видами сировини і випускної продукції;
- програма реалізації продукції у свіжому і переробленому вигляді з урахуванням сезонної зміни цін по каналах торгівлі;
- основні економічні показники функціонування в цілому і на кожній стадії руху продукції в агропромисловому підприємстві.

Оптимальний план повинен задовольняти слідуючі умови:

- обмеженість трудових і матеріально-грошових ресурсів;
- технологічні взаємозв'язки при розробці насаджень і руху продукції;
- обмеженість виробничих потужностей по зберігання, переробці та реалізації продукції;
- виконання зобов'язань по реалізації продукції у свіжому і переробленому вигляді .

Метою визначення оптимальної виробничої структури є досягнення максимальної економічної ефективності функціонування агропромислового підприємства .

2. Початкова інформація

Агропромислове підприємство, головною галуззю якого є садівництво, має площу багатолітніх насаджень, яка дорівнює 2200 га .Виходячи з рельєфу місцевості, рівня ґрунтових вод, співвідношення сортів обпилювачів та інших умов, яблуні на карликових прищепах не можуть перевищувати 50 % від загальної площі плодючого саду, а яблуні осінніх сортів повинні становити не менше 20 % від загальної площі багатолітніх насаджень. Загальний об'єм трудових ресурсів, які можуть бути використані на догляд за плодючим і молодим садом, збирання яблук, їх зберігання і переробку, складає в цілому 3550 тис. люд.-год. Показники по продуктивності багатолітніх насаджень, витратам на одиницю площі, а також строки вступу в плодоносіння,

експлуатація яблунь і період дозрівання та зберігання продукції в розрізі сортів і прищеп показані у таблиці 1.

Таблиця 1.

Початкова інформація про багаторічні насадження (яблуневий сад)

Багаторічні насадження	Урожайність, т/га	Витрати на 1 га		Строки	
		грн.	люд.-год.	початок плодоносіння та експлуатації, років	дозрівання та зберігання, місяців
Сад , що плодоносить					
на сильнорослому підщепі:					
зимові сорти	15	2535	1225	15	10/4
осінні сорти	20	2140	1050	12	9/2
на низькорослому прищепі:					
зимові сорти	18	1825	1060	8	10/3
осінні сорти	25	1560	875	8	9/1
Молодий сад					
на сильнорослому підщепі:					
зимові сорти	-	345	210	7	-
осінні сорти	-	345	210	6	-
на низькорослому підщепі:					
зимові сорти	-	320	200	4	-
осінні сорти	-	320	200	3	-

Господарство має овочесховище загальною потужністю 17000 тон, а також цехами по виробництву джему та варення місячною потужністю відповідно 200 і 80 тон.

Таблиця 2.

Початкова інформація по каналах руху та розподілу продукції.

Показники	Канали розподілу і руху продукції			
	зберігання (в місяць)	реалізація в свіжому вигляді	виробництво з яблук	
			джем	варення
Витрати на 1 т:				
грн.	50	9	640	750
люд.-год.	25	10	80	85
Розподіл сировини на виробництво 1 ц, кг	-	-	95	73
Норма природної убилі, %	2	-	-	-

Реалізаційні ціни на однієї тони джему і варення складають відповідно 1240 грн. і 1030 грн. за 1 тону.

Таблиця 3.

Ціни та об'єми реалізації яблук

Місяць	Середня ціна реалізації 1 т, грн.	Максимально можливі об'єми реалізації, т
Вересень-жовтень	500	12000
Листопад	700	600
Грудень-січень	1000	6000
Лютий	1500	3000

План по реалізації яблук в свіжому вигляді складає 26 тис. тон, а виробництву джему та варення, відповідно 1 та 0.4 тис. тон.

3. Система змінних моделі

По багаторічних насадженнях:

площа саду, який плодоносить, га

на сильнорослій підщепі

 X_1 – зимові сорти X_2 – осінні сорти

на низькорослій підщепі (підвої)

 X_3 – зимові сорти X_4 – осінні сорти

площа молодого (не плодоносить) саду, га

 X_5 – зимові сорти

X₆ – осінні сорти
на низькорослій підщепі (подвое)

X₇ – зимові сорти

X₈ – осінні сорти

По зберіганню продукції:

об'єм зберігання яблук, зимових сортів на сильнорослій підщепі, т

X₉ – у жовтні

X₁₀ – у листопаді

X₁₁ – у грудні

X₁₂ – у січні

X₁₃ – у лютому

об'єм зберігання яблук осінніх сортів на сильнорослій підщепі (подвое)

X₁₄ – у вересні

X₁₅ – у жовтні

X₁₆ – в листопаді

Об'єм зберігання яблук зимових сортів на низькорослій підщепі (подвое),

тон

X₁₇ – у жовтні

X₁₈ – в листопаді

X₁₉ – у грудні

X₂₀ – у січні

Об'єм зберігання яблук осінніх сортів на низькорослій підщепі (подвое)

тон

X₂₁ – у вересні

X₂₂ – у жовтні

Реалізація у свіжому вигляді, тон:

Об'єм прямої реалізації яблук осінніх сортів у вересні

X₂₃ – з сильною підщепою

X₂₄ – з низькорослою підщепою

Об'єм прямої реалізації яблук зимових сортів у жовтні

X₂₅ – з сильною підщепою

X₂₆ – з низькорослою підщепою

Об'єм реалізації яблук з сховищ

X₂₇ – в жовтні

X₂₈ – в листопаді

X₂₉ – в грудні

X₃₀ – в січні

X₃₁ – в лютому

По виробництву продукції переробки яблук:

Об'єм виробництва джему

- з яблук прямих поставок осінніх сортів у вересні

X_{32} – на сильнорослій підщепі

X_{33} – на низькорослій підщепі

- зимових сортів у жовтні

X_{34} – на сильнорослій підщепі

X_{35} – на низькорослій підщепі

- з яблук, що надійшли після зберігання

X_{36} – в жовтні

X_{37} – в листопаді

X_{38} – в грудні

X_{39} – в січні

X_{40} – в лютому

Об'єми виробництва варення

- з яблук прямих поставок осінніх сортів у вересні

X_{41} – на сильнорослих підщепах

X_{42} – на карликових підщепах

- зимових сортів у жовтні

X_{43} – на сильнорослих підщепах

X_{44} – на низькорослих підщепах

- з яблук, що надійшли після зберігання

X_{45} – в жовтні

X_{46} – в листопаді

X_{47} – в грудні

X_{48} – в січні

X_{49} – в лютому

Загальні економічні показники, грн.:

X_{50} – сума матеріально-грошових затрат в цілому по господарству

X_{51} – виручка від реалізації продукції.

Розробка моделі виконується по групам обмежень на основі прийнятої системи змінних та заданої початкової інформації.

4. Система обмежень моделі

По виробничим ресурсам:

1. Площа багаторічних посадок, га

$$X_1 + X_2 + \dots + X_8 \leq 2200.$$

Трудові ресурси, людино-год.

$$1225X_1+1050X_2+1060X_3+875X_4+210X_5+210X_6+200X_7+ \\ 200X_8+25X_9+...+25X_{22}+10X_{23}+...+10X_{31}+80X_{32}+... \\ +80X_{40}+ +85X_{41}+...+85X_{49} \leq 3\,550\,000.$$

Техніко-економічні коефіцієнти при змінних являють собою питомі затрати трудових ресурсів на одиницю розміру відповідної галузі (виду діяльності) і відображені в таблицях 1 і 2.

У загальному вигляді данна група обмежень може бути представлена таким чином:

$$\left(\sum_{j \in J} \right) \left[a_{ij} x_j + \sum_{s \in S} a_{ijs} x_{js} + \sum_{t \in T} \left(\sum_{h \in H} a_{hijt} x_{hjt} + \sum_{p \in P} a_{pijt} x_{pjt} + \sum_{r \in R} a_{rijt} x_{rjt} \right) \right] \leq B_i, (i \in I)$$

де індекси :

i – видів ресурсів ;

j – видів культур, насаджень ;

h – типів складів;

s – стадій розвитку культур, насаджень (до вступу у плодоносіння) ;

t - період

p - видів виготовлення продукції переробки

r - каналів реалізації продукції;

Множини:

i - видів ресурсів;

I - видів (груп) культур насаджень;

H - типів сховищ;

S - стадій розвитку культур насаджень;

T - періодів;

P - видів перероблюючи міцностей;

R - каналів реалізації продукції;

Шукані змінні:

X_j - площа j -го виду культури, в стадії плодоношення;

X_{js} - площ j -го виду культури, насаджень до вступу в плодоношення (на s -ой стадії розвитку);

X_{hjt} - об'єм зберігання в сховищах h -го типа продукції від культури, насаджень j -го вида t -ом періоду;

X_{pijt} - об'єм випуску p -го продукції переробки сировини від j -го виду культури, насаджень в t -ом періоді;

X_{rijt} - об'єм реалізації продукції r -го виду продукції від j -го виду культури, насаджень в t -ом періоді;

Техніко-економічні коефіцієнти при змінних і вільні члени:

a_{ij} a_{ijs} - витрати i -го виду ресурсів на одиниці площі j -го виду культури, насаджень в фазі плодоношення і до вступу в плодоношення (на s стадії розвитку);

a_{hijt} , a_{pijt} , a_{rjt} - затрати i -го виду ресурсів на зберігання (h), переробка (P) і реалізація (r) одиниці продукції j -й культури, насадження за t -й період;

V_i – об'єм i -го виду ресурса.

По технологічним взаємостосункам:

3. Співвідношення між плодоносним та молодим (неплодоносним) садом на сильнорослому подвоїв яблунь зимових сортів, *га*:

$$15X_5 \times 5 \geq 7X_1 \quad \text{або} \quad 7X_1 - 15X_5 \leq 0.$$

Техніко-економічними коефіцієнтами при перемінних строки вступання в плодоношення та експлуатацію багатолітніх насаджень (табл. 1).

Для осінніх сортів запис буде мати вид:

$$4. 6X_2 - 12X_6 \leq 0.$$

Аналогічно записуються обмеження 5-6 для яблунь на низькорослій підщепі.

7. Питома вага яблунь на низькорослій підщепі не більш 50% від площі плодоношення саду, *га*:

$$\begin{aligned} X_3 + X_4 &\leq 0,5(X_1 + X_2 + X_3 + X_4) \quad \text{або} \\ -0,5X_1 + 0,5X_2 + 0,5X_3 + 0,5X_4 &\leq 0 \quad \text{або} \\ -X_1 - X_2 + X_3 + X_4 &\leq 0. \end{aligned}$$

Обмеження по питомій вазі осінніх сортів (не менше 20% від всієї площі багаторічних насаджень; має аналогічний характер.

Математична запис даної групи обмежень має вид пропорціонального зв'язку:

$$W_j X_{j \geq} \leq W_{j1} X_{j1}, \quad (J \in I),$$

де W_j, W_s - коефіцієнт пропорційного зв'язку.

9. По виробництву та руху продукції зимових сортів яблунь на сильнорослій підщепі (подвое), тон.

$$\begin{aligned} 15X_1 &\geq X_9 + X_{25} + 0,95X_{34} + 0,73X_{43} \quad \text{або} \\ -15X_1 + X_9 + X_{25} + 0,95X_{34} + 0,73X_{43} &\leq 0, \end{aligned}$$

де техніко-економічні коефіцієнти при перемінній X_1 - урожайність (тон з 1 га); при перемінних X_{34} і X_{43} – витрати сировини на виробництво 1 тони джему і варення, тон.

Відповідне обмеження для осінніх сортів буде мати вигляд:

$$10. -20X_2 + X_{14} + X_{23} + 0,95X_{32} + 0,73X_{41} \leq 0$$

Обмеження 11-12 по виробництву та руху продукції зимових та осінніх сортів яблунь на низькорослій підщепі записуються аналогічно.

В тематичній формі дана група обмежень має вигляд:

$$V_j X_j \geq X_{hjt_0} + Q_{pj} X_{pj t_0} + X_{rjt_0}, (j \in J, h \in H, r \in R, p \in P)$$

де Q_{pj} - витрати продукції від j-го виду культури, насаджень на виробництво одиниці p-ої продукції переробки (витрати сировини);

V_j - вихід продукції з одиниці площі j-ї культури, насаджень (урожайність).

13. Розподіл яблук з сховищ у вересні:

$$\begin{aligned} 0,98X_{14} + 0,98X_{21} &\geq X_{15} + X_{22} \text{ або} \\ -0,98X_{14} + X_{15} - 0,98X_{21} + X_{22} &\geq 0, \end{aligned}$$

де 0,98 - коефіцієнт збереженості продукції (1-0,02).

14. Розподіл яблук з сховищ в жовтні:

$$-0,98X_{14} + X_{10} - 0,98X_{15} + X_{16} - 0,98X_{17} + X_{18} - 0,98X_{22} + X_{27} + 0,95X_{36} + 0,73X_{45} \leq 0$$

У жовтні починається надходження яблук з сховищ на переробку. Розподіл в листопаді-лютому відображається в обмеженнях 15-18 аналогічним шляхом.

Математичний запис цієї групи обмежень може бути представлена таким чином:

$$\alpha_{hjt} X_{hjt} \geq X_{hjt+1} + q_{pj} X_{pj t+1} + X_{rjt+1} + X_{rjt+1}, (t \in T, j \in J, h \in H, r \in R, p \in P)$$

де α_{hjt} - коефіцієнт збереженості продукції j-го виду культури, насаджень при зберіганні в t-му періоді.

19. Обмеження по узгодженню об'ємів зберігання яблук на сильнорослій підщепі зимових сортів в жовтні-листопаді

$$-0,98X_9 + X_{10} \leq 0.$$

Обмеження 20-22 відображають співвідношення об'ємів зберігання яблук на сильнорослій підщепі зимових сортів у листопаді-грудні-січні-лютому.

Аналогічно записуються обмеження по співвідношенню об'ємів зберігання яблук осінніх сортів на сильнорослій підщепі у вересні-жовтні-листопаді (обмеження 23-24), зимових сортів на низькорослій підщепі (подвое) у жовтні-листопаді-грудні-січні (обмеження 25-27), осінніх сортів на низькорослій підщепі у вересні-жовтні (обмеження 28).

В математичній формі дані обмеження можуть бути виражені:

$$\alpha_{hjt} X_{hjt} \geq X_{hjt+1}, (t \in T, j \in J, h \in H,)$$

29. Обмеження по виробничим потужностям зберігання продукції у вересні, тон (в період закладки продукції на зберігання):

$$X_{14} + X_{21} \leq 17000$$

30. Обмеження по потужності зберігання продукції у жовтні, тон:

$$X_9 + X_{15} + X_{17} + X_{22} \leq 17000$$

31. Обмеження по потужності переробки яблук в джем у вересні, тон:

$$X_{32} + X_{33} \leq 200$$

Аналогічний запис мають обмеження 32-36 по потужностях переробки яблук в джем у жовтні-лютому з урахуванням відповідних каналів надходження сировини, а також обмеження 37-42 по потужностях переробки яблук у варення у вересні-лютому.

43. Обмеження по об'ємах реалізації продукції у свіжому вигляді у вересні та жовтні:

$$X_{23} + X_{24} + X_{25} + X_{26} + X_{27} \leq 12000$$

Аналогічний запис обмежень 44 (для листопада), 45 (для грудня та січня) і 46 (для лютого).

Вільні члени обмежень представляють собою наявні потужності по зберіганню, переробці та реалізації продукції та беруться з табл. 3.

В загальному вигляді обмеження можуть бути записані:

- по зберіганню продукції

$$\left(\sum_{j \in J} \right) X_{hjt} \leq Q_{hjt} (h \in H, j \in J, t \in T);$$

- по переробці продукції

$$\left(\sum_{j \in J} \right) X_{pjt} \leq Q_{pjt} (p \in P, j \in J, t \in T);$$

- по реалізації продукції

$$\left(\sum_{j \in J} \right) X_{rjt} \leq Q_{rjt} (r \in R, j \in J, t \in T);$$

де: $Q_{hjt}, Q_{pjt}, Q_{rjt}$ - потужність зберігання, переробки та реалізації продукції j-го виду культури, посаджень у часовому розрізі.

Обмеження по виконанню плану-замовленню:

47. Реалізація яблук у свіжому вигляді, тон

$$X_{23} + X_{24} + \dots + X_{31} \geq 26000$$

48. Виробництво джему, тон

$$X_{32} + X_{33} + \dots + X_{40} \geq 1000$$

49. Виробництво варення, тон

$$X_{41} + X_{42} + \dots + X_{49} \geq 400$$

Математичний запис цієї групи обмежень має наступний вигляд:

- по реалізації продукції в свіжому вигляді

$$\left(\sum_{j \in J} \right) \sum_{t \in T} X_{rjt} \geq A_r (r \in R)$$

- по реалізації продукції переробки

$$\left(\sum_{j \in J} \right) \sum_{t \in T} X_{pjt} \geq A_p \quad (p \in P)$$

де A_r, A_p - план-заказ по реалізації продукції в свіжому і переробленому вигляді.

Обмеження по розрахунку загальних економічних показників:

50. Сума затрат по виробництву, зберіганню, переробці і реалізації продукції, грн.

$$253X_1 + 2140X_2 + 1825X_3 + 1560X_4 + 345X_5 + 345X_6 + 320X_7 + 320X_8 + 50X_9 + \dots + 50X_{22} + 9X_{23} + \dots + 9X_{31} + 640X_{32} + \dots + 640X_{40} + 750X_{41} - X_{50} = 0$$

51. Виручка від реалізації, грн.

$$500X_{23} + 500X_{24} + 500X_{25} + 300X_{26} + 500X_{27} + 700X_{28} + 1000X_{29} + 1000X_{30} + 1500X_{31} + \dots + 1240X_{32} + \dots + 1240X_{40} + 1030X_{41} + \dots + 1030X_{49} - X_{51} = 0$$

В загальному вигляді розрахунок основних економічних показників може бути записаний слідуючим чином:

$$\left(\sum_{j \in I} \right) \left[C_{kj} X_{j0} + C_{kjs} X_{js} + \sum_{e \in T} \left(\sum_{h \in H} C_{hkjt} X_{hjt} + \sum_{p \in P} C_{pkjt} X_{pjt} + \sum_{r \in R} C_{rkjt} X_{rjt} \right) \right]$$

Де : k - індекс економічного показника ($k \in K$);

K - множина видів економічних показників;

$C_{kj}, C_{kjs}, C_{hkjt}, C_{pkjt}, C_{rkjt}$ - питомі коефіцієнти витрат трудових та грошово-матеріальних ресурсів ($k \in K1$) та виходу продукції в грошовому виразі ($k \in K2$) на одиницю галузі (виду діяльності); $K \in K1, K2$.

Цільова функція:

Максимум прибутку від реалізації продукції, грн.

$$Z = X_{51} - X_{50} \rightarrow \max$$

Або в загальному вигляді:

$$Z = X_K - X_K \rightarrow \max$$

$K \in K_2 \quad K \in K_1$

5. Аналіз оптимального рішення

Оптимальне рішення економіко-математичної моделі оптимізації виробничої структури агропромислового підприємства дозволяє відповісти на всі питання, що були використані в постановці задачі. Аналіз рішення доцільно проводити по технологічним стадіям виробництва, зберігання, переробки та реалізації продукції агропромислового підприємства.

Виробничу структуру садівництва характеризує структура багаторічних насаджень.

Таблиця 4

Структура багаторічних насаджень

Багаторічні насадження	Площа	
	га	в % до підсумку
Плодоносний сад		
Яблуні на сильнорослій підщепі:		
- зимові сорти	235,27	10,69
- осінні сорти	514,14	23,37
Яблуні на низькорослій підщепі:		
- зимові сорти	426,39	19,38
- осінні сорти	323,02	14,68
Молодий сад		
Яблуні на сильнорослій підщепі:		
- зимові сорти	109,79	5,00
- осінні сорти	257,07	11,69
Яблуні на низькорослій підщепі:		
- зимові сорти	213,19	9,69
- осінні сорти	121,13	5,50
Всього:	2200,00	100,00

Площу та структуру саду розглядають в розрізі відображених в моделі змінних. Співвідношення між молодим та плодоносним садом в розрізі подвоїв та сортів забезпечує просте відновлення багаторічних насаджень з врахуванням строків вступу в плодоносність та експлуатацію яблунь.

Валовий збір яблук з плодоносного саду може бути розподілений по трьох каналах: пряма реалізація у свіжому вигляді, переробка на джем та варення і закладка на зберігання. Розподіл та рух яблук при зберіганні в оптимальному рішенні визначає програму зберігання продукції.

Місяці	Сильноросла підщепа		Низькоросла підщепа	
	зимові сорти	осінні сорти	зимові сорти	осінні сорти
вересень	х	6110	х	-
жовтень	3529	5988	7427	-
листопад	3458	5868	7278	х
грудень	3389	х	6635	х

Місяці	Сильноросла підщепа		Низькоросла підщепа	
	зимові сорти	осінні сорти	зимові сорти	осінні сорти
січень	3322	х	253	х
лютий	3255	х	х	х

Яблука від осінніх сортів на низькорослій підщепі (подвое) поступають в пряму реалізацію.

Рух продукції розглядається по місяцях можливого зберігання яблук різних сортів з врахуванням підщеп.

Зменшення продукції, заложеної на зберігання, по місяцям визвано, по-перше, природною втратою, по-друге, надходженням яблук для реалізацію у свіжому вигляді і на переробку для виробництва джему та варення.

Надійшовші об'єми продукції для реалізації у свіжому вигляді і на переробку необхідно розглядати як із сховищ, так і обминаючи їх в місяці масового збору яблук.

Таблиця 6.

Розподіл яблук на реалізацію у свіжому вигляді і переробку по місяцях надходження та зберігання, тон

Джерела надходження продукції по місяцях	Витрата продукцію на, т		
	реалізацію у свіжому вигляді	виробництво джему	виробництво варення
Пряме надходження:			
у вересні осінні сорти			
- на сильнорослій підщепі	4173	-	-
- на низькорослій підщепі	4827	190	58,4
у жовтні зимові сорти			
- на сильнорослій підщепі	-	-	-
- на низькорослій підщепі (подвое)	-	190	58,4
надходження зі сховищ:			
у жовтні	-	-	-
у листопаді	6000	190	58,4
у грудні	6000	190	58,4
у січні	-	190	58,4
у лютому	3000	190	-

Розгляд розподілу яблук по каналах руху продукції дозволяє скласти програму реалізації у свіжому вигляді, надходження на виробництво джему і варення. В отриманому оптимальному рішенні значну частину яблук у свіжому вигляді необхідно реалізувати у вересні (12000 тон), максимально використовуючи наявні потужності торгівлі і реалізовуючи прямі канали “сад-магазин”. Рівномірне надходження сировини для виробництва джему дозволить забезпечити ритмічну роботу переробних цехів.

Однією з основних умов постановки задачі є обмеження по гарантованому виробництву та реалізації продукції (план-замовлення).

Таблиця 7.

Продукція	Реалізація продукції, т		
	План-заказ	По оптимальному рішенню	Надпланова продукція
Яблука у свіжому вигляді	26000	27000	1000
Джем	1000	1200	200
Варення	400	400	-

З точки зору критерію оптимальності деякі види виробленої продукції в агропромисловому підприємстві продукції можуть мати велику ефективність і при наявності достатньої кількості виробничих ресурсів виробництво цих видів продукції перевищує установлені зобов'язання. Малоефективна продукція в оптимальному рішенні виходить відповідно до заданого об'єму.

У завершені аналізу оптимального рішення проводять аналіз об'єктивно обумовлених оцінок (двоїстих оцінок) і розглядають основні економічні показники функціонування агропромислового підприємства.

Аналіз двоїстих оцінок проводять по основним групам обмежень, ілюструючи кожен з них двоїстою оцінкою з характерного обмеження (висновки по коректуванню оптимального плану на основі двоїстих оцінок дійсні в рамках його стабільності).

Таблиця 8.

Двоїсті оцінки в окремих групах обмежень моделі

№ пп	Назва групи обмежень	Одиниця виміру	Значення додаткової змінної	Тип і об'єм обмежень	Двоїсті оцінки
1	2	3	4	5	6
1	По основним виробничим ресурсам				
	- площа багаторічних насаджень	тис. га	0	≤2,200	-45,38

№ пп	Назва групи обмежень	Одиниця виміру	Значення додаткової змінної	Тип і об'єм обмежень	Двоїсті оцінки
	- об'єм трудових ресурсів	тис. люд.- год	30,65	≤ 3550	0
2	По технологічним зв'язкам				
	- співвідношення між плодоносним і молодим садом зимових сортів яблунь на сильнорослій підщепі	тис.га	0	≤ 0	-26,03
	- співвідношення між яблунями на низькорослій і сильнорослій підщепі	тис. га	0	≤ 0	-548,14
3	По розподілу і руху продукції				
	- початковий розподіл яблук зимових сортів на сильнорослій підщепі	тис. га	0	≤ 0	-147,63
	- розподіл яблук зимових сортів на сильнорослій підщепі у жовтні	тис. т	0	≤ 0	-198,48
4	По виконанню плану- заказу виробництва продукції				
	- реалізація яблук у свіжому вигляді	тис. т	1,0	$\geq 26,0$	0
	- виробництво джему	тис. т	0,2	$\geq 1,0$	0
	- виробництво варення	тис. т	0	$\geq 0,4$	33,21
5	По виробничим потужностям				
	- реалізація яблук у вересні і жовтні	тис.шт	0	$\leq 12,00$	339,38

№ пп	Назва групи обмежень	Одиниця виміру	Значення додаткової змінної	Тип і об'єм обмежень	Двоїсті оцінки
	- зберігання у жовтні	тис.шт	0,057	$\leq 17,00$	0
	- виробництво джему у жовтні	тис.шт	0	$\leq 0,2$	462,71

Знак при подвійній оцінці залежить від способу їх у використаній програмі та типу екстремуму. В програмі, по якій вирішувалася задача, на максимум від'ємне значення подвійної оцінки показує збільшення цільової функції при збільшенні об'єму обмежень на одиницю і, навпаки, позитивне значення подвійної оцінки – зменшення цільової функції.

Нульові подвійні оцінки свідчать в обмеженнях “ $\leq$ ” про невикористання об'єма ресурсу, а в обмеженнях типа “ $\geq$ ” про перевищення об'єму продукції. Змінення об'ємів в сторону зусилання обмежень на величину, не перевищуюче значення доповнюючої змінної (тобто в межах стійкості) не спричинить зміни цільової функції.

В групі обмежень по основним виробничим ресурсам збільшення площі багатолітніх насаджень на 1 га, який призведе до збільшення прибутку на 45,38грн, а збільшення об'єму трудових ресурсів не спричинить змін цільової функції оскільки у рішенні одержали недовикористання трудових ресурсів і можливо зменшення їх об'єму на величину, яка не перевищує значення доповнюючої змінної (30,65тис.чол.ч.)

При аналізі подвійних оцінок в обмеженнях по відношенню окремих видів і груп змінних або обмеженню пропорційних зв'язків між змінними, де об'єми обмежень дорівнюють 0, необхідно змінити обмеження шляхом переносу від'ємних змінних за знак нерівності. Значення подвійної оцінки покажуть змінення цільової функції і при збільшенні правої частини на одиницю. Наприклад, обмеження по відношенню між яблунами на карликових та сильнорослих підщепах має вигляд:

$$-x_1 - x_2 + x_3 + x_4 \leq 0 \quad \text{або} \quad x_3 + x_4 \leq x_1 + x_2,$$

тобто площа яблуневого саду на низькорослій підщепі не більш площі яблуневого саду на швидкорослій підщепі, і збільшення площі яблуневого саду на низькорослій підщепі на 1га приведе к збільшенню прибутку на величину подвійної оцінки (548,14 грн.). В обмеженнях по виконанню плана-заказу виробляється продукції нульові подвійні оцінки свідчать про перевиконання плану-заказу на величину доповнюючої змінної, а не нульова оцінка в обмеженнях по виробництву варення про відносну неефективність виробництва

даної продукції і збільшення плану-заказу по виробництву варення на 1 тону приведе до зменшення прибутку на 33,21 грн.

Аналіз подвійних оцінок по залишившихся групам обмежень проводять аналогічно загальним положенням і приведеним прикладам.

Після розглядання подвійних оцінок приводяться показники економічної ефективності функціонування агропромислового підприємства: вартість товарної продукції, матеріально-грошові і трудові затрати, прибуток, як в цілому, так і по технологічним стадіям виробництва продукції.

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Іглін С. П., Зайцев Ю. І., Решетняк Ю. Б. Лінійне програмування : навч. посіб. Харків : НТМТ, 2023. 120 с. URL: <https://repository.kpi.kharkov.ua/server/api/core/bitstreams/384866bb-5028-4546-996c-eb0a5cb346f6/content>
2. Жалдак М. І., Кузьміна Н. М., Михалін Г. О. Теорія ймовірностей і математична статистика : підручник. Київ : НПУ імені М.П. Драгоманова, 2020. 750 с. URI: <http://enpuir.npu.edu.ua/handle/123456789/35207>
3. Касьяненко В.О., Старченко Л.В. Моделювання та прогнозування економічних процесів : конспект лекцій. Суми : Університетська книга, 2023. 185 с.
4. Конверський А. Основи методології та організація наукових досліджень : підруч. Київ : Центр учбової літератури, 2021. 352 с.
5. Кузьменко О. В., Козьменко О. В. Економіко-математичні методи і моделі: економетрика : навчальний посібник. Суми : Університетська книга, 2025. 408 с.
6. Кучма М. І. Математичне програмування: приклади і задачі : навчальний посібник. Львів : Новий світ-2000, 2023. 344 с.
7. Медведєв М. Г. Теорія ймовірності та математична статистика : підручник. Київ : Ліра-К, 2024. 536 с.
8. Найко Д. А. Шевчук О. Ф. Теорія ймовірностей та математична статистика : навч. посіб. Вінниця : ВНАУ, 2020. 382 с. URL: <http://repository.vsau.org/getfile.php/24513.pdf>
9. Новаківський І. І., Войцеховська Ю. В. Основи статистичного прогнозування економічних процесів : навчальний посібник. Львів : Новий світ-2000, 2025. 200 с.
10. Озеранський В. С., Крилик Л. В., Шевчук О. Ф. Теорія ймовірностей та математична статистика : лабораторний практикум. Вінниця : ВНТУ, 2025. 113 с.

- 11.Пушак Я.С., Лозовий Б.Н. Теорія ймовірностей і елементи математичної статистики : навчальний посібник. Львів : Магнолія 2006, 2024. 276 с.
- 12.Рабик В. М. Основи теорії ймовірностей : навчальний посібник. Львів : ПП "Магнолія 2006", 2025. 176 с.
- 13.Рижок З. Р., Поляковська Л. Л., Ступень Р. М., Колодій П. П. Математична обробка геодезичних вимірів : навчальний посібник. Львів : Галицька видавнича спілка, 2020. 179 с. URL: <https://surl.li/vbcoet>
- 14.Сидоренко Ю. В. Теорія ймовірностей : навч. посібник. Київ : НТУУ «КПІ імені Ігоря Сікорського», 2020. 81 с.
- 15.Теорія ймовірностей та математична статистика : навчальний посібник. 2-ге видання / Копич І.М., Сороківський В.М., Кісілевич О.В., Пенцак О.В. Львів : Новий світ-2000, 2025. 382 с.

Навчальне видання

СТАТИСТИЧНИЙ АНАЛІЗ І МОДЕЛЮВАННЯ ПРОЦЕСІВ В АГРОНОМІЇ

Конспект лекцій

Укладачі:

Шебаніна Олена В'ячеславівна
Тищенко Світлана Іванівна
Пархоменко Олександр Юрійович та ін.

Формат 60x84 1/16. Ум. друк. арк. 11,81.

Тираж 50 прим. Зам. № __

Надруковано у видавничому відділі
Миколаївського національного аграрного університету
54020, м. Миколаїв, вул. Георгія Гонгадзе, 9

Свідоцтво суб'єкта видавничої справи ДК № 4490 від 20.02.2013 р.