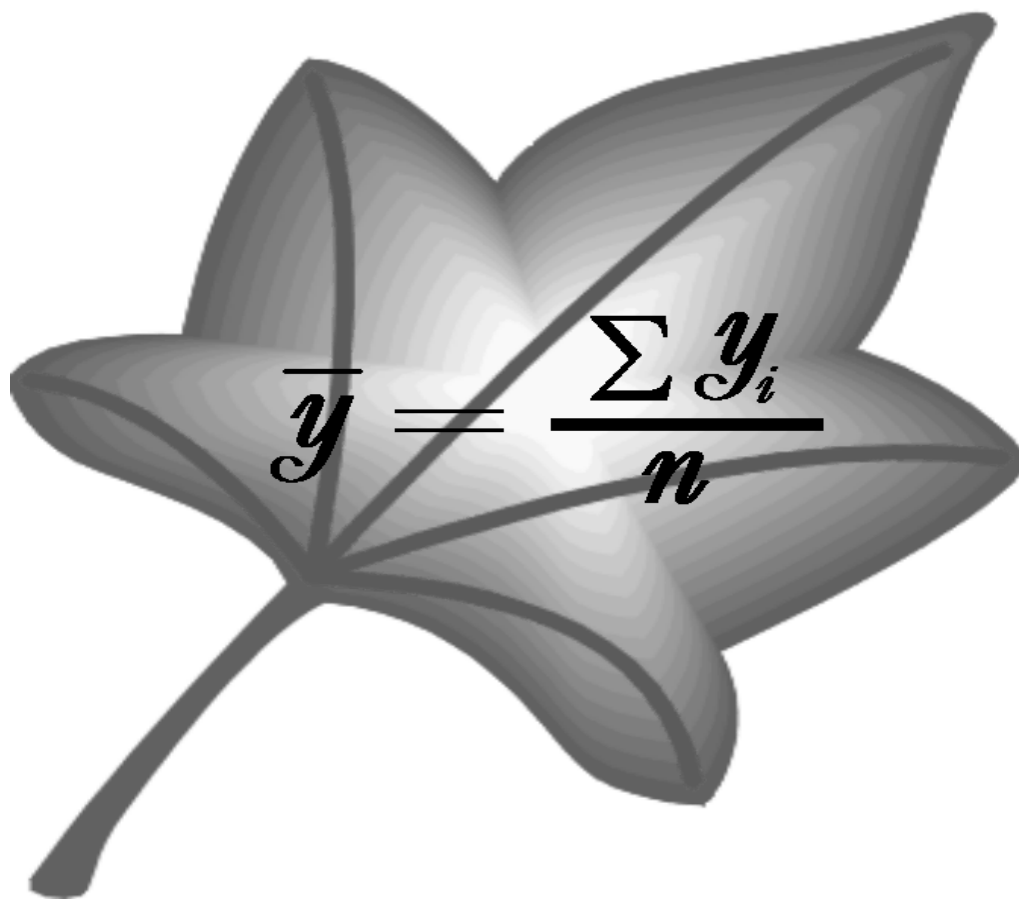


Приседський Ю.Г.

**СТАТИСТИЧНА ОБРОБКА  
РЕЗУЛЬТАТІВ  
БІОЛОГІЧНИХ  
ЕКСПЕРИМЕНТІВ**



Донецьк – 1999

**МІНІСТЕРСТВО ОСВІТИ УКРАЇНИ  
ДОНЕЦЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ**

**ПРИСЕДСЬКИЙ Ю.Г.**

**СТАТИСТИЧНА ОБРОБКА  
РЕЗУЛЬТАТІВ БІОЛОГІЧНИХ  
ЕКСПЕРИМЕНТІВ**

**НАВЧАЛЬНИЙ ПОСІБНИК**

**Донецьк - 1999**

ББК В.172  
П77  
УДК 578.087.1:519.24.001

**Приседський Ю.Г.**

Статистична обробка результатів біологічних експериментів. – Донецьк: Юго-Восток, 1999. – 210 с. (рис. 13, табл. 91, бібліогр. список 20 наймен.).

**ISBN 966-7418-2403**

В учбовому посібнику наводяться методи первинної статистичної обробки результатів біологічних експериментів; дисперсійного, кореляційного та регресійного аналізів; непараметричні методи статистичної обробки. Подані критерії оцінки вірогідності різниці між середніми, в тому числі і для множинних порівнянь, та вірогідності впливу факторів на об'єкти досліджень.

Описані основні методи планування експериментів, які використовуються для створення математичних моделей та оптимізації біологічних процесів.

Навчальний посібник призначено для студентів, аспірантів, наукових працівників, які ведуть дослідження в різних галузях біології.

**ББК В.172**

Рекомендовано Міністерством освіти України як навчальний посібник для студентів біологічних факультетів вищих навчальних закладів освіти.

Відповідальний редактор А.О.Загнітко, д-р філологічних наук, проф.

Рецензенти: А.З.Глухов, д-р біологічних наук, проф.,  
(Донецький ботанічний сад НАН України)  
Б.В.Бондарєв, д-р фізико-математичних наук, проф.,  
(Донецький державний університет)

**ISBN 966-7418-2403**

© Ю.Г.Приседський, 1999

© Юго-Восток

## ВСТУП

Навчальний посібник присвячено деяким розділам біометрії (науки про статистичний аналіз групових властивостей у біології) в її практичному вживанні. Посібник розраховано на те, що студенти вже вивчали курс біометрії і знайомі з основними положеннями статистичної обробки експериментальних даних. Нагадаємо, що термін “статистична обробка” у цьому випадку означає сукупність постулатів та методів теорії ймовірностей та математичної статистики, пристосованих до специфіки біологічних об’єктів та особливостей біологічного досліджу.

Сучасна біологія не може розвиватися без допомоги математики. Математичні методи використовуються передусім для опису біологічних множин (таксономічних розділів, популяцій, штамів, сортів та ін.).

Математика необхідна для вичерпного аналізу інформації про типові об’єкти, їх різноманітність, структуру цієї різноманітності; про системи біологічних взаємовідносин та взаємодій у різних біоценозах; про вплив різних факторів середовища на біологічні об’єкти. Чим глибше вивчаються ці явища, тим більш обґрунтовані висновки можна зробити за допомогою математичного апарату.

Більшість з наведених завдань взагалі не можуть бути вирішені без застосування спеціальних математичних методів. Наприклад, порівняння вибірових груп та визначення вірогідності такого порівняння з певною ймовірністю, визначення достатньої чисельності дослідних об’єктів, сили впливу факторів на біологічні процеси та явища.

В навчальному посібнику наведені основні методи статистичної обробки даних, які використовуються в практиці експерименту в галузі фізіології рослин: первинна статистична обробка даних, дисперсійний аналіз кількісних та якісних ознак, методи кореляційного та регресійного аналізу та деякі методи непараметричної обробки даних. Окремий розділ присвячено основам планування експериментів.

Досліди з фізіології рослин мають певні особливості. Вони потребують тривалого виконання, великої кількості реактивів (інколи досить дорогих) і складного обладнання. Дуже часто дослідник має також обмежену кількість дослідного матеріалу. Все це визначає невеликі об’єми вибірок (рідко вони перевищують 10 вимірів). Тому в посібнику основну увагу приділено методам, які дозволяють проводити статистичний аналіз малочисельних вибірок.

Навчальний посібник розрахований на студентів та аспірантів біологічних факультетів спеціальності "Фізіологія рослин". Він також буде корисним науковим співробітникам та викладачам, які займаються науковою діяльністю.

Автор висловлює щире подяку кандидату математичних наук Є.І.Величко, доктору філологічних наук А.П.Загнітко, доктору біологічних наук М.І.Бойко та кандидату біологічних наук, доценту Г.П.Липницькій за велику допомогу в створенні та виданні посібника.

# РОЗДІЛ 1. СТАТИСТИЧНА ОБРОБКА РЕЗУЛЬТАТІВ БІОЛОГІЧНИХ ЕКСПЕРИМЕНТІВ

## 1. ЗАВДАННЯ, ЯКІ ВИРІШУЮТЬСЯ ПІД ЧАС СТАТИСТИЧНОЇ ОБРОБКИ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ. ГЕНЕРАЛЬНА СУКУПНІСТЬ. ВИБІРКА

Властивості, які вивчаються біометрією, є груповими і мають прояв тільки за умов об'єднання біологічних об'єктів у групи. Групові властивості не можуть бути властивостями окремих особин. Тому ці властивості відрізняють групу як ціле від простої суми включених до неї особин. Якщо отримано вимір ознаки у однієї істоти, то такий вимір може мати велике значення і бути об'єктом вивчення суто біологічних наук (анатомії, гістології, фізіології та ін.). Але одне значення не може стати об'єктом вивчення біометрії, бо один об'єкт не є групою.

Біометричний аналіз можна застосовувати до будь-яких неединичних явищ, об'єднаних у групи будь-якої чисельності, починаючи з  $n=2$ . Але з цього не випливає, що сучасна біологія спроможна робити вірогідні висновки за будь-якої кількості об'єктів. У більшості випадків прогнози, які мають практичне значення у біології, отримані на основі вивчення досить великих груп. Однак, не виключена можливість отримання вірогідних результатів і при вивченні малочисельних груп.

Групові властивості можна розділити на дві категорії – *основні* та *супряжені*.

До *основних* відносяться чотири категорії групових властивостей:

- 1 – *середній рівень ознаки*, який характеризує групу загалом;
- 2 – *різноманітність ознаки*, яка характеризує варіабельність особин у групі;
- 3 – *розподілення ознаки*, яке характеризує певні співвідношення у кількості особин, що мають різні значення ознаки;
- 4 – *репрезентативність вибіркової групи*, яка дає можливість на основі вивчення відносно невеликої групи об'єктів (вибірки) отримати достатньо точну характеристику всієї групи об'єктів (генеральної сукупності).

До *супряжених* належать такі групові властивості, які виявляються внаслідок зв'язку із розвитком основних властивостей.

Якщо супряження двох або більше ознак, які варіюють полягає в тому, що зміни значення середньої величини однієї ознаки відбувається в більшій або меншій залежності від змін середніх рівнів другої (інших) ознаки, то таке групове супряження зветься *регресією*.

Залежність ступеня та структури різноманітності двох або більше ознак, які вивчаються одночасно отримало назву *кореляції* і вимірюється коефіцієнтами парної, окремої, множинної, прямолінійної та нелінійної кореляції.

Окремий вид супряженого різноманіття виявляється у формі впливу факторів, які визначають розвиток ознак, що вивчаються, як функції цих факторів. Сила і вірогідність такого впливу вимірюються показниками *дисперсійного аналізу*, який має для біологічних досліджень особливо велике значення.

Саме цим методам і буде приділено найбільшу увагу в посібнику.

### **Генеральна сукупність та вибірка. Способи добору об'єктів до вибірки.**

Залучення біологічних об'єктів до досліджень можна вести двома основними методами. По-перше, це дослідження всіх особин масиву, який вивчається. По-друге – вивчення певним чином відібраної їх частини. У першому випадку проводиться загальне обстеження всієї *генеральної сукупності*, в другому випадку – *вибіркове дослідження*.

#### **Генеральна сукупність**

Весь масив особин певної категорії, які підлягають вивченню, зветься *генеральною сукупністю*. Об'єм генеральної сукупності визначається завданнями досліджень.

Якщо вивчається який-небудь вид диких рослин або тварин, то генеральною сукупністю будуть всі особини цього виду. У більшості випадків (виключаючи ендемічні або рідкісні види, коли кількість особин може складати декілька сотень або десятків екземплярів) такі генеральні сукупності будуть дуже великими і під час розрахунків приймаються за нескінченно велику величину.

Якщо вивчається дія певного агента (хімічного, фізичного, природного та ін.) на рослини або тварини певної категорії, то генеральну сукупність складають всі особини тієї категорії (виду, статі, господарського призначення), до якої віднесені досліджувані об'єкти. В цьому випадку генеральна сукупність буде більш обмеженою, ніж у першому (скінченою). Але все одно вона залишається недосяжною для суцільного вивчення.

Інколи вивчаються невеликі сукупності, наприклад, череда тварин, що закріплена за певним працівником, або деревні рослини одного виду, які ростуть у парку. Невеликі генеральні сукупності зустрічаються і під час вивчення рослин та тварин, які належать до колекцій, з метою характеристики певної групи цієї колекції. У цьому випадку генеральна сукупність скінчена і може підлягати суцільному вивченню.

Характеристики групових властивостей (середнє значення, варіабільність та ін.), які належать до всієї генеральної сукупності, називаються *генеральними параметрами*. Ці параметри звичайно позначаються літерами грецького алфавіту ( $\gamma$ ,  $\mu$ ,  $\sigma$ )

## Вибірка

**Вибірка** – група об'єктів, які відрізняються трьома особливостями:

а) вибірка – це частина генеральної сукупності;

б) до вибірки відбираються особини у випадковому порядку, але за певними правилами;

в) особини, які відбираються до вибірки, досліджуються з метою характеристики генеральної сукупності.

Для того, щоб за результатами дослідження вибірки можна було отримати достатньо точну характеристику всієї генеральної сукупності, необхідно правильно організувати добір особин з генеральної сукупності.

### Способи добору об'єктів до вибірки

Теорією та практикою розроблено декілька систем добору особин. В основу всіх цих методів покладено прагнення забезпечити максимальну можливість вибору будь-якого об'єкта з генеральної сукупності. Тенденційність, упередженість при доборі об'єктів для вибіркового дослідження заважають отриманню правильних загальних висновків, роблять результати вибіркового дослідження непридатними для характеристики всієї генеральної сукупності, тобто нерепрезентативними.

Наприклад, коли вивчаються властивості виду, сорту, породи з метою виявлення характерних особливостей масивів, не можна добирати лише кращих представників з найліпших умов помешкання; також неправильно для вказаної мети добирати тільки середніх або тільки гірших, або третину кращих, третину середніх та третину гірших особин. За такої організації добору завжди будуть отримані спотворені результати, які не будуть відповідати розподілу ознак у генеральній сукупності. Для отримання неспотвореної характеристики всієї генеральної сукупності необхідно прагнути забезпечити можливість добору у вибірку будь-якої особини з будь-якої частини генеральної сукупності. Ця вимога повинна виконуватися тим суворіше, чим більша мінливість ознаки, яка підлягає вивченню. У різних дослідженнях застосовується п'ять основних методів добору об'єктів у вибірку.

**Випадковий повторний добір.** Об'єкти для вивчення відбираються з генеральної сукупності без попереднього врахування розвитку у них ознаки, тобто у випадковому порядку. Після добору об'єкт вивчається і повертається у свою генеральну сукупність. У цьому випадку будь-який об'єкт досліджень може знову потрапити у вибірку.

**Випадковий безповторний добір.** Об'єкти, які відібрані до групи випадковим чином, не повертаються у генеральну сукупність і не можуть знову потрапити у вибірку. Цей метод найбільш розповсюджений у фізіології рослин.

**Механічний добір.** При цьому методі проводиться добір з окремих частин генеральної сукупності. Ці частини попередньо намічаються меха-

нічно (за ділянками дослідного поля, випадковими групами тварин, які добираються з різних ареалів популяції, та ін.). Використовується стільки таких частин, скільки об'єктів передбачається взяти для вивчення, тому кількість частин завжди дорівнює об'єму вибірки.

**Типовий пропорційний добір.** Цей метод передбачає попереднє вивчення генеральної сукупності. За результатами такого вивчення уся генеральна сукупність поділяється на частини за типом рослинних угруповань, або за рельєфом місцевості та ін. З кожної частини для вивчення відбирається у випадковому порядку кількість об'єктів, пропорційна густоті мешкання об'єкта в окремих частинах.

**Серійний добір.** Генеральна сукупність поділяється на частини – серії, певні з яких досліджуються методом суцільного добору. Цей метод може бути застосованим у тих випадках, коли об'єкти досліджень достатньо рівномірно поширені на певному просторі або на певній території.

Характеристики групових властивостей, отримані під час вивчення вибірок (середнє значення, середнє квадратичне відхилення, помилка репрезентативності та ін.), називаються ***вибірковими показниками*** (вони звичайно позначаються літерами латинського алфавіту типу  $\bar{y}$ ,  $s$ ,  $s^2$  та ін.).

### Властивості вибірок. Їх репрезентативність

При безпосередньому вивченні відібраних об'єктів буде отримано матеріал, який характеризує саму вибірку і має значення первинних фактів. Ці дані підлягають ретельному аналізу та порівнянню з результатами інших робіт. Водночас об'єкти добиралися у вибірку спеціальними методами у достатній кількості, що робить результати вивчення вибірки показовими не лише для самої вибірки, а для усієї генеральної сукупності. Вибірка за певних умов стає відбиттям генеральної сукупності. Це найважливіша властивість вибірки, яку називають ***репрезентативністю***.

Як і будь-яка інша властивість, репрезентативність може бути достатньою або недостатньою. У першому випадку на основі вивчення вибірки будуть отримані вірогідні оцінки генеральних параметрів, що дозволяє розширити коло застосування результатів досліджень на всю генеральну сукупність.

Разом з тим, оцінювання генеральних параметрів за вибірковими показниками має свої особливості. Частина ніколи не може повністю охарактеризувати ціле. Тому оцінка генеральної сукупності за результатами вибіркового дослідження завжди буде мати певну більшу або меншу похибку. Остання складається з кількох складових. Частина з них може бути визначена під час статистичної обробки, а інша частина здебільшого залежить від дослідника, його вміння правильно організувати вибірку та його сумління.

До помилок, яких можна позбутися або звести до мінімального рівня завдяки правильній організації досліджу, належать наступні:



1. **Методичні помилки** – застосування неправильної методики добору та обробки експериментального матеріалу, недосконале проведення хімічних аналізів, відмінність умов експерименту для різних груп об'єктів дослідження.

2. **Помилки точності** – помилки первинної реєстрації фактів, вимірювання неперевіреними або зіпсованими приладами, розрахунки з недостатньою або надлишковою точністю (неправильне округлення проміжних результатів).

3. **Помилки уваги** – описки, прорахунки, пропуски даних, переплутування результатів досліджень, друкарські помилки.

4. **Помилки типовості** з'являються головним чином на початкових етапах досліджень. Вони пов'язані з добором до групи нетипових для генеральної сукупності об'єктів. Це особливо небезпечний вид помилок. Їх можна припуститися несвідомо через нерозуміння правил добору об'єктів до групи (рандомізації). Іншою причиною появи цих помилок (найнебезпечнішою) є тенденційність дослідника, бажання за будь-яку ціну підтвердити свої припущення.

Статистично визначити під час математичної обробки даних можна **помилку репрезентативності**. Ця помилка з'являється завжди, коли за результатами дослідження частини (вибірки) характеризують ціле (генеральну сукупність). Цієї помилки не можливо уникнути, але можна звести до мінімальних значень завдяки дослідженню вибірки необхідного об'єму.

Розуміння сутності помилок репрезентативності застереже від необгрунтованого їх застосування. Визначати величину помилок репрезентативності треба тільки для вибірових показників. Генеральні параметри не мають помилок репрезентативності. Визначення цієї помилки має сенс тільки у тому випадку, коли всі інші помилки або усунуті, або зведені до мінімуму.

При поширенні значень вибірових показників на генеральну сукупність треба враховувати, що величина генерального параметра не може бути виражена одним числом. Існують мінімальна та максимальна межі, в яких можуть знаходитися значення генеральних параметрів. За вибіровими результатами ці межі можна визначити тільки з кінцевою точністю або вірогідністю. Вірогідність, з якою визначаються параметри, залежить від мети та відповідальності досліджень. У біології використовується три рівні вірогідності відповідності вибірового показника генеральному параметру – 95%, 99% та 99.9% (або, використовуючи терміну “рівень значущості”, – 0.95, 0.99 та 0.999). Цим рівням значущості відповідають певні числові значення допусків та поправок, які наводяться у відповідних таблицях стандартних значень (критерії Ст'юдента, Фішера та ін.). Критеріями оцінки генерального параметра за вибіровими значеннями є помилка вибірового середнього (помилка репрезентативності) та надійний інтервал.

## **Обчислення мінімального об'єму вибірок за визначеною помилкою репрезентативності.**

Як уже відзначалося, для отримання прийнятної помилки передбачення генерального параметра за вибірковими показниками необхідно, щоб вибірка була репрезентативною. Яким чином можна домогтися цього, заздалегідь передбачивши репрезентативність вибірки? Перш за все треба так провести добір вибірки, щоб оцінки вибіркових параметрів з достатньою надійністю прогнозували значення параметрів генеральної сукупності. Багато в чому якість такого прогнозування залежить від об'єму вибірки. Чим більша вибірка, тим точніший прогноз. Разом з тим якщо вибірка буде занадто великою, її не можна буде дослідити (провести заміри, визначення хімічного складу та ін.). У такому разі виникає протиріччя між якістю прогнозу та обсягом “фізичної” роботи дослідника.

Розв'язати це протиріччя можна, застосовуючи метод обчислення мінімального об'єму вибірки. Для розрахунку цього показника треба визначити бажану точність спостережень (прийнятне розходження між середньою арифметичною вибірки та середньою арифметичною генеральної сукупності); ступінь надійності (рівень значущості) та ступінь однорідності сукупності об'єктів дослідження (оцінюється за середнім квадратичним відхиленням). Бажана точність – це прийнятне з визначеною ймовірністю відхилення вибіркового середнього від генерального параметра. Цей показник (позначимо його літерою  $D$ ) можна визначити за формулою:

$$D = t \cdot S_{\bar{y}}$$

де:  $t$  – стандартне значення, яке залежить від обраного рівня значущості (при  $\alpha=0.99$   $t$  повинно дорівнювати 3, при  $\alpha=0.95$  –  $t$  дорівнює 2);  $S_{\bar{y}}$  – помилка репрезентативності.

У зв'язку з тим, що

$$S_{\bar{y}} = \frac{s}{\sqrt{n}},$$

можна записати

$$D = t \cdot \frac{s}{\sqrt{n}}.$$

Звідси:

$$n = \frac{t^2 \cdot s^2}{D^2}.$$

Під час розрахунків  $D$  визначається дослідником,  $t$  обирається для необхідного рівня значущості. Окремі труднощі виникають при встановленні середнього квадратичного відхилення. Доведено, що значення цього параметра практично не залежить від об'єму вибірки. Тому (якщо це можливо) його можна визначити досліджуючи вибірки малого об'єму. Якщо це неможливо, потрібно виявити мінімальне та максимальне значення ознаки, яка вивчається, та обчислити розмах її варіації (різницю між цими показниками). Далі, виходячи з того, що варіаційний розмах дорівнює приблизно шести середнім квадратичним відхиленням, розраховують цей показник.

### ***Основні принципи вибору методів статистичної обробки експериментальних даних***

Усі методи статистичної обробки експериментальних даних поділяються на параметричні та непараметричні. До параметричних методів належать, наприклад, обчислення середнього арифметичного та помилки репрезентативності, порівняння середніх за допомогою критеріїв Ст'юдента, Фішера, Дункана та ін., методи дисперсійного, кореляційного та регресійного аналізів. Ці методи пов'язані з певними параметрами розподілу вибірових значень і потребують попереднього визначення цих параметрів. До непараметричних (рангових або порядкових) методів належать рангові оцінки відмінності між вибірками. Серед них є аналоги дисперсійного та кореляційного аналізу, методів множинних порівнянь та ін. Загальною властивістю цих методів є відсутність необхідності перевірки вибірки на нормальність розподілу, розрахунку середніх значень та показників варіації вибірки. Водночас непараметричні методи відрізняються від параметричних меншими точністю та статистичною потужністю. Відтак, поширення висновків, зроблених за результатами такого аналізу, на генеральну сукупність ускладнюється.

Існують певні критерії застосування параметричних та непараметричних методів статистичної обробки. Основним з цих критеріїв є нормальність розподілу даних у вибірці. Непараметричні методи використовуються при відхиленні розподілу від нормального і його не можливо привести до такого шляхом застосування перетворень результатів аналізу. Непараметричні методи можна також використовувати у тих випадках, коли відповідальність досліджень не дуже висока або коли потрібно оцінити тільки тенденцію змін параметрів досліджень. В усіх інших випадках застосовуються параметричні методи.

## 2. ПЕРВИННА СТАТИСТИЧНА ОБРОБКА ДАНИХ

### **Методика обчислення значення вибіркового середнього арифметичного**

Однією з найважливіших характеристик, які варіюють, є середня величина ознаки. Якщо характеризують одну або іншу дослідні групи, говорять про їх середню продуктивність (врожайність), середню інтенсивність фотосинтезу, дихання і так далі. Можна вести мову про середню успішність учнів у класі, школі, середню швидкість біохімічних реакцій та про багато інших середніх значень. Особливість середніх полягає в тому, що вони невілюють індивідуальну різницю об'єктів досліджень, внаслідок чого з'являється більш або менш стійка числова характеристика ознаки – не окремих представників, а всієї статистичної групи. Таким чином, середня величина характеризує групові властивості. В середньому значенні знаходить своє відбиття внутрішній зв'язок, який існує між окремими об'єктами та всією їх сукупністю в цілому. Середнє значення – це центр розподілу; воно займає центральне місце в загальній масі варіативних ознак.

Існує декілька видів середніх. Середні, які використовуються у біології, поділяються на параметричні (ступеневі) та порядкові (непараметричні). До першої групи відносяться середні: арифметична, гармонічна, квадратична та інші. Ці показники мають функціональний зв'язок з розподілом ознак, які варіюють. Порядкові середні такого зв'язку не мають. Вони характеризують лише структурні особливості варіації. До непараметричних середніх належать мода та медіана.

Крім того, розрізняють генеральну середню, яка характеризує усю генеральну сукупність, та вибіркові середні, які відносяться до характеристики вибірових параметрів. У практиці фізіологічного експерименту найчастіше використовується вибіркова середня арифметична величина. Через цей показник опосередковано характеризується генеральна середня.

Обчислення значення будь-якої параметричної середньої проводиться за наступною загальною формулою:

$$\bar{y} = k \sqrt{\frac{\sum y_i^k}{n}},$$

де  $n$  – об'єм вибірки,  $k$  – ступінь середнього значення.

Вибираючи середню, треба враховувати розмірність одиниць виміру ознаки. Якщо це лінійні розміри, або вага, або інші одиниці першого ступеня треба використовувати середню арифметичну. При вимірюванні площі застосовується середня квадратична; об'єму – середня кубічна і так далі. Як бачимо, більшість завдань фізіологічних дослідів може бути

розв'язана за допомогою середньої арифметичної. Її обчислення може здійснюватися кількома способами, наприклад сумуванням дат, зваженої середньої, умовної середньої. Коли є достатня забезпеченість обчислювальною технікою, найчастіше використовується метод сумування дат. Формула для обчислення середньої цим методом має наступний вигляд:

$$\bar{y} = \frac{y_1 + y_2 + y_3 + \dots + y_n}{n} = \frac{\sum y_i}{n},$$

де:  $y_i$  – значення дат у вибірці;  $y$  – значення середньої арифметичної;  $n$  – об'єм вибірки;  $i$  – номер дати у вибірці, який може приймати значення від 1 до  $n$ .

Якщо у вибірці зустрічається по декілька дат з однаковими значеннями ознаки, то можна згрупувати їх за частотою зустрічальності. У цьому випадку формула дещо змінюється:

$$\bar{y} = \frac{\sum p_u \cdot y_u}{n},$$

де:  $y_u$  –  $u$ -те значення ознаки;  $p_u$  – частота зустрічальності  $u$ -того значення ознаки (кількість дат у вибірці, які мають  $u$ -те значення);  $n$  – об'єм вибірки.

### **Розрахунок сум квадратів відхилень та середнього квадратичного відхилення**

Середня величина є найважливішою статистичною характеристикою ознаки. Але вона не подає ніякої уяви про варіювання ознаки у різних об'єктів дослідження. Однак без урахування ступеня варіювання не можна скласти повну характеристику ознаки біологічного об'єкта. Звідси витікає, що одночасно з використанням середніх величин потрібні ще й показники варіації ознак.

Для приблизної оцінки варіювання можна використовувати розмах варіювання, тобто різницю між мінімальним та максимальним значеннями ознаки у досліджуваній сукупності. Найчастіше для оцінки варіювання використовується середній квадрат відхилень (дисперсія, центральний момент другого порядку, варіанса). Цей показник обчислюється за такою загальною формулою:

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{n - 1}.$$

У практичній роботі зручніше користуватися робочими формулами для обчислення цього показника:

$$H = \sum y_i^2 - \frac{(\sum y_i)^2}{n};$$

$$s^2 = \frac{H}{n-1};$$

$$s = \sqrt{s^2}.$$

Величина  $s$  називається середнім квадратичним відхиленням, а величина  $n-1$  має назву кількості ступенів вільності (загальне позначення для цього показника –  $f$ ). Ступінь вільності характеризує кількість дат, які вільно варіюють у вибірці. Для пояснення цього можна скористатися наступним прикладом. Припустимо, що з певної сукупності необхідно відібрати у випадковому порядку три варіанти з тією умовою, що їх сума повинна дорівнювати 30. Зрозуміло, що перші дві варіанти можуть мати будь-яке значення (вибірка підібрана так, що сума будь-яких двох варіант буде менше 30). Вільність їх варіювання нічим не обмежена. Третя варіанта може набути тільки одного значення, яке дорівнює різниці між 30 та сумою перших двох варіант. У такому разі остання варіанта не буде мати ступеня вільності.

У математичній статистиці доведено, що під час розрахунку середньої величини ніяких обмежень варіювання дат не існує. При обчисленні показників варіації та інших статистичних показників один або декілька членів вибірки завжди не мають ступеня вільності. Якщо ці обмеження позначити літерою  $v$ , тоді у загальному вигляді формулу для розрахунку кількості ступенів вільності можна записати у такому вигляді:

$$f = n - v.$$

### ***Перевірка вибірки на нормальність розподілу***

Перевірка вибірки на нормальність розподілу може бути проведена кількома методами. Одні з них використовуються тільки для перевірки великих вибірок, які містять більше, ніж 50 варіант. Окремі з них засновані на обчисленні асиметрії та ексцесу, сутність інших методів полягає у перевірці відповідності розподілу вибіркових дат функції нормального розподілу. Однак в практиці фізіологічного експерименту рідко зустрічаються вибірки з об'ємом більше 20 варіант. Тому в посібнику будуть наведені тільки ті методи, які застосовуються для перевірки нормальності малих за об'ємом вибірок.

#### **Метод середнього абсолютного відхилення**

Цей метод застосовується коли кількість дат не перевищує 120. Послідовність виконання розрахунків наступна:

- обчислюється середнє арифметичне значення та незсунене середнє квадратичне відхилення;
- розраховується відхилення усіх дат від середнього значення та обчислюється їх сума за абсолютною величиною;
- обчислюється середнє абсолютне відхилення:

$$CAB = \frac{\sum |y_i - \bar{y}|}{n},$$

- перевіряється виконання умови

$$\left| \frac{CAB}{s} - 0.7979 \right| < \frac{0.4}{n}.$$

Якщо нерівність виконується, то розподіл вибірки нормальний, якщо ліва частина дорівнює або перевищує праву, тоді гіпотеза про нормальність розподілу вибірки відкидається.

### Метод розмаху

Швидко перевірку нормальності розподілу порівняно широкого класу вибірок ( $3 < n < 1000$ ) можна виконати з допомогою методу розмаху варіювання  $R$ . Для цього:

- знаходять максимальний розмах у вибірці за формулою

$$R = y_{\max} - y_{\min},$$

- розраховується відношення розмаху до зсунутої оцінки середнього квадратичного відхилення, обчислюваного за формулою

$$\bar{S} = \sqrt{\frac{1}{n} \cdot \sum (y_i - \bar{y})^2} = \sqrt{\frac{H}{n}}$$

відношення обчислюється за формулою:

$$L = \frac{R}{\bar{S}}.$$

Далі отримане відношення порівнюють зі стандартними мінімальним та максимальним значеннями, які подаються в спеціальних таблицях (табл. 2а та 2б додатків). Якщо обчислене значення вкладається в межі між мінімальним та максимальним стандартними значеннями, гіпотеза нормальності розподілу зберігається. В іншому випадку – відкидається. Важливо, щоб розрахований показник увійшов до стандартного інтервалу для 10 % рівня безпомилкових прогнозів.

## Критерій Шапіро–Уїлка

У випадку обмеженої кількості повторів для оцінки гіпотези про нормальність розподілу можна рекомендувати використання критерію Шапіро–Уїлка. Для визначення нормальності розподілу в цьому випадку достатньо мати від трьох до п'ятдесяти дат у вибірці. Для розрахунку використовується така послідовність:

– дати у вибірці розташовують у порядку збільшення їх величин, тобто отримують ряд:

$$y_1 \leq y_2 \leq y_3 \leq \dots \leq y_n$$

– обчислюють кількість пар даних у вибірці ( $u$ ):  
якщо кількість парна:

$$u = \frac{n}{2}$$

якщо непарна:

$$u = \frac{n-1}{2}$$

– розраховується допоміжна величина  $b$ :

$$b = a_n \cdot (y_n - y_1) + a_{n-1} \cdot (y_{n-1} - y_2) + \dots + a_{n-u+1} \cdot (y_{n-u+1} - y_u) = \sum a_{n-i+1} \cdot (y_{n-i+1} - y_i)$$

де значення  $a_{n-i+1}$  для  $i=1, \dots, u$  беруться з таблиць стандартних значень множників (табл. 3 додатків). Якщо  $n$  непарне число, дата  $y_{u+1}$  не використовується у розрахунках

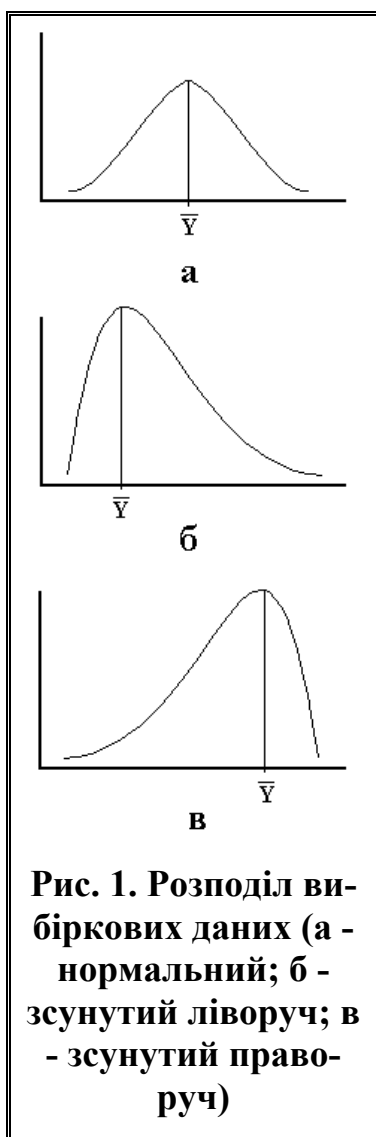
– обчислюється  $W$ -критерій

$$W = \frac{b^2}{H}$$

Цей показник порівнюється зі стандартним значенням критерію Шапіро–Уїлка (табл. 4 додатків) для обраного рівня значущості  $\alpha$  та кількості повторів  $n$  ( $W_{\alpha n}$ ). Якщо отримане значення менше стандартного, гіпотеза про нормальність розподілу відкидається. У біології у більшості випадків за стандартний приймається 5 % рівень значущості.



## Прийняття рішень при відхиленні розподілу від нормального



Якщо гіпотеза про нормальність розподілу не підтвердилася, існує декілька способів вирішення цієї проблеми. По-перше, (це можна зробити ще до початку розрахунків) перевірити існують чи ні обмеження на варіювання ознаки. Якщо такі обмеження існують, то є велика ймовірність того, що розподіл вибірки буде відхилятися від нормального. По-друге, треба спробувати різні способи перетворення дат. У більшості випадків з допомогою перетворення можна привести розподіл вибірки до нормального. По-третє, (це винятковий варіант) переробити дослід з новою більшою кількістю повторювань. Цим способом користуються тоді, коли інші методи не дали позитивних результатів. І останній, четвертий, спосіб – застосування непараметричних методів статистичної обробки. Він застосовується у випадках, коли і збільшення об'єму вибірки, і перетворення не дали позитивного результату.

На початку перетворення велику допомогу у виборі його методу можуть надати гістограма та полігон частот розподілу, тобто графічне відбиття частот дат у вибірці. Аналіз відхилення реальної гістограми від теоретичної може підказати спосіб перетворення. Наприклад, у біологічних дослідженнях часто зустрічається логарифмічна залежність і логарифмічний розподіл. У таких випадках ліва частина полігону крута, а права – похила (рис.1). Логарифмування таких вихідних даних призводить до того, що ліва частина полігону розтягується і розподіл перетворюється на приблизно нормальний. Крім цього методу, у випадках крутої лівої частини рекомендуються деякі інші методи, які здійснюються за формулами:

$$y' = \lg y; \quad y' = \ln y; \quad y' = \lg y \cdot 10^p;$$

$$y' = \lg(y \pm p); \quad y' = 1/y; \quad y' = 1/\sqrt{y}.$$

Перетворення «обернена величина» є найбільш «сильним», а найслабшим є логарифмічне перетворення.

Для нормалізації розподілу, зсунутого праворуч, застосовуються тригонометричні перетворення та перетворення типу

$$y = y^p$$

Значення  $p$  дорівнює 2, якщо розподіл сильно зсунутий праворуч, і 1,5 при помірному зсуві.

У випадках, коли перетворення не допомогло, використовують непараметричні методи або повторюють експеримент з більшою вибіркою.

### **Перевірка вибірки на однорідність**

Іншим важливим показником, щодо визначення якості вибірки, є її однорідність або наявність «випадаючих» варіант. У цьому випадку максимальна або мінімальна дати значно відрізняються від усіх інших і виникає сумнів щодо їх належності до цієї генеральної сукупності. Причини неоднорідності вибірки можуть бути різними. По-перше, можливі технічні помилки, які допускаються під час добору об'єктів у вибірку; по-друге, помилки, пов'язані з якістю проведення експерименту та похибки обладнання; по-третє, наслідок великої варіабельності ознаки, тобто явище цілком природне, яке визначається властивостями ознаки. Якщо дати потрапили у вибірку випадково (перша та друга причини) і до певної генеральної сукупності не належать, їх потрібно відкинути. В останньому випадку цього робити не можна навіть тоді, коли відхилення досить великі.

Існують різні способи статистичної оцінки «випадаючих» дат. Достатньо простим і надійним є метод нормування сумнівних дат по відносно середньої арифметичної вибірки. Нульовою гіпотезою вважається припущення, що підозрілі дати належать до тієї ж генеральної сукупності, що і всі інші дати вибірки. Критерієм оцінки нульової гіпотези є нормоване відхилення:

$$t = \frac{y_i - \bar{y}}{s},$$

де  $t$  – нормоване відхилення;  $y_i$  – підозріла дата;  $\bar{y}$  – вибіркова середня арифметична;  $s$  – незсунуте середньоквадратичне відхилення.

Дата відкидається, якщо її значення виходить за межі надійного інтервалу, встановлюваного для певного рівня ймовірності безпомилкових прогнозів (за правилом «плюс-мінус трьох сигм»). Обчислення надійного інтервалу проводять після розрахунку середньої арифметичної та незсунутого середньоквадратичного відхилення у такій послідовності.

1. У таблиці стандартних значень критерію Ст'юдента для відкидання «випадаючих» варіант (табл. 5 додатків) знаходять значення, яке відповідає обраному рівню значущості та об'єму вибірки.
2. Обчислюють значення критерію за формулою:

$$t = t_{st} \cdot s.$$

3. Розраховують мінімальне та максимальне значення варіаційного ряду:

$$y_{\min} = \bar{y} - t, \quad y_{\max} = \bar{y} + t.$$

4. Крайні значення вибірки порівнюються зі значеннями надійного інтервалу. Якщо їх значення виходять за межі інтервалу, нульова гіпотеза відкидається (дати повинні бути виключені з вибірки), у противному разі нульова гіпотеза зберігається, тобто всі дати вибірки належать до однієї генеральної сукупності, а велика розбіжність між ними зумовлена природним варіюванням.

Ще одним методом перевірки на приналежність дати до вибірки є метод обчислення показника  $\tau$  яке проводиться за такими формулами:

$$\tau_{\max} = \frac{y_{\max} - \bar{y}}{s} \quad \text{та} \quad \tau_{\min} = \frac{\bar{y} - y_{\min}}{s}.$$

Обчислені значення порівнюються зі стандартними (табл. 6 додатків) для обраного рівня значущості та об'єму вибірки. Якщо обчислене значення перевищує стандартне, нульова гіпотеза відхиляється і варіанта повинна бути видалена з вибірки. Для малих вибірок можна рекомендувати спрощений метод оцінки приналежності дати до вибірки. Порівняння ведеться за значеннями крайніх варіант у впорядкованому за зростанням величин ряді  $y(1) > y(2) > \dots > y(n-1) > y(n)$ . Обчислення для перевірки найбільших варіант ведеться за формулою:

$$\tau' = \frac{y(n) - y(n-1)}{y(n) - y(2)},$$

а для перевірки найменших варіант –

$$\tau'' = \frac{y(2) - y(1)}{y(n-1) - y(1)}$$

Отримані значення порівнюються з табличними (табл. 7 додатків). Умовою відхилення нульової гіпотези (дата, що перевіряється, належить до вибірки) є перевищення обчисленого значення над стандартним. Слід врахувати, що останній метод дає надійні результати тільки з малими об'ємами вибірки, коли  $n$  не перевищує 30.

Якщо у вибірці були визначені і вилучені «випадаючі» варіанти, то потрібно знову провести всю первинну статистичну обробку вибірки.

Після того, як отримані позитивні результати тестування вибірок на нормальність розподілу та на однорідність, можна продовжити розрахунки для з'ясування основних показників варіювання вибірки.

## **Розрахунок коефіцієнта варіації, помилки репрезентативності та надійного інтервалу**

Середнє квадратичне відхилення є основною статистичною характеристикою варіабельності ознак. Цей показник не залежить від кількості спостережень і може бути використаним для порівняльної оцінки варіювання ознак. Разом з тим цей показник має один суттєвий недолік для такого використання. Середнє квадратичне відхилення має таку ж розмірність, як середнє значення ознаки. Це обмежує можливість порівняння різних ознак з різною розмірністю. Так, практично не можливо за допомогою середнього квадратичного відхилення порівняти ступінь варіювання, наприклад, довжини тіла та його ваги навіть у особин одного виду. Для того, щоб середнє квадратичне відхилення можна було застосовувати для порівняння варіювань ознак, його потрібно перетворити у відносну величину, якій не властива розмірність. Такою величиною є *коефіцієнт варіації*, який може виражатися у частках одиниці або у відсотках і обчислюється за формулами:

$$CV = \frac{s}{\bar{y}}, \text{ часток од. або } V = \frac{s}{\bar{y}} \cdot 100 \%$$

Наступним важливим показником, який характеризує відповідність вибіркового середнього генеральному параметру, є помилка репрезентативності. Точніше, помилка репрезентативності показує, в яких межах знаходиться генеральна середня відносно вибіркової середньої за обраного рівня значущості. У математичній статистиці доведено, що вибіркові середні варіюють в  $n$  разів менше, ніж окремі дані у вибірці. Звідси витікає, що помилка репрезентативності дорівнює:

$$S_{\bar{y}} = \frac{s}{\sqrt{f}} = \sqrt{\frac{s^2}{f}},$$

де  $f$  – кількість ступенів вільності (для вибірок об'ємом менше 30  $f$  дорівнює  $n-1$ , для вибірок більшого об'єму можна прийняти, що  $f=n$ ).

Разом з тим у математичній статистиці доведено, що помилка репрезентативності забезпечує приблизно 67% ймовірності того, що середня арифметична генеральної сукупності буде знаходитися у розрахованому інтервалі

$$\mu = \bar{y} \pm S_{\bar{y}}.$$

Якщо необхідна точніша оцінка, треба розрахувати надійний інтервал ( $\Delta$ ) для визначеного рівня значущості:

$$\Delta = t_{\alpha, f} \cdot S_{\bar{y}},$$

де  $t_{\alpha, f}$  – стандартне значення критерію Ст'юдента (табл. 1 додатків) для обраного рівня значущості ( $\alpha$ ) та кількості ступенів вільності ( $f$ ). Якщо тепер записати результати розрахунків у вигляді:

$$\bar{y} - \Delta \leq \mu \leq \bar{y} + \Delta \quad \text{або} \quad \mu = \bar{y} \pm \Delta$$

можна буде з обраною ймовірністю стверджувати, що генеральна середня буде знаходитися в обчисленому інтервалі.

### Приклад первинної статистичної обробки

Під час досліджень довжини десяти паростків горобини звичайної були отримані такі результати (см):

*10,1; 11,1; 11,1; 11,9; 14,4; 12,7; 12,2; 9,6; 11,5; 11,8.*

Потрібно обчислити середнє арифметичне значення та показники варіації вибірки. Для цього спочатку обчислюються сума дат та сума квадратів дат:

$$\begin{aligned} \sum y_i &= y_1 + y_2 + \dots + y_n = 10,1 + 11,1 + \dots + 11,8 = 116,4, \\ \sum y_i^2 &= y_1^2 + y_2^2 + \dots + y_n^2 = 102,01 + 123,21 + \dots + 139,24 = 1371,18 \end{aligned}$$

Далі обчислюються середня арифметична, сума квадратів відхилень, середній квадрат відхилень та середнє квадратичне відхилення:

$$\begin{aligned} \bar{y} &= \frac{\sum y_i}{n} = \frac{116,4}{10} = 11,64, \\ H &= \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 1371,18 - \frac{116,4^2}{10} = 1371,18 - 1354,90 = 16,28, \\ s^2 &= \frac{H}{n-1} = \frac{16,28}{10-1} = 1,81, \quad s = \sqrt{s^2} = \sqrt{1,81} = 1,35. \end{aligned}$$

Тепер потрібно перевірити вибірку на нормальність розподілу та однорідність (наявність "випадаючих" варіант).

### Перевірка на нормальність розподілу

Скористаємося методом Шапіро–Уїлка. Для цього вибірка упорядковується за зростанням значень та обчислюється кількість пар:

*9,6; 10,1; 11,1; 11,1; 11,5; 11,8; 11,9; 12,2; 12,7; 14,4*

Оскільки кількість даних парна то

$$u = \frac{n}{2} = \frac{10}{5} = 5$$

Розраховується значення  $b$  та критерію нормальності розподілу  $W$  (коefficientи беруться з таблиць):

$$b = a_n \cdot (y_n - y_1) + a_{n-1} \cdot (y_{n-1} - y_{1=2}) + \dots = 0,5601 \cdot (14,4 - 9,6) + 0,3315 \cdot (12,7 - 10,1) + 0,2260 \cdot (12,2 - 11,1) + 0,1429 \cdot (11,1 - 11,1) + 0,0695 \cdot (11,8 - 11,5) = 3,9418$$

$$W = \frac{b^2}{H} = \frac{3,9418^2}{16,28} = 0,954$$

Оскільки обчислене значення критерію перевищує стандартне (0,842), розподіл вибірки визнається нормальним.

### *Перевірка вибірки на однорідність*

Спочатку обчислюється допуск:

$$t = t_{st} \cdot s = 2,37 \cdot 1,35 = 3,20,$$

де  $t_{st}$  – стандартне значення з таблиці Ст'юдента для перевірки вибірки на однорідність. Далі розраховується надійний інтервал:

$$\Delta_{\min} = \bar{y} - t = 11,64 - 3,20 = 8,44, \quad \Delta_{\max} = \bar{y} + t = 11,64 + 3,20 = 14,84$$

Використаємо також й інший метод перевірки. Впорядковується ряд за зростанням значень:

$$9,6; 10,1; 11,1; 11,1; 11,5; 11,8; 11,9; 12,2; 12,7; 14,4$$

Перевіряється приналежність крайніх варіант (9,6 та 14,4) до вибірки. Для цього обчислюються значення:

$$\tau' = \frac{y_{(n)} - y_{(n-1)}}{y_{(n)} - y_{(2)}} = \frac{14,4 - 12,7}{14,4 - 10,1} = \frac{1,7}{4,3} = 0,395 \text{ та}$$

$$\tau'' = \frac{y_{(2)} - y_{(1)}}{y_{(n-1)} - y_{(1)}} = \frac{10,1 - 9,6}{12,7 - 9,6} = \frac{0,5}{3,1} = 0,161$$

Отримані значення порівнюються зі стандартними для рівня значущості  $\alpha=0,05$  та об'єму вибірки  $n=10$ . Це значення (табл. 7 додатків) дорівнює **0,597**. Обидва обчислені критерії менші за стандартні значення. Тому нульова гіпотеза зберігається (крайні варіанти належать до вибірки).

Таким чином, вибірка однорідна і можна переходити до подальшої обробки, яка включає розрахунок помилки репрезентативності, надійного інтервалу і, якщо потрібно, коефіцієнту варіації:

$$S_{\bar{y}} = \sqrt{\frac{s^2}{n-1}} = \sqrt{\frac{1,81}{10-1}} = 0,45 ,$$

$$\Delta = S_{\bar{y}} \cdot t_{st} = 0,45 \cdot 2,26 = 1,02 ,$$

$$CV = \frac{s}{\bar{y}} = \frac{1,35}{11,64} = 0,16 , \text{ або } CV = \frac{s}{\bar{y}} \cdot 100\% = \frac{1,35}{11,64} \cdot 100 = 16\% .$$

Оскільки об'єм вибірки менше 30, для обчислення помилки репрезентативності береться не кількість вимірів, а кількість ступенів вільності ( $n-1$ ). Тепер результати аналізу можна записати:

$$\mu = \bar{y} \pm \Delta = 11,64 \pm 1,02 ,$$

тобто середня арифметична генеральної сукупності з ймовірністю 95% буде знаходитися в межах від 10,62 см до 12,66 см.

### 3. ДИСПЕРСІЙНИЙ АНАЛІЗ

#### *Основи дисперсійного аналізу.*

Дисперсійний аналіз – метод статистичної обробки експериментальних даних, який проводиться з метою визначення існування впливу різних факторів на ознаку, яка вивчається.

У біологічній статистиці, як і взагалі у біології, *факторами* називають один або декілька можливих способів впливу на об'єкт досліджень. За природою всі фактори поділяються на кілька категорій. Перш за все вирізняють *біотичні* та *абіотичні* фактори. До абіотичних належать фактори, пов'язані з дією неживих об'єктів або явищ. Це можуть бути кліматичні умови (температура, вологість, освітлення та ін.), забруднення навколишнього середовища, хімічні та фізичні властивості живильних середовищ та інші фактори. До біотичних належать фактори, пов'язані з впливом життєдіяльності інших організмів на об'єкти, які підлягають вивченню. Це може бути ураження патогенами, симбіотичні, алелопатичні взаємовідносини між живими організмами, хижацтво та інше. Фактори також поділяються на *кількісні* та *якісні*. Кількісними є такі фактори, значення яких підлягає вимірюванню і може бути охарактеризовано певним числовим значенням. Прикладом таких факторів є температура, освітленість, рН середовища та інші. Якісні фактори не підлягають числовому опису, як, наприклад, лісництво, поле, характер ландшафту і таке інше. Важливою для дослідника властивістю факторів є їх контрольованість. До *контрольованих* факторів належать такі, значення яких дослідник може контролювати за допомогою вимірювальних приладів та встановлювати і підтримувати на певному рівні протягом тривалого часу. Так, наприклад, можна підтримувати визначену температуру в термостаті. У кліматичних камерах регулюються і підтримуються температура, вологість повітря, інтенсивність освітлення. У *неконтрольованих* факторів можна тільки встановити їх значення, але дослідник не може їх регулювати і підтримувати. Прикладом останніх можуть слугувати погодні умови. У дослідках такі фактори найчастіше є заважаючими, оскільки збільшують випадкову дисперсію. Для зменшення їхньої дії на об'єкт досліджень використовують спеціальні плани експериментів. Вони дозволяють під час проведення обчислень урахувати вплив неконтрольованих факторів і вилучити його з випадкової дисперсії.

Ознаки, які вивчаються дослідниками, також поділяються на *кількісні* і *якісні*. Кількісні ознаки виражаються числовим значенням, яке може бути отримане в результаті вимірів або обчислень за допомогою формул за результатами вимірів, і мають певну розмірність. До таких ознак належать показники росту (довжина, вага та ін.), фізіологічні та біохімічні показники (інтенсивність фотосинтезу, транспірації, активність ферментів та ін.).



Якісні показники не підлягають виміру і є числами, які становлять частки або відсотки від цілого. До перших, наприклад, можуть бути віднесені смакові якості плодів, грибів, пахощі різних квітів та інші. Для проведення розрахунків з такими ознаками розробляються спеціальні методи числового їх вираження. Найчастіше використовуються бальні системи, коли за певними градаціями якості закріплюються певні бали (смачний – 5, не дуже смачний – 4, несмачний – 3; дуже несмачний – 2 і т.д.). До других відносять такі показники, як кількість хворих людей із загальної кількості обстежених, кількість пошкоджених рослин в масиві і таке інше. Відповідно до цього дисперсійний аналіз поділяється на аналіз кількісних ознак та аналіз якісних ознак.

За кількістю факторів, які вивчаються у дисперсійному аналізі, його поділяють на однофакторний та багатфакторний (двофакторний, трифакторний). Слід враховувати, що зі збільшенням кількості факторів обчислення швидко ускладнюються, а об'єм розрахунків значно зростає. Тому на практиці дисперсійний аналіз з кількістю факторів більше трьох замінюють спеціальним багатфакторним регресійним аналізом, який проводиться за спеціальними планами. Складність обчислень багато в чому залежить також від співвідношення об'ємів, вибірок, які складають дисперсійний комплекс. Простіше проводити дисперсійний аналіз у тому випадку, коли об'єми (кількість дат) всіх вибірок комплексу однакові – *рівномірний комплекс*. Такі комплекси називають ще *ортогональними*, тому що в них виконуються деякі дуже важливі умови, (див. пізніше). У нерівномірних (неортогональних) комплексах ці умови не виконуються і обчислення значно ускладнюються. Крім того, для ортогональних комплексів існує більша кількість методів перевірки гіпотез про однорідність дисперсій та порівняння середніх. Тому, якщо це можливо, треба планувати досліди так, щоб можна було використовувати статистичну обробку ортогональних комплексів.

Для проведення найпростішого дисперсійного аналізу потрібні хоча б дві вибірки, які будуть складати дисперсійний комплекс. Кожна з цих вибірок складається з даних, отриманих при вивченні впливу на ознаку двох різних значень фактора. Але більш значну статистичну інформацію можна отримати при кількості значень фактора, яка перевищує три. У дисперсійному аналізі кожне значення фактора прийнято називати *рівнем* або *градацією*. Метою дисперсійного аналізу є встановлення вірогідності відмінностей між значеннями ознаки в межах кожної градації фактора або факторів в багатфакторному аналізі. Порівняння вибірок здійснюється за допомогою обчислення показників варіювання, які залежать від впливу фактора (факторний середній квадрат відхилень), впливу випадкових умов та мінливості самого об'єкта досліджень (випадковий квадрат відхилень). Якщо факторне варіювання вірогідно перевищує випадкове варіювання,

вплив фактора вважається вірогідним. В іншому випадку робиться висновок про відсутність дії фактора на об'єкт дослідження.

Під час проведення дисперсійного аналізу повинні виконуватися певні умови варіювання ознак. Перш за все, у всіх вибірках варіювання значень ознак повинно підкорятися закону нормального розподілу. В іншому випадку можуть бути отримані хибні результати, які не дозволяють зробити правильний висновок. Треба підкреслити, що показники дисперсійного аналізу дуже чутливі до нормальності розподілу значень у вибірках. Методи перевірки вибірок на нормальність розподілу наведені у попередньому розділі.

Іншим важливим показником, який може обмежити застосування дисперсійного аналізу, є однорідність вибірових дисперсій (вибірових квадратів відхилень). Якщо різниця між вибіровими дисперсіями перевищує певний рівень, це означає, що вибірки належать до різних генеральних сукупностей і їх порівняння не є можливим. Тому проведенню дисперсійного аналізу обов'язково повинні передувати перевірки вибірок на нормальність розподілу та всього комплексу на однорідність – на однорідність дисперсій.

### ***Перевірка гіпотези про однорідність дисперсій***

Перевірка гіпотези про *однорідність дисперсій* проводиться для встановлення приналежності вибірок до однієї генеральної сукупності і можливості об'єднання їх в дисперсійний комплекс. У літературі цей тест ще називають перевіркою на *рівність дисперсій*. Але у цьому разі термін "рівність дисперсій" не означає їх математичної рівності. Різниця між вибіровими середніми квадратами може бути досить суттєвою, але вона не повинна перевищувати певного критичного значення. Тому термін "однорідність" у цьому випадку є прийнятнішим. Вибір методу перевірки залежить від ортогональності комплексу. Для тестування ортогональних комплексів найприйнятнішими є методи Кохрена, Хартлі, Леслі і Брауна.

#### **Метод Кохрена**

Критерій Кохрена обчислюється за формулою:

$$G = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \dots + s_k^2} = \frac{s_{\max}^2}{\sum s_i^2},$$

де:  $\sum S_i^2$  – сума всіх  $k$  вибірових середніх квадратів порівнюваних відхилень;

$s_{\max}^2$  – значення максимального за величиною середнього квадрата відхилень;

$k$  – кількість градацій фактора.

Отримане значення  $G$  порівнюють зі стандартним критичним значенням  $G_{\alpha,k,f}$  ( $\alpha$  – обраний рівень значущості,  $k$  – кількість порівнюваних середніх квадратів відхилень,  $f$  – кількість ступенів вільності для кожної вибірки  $(n-1)$ ). Якщо розраховане значення критерію Кохрена більше за стандартне (табл. 9 додатків), гіпотеза про однорідність дисперсій відкидається.

### Критерій Хартлі

Обчислюється відношення:

$$F = \frac{s_{\max}^2}{s_{\min}^2}.$$

Якщо обчислене значення перевищує стандартну критичну величину (табл. 10 додатків)  $F_{\alpha,k,f}$  ( $s_{\min}^2$  – мінімальний за значенням вибіркової середньої квадрат відхилень, інші позначення ті ж самі, що і у попередньому методі), гіпотеза про однорідність середніх квадратів відхилень відхиляється.

### Критерій Леслі і Брауна

Для оцінки однорідності дисперсій за методом Леслі і Брауна розраховують розмахи варіювання усіх  $k$  вибірок комплексу ( $R_k$ )

$$R_k = y_{k,\max} - y_{k,\min}$$

Далі знаходять максимальний та мінімальний розмахи і обчислюють їх відношення:

$$Q = \frac{R_{\max}}{R_{\min}}$$

Якщо обчислене значення  $Q$  перевищує критичне значення  $Q_{\alpha,k,f}$  (табл. 11 додатків), гіпотеза про однорідність вибірових дисперсій відхиляється. Цей тест зручний тим, що дозволяє перевірити однорідність дисперсій ще до початку статистичної обробки і може допомогти з'ясувати доцільність подальших обчислень.

Для перевірки гіпотези про однорідність вибірових середніх квадратів відхилень у неортогональних комплексах використовується критерій Бартлета.

## Критерій Бартлета

Результати обчислень критерію Бартлету значною мірою залежать від функції розподілу вибірових дат. Тому перед проведенням розрахунків критерію обов'язково проводиться перевірка вибірок на нормальність розподілу. Обчислення критерію виконують за формулою:

$$\chi^2 = \frac{1}{C} \cdot \left[ 2,3026 \cdot \left( f \cdot \lg s^2 - \sum_{i=1}^k f_i \cdot \lg s_i^2 \right) \right] = \frac{1}{C} \cdot \left( f \cdot \ln s^2 - \sum_{i=1}^k f_i \cdot \ln s_i^2 \right),$$

$$C = \frac{\sum_{i=1}^k \frac{1}{f_i} - \frac{1}{f}}{3 \cdot (k-1)} + 1, \quad s^2 = \frac{\sum_{i=1}^k f_i \cdot s_i^2}{f}, \quad f = N - k, \quad N = \sum_{i=1}^k n_i$$

де:  $N$  – загальний об'єм комплексу;

$k$  – кількість вибірок, кожна з яких повинна мати не менше п'яти повторностей;

$s^2$  – зважене середнє квадратичне відхилення;

$f_i$  – кількість ступенів вільності  $i$ -тої вибірки;

$s_i^2$  – середній квадрат відхилень  $i$ -тої вибірки.

Нульова гіпотеза (всі дисперсії в комплексі однорідні) відкидається у випадках, коли обчислене значення  $\chi^2$  більше критичного значення  $\chi^2_{\alpha, k-1}$  (табл. 12 додатків).

Якщо за результатами одного з тестів зроблено висновок про неоднорідність середніх квадратів відхилень, подальші розрахунки втрачають сенс. У цьому випадку треба проаналізувати отримані результати досліджень. Особливо ретельний аналіз потрібен для вибірок з максимальною дисперсією або середніми квадратами відхилень, що дуже відрізняються від усіх інших. Після такого аналізу приймається рішення про повторне проведення або всіх дослідів, або їх частини.

### **Однофакторний дисперсійний аналіз**

У однофакторних комплексах вивчається дія на об'єкт дослідження лише одного фактора. При цьому дослідником визначаються градації фактора, які підтримуються протягом усього терміну дослідження. Важливою умовою проведення дослідження є виключення або зведення до мінімуму впливу факторів, які заважають. Для цього всі градації фактора досліджуються одночасно і за однакових умов. Слід враховувати, що інтерпретація результатів дисперсійного аналізу можлива тільки в межах варіювання значень фактора. Спроби поширення результатів впливу за межі мінімального або максимального значення фактора дуже часто приводить до похибки. Напри-

клад, вивчалася дія температури на ріст міцелію гриба. У межах від 15 до 35°C спостерігався позитивний вплив температури, тобто ріст міцелію прискорювався. Створюється враження, що чим вище температура, тим краще росте гриб. Але якщо підняти температуру до 45–55°C, ріст міцелію пригнічується, а при 60–70°C гриб взагалі гине. Рівні багатьох факторів, які діють на біологічні об'єкти, можна поділити на *субоптимальні, оптимальні, супероптимальні, сублетальні та летальні*. При збільшенні рівня фактора в межах субоптимальних значень спостерігається їх позитивний вплив на об'єкт до досягнення оптимального рівня. Супероптимальні рівні спочатку практично не діють на значення ознак, а в подальшому можуть пригнічувати об'єкт досліджень. Летальні рівні викликають смерть організму. Ці властивості факторів треба обов'язково враховувати при визначенні рівнів фактора, виходячи з мети досліджу.

### Схема однофакторного дисперсійного аналізу

Однофакторний дисперсійний аналіз проводиться за загальною для ортогональних та неортогональних комплексів схемою. Завдання полягає в обчисленні загального, факторного та випадкового варіювання ознаки і встановленні вірогідності різниці між факторним та випадковим варіюванням. В однофакторному аналізі факторне варіювання залежить тільки від впливу фактора, дія якого на об'єкт досліджень вивчається. Випадкове варіювання складається з природної мінливості об'єкта та впливу заважаючих факторів. Для визначення цих показників необхідно розрахувати певні допоміжні величини (суми квадратів, кількості ступенів вільності, вибіркові показники варіації). На першому етапі дисперсійного аналізу проводиться первинна статистична обробка вибірових даних з кожної градації фактора. Далі перевіряється однорідність дисперсій у комплексі. Після цього обчислюються суми квадратів з різних видів варіювання. Для цього можна використовувати деякі проміжні результати первинної статистичної обробки – суми вибірових дат, суми квадратів вибірових дат та квадрати сум дат. Їх також можна обчислювати і безпосередньо в дисперсійному аналізі. Також встановлюється об'єм комплексу, тобто сума всіх об'ємів вибірок, включених в комплекс. Формули для обчислення цих показників такі:

$$N = n_1 + n_2 + \dots + n_k = \sum n_i ,$$

(для рівномірних комплексів для обчислення загального об'єму можна також користуватися формулою:  $N = k \cdot n$ ),

$$D_y = y_{11} + y_{12} + \dots + y_{kn_k} = \sum_{i=1}^k \sum_{j=1}^{n_k} y_{ij} ,$$

$$D_{y^2} = y_{11}^2 + y_{12}^2 + \dots + y_{kn_k}^2 = \sum_{i=1}^k \sum_{j=1}^{n_k} y_{ij}^2,$$

$$H_k = \frac{\left( \sum_{i=1}^{n_k} y_{ki} \right)^2}{n_k}, \quad \sum_{i=1}^k H_i = H_1 + H_2 + \dots + H_k.$$

Після обчислення цих показників, розраховуються загальна, факторна та випадкова суми квадратів відхилень:

$$SS_y = D_{y^2} - \frac{D_y^2}{N} \text{ – загальна сума квадратів;}$$

$$SS_x = \sum_{i=1}^k H_i - \frac{D_y^2}{N} \text{ – факторна сума квадратів;}$$

$$SS_z = D_{y^2} - \sum_{i=1}^k H_i = SS_y - SS_x \text{ – випадкова сума квадратів.}$$

Для подальших розрахунків необхідно обчислити відповідні ступені вільності:

$$\begin{aligned} \text{для загального варіювання – } f_y &= N - 1; \\ \text{для факторного варіювання – } f_x &= k - 1; \\ \text{для випадкового варіювання – } f_z &= N - k. \end{aligned}$$

І, нарешті, обчислюють середні квадрати відхилень для факторного та випадкового варіювань:

$$S_x^2 = \frac{SS_x}{f_x} \text{ – факторний середній квадрат відхилень та}$$

$$S_z^2 = \frac{SS_z}{f_z} \text{ – випадковий середній квадрат відхилень.}$$

Після проведення всіх цих обчислень можна визначити *показник сили впливу*, який вказує на частку або відсоток факторного варіювання в загальному варіюванні і розраховується за формулою:

$$\eta_x^2 = \frac{SS_x}{SS_y} = 1 - \frac{SS_z}{SS_y}.$$

У такому вигляді буде отримано показник сили впливу, який виражається частками одиниці, а якщо його помножити на 100, ми отримаємо показник

сили впливу, виражений у відсотках. Як і будь-який біометричний показник, показник сили впливу обчислюється з певною помилкою, яка може бути розрахована за формулою:

$$S_{\eta}^2 = (1 - \eta^2) \cdot \frac{k - 1}{N - k}$$

Можна також обчислити надійний інтервал для показника сили впливу. Для цього потрібно отриману помилку показника сили впливу помножити на стандартне значення критерію Фішера ( $F_{\alpha, f_x, f_z}$ ).

При обчисленні показників дисперсії зручно користуватися таблицею, до якої заносяться експериментальні дані та проміжні результати обчислень (табл. 1).

Таблиця 1.

Однофакторний дисперсійний комплекс

Дати ( $y_{ki}$ )	Градації фактора					Кількість градацій	Суми квадратів
	1	2	3	...	k		
	$y_{11}$	$y_{21}$	$y_{31}$	...	$y_{k1}$	$k=4$	$SS_y$
	$y_{12}$	$y_{22}$	$y_{32}$	...	$y_{k2}$		$SS_x$
	...	...	...	...	...		$SS_z$
	...	...	...	...	...		$S_x^2$
	$y_{1n1}$	$y_{2n2}$	$y_{3n3}$	...	$y_{knk}$		$S_z^2$
об'єми вибірок	$n_1$	$n_2$	$n_3$	...	$n_k$	Об'єм комплексу $N = \sum n_k$	
$\sum y_{ki}$	$\sum y_{1i}$	$\sum y_{2i}$	$\sum y_{3i}$	...	$\sum y_{ki}$	$D_y = \sum \sum y_{ki}$	$F = S_x^2 / S_z^2$
$H_i$	$H_1$	$H_2$	$H_3$	...	$H_k$	$\sum H_i$	
$\sum y_{ki}^2$	$\sum y_{1i}^2$	$\sum y_{2i}^2$	$\sum y_{3i}^2$	...	$\sum y_{ki}^2$	$D_{y^2} = \sum \sum y_{ki}^2$	

Оцінювання показників дисперсії. Критерій Фішера

Дисперсійний аналіз дозволяє розкласти загальне варіювання дисперсійного комплексу ( $SS_y$ ) на складові частини – факторне варіювання ( $SS_x$ ), яке залежить від дії фактора, та випадкове варіювання ( $SS_z$ ), яке залежить від мінливості об'єкта досліджень і впливу неконтрольованих факторів оточуючого середовища. Після обчислення показника сили впливу можна оцінити, яку частку в загальному варіюванні займає дія фактора. Але всі ці показники не містять інформації про вірогідність впливу фактора. Для то-

го, щоб довести, що фактор дійсно впливає на значення *результативної ознаки* (ознака, яка підлягає вивченню при проведенні досліду), потрібно розрахувати показник вірогідності впливу. Вплив фактора вважається вірогідним, якщо факторний середній квадрат відхилень вірогідно перевищує випадковий квадрат відхилень або показник сили впливу вірогідно перевищує свою помилку. Попередню оцінку можна зробити, порівнявши факторний середній квадрат відхилень ( $S_x^2$ ) з випадковим середнім квадратом відхилень ( $S_z^2$ ). Якщо перший показник менший або дорівнює другому, можна відразу зробити висновок про відсутність вірогідного впливу фактора. В іншому випадку потрібно провести перевірку вірогідності впливу з допомогою *критерію Фішера*, який використовується в біологічній статистиці для встановлення вірогідності різниці між двома дисперсіями дуже часто і обчислюється за формулою:

$$F = \frac{S_x^2}{S_z^2}.$$

Якщо для встановлення вірогідності впливу використовується показник сили впливу, то розрахунки проводяться за формулою:

$$F = \frac{\eta^2}{S_\eta^2}.$$

Отримане значення  $F$  порівнюється з критичним значенням  $F_{\alpha, f_x, f_z}$  (табл. 14 додатків) для визначеного рівня значущості  $\alpha$  та кількості ступенів вільності факторного варіювання ( $f_x$ ), і випадкового варіювання ( $f_z$ ). Якщо обчислене значення перевищує або дорівнює критичному стандартному значенню, вплив фактора визнається вірогідним на обраному рівні значущості ( $\alpha$ ). В іншому випадку визнається відсутність впливу фактора на об'єкт досліджень.

## Приклад

Досліджувалася дія фізіологічно активної речовини на накопичення органічної речовини паростками робінії псевдоакації. Отримані результати, наведені в табл. 2. Потрібно з'ясувати наявність впливу дослідної речовини на результативну ознаку. До цієї ж таблиці записуються і проміжні результати.

Аналіз середніх значень (останній рядок таблиці) дає змогу припустити, що існує позитивний вплив фізіологічно активної речовини на накопичення біомаси рослинами. Для того, щоб підтвердити або спростувати це припущення, обчислимо показники дисперсійного аналізу:



$$SS_y = \sum y^2 - \frac{D_y^2}{N} = 9060 - \frac{(418)^2}{24} = 9060 - 7280,17 = 1779,83,$$

$$SS_x = \sum H_i - \frac{D_y^2}{N} = 8861,60 - 7280,17 = 1581,43,$$

$$SS_z = SS_y - SS_x = 1779,83 - 1581,43 = 198,50,$$

Таблиця 2

Вплив фізіологічно активної речовини на накопичення маси паростками робінії псевдоакації

Дати ( $y_{ki}$ )	Градації фактора				Кількість градацій
	1	2	3	4	
1	13,0	7,0	21,0	29,0	$k=4$
2	7,0	11,0	19,0	27,0	
3	8,0	9,0	18,0	24,0	
4	9,0	9,0	14,0	34,0	
5	10,0	10,0	23,0	27,0	
6	8,0		27,0	31,0	
7			23,0		
Об'єми вибі- рок ( $n_k$ )	6	5	7	6	$N=6+5+7+6$ $=24$
$\sum y_{ki}$	55	46	145	172	$D_y=418$
$H_i=(\sum y_{ki})^2/n_k$	504,16	423,20	3003,57	4930,67	$\sum H_i=8861,60$
$\sum y_{ki}^2$	527	432	3109	4992	$D_y^2=9060$
$\bar{y}$	9,17	9,20	20,71	28,67	

$$f_x = k - 1 = 4 - 1 = 3, \quad f_z = N - k = 24 - 4 = 20,$$

$$S_x^2 = \frac{SS_x}{f_x} = \frac{1581,43}{3} = 527,14, \quad S_z^2 = \frac{SS_z}{f_z} = \frac{198,50}{20} = 9,93,$$

$$F_x = \frac{S_x^2}{S_z^2} = \frac{527,14}{9,93} = 53,08$$

Обчислене значення критерію вірогідності впливу дорівнює 53,08, що значно перевищує стандартне значення критерію Фішера (3,1). Таким чином, вплив фізіологічно активної речовини є вірогідним, тобто ця речовина дійсно позитивно впливає на накопичення органічної речовини у рослин робінії псевдоакації.

## **Двофакторний дисперсійний аналіз**

Як вже повідомлялося раніше, якщо вивчається одночасна дія на об'єкт досліджень кількох факторів, то такі комплекси називаються багатофакторними. Простішим випадком багатофакторних комплексів є двофакторні. В них досліджується дія двох факторів на результативну ознаку об'єкта досліджень. Під час статистичної обробки двофакторних комплексів загальний факторний вплив розкладається на вплив окремих факторів, а також визначається вплив *взаємодії* факторів. Під взаємодією розуміють залежність дії одного фактора від рівня, на якому знаходиться інший фактор. Якщо позначити фактори літерами **A** та **B**, то можна записати відповідну формулу варіювання ознак:

$$SS_y = SS_x + SS_z = SS_A + SS_B + SS_{AB} + SS_z.$$

Для того, щоб обчислити ці показники необхідно дослідити всі рівні фактору **A** при всіх рівнях фактора **B**. Тобто план досліджень повинен бути складений за формою, поданою в табл. 3.

Таблиця 3

План досліджень двофакторного комплексу

<b>Градації фактору A</b>	<b>A<sub>1</sub></b>			<b>A<sub>2</sub></b>			<b>A<sub>3</sub></b>		
<b>Градації фактору B</b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>

Наприклад, вивчалася дія рН поживного середовища на ріст міцелію гриба коренева губка при різних температурах. Для проведення досліду були обрані такі значення рН (фактор **A**): 6,0 (**A<sub>1</sub>**); 7,0 (**A<sub>2</sub>**); 8,0 (**A<sub>3</sub>**) та температури (фактор **B**): 25°C (**B<sub>1</sub>**); 30°C (**B<sub>2</sub>**) та 35°C (**B<sub>3</sub>**). Для того, щоб можна було провести двофакторний дисперсійний аналіз, треба виконати 9 дослідів з такими умовами: 1) *pH*=6,0, *t*=25 °C; 2) *pH*=6,0, *t*=30 °C; 3) *pH*=6,0, *t*=35 °C; 4) *pH*=7,0, *t*=25 °C; 5) *pH*=7,0, *t*=30 °C; 6) *pH*=7,0, *t*=35 °C; 7) *pH*=8,0, *t*=25 °C; 8) *pH*=8,0, *t*=30 °C; 9) *pH*=8,0, *t*=35 °C. У цьому випадку загальне факторне варіювання можна розкласти на складові частини, наведені у формулі. Тобто можна з'ясувати, як впливають на ріст гриба температура та рН поживного середовища, а також визначити вплив температури на напрям дії рН середовища.

Інколи таке планування досліду неможливе. У цьому разі рівням одного фактора (наприклад, фактора **A**) відповідають неоднакові рівні іншого фактора (фактора **B**). Такі комплекси зветься *ієрархічними*. Схема такого експерименту подана в табл. 4.

За результатами статистичної обробки ієрархічних комплексів можна встановити вплив на результативну ознаку фактора  $A$  та фактора  $B$ , але не є можливим встановлення їх взаємодії, тобто формула варіювання буде такою:

$$SS_y = SS_x + SS_z = SS_A + SS_B + SS_z.$$

Таблиця 4

Схема ієрархічного двофакторного дисперсійного комплексу

<b>Градації фактору <math>A</math></b>	<b>A<sub>1</sub></b>			<b>A<sub>2</sub></b>		<b>A<sub>3</sub></b>		<b>A<sub>4</sub></b>
<b>Градації фактору <math>B</math></b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>B<sub>6</sub></b>	<b>B<sub>7</sub></b>	<b>B<sub>8</sub></b>

Такі комплекси дуже часто зустрічаються під час дослідження біологічних об'єктів у природних умовах. Наприклад, вивчався вміст білка в насінні сосни різного кольору (темне та світле забарвлення насінин) з різних лісництв. У цьому випадку фактор  $A$  (колір насінин) має дві градації, а фактор  $B$  (лісництво) буде мати стільки градацій, скільки лісництв досліджено.

Методики проведення дисперсійного аналізу різняться для ортогональних та неортогональних комплексів. Далі будуть наведені методи проведення обчислень для двофакторних ортогональних, неортогональних та ієрархічних комплексів.

#### Дисперсійний аналіз двофакторних ортогональних комплексів

До ортогональних комплексів належать рівномірні та пропорційні комплекси. У рівномірних комплексах усі вибірки, які складають комплекс, мають однакову кількість повторювань (однакові об'єми). У пропорційних комплексах існує певне постійне співвідношення між об'ємами вибірок, які відносяться до однакових градацій факторів. У практиці біологічних досліджень пропорційні комплекси зустрічаються рідко, тому як ортогональних комплекси розглядатимуться тільки рівномірні комплекси. Важливою особливістю ортогональних комплексів є рівність загального факторного варіювання сумі факторних варіювань і їх взаємодії:

$$SS_x = SS_A + SS_B + SS_{AB},$$

де:  $SS_A$  та  $SS_B$  – суми квадратів відхилень, що відбивають дію контрольованих факторів ( $A$  та  $B$  відповідно),  $SS_{AB}$  – сума квадратів відхилень, що характеризує взаємодію факторів.

Для обчислення показників двофакторного комплексу необхідно спочатку розрахувати показники загального, факторного та випадкового

варіювання для всього комплексу за методикою, наведеною у попередньому розділі. Далі обчислюються деякі допоміжні суми для варіювання за факторами  $A$  та  $B$ . Для проведення розрахунків складають таблицю, до якої заносяться результати аналізів (табл. 5).

Таблиця 5

Двофакторний дисперсійний комплекс

Градації фактора $B$	Градації фактора $A$			$\Sigma B$
	$A_1$	$A_2$	$A_3$	
$B_1$	$y_{1(a1b1)},$ $y_{2(a1b1)}, \dots,$ $y_{n(a1b1)}$	$y_{1(a2b1)},$ $y_{2(a2b1)}, \dots,$ $y_{n(a2b1)}$	$y_{1(a3b1)},$ $y_{2(a3b1)}, \dots,$ $y_{n(a3b1)}$	$\Sigma B_1$
$B_2$	$y_{1(a1b2)},$ $y_{2(a1b2)}, \dots,$ $y_{n(a1b2)}$	$y_{1(a2b2)},$ $y_{2(a2b2)}, \dots,$ $y_{n(a2b2)}$	$y_{1(a3b2)},$ $y_{2(a3b2)}, \dots,$ $y_{n(a3b2)}$	$\Sigma B_2$
$\Sigma A$	$\Sigma A_1$	$\Sigma A_2$	$\Sigma A_3$	

Для фактора градації  $A_1$  ці суми обчислюються наступним чином:

$$\Sigma A_1 = y_{1(a1b1)} + y_{2(a1b1)} + \dots + y_{n(a1b2)}$$

тобто сумуються всі дані для градації фактора  $A_1$  за градаціями фактору  $B_1$  та  $B_2$  ( $n$  – об'єми вибірок). Так само знаходяться суми за градаціям фактора  $A_2$  та  $A_3$ :

$$\begin{aligned} \Sigma A_2 &= y_{1(a2b1)} + y_{2(a2b1)} + \dots + y_{n(a2b2)}, \\ \Sigma A_3 &= y_{1(a3b1)} + y_{2(a3b1)} + \dots + y_{n(a2b2)} \end{aligned}$$

Аналогічно обчислюються і суми за градаціями фактора  $B$  ( $B_1$  та  $B_2$ ):

$$\begin{aligned} \Sigma B_1 &= y_{1(a1b1)} + y_{2(a1b1)} + \dots + y_{n(a3b1)}, \\ \Sigma B_2 &= y_{1(a1b2)} + y_{1(a1b2)} + \dots + y_{n(a3b2)} \end{aligned}$$

Отримані суми зводять до другого ступеня і знаходять суми квадратів окремо для фактора  $A$  та для фактора  $B$ :

$$\Sigma (\Sigma A_i)^2 = (\Sigma A_1)^2 + (\Sigma A_2)^2 + (\Sigma A_3)^2, \quad \Sigma (\Sigma B_i)^2 = (\Sigma B_1)^2 + (\Sigma B_2)^2$$

Обчислюються кількості дат за фактором  $A$  та фактору  $B$  і загальний об'єм комплексу:

$$n_A = n \cdot g_B, \quad n_B = n \cdot g_A, \quad N = n \cdot g_A \cdot g_B,$$

де  $N$  – загальний об'єм комплексу (кількість усіх повторювань у всіх вибірках),  $g_A$  – кількість градацій фактора  $A$ ,  $g_B$  – кількість градацій фактора  $B$ . Розраховують суми квадратів відхилень за факторами:

$$SS_A = \frac{\sum (\sum A_i)^2}{n_A} - \frac{(\sum y_{i,j})^2}{N}, \quad SS_B = \frac{\sum (\sum B_i)^2}{n_B} - \frac{(\sum y_{i,j})^2}{N},$$

$$SS_{AB} = SS_x - (SS_A + SS_B)$$

Далі обчислюються ступені вільності для всіх варіювань:

$$f_y = N - 1, \quad f_x = k - 1, \quad f_z = N - k, \quad k = g_A \cdot g_B,$$

$$f_A = g_A - 1, \quad f_B = g_B - 1, \quad f_{AB} = f_A \cdot f_B$$

і середні квадрати відхилень для факторного та випадкового впливу:

$$S_x^2 = \frac{SS_x}{f_x}, \quad S_z^2 = \frac{SS_z}{f_z}, \quad S_A^2 = \frac{SS_A}{f_A}, \quad S_B^2 = \frac{SS_B}{f_B}, \quad S_{AB}^2 = \frac{SS_{AB}}{f_{AB}}.$$

Наприкінці статистичної обробки обчислюються значення критерію Фішера і визначається вірогідність впливу:

$$F_x = \frac{S_x^2}{S_z^2}, \quad F_A = \frac{S_A^2}{S_z^2}, \quad F_B = \frac{S_B^2}{S_z^2}, \quad F_{AB} = \frac{S_{AB}^2}{S_z^2}.$$

Вплив фактора визнається вірогідним на обраному рівні значущості, якщо обчислене значення дорівнює або перевищує стандартне критичне значення для кількості ступенів вільності для відповідних факторів. Можна також розрахувати показники сили впливу окремих факторів.

## Приклад

Вивчався вплив забруднення ґрунту сульфідом натрію (фактор А) та фторидом натрію (фактор В) на ріст паростків горобини звичайної. Результати досліджень наведені в табл. 6. Потрібно з'ясувати вірогідність впливу забруднення на ростові процеси паростків. До таблиці також занесені проміжні результати обчислень.

Таблиця 6

Вплив забруднення ґрунту на ріст паростків горобини звичайної

Градації Фактора В	Градації фактора А			Суми за фактором В
	А <sub>1</sub>	А <sub>2</sub>	А <sub>3</sub>	
<b>В<sub>1</sub></b>	23,2; 18,4; 20,3; 28,6; 13,4 24,5	10,8; 13,1; 12,8; 12,0; 15,7; 13,5	10,1; 10,6; 8,6; 12,1; 8,6; 9,4	
$\sum y_{ki}$	<b>128,4</b>	<b>77,9</b>	<b>59,4</b>	<b>265,7</b>
$\sum y_{ki}^2$	<b>2886,66</b>	<b>1024,83</b>	<b>597,06</b>	<b>70596,49</b>
$H_k$	<b>2747,76</b>	<b>1011,40</b>	<b>588,06</b>	
$\bar{y}$	<b>21,4</b>	<b>12,9</b>	<b>9,9</b>	
<b>В<sub>2</sub></b>	18,4; 10,4; 14,6; 9,5; 17,6; 10,8	13,0; 15,7; 12,7; 12,6; 12,6; 14,6	10,6; 8,6; 11,4; 12,3; 11,1; 8,6	
$\sum y_{ki}$	<b>81,3</b>	<b>81,2</b>	<b>62,6</b>	<b>225,1</b>
$\sum y_{ki}^2$	<b>1176,53</b>	<b>1107,46</b>	<b>664,74</b>	<b>50670,01</b>
$H_k$	<b>1101,61</b>	<b>1098,91</b>	<b>653,13</b>	
$\bar{y}$	<b>13,6</b>	<b>13,5</b>	<b>10,4</b>	
<b>В<sub>3</sub></b>	7,0; 11,4; 11,9; 11,8; 10,4; 5,1	6,5; 10,7; 12,2; 11,8; 6,2; 8,3	4,5; 4,3; 3,7; 5,7 5,1; 4,9	
$\sum y_{ki}$	<b>57,6</b>	<b>55,7</b>	<b>28,2</b>	<b>141,5</b>
$\sum y_{ki}^2$	<b>593,98</b>	<b>552,15</b>	<b>134,84</b>	<b>20022,25</b>
$H_k$	<b>552,96</b>	<b>517,08</b>	<b>132,54</b>	
$\bar{y}$	<b>9,6</b>	<b>9,3</b>	<b>4,7</b>	
<b>Суми за фактором А</b>	<b>267,3</b>	<b>214,8</b>	<b>150,2</b>	
<b>Суми квадратів за факто- ром А</b>	<b>71449,29</b>	<b>46139,04</b>	<b>22560,04</b>	

Після проведення первинної статистичної обробки та перевірки дисперсій на однорідність обчислюються суми дат, суми квадратів дат та показники варіювання:

$$SS_y = \sum y_{ki}^2 - \frac{(\sum y_{ki})^2}{N} = 8738,35 - \frac{632,3^2}{54} = 8738,36 - 7403,76 = 1334,59,$$

$$SS_x = \sum H_k - \frac{(\sum y_{ki})^2}{N} = 8403,45 - 7403,76 = 999,69,$$

$$\begin{aligned}
SS_z &= SS_y - SS_x = 1334,59 - 999,69 = 334,90, \\
k &= g_A \cdot g_B = 3 \cdot 3 = 9, \quad N = n \cdot k = 6 \cdot 9 = 54, \quad f_x = k - 1 = 9 - 1 = 8, \\
f_z &= N - k = 54 - 9 = 45, \quad f_A = g_A - 1 = 3 - 1 = 2, \\
f_B &= g_B - 1 = 3 - 1 = 2, \\
f_{AB} &= f_A \cdot f_B = 2 \cdot 2 = 4, \quad n_A = n \cdot g_B = 6 \cdot 3 = 18, \quad n_B = n \cdot g_A = 6 \cdot 3 = 18, \\
SS_A &= \frac{\sum A_k^2}{n_A} - \frac{(\sum y_{ki})^2}{N} = \frac{71449,29 + 46139,04 + 22560,04}{18} - 7403,76 = 382,26, \\
SS_B &= \frac{\sum B_k^2}{n_B} - \frac{(\sum y_{ki})^2}{N} = \frac{70596,49 + 50670,012 + 20022,25}{18} - 7403,76 = 445,61, \\
SS_{AB} &= SS_x - SS_A - SS_B = 999,69 - 382,26 - 445,61 = 171,82 \\
S_x^2 &= \frac{SS_x}{f_x} = \frac{999,69}{8} = 124,96, \quad S_z^2 = \frac{SS_z}{f_z} = \frac{334,90}{45} = 7,44 \\
S_A^2 &= \frac{SS_A}{f_A} = \frac{382,26}{2} = 191,13, \quad S_B^2 = \frac{SS_B}{f_B} = \frac{445,61}{2} = 222,81, \\
S_{AB}^2 &= \frac{SS_{AB}}{f_{AB}} = \frac{171,82}{4} = 42,96, \quad F_x = \frac{S_x^2}{S_z^2} = \frac{124,96}{7,44} = 16,80, \\
F_A &= \frac{S_A^2}{S_z^2} = \frac{191,13}{7,44} = 25,68, \quad F_B = \frac{S_B^2}{S_z^2} = \frac{222,81}{7,44} = 29,95, \\
F_{AB} &= \frac{S_{AB}^2}{S_z^2} = \frac{42,96}{7,44} = 5,78
\end{aligned}$$

Результати дисперсійного аналізу показали вірогідність впливу забруднення ґрунту на ростові процеси паростків. Можна з ймовірністю 95% ( $\alpha=0,05$ ) зробити висновок про негативну дію сульфїту натрію (обчислене значення вірогідності  $F=25,68$ , стандартне  $F_{st}=3,2$ ), фториду натрію (обчислене  $F=29,95$ , стандартне значення  $F_{st}=3,2$ ) та їх взаємодії (обчислене  $F=5,78$ , стандартне  $F_{st}=2,6$ ).

#### Дисперсійний аналіз двофакторних неортогональних комплексів

Під час проведення дисперсійного аналізу неортогональних (нерівномірних) комплексів слід враховувати ту обставину, що не виконується рівність між факторною сумою квадратів відхилень та сумою сум квадратів відхилень з окремих факторів:

$$SS_x \neq SS_A + SS_B + SS_{AB}.$$

Тому проводячи обчислення потрібно спочатку розрахувати відповідні суми квадратів відхилень, знайти поправочний коефіцієнт і обчислити виправлені суми квадратів відхилень.

Спочатку розраховують загальну, факторну та випадкову суми квадратів відхилень і вибіркові середні значення для всіх вибірок за градаціями факторів. Обчислені середні значення заносять до таблиці для зручності подальших розрахунків (табл. 7).

Таблиця 7

Значення вибіркових середніх і проміжні показники двофакторного неортогонального комплексу

Градації фактора В	Градації фактора А				$\Sigma y_{Bj}$	$h_B$	$h_B^2$
	А1		А2				
	$y_{aibj}$	$n_{aibj}$	$y_{aibj}$	$n_{aibj}$			
<b>В1</b>	$y_{a1b1}$	$n_{a1b1}$	$y_{a2b1}$	$n_{a2b1}$	$\Sigma y_{ajb1}$	$h_{B1} = \Sigma y_{ajb1} / g_A$	$h_{B1}^2$
<b>В2</b>	$y_{a1b2}$	$n_{a1b2}$	$y_{a2b2}$	$n_{a2b2}$	$\Sigma y_{ajb2}$	$h_{B2} = \Sigma y_{ajb2} / g_A$	$h_{B2}^2$
<b>В3</b>	$y_{a1b3}$	$n_{a1b3}$	$y_{a2b3}$	$n_{a2b3}$	$\Sigma y_{ajb3}$	$h_{B3} = \Sigma y_{ajb3} / g_A$	$h_{B3}^2$
$\Sigma y_A$	$\Sigma y_{a1bi}$		$\Sigma y_{a2bi}$				
$h_A$	$h_{A1} = \Sigma y_{a1bi} / g_B$		$h_{A2} = \Sigma y_{a2bi} / g_B$				
$h_A^2$	$h_{A1}^2$		$h_{A2}^2$				

Далі обчислюють суми показників  $h_A$ ,  $h_B$ ,  $h_A^2$ ,  $h_B^2$  і за формулами:

$$SS'_A = g_B \cdot \left( \sum h_A^2 - \frac{(\sum h_A)^2}{g_A} \right) \text{ та}$$

$$SS'_B = g_A \cdot \left( \sum h_B^2 - \frac{(\sum h_B)^2}{g_B} \right),$$

розраховують не виправлені суми квадратів відхилень для факторів А та В. За формулою

$$SS'_x = \sum \bar{y}_{ij}^2 - \frac{(\sum \bar{y}_{ij})^2}{g_A \cdot g_B}$$

обчислюють не виправлену факторну суму квадратів відхилень. Далі знаходять суму квадратів відхилень за взаємодією факторів:

$$SS'_{AB} = SS'_x - SS'_A - SS'_B = SS'_x - (SS'_A + SS'_B)$$



та поправочний коефіцієнт:

$$ko = \frac{SS_x}{SS_x}$$

Помножуючи суми квадратів відхилень відповідних факторів на поправочний коефіцієнт, знаходять виправлені суми квадратів відхилень:

$$SS_A = ko \cdot SS_A', \quad SS_B = ko \cdot SS_B', \quad SS_{AB} = ko \cdot SS_{AB}'.$$

Далі знаходять кількості ступенів вільності:

$$f_y = N - 1, \quad f_x = k - 1, \quad f_z = N - k, \quad f_A = g_A - 1 \\ f_B = g_B - 1, \quad f_{AB} = g_A \cdot g_B - 1,$$

середні квадрати відхилень:

$$S_z^2 = \frac{SS_z}{f_z}, \quad S_x^2 = \frac{SS_x}{f_x}, \quad S_A^2 = \frac{SS_A}{f_A}, \quad S_B^2 = \frac{SS_B}{f_B}, \quad S_{AB}^2 = \frac{SS_{AB}}{f_{AB}}$$

та вірогідність впливу факторів:

$$F_x = \frac{S_x^2}{S_z^2}, \quad F_A = \frac{S_A^2}{S_z^2}, \quad F_B = \frac{S_B^2}{S_z^2}, \quad F_{AB} = \frac{S_{AB}^2}{S_z^2}.$$

Нульова гіпотеза (вплив фактора невірогідний) відкидається у тому випадку, якщо обчислене значення дорівнює або перевищує стандартне критичне значення, знайдене в таблиці Фішера, для обраного рівня значущості і відповідних кількостей ступенів вільності.

## Приклад

Вивчався вплив азотних та калійних добрив на ріст паростків робінії псевдоакації. Кількість доз азотних добрив (фактор А) становила 2, калійні добрива випробовувалися у трьох концентраціях. Вихідні дані дослідів та окремі проміжні результати обчислень наведені у табл. 8. З допомогою цих даних проводиться обчислення загальної, факторіальної та випадкової сум квадратів відхилень, факторіального та випадкового середніх квадратів відхилень та вірогідності впливу факторів.

Таблиця 8

Результати вивчення дії азотних та калійних добрив на ріст паростків робі-  
нії псевдоакації

Градації фактора В		Градації фактора А	
		А <sub>1</sub>	А <sub>2</sub>
В <sub>1</sub>	$y_{ij}$	14,3; 15,9; 17,5	16,3; 15,4; 14,6; 18,4; 17,8
	$n_k$	3	5
	$y_k$	15,9	16,5
	$\sum y_{ij}$	47,7	82,5
	$\sum y_{ji}^2$	763,55	1371,41
	$(\sum y_{ij})^2/n_k=H_k$	758,43	1361,25
В <sub>2</sub>	$y_{ij}$	15,5; 20,2; 18,1; 17,3	15,4; 20,5; 20,2; 23,2
	$n_k$	4	4
	$y_k$	17,8	19,8
	$\sum y_{ij}$	71,1	79,3
	$\sum y_{ji}^2$	1275,19	1603,69
	$(\sum y_{ij})^2/n_k=H_k$	1263,80	1572,12
В <sub>3</sub>	$y_{ij}$	22,6; 17,8; 21,9; 17,6; 18,6	21,9; 24,7; 26,4
	$n_k$	5	3
	$y_k$	19,7	24,3
	$\sum y_{ij}$	98,5	73,0
	$\sum y_{ji}^2$	1962,93	1786,66
	$(\sum y_{ij})^2/n_k=H_k$	1940,45	1776,30

Для обчислення загальної, факторіальної та випадкової сум квадратів та факторіального і випадкового середніх квадратів відхилень використовуються загальні формули:

$$SS_y = \sum y_{ij}^2 - \frac{(\sum y_{ij})^2}{N} = 763,55 + 1275,19 + \dots + 1786,66 -$$

$$- \frac{(47,7 + 71,1 + \dots + 73,0)^2}{24} = 8763,42 - 8516,43 = 246,99$$

$$SS_x = \sum H_k - \frac{(\sum y_{ij})^2}{N} = 758,43 + 1572,12 + \dots + 1940,45 - 1786,66 =$$

$$= 8672,35 - 8516,43 = 155,92$$

$$SS_z = SS_y - SS_x = 246,99 - 155,92 = 91,07$$

Далі обчислюються ступені вільності та середні квадрати відхилень:

$$N = \sum n_k = 3 + 4 + 5 + 5 + 4 + 3 = 25, \quad k = g_A \cdot g_B = 2 \cdot 3 = 6,$$

$$f_x = k - 1 = 6 - 1 = 5, \quad f_z = N - k = 24 - 6 = 18, \quad f_A = g_A - 1 = 2 - 1 = 1,$$

$$f_B = g_B - 1 = 3 - 1 = 2, \quad f_{AB} = f_A \cdot f_B = 1 \cdot 2 = 2,$$

$$S_x^2 = \frac{SS_x}{f_x} = \frac{155,92}{5} = 31,184, \quad S_z^2 = \frac{SS_z}{f_z} = \frac{91,07}{18} = 5,06,$$

$$F_x = \frac{S_x^2}{S_z^2} = \frac{31,184}{5,06} = 6,17$$

Як бачимо, вплив факторів вірогідний, оскільки обчислене значення критерію вірогідності (6,17) значно перевищує стандартне (2,77). Тепер можна починати обчислення показників варіювання за окремими факторами та їх взаємодією. Для цього середні значення заносяться до таблиці, яка також буде використана для внесення проміжних сум (табл. 9).

Таблиця 9

Обчислення показників двофакторного дисперсійного аналізу

Градації фактора В	Градації фактору А		$\sum y_{Bj}$	$h_B = \frac{\sum y_{ajbi}}{g_A}$	$h_B^2$
	А1	А2			
	$y_{aibj}$	$y_{aibj}$			
<b>В1</b>	<b>15,9</b>	<b>16,5</b>	<b>32,4</b>	<b>16,2</b>	<b>262,44</b>
<b>В2</b>	<b>17,8</b>	<b>19,8</b>	<b>37,6</b>	<b>18,8</b>	<b>353,44</b>
<b>В3</b>	<b>19,7</b>	<b>24,4</b>	<b>44,0</b>	<b>22,0</b>	<b>484,00</b>
$\sum y_A$	<b>53,4</b>	<b>60,6</b>			
$h_A = \frac{\sum y_{aibj}}{g_B}$	<b>17,8</b>	<b>20,2</b>			
$h_A^2$	<b>316,84</b>	<b>408,04</b>			

Далі обчислюються некореговані суми квадратів відхилень за фактором А, фактором В та взаємодії факторів, середні квадрати відхилень за факторами та вірогідність їх впливу.

$$SS'_A = g_B \cdot \left( \sum h_A^2 - \frac{(\sum h_A)^2}{g_A} \right) = 3 \cdot \left( 316,84 + 408,04 - \frac{(17,8 + 20,2)^2}{2} \right) = 8,64,$$

$$SS'_B = g_A \cdot \left( \sum h_B^2 - \frac{(\sum h_B)^2}{g_B} \right) = 2 \cdot \left( 262,44 + 353,44 + 484,00 - \frac{(16,2 + 18,8 + 22,0)^2}{3} \right) = 33,76,$$

$$SS'_x = \sum \bar{y}^2 - \frac{(\sum \bar{y})^2}{g_A \cdot g_B} = 2212,52 - \frac{114^2}{6} = 46,52,$$

$$SS'_{AB} = SS'_x - SS'_A - SS'_B = 46,52 - 8,64 - 33,76 = 4,12,$$

$$Ko = \frac{SS'_x}{SS'_x} = \frac{155,92}{46,52} = 3,35.$$

Розраховуються кореговані суми квадратів відхилень, середні квадрати відхилень та вірогідності впливу факторів:

$$SS_A = SS'_A \cdot Ko = 8,64 \cdot 3,35 = 28,96, \quad SS_B = SS'_B \cdot Ko = 33,76 \cdot 3,35 = 113,15,$$

$$SS_{AB} = SS'_{AB} \cdot Ko = 4,12 \cdot 3,35 = 13,81, \quad S_A^2 = \frac{SS_A}{f_A} = \frac{28,96}{1} = 28,96,$$

$$S_B^2 = \frac{SS_B}{f_B} = \frac{113,15}{2} = 56,58, \quad S_{AB}^2 = \frac{SS_{AB}}{f_{AB}} = \frac{13,81}{2} = 6,91,$$

$$F_A = \frac{S_A^2}{S_z^2} = \frac{28,96}{5,06} = 5,72, \quad F_B = \frac{S_B^2}{S_z^2} = \frac{56,58}{5,06} = 11,18,$$

$$F_{AB} = \frac{S_{AB}^2}{S_z^2} = \frac{6,91}{5,06} = 1,36$$

Результати аналізу дозволяють зробити висновок про вірогідність впливу азотних та калійних добрив на ріст паростків робінії псевдоакації (за фактором **A** обчислене значення критерію вірогідності становить 5,72, а стандартне – 4,41, за фактором **B** – 11,18 та 3,55 відповідно). Одночасно, взаємодія факторів виявилася невірогідною, тобто дія одного з добрив не залежить від дії іншого (обчислене значення критерію вірогідності складає 1,36, стандартне – 3,55).

#### Дисперсійний аналіз двофакторних ортогональних ієрархічних комплексів

В ієрархічних комплексах можна встановити тільки вплив факторів на ознаки об'єкта досліджень. Взаємодія факторів виявляється змішаною з дією другого фактора. Математично це можна записати за допомогою формул:

$$\sigma_A^2 = S_A^2 \pm \Delta_{S_A^2}, \quad \sigma_B^2 = (S_B^2 \pm \Delta_{S_B^2}) + (S_{AB}^2 \pm \Delta_{S_{AB}^2})$$

Звідси витікає, що під час проведення ієрархічного аналізу можна обчислити вплив фактора **A** та змішаний вплив фактора **B** і взаємодії факторів **A** та **B**. Для проведення аналізу, як і у попередніх випадках, обчислюються

показники варіювання всіх вибірок (вибіркові середні арифметичні, помилки репрезентативності, надійні інтервали) і загальна, факторна та випадкова суми квадратів відхилень для всього комплексу. Для зручності експериментальні дані заносяться до таблиці (табл. 10).

Таблиця 10

Двофакторний ієрархічний комплекс

Градації фактора А	Градації фактора В			
	В <sub>1</sub>	В <sub>2</sub>	В <sub>3</sub>	В <sub>4</sub>
А <sub>1</sub>	$y_{1(a1b1)}, y_{2(a1b1)}, \dots, y_{n(a1b1)}$	$y_{1(a1b2)}, y_{2(a1b2)}, \dots, y_{n(a1b2)}$		
А <sub>2</sub>			$y_{1(a2b3)}, y_{2(a2b3)}, \dots, y_{n(a2b3)}$	$y_{1(a2b4)}, y_{2(a2b4)}, \dots, y_{n(a2b4)}$
$\sum y_{AiBj}$	$\sum y_{i(a1b1)}$	$\sum y_{i(a1b2)}$	$\sum y_{i(a2b3)}$	$\sum y_{i(a2b4)}$
$\sum y_{Aj}$	$\sum y_{A1} = \sum y_{i(a1b1)} + \sum y_{i(a1b2)}$		$\sum y_{A2} = \sum y_{i(a2b3)} + \sum y_{i(a2b4)}$	

Обчислюються суми дат та кількості спостережень (табл. 10) за градаціями фактора А ( $n_a$ ). Для цього об'єм вибірок помножують на кількість градацій фактора В в межах відповідної градації фактора А:

$$n_{a_i} = n \cdot g_{B(a_i)}$$

Розраховується сума квадратів відхилень за фактором А:

$$SS_A = \sum \frac{(\sum y_{Aj})^2}{n_{a_i}} - \frac{(\sum y_{ij})^2}{N},$$

де  $N$  – загальний об'єм комплексу ( $N = g_A \cdot g_B$ ), та сума квадратів відхилень за фактором В:

$$SS_B = \sum \frac{\left(\sum_{i=1}^n y_{ij}\right)^2}{n} - \sum \frac{(\sum y_{Ai})^2}{n_{a_i}} = \sum H_i - \sum \frac{(\sum y_{Ai})^2}{n_{a_i}}.$$

На останньому етапі обчислюються кількості ступенів вільності для факторів, середні квадрати відхилень та вірогідності впливу за факторами за такими формулами:

$$f_x = k - 1, \quad f_z = N - k, \quad k = g_A \cdot g_B, \quad f_A = g_A - 1, \quad f_B = g_B - g_A,$$

$$S_x^2 = \frac{SS_x}{f_x}, \quad S_z^2 = \frac{SS_z}{f_z}, \quad S_A^2 = \frac{SS_A}{f_A}, \quad S_B^2 = \frac{SS_B}{f_B},$$

$$F_x = \frac{S_x^2}{S_z^2}, \quad F_A = \frac{S_A^2}{S_B^2}, \quad F_B = \frac{S_B^2}{S_z^2}.$$

Потрібно звернути увагу на відмінності в обчисленні вірогідності впливу фактора  $A$  від інших вірогідностей. Для обчислення вірогідності впливу фактора  $A$  береться відношення середнього квадрата відхилень за цим фактором до середнього квадрата відхилень за фактором  $B$ , а не до випадкового середнього квадрата відхилень. Вплив фактора визнається вірогідним, якщо обчислене значення дорівнює або перевищує стандартне критичне значення критерію Фішера для обраного рівня значущості та відповідних кількостей ступенів вільності.

### Приклад

Вивчалася дія забруднення повітря на активність ферменту супероксиддисмутази на території ботанічного ( $A_1$ ) саду та коксохімічного заводу ( $A_2$ ). Листя для аналізу відбирали з шести дерев одного віку: три на території ботанічного саду ( $B_1, B_2, B_3$ ) та три на території коксохімічного заводу ( $B_4, B_5$  та  $B_6$ ). Кожна з шести вибірок складалася з чотирьох повторювань. Як і в попередніх прикладах, результати досліджень та деякі проміжні результати обчислень для зручності заносяться до таблиці (табл. 11).

Проводиться обчислення загальних показників варіювання комплексу:

$$SS_y = \sum y_{ki}^2 - \frac{(\sum y_{ki})^2}{N} = 198 + 309 + \dots + 425 - \frac{(28 + 35 + \dots + 41)^2}{24} = 1867 - 1820,04 = 46,96,$$

$$SS_x = \sum H_k - \frac{(\sum y_{ki})^2}{N} = 196,00 + 306,25 + \dots + 420,25 - 1820,04 = 27,71$$

$$SS_z = SS_y - SS_x = 46,96 - 27,71 = 19,25,$$

$$N = \sum n_k = 4 + 4 + \dots + 4 = 24, \quad f_x = k - 1 = 6 - 1 = 5,$$

$$f_z = N - k = 24 - 6 = 18, \quad f_A = g_A - 1 = 2 - 1 = 1,$$

$$f_B = g_B - g_A = 6 - 2 = 4,$$

$$S_x^2 = \frac{SS_x}{f_x} = \frac{27,71}{5} = 5,54, \quad S_z^2 = \frac{SS_z}{f_z} = \frac{19,25}{18} = 1,07,$$

$$F_x = \frac{S_x^2}{S_z^2} = \frac{5,54}{1,07} = 5,18,$$

**Таблиця 11. Результати визначення активності супероксиддісму-тази в листках дерев**

Градації Фактора А	Градації фактора В					
	В <sub>1</sub>	В <sub>2</sub>	В <sub>3</sub>	В <sub>4</sub>	В <sub>5</sub>	В <sub>6</sub>
А <sub>1</sub>	7; 6; 8; 7	8; 9; 8; 10	9; 7; 9; 9			
А <sub>2</sub>				10; 8; 10; 11	9; 8; 7; 8	12; 10; 10; 9
$\sum y_{AiBj}$	28	35	34	39	32	41
$(\sum y_{AiBj})^2/n=H_k$	196,00	306,25	289,00	380,25	256,00	240,25
$\sum y^2_{AiBj}$	198	309	292	385	258	425
$\sum y_{Aj}$	28+35+34=97			39+32+41=112		
$n_{Aj}$	4·3=12			4·3=12		
$\sum y_{Aj}/n_A$	97/12=8,08			112/12=9,33		
$\sum y^2_{Aj}$	9409			12544		

Далі обчислюються показники варіювання за факторами *A* та *B*:

$$SS_A = \sum \frac{(\sum y_{Aj})^2}{n_{a_i}} - \frac{(\sum y_{ij})^2}{N} = \frac{9409 + 12544}{12} - 1820,04 = 9,38,$$

$$SS_B = \sum H_i - \sum \frac{(\sum y_{Ai})^2}{n_{a_i}} = 1847,75 - 1829,42 = 18,33.$$

$$S_A^2 = \frac{SS_A}{f_A} = \frac{9,38}{1} = 9,38, \quad S_B^2 = \frac{SS_B}{f_B} = \frac{18,33}{4} = 4,58,$$

$$F_A = \frac{S_A^2}{S_B^2} = \frac{9,38}{4,58} = 2,05, \quad F_B = \frac{S_B^2}{S_z^2} = \frac{4,58}{1,07} = 4,28.$$

Отримані результати свідчать про наявність різниці в активності фермента у різних рослин (фактор *B*), а місце зростання (фактор *A*) на цей показник не впливає. Але слід пам'ятати, що у статистичному показнику впливу фактора *B* міститься частка, яка визначається взаємодією двох факторів і яку неможливо відокремити.

### **Дисперсійний аналіз трифакторних ортогональних комплексів**

У трифакторних комплексах вивчається вплив трьох факторів, які діють на об'єкт досліджень одночасно. Алгоритм аналізу практично не відрізняється від алгоритмів двофакторного аналізу, але потребує набагато

більше обчислювальної роботи. Факторіальна сума квадратів відхилень ( $SS_x$ ) у трифакторних комплексах розкладається на сім складових частин, що може бути так:

$$SS_x = SS_A + SS_B + SS_{AB} + SS_C + SS_{AC} + SS_{BC} + SS_{ABC}$$

Під час проведення статистичної обробки трифакторних комплексів спочатку обчислюються вибіркові показники варіації всіх вибірок, які входять до комплексу, та розраховуються загальна, факторіальна та випадкова сума квадратів відхилень (як в однофакторному дисперсійному аналізі). Складаються допоміжні таблиці комплексу: загальна та за парами факторів (табл. 12, 13, 14, 15). До загальної таблиці заносяться експериментальні дані та обчислюються суми дат за градаціями факторів. До таблиці за парами факторів заносяться вже суми дат.

Спочатку обчислюються суми квадратів відхилень за факторами  $A$  та  $B$ , використовуючи згруповані за цими факторами дані (табл. 13). Для розрахунку сум квадратів відхилень за факторами  $A$  та  $B$  застосовують формули:

$$SS_A = \sum \frac{(\sum y_{Ai})^2}{n \cdot g_b \cdot g_c} - \frac{(\sum y_{ijd})^2}{N} \quad \text{та} \quad SS_B = \sum \frac{(\sum y_{Bi})^2}{n \cdot g_a \cdot g_c} - \frac{(\sum y_{ijd})^2}{N}.$$

Далі, використовуючи дані таблиці взаємодії факторів  $A$  та  $C$ , обчислюють значення суми квадратів відхилень для фактору  $C$ :

$$SS_C = \sum \frac{(\sum y_{C(a)i})^2}{n \cdot g_A \cdot g_C} - \frac{(\sum y_{ijd})^2}{N}.$$

Тепер розраховуються допоміжні величини  $H_{AB}$ ,  $H_{AC}$ ,  $H_{BC}$  та суми квадратів відхилень за взаємодіями факторів:

$$H_{AB} = \frac{\sum (\sum y_{AiBj})^2}{n \cdot g_c} - \frac{(\sum y_{ijd})^2}{N},$$

**Таблиця 12.** Загальна таблиця трифакторного комплексу для обчислення показників загального, факторного та випадкового варіювання

Градації факторів			Значення $y_i$	$\sum y_i$
A	B	C		
A1	B1	C1	$y_{1(a1b1c1)}, y_{2(a1b1c1)}, \dots, y_{n(a1b1c1)}$	$\sum y_{i(a1b1c1)}$
		C2	$y_{1(a1b1c2)}, y_{2(a1b1c2)}, \dots, y_{n(a1b1c2)}$	$\sum y_{i(a1b1c2)}$
		C3	$y_{1(a1b1c3)}, y_{2(a1b1c3)}, \dots, y_{n(a1b1c3)}$	$\sum y_{i(a1b1c3)}$



A2	B2	C1	$y_1(a1b2c1), y_2(a1b2c1), \dots, y_n(a1b2c1)$	$\sum y_i(a1b2c1)$
		C2	$y_1(a1b2c2), y_2(a1b2c2), \dots, y_n(a1b2c2)$	$\sum y_i(a1b2c2)$
		C3	$y_1(a1b2c3), y_2(a1b2c3), \dots, y_n(a1b2c3)$	$\sum y_i(a1b2c3)$
	B1	C1	$y_1(a2b1c1), y_2(a2b1c1), \dots, y_n(a2b1c1)$	$\sum y_i(a2b1c1)$
		C2	$y_1(a2b1c2), y_2(a2b1c2), \dots, y_n(a2b1c2)$	$\sum y_i(a2b1c2)$
		C3	$y_1(a2b1c3), y_2(a2b1c3), \dots, y_n(a2b1c3)$	$\sum y_i(a2b1c3)$
B2	C1	$y_1(a2b2c1), y_2(a2b2c1), \dots, y_n(a2b2c1)$	$\sum y_i(a2b2c1)$	
	C2	$y_1(a2b2c2), y_2(a2b2c2), \dots, y_n(a2b2c2)$	$\sum y_i(a2b2c2)$	
	C3	$y_1(a2b2c3), y_2(a2b2c3), \dots, y_n(a2b2c3)$	$\sum y_i(a2b2c3)$	

$$H_{AC} = \frac{\sum (\sum y_{AiCj})^2}{n \cdot g_B} - \frac{(\sum y_{ijd})^2}{N}$$

$$H_{BC} = \frac{\sum (\sum y_{BiCj})^2}{n \cdot g_A} - \frac{(\sum y_{ijd})^2}{N},$$

$$SS_{AB} = H_{AB} - (SS_A + SS_B), \quad SS_{AC} = H_{AC} - (SS_A + SS_C),$$

$$SS_{BC} = H_{BC} - (SS_B + SS_C) \text{ та}$$

$$SS_{ABC} = SS_x - (SS_A + SS_B + SS_{AB} + SS_C + SS_{AC} + SS_{BC})$$

Останнім етапом є обчислення кількостей ступенів вільності, середніх квадратів відхилень та показників вірогідності впливу факторів:

$$f_x = k - 1, \quad f_z = N - k, \quad k = g_A \cdot g_B \cdot g_C, \quad N = n \cdot g_A \cdot g_B \cdot g_C,$$

$$f_A = g_A - 1, \quad f_B = g_B - 1, \quad f_C = g_C - 1, \quad f_{AB} = f_A \cdot f_B$$

$$f_{AC} = f_A \cdot f_C, \quad f_{BC} = f_B \cdot f_C, \quad f_{ABC} = f_A \cdot f_B \cdot f_C$$

$$S_x^2 = \frac{SS_x}{f_x}, \quad S_z^2 = \frac{SS_z}{f_z}, \quad S_A^2 = \frac{SS_A}{f_A}, \quad S_B^2 = \frac{SS_B}{f_B}, \quad S_C^2 = \frac{SS_C}{f_C}$$

$$S_{AB}^2 = \frac{SS_{AB}}{f_{AB}}, \quad S_{AC}^2 = \frac{SS_{AC}}{f_{AC}}, \quad S_{BC}^2 = \frac{SS_{BC}}{f_{BC}}, \quad S_{ABC}^2 = \frac{SS_{ABC}}{f_{ABC}},$$

**Таблиця 13.** Дані для обчислення сум квадратів відхилень за факторами А та В

Градації фактора А	Градації фактора В		$\sum y_A$
	B <sub>1</sub>	B <sub>2</sub>	
A <sub>1</sub>	$\sum y_{a1b1c1}, \sum y_{a1b1c2},$ $\sum y_{a1b1c3}$	$\sum y_{a1b2c1}, \sum y_{a1b2c2},$ $\sum y_{a1b2c3}$	
$\sum y_{AiBj}$	$\sum y_{a1b1}$	$\sum y_{a1b2}$	$\sum y_{A1} = \sum y_{a1b1} + \sum y_{a1b2}$

$A_2$	$\sum y_{a2b1c1}, \sum y_{a2b1c2},$ $\sum y_{a1b1c3}$	$\sum y_{a2b2c1}, \sum y_{a2b2c2},$ $\sum y_{a2b2c3}$	
$\sum y_{AiBj}$	$\sum y_{a2b1}$	$\sum y_{a2b2}$	$\sum y_{A2} = \sum y_{a2b1} + \sum y_{a2b2}$
$\sum y_B$	$\sum y_{B1} = \sum y_{a1b1} + \sum y_{a2b1}$	$\sum y_{B2} = \sum y_{a1b2} + \sum y_{a2b2}$	

Таблиця 14. Дані для обчислення суми квадратів відхилень за фактором С

Градації фактора А	Градації фактора С		
	$C_1$	$C_2$	$C_3$
$A_1$	$\sum y_{a1b1c1}, \sum y_{a1b2c1}$	$\sum y_{a1b1c2}, \sum y_{a1b2c2}$	$\sum y_{a1b1c3}, \sum y_{a1b2c3}$
$\sum y_{aicj}$	$\sum y_{a1c1}$	$\sum y_{a1c2}$	$\sum y_{a1c3}$
$A_2$	$\sum y_{a2b1c1}, \sum y_{a2b2c1}$	$\sum y_{a2b1c2}, \sum y_{a2b2c2}$	$\sum y_{a2b1c3}, \sum y_{a2b2c3}$
$\sum y_{aicj}$	$\sum y_{a2c1}$	$\sum y_{a2c2}$	$\sum y_{a2c3}$
$\sum y_{C(A)i}$	$\sum y_{C(A)1} = \sum y_{a1c1} +$ $\sum y_{a2c1}$	$\sum y_{C(A)2} = \sum y_{a1c2} +$ $\sum y_{a2c2}$	$\sum y_{C(A)3} = \sum y_{a1c3} +$ $\sum y_{a2c3}$

Таблиця 15. Дані для обчислення допоміжних показників за взаємодією факторів

Градації фактора В	Градації фактора С		
	$C_1$	$C_2$	$C_3$
$B_1$	$\sum y_{a1b1c1}, \sum y_{a2b1c1}$	$\sum y_{a1b1c2}, \sum y_{a2b1c2}$	$\sum y_{a1b1c3}, \sum y_{a2b1c3}$
$\sum y_{bicj}$	$\sum y_{b1c1}$	$\sum y_{b1c2}$	$\sum y_{b1c3}$
$B_2$	$\sum y_{a1b2c1}, \sum y_{a2b2c1}$	$\sum y_{a1b2c2}, \sum y_{a2b2c2}$	$\sum y_{a1b2c3}, \sum y_{a2b2c3}$
$\sum y_{bicj}$	$\sum y_{b2c1}$	$\sum y_{b2c2}$	$\sum y_{b2c3}$

$$F_x = \frac{S_x^2}{S_z^2}, \quad F_A = \frac{S_A^2}{S_z^2}, \quad F_B = \frac{S_B^2}{S_z^2}, \quad F_C = \frac{S_C^2}{S_z^2}, \quad F_{AB} = \frac{S_{AB}^2}{S_z^2},$$

$$F_{AC} = \frac{S_{AC}^2}{S_z^2}, \quad F_{BC} = \frac{S_{BC}^2}{S_z^2}, \quad F_{ABC} = \frac{S_{ABC}^2}{S_z^2}$$

Вплив факторів, як і в інших випадках, визнається вірогідним, якщо обчислене значення критерію вірогідності дорівнює або перевищує стандартне критичне значення критерію Фішера для обраного рівня значущості та відповідних значень ступенів вільності.

### Приклад

Досліджувався вплив забруднення ґрунту фтором (фактор *A*), сіркою (фактор *B*) та внесення стимулятора росту (фактор *C*) на ріст паростків дубу звичайного. Кожний фактор випробовувався на двох рівнях. Результати експерименту та проміжні результати обчислень подані в табл. 16.

Спочатку обчислюються показники загального варіювання:

$$\begin{aligned}
 SS_y &= \sum y_{ki}^2 - \frac{(\sum y_{ki})^2}{N} = 2737 - \frac{247^2}{24} = 2937,0 - 2542,04 = 394,96 \\
 SS_x &= \sum H_i - \frac{(\sum y_{ki})^2}{N} = 2878,10 - 2542,04 = 336,06, \\
 SS_z &= SS_y - SS_x = 394,94 - 336,06 = 58,90, \\
 k &= g_A \cdot g_B \cdot g_C = 2 \cdot 2 \cdot 2 = 8, \quad N = \sum n_k = k \cdot n = 8 \cdot 3 = 24, \\
 f_x &= k - 1 = 8 - 1 = 7, \quad f_z = N - k = 24 - 8 = 16, \quad f_A = g_A - 1 = 2 - 1 = 1, \\
 f_B &= g_B - 1 = 2 - 1 = 1, \quad f_C = g_C - 1 = 2 - 1 = 1, \quad f_{AB} = f_A \cdot f_B = 1 \cdot 1 = 1, \\
 f_{AC} &= f_A \cdot f_C = 1 \cdot 1 = 1, \quad f_{BC} = f_B \cdot f_C = 1 \cdot 1 = 1, \\
 f_{ABC} &= f_A \cdot f_B \cdot f_C = 1 \cdot 1 \cdot 1 = 1, \quad S_x^2 = \frac{SS_x}{f_x} = \frac{336,06}{7} = 38,01, \\
 S_z^2 &= \frac{SS_z}{f_z} = \frac{58,90}{16} = 3,68, \quad F_x = \frac{S_x^2}{S_z^2} = \frac{48,01}{3,68} = 13,01
 \end{aligned}$$

Далі обчислюються суми квадратів відхилень за окремими факторами. Для цього відповідні суми заносяться до таблиць (табл. 17, 18, 19).

За факторам *A* та *B* необхідні суми занесені до табл. 17. Використовуючи ці суми обчислюються показники  $SS_A$ ,  $SS_B$  та  $SS_{AB}$ .

$$\begin{aligned}
 SS_A &= \sum \frac{(\sum y_{Ai})^2}{n \cdot g_b \cdot g_c} - \frac{(\sum y_{ki})^2}{N} = \frac{136^2 + 111^2}{3 \cdot 2 \cdot 2} - 2542,04 = 2568,08 - 2542,04 = 26,04, \\
 SS_B &= \sum \frac{(\sum y_{Bi})^2}{n \cdot g_a \cdot g_c} - \frac{(\sum y_{ki})^2}{N} = \frac{157^2 + 90^2}{3 \cdot 2 \cdot 2} - 2542,04 = 2729,08 - 2542,04 = 187,04, \\
 H_{AB} &= \frac{\sum (\sum y_{AiBj})^2}{n \cdot g_c} - \frac{(\sum y_{ki})^2}{N} = \frac{84^2 + 52^2 + 73^2 + 38^2}{3 \cdot 2} - 2542,04 = 2755,50 - 2542,04 = 213,46, \\
 SS_{AB} &= H_{AB} - (SS_A + SS_B) = 213,46 - (26,04 + 187,04) = 0,38.
 \end{aligned}$$

Таблиця 16

Результати досліджень впливу стимулятора росту на ріст паростків в умовах забруднення ґрунту сіркою та фтором

Градації факторів			Результати експерименту	$\sum y_{ki}$	$\frac{(\sum y_{ki})^2}{n} = H_i$	$\sum y_{ki}^2$
A	B	C				
A <sub>1</sub>	B <sub>1</sub>	C <sub>1</sub>	10,0; 11,0; 13,0	34,0	385,3	390,0
		C <sub>2</sub>	16,0; 16,0; 19,0	50,0	833,3	842,0
	B <sub>2</sub>	C <sub>1</sub>	7,0; 8,0; 5,0	20,0	133,3	138,0
		C <sub>2</sub>	12,0; 11,0; 9,0	32,0	341,3	346,0
A <sub>2</sub>	B <sub>1</sub>	C <sub>1</sub>	9,0; 11,0; 12,0	32,0	341,0	346,0
		C <sub>2</sub>	13,0; 17,0; 11,0	41,0	560,3	579,0
	B <sub>2</sub>	C <sub>1</sub>	4,0; 2,0; 5,0	11,0	40,3	45,0
		C <sub>2</sub>	9,0; 11,0; 7,0	27,0	243,0	251,0
Суми			$N = \sum n_k = 24$	247,0	2878,1	2937,0

Таблиця 17

Дані для обчислення сум квадратів відхилень за факторами *A* та *B*

Градації фактора A	Градації фактора B		$\sum y_A$
	B <sub>1</sub>	B <sub>2</sub>	
A <sub>1</sub>	54; 50	20; 32	
$\sum y_{A_i B_j}$	84	52	136
A <sub>2</sub>	32; 41	11; 27	
$\sum y_{A_i B_j}$	73	38	111
$\sum y_B$	157	90	

Далі, за даними, наведеними в табл. 18, обчислюються суми квадратів за фактором *C* та взаємодією *AC*.

Таблиця 18

Дані для обчислення суми квадратів відхилень по фактору *C*

Градації фактора A	Градації фактора C	
	C <sub>1</sub>	C <sub>2</sub>
A <sub>1</sub>	34; 20	50; 32
$\sum y_{A_i C_j}$	54	82

$A_2$	32; 11	41; 27
$\sum y_{Aicj}$	43	68
$\sum y_{C(A)i}$	97	150

$$SS_C = \sum \frac{(\sum y_{C(a)i})^2}{n \cdot g_A \cdot g_C} - \frac{(\sum y_{ki})^2}{N} = \frac{97^2 + 150^2}{3 \cdot 2 \cdot 2} - 2542,04 = 2659,08 - 2542,04 = 117,04,$$

$$H_{AC} = \frac{\sum (\sum y_{Aicj})^2}{n \cdot g_B} - \frac{(\sum y_{ki})^2}{N} = \frac{54^2 + 43^2 + 82^2 + 68^2}{3 \cdot 2} - 2542,04 = 2685,50 - 2542,04 = 143,46,$$

$$SS_{AC} = H_{AC} - (SS_A + SS_C) = 143,46 - (26,04 + 117,04) = 0,38.$$

Наступним кроком є обчислення сум квадратів відхилень за взаємодією факторів  $B$  та  $C$ . Для цього використовуються дані, подані в табл. 19.

$$H_{BC} = \frac{\sum (\sum y_{Bicj})^2}{n \cdot g_A} - \frac{(\sum y_{ki})^2}{N} = \frac{66^2 + 31^2 + 91^2 + 59^2}{3 \cdot 2} - 2542,04 = 2846,50 - 2542,04 = 304,46,$$

$$SS_{BC} = H_{BC} - (SS_B + SS_C) = 304,46 - (187,04 + 117,04) = 0,38.$$

Таблиця 19

Дані для обчислення допоміжних показників за взаємодією факторів

Градації фактора $B$	Градації фактора $C$	
	$C_1$	$C_2$
$B_1$	34; 32	50; 41
$\sum y_{Bicj}$	66	91
$B_2$	20; 11	32; 27
$\sum y_{Bicj}$	31	59

Далі обчислюється сума квадратів відхилень за взаємодією всіх факторів

$$SS_{ABC} = SS_x - (SS_A + SS_B + SS_{AB} + SS_C + SS_{AC} + SS_{BC}) =$$

$$= 336,06 - (26,04 + 187,04 + 117,04 + 0,38 + 0,38 + 0,38) = 4,80$$

Наприкінці аналізу розраховуються середні квадрати відхилень та вірогідності впливу окремих факторів та їх взаємодії:

$$S_A^2 = \frac{SS_A}{f_A} = \frac{26,04}{1} = 26,04, \quad S_B^2 = \frac{SS_B}{f_B} = \frac{187,04}{1} = 187,04,$$

$$S_C^2 = \frac{SS_C}{f_C} = \frac{117,04}{1} = 117,04, \quad S_{AB}^2 = \frac{SS_{AB}}{f_{AB}} = \frac{0,38}{1} = 0,38,$$

$$S_{AC}^2 = \frac{SS_{AC}}{f_{AC}} = \frac{0,38}{1} = 0,38, \quad S_{BC}^2 = \frac{SS_{BC}}{f_{BC}} = \frac{0,38}{1} = 0,38,$$

$$S_{ABC}^2 = \frac{SS_{ABC}}{f_{ABC}} = \frac{4,80}{1} = 4,80,$$

$$F_A = \frac{S_A^2}{S_z^2} = \frac{26,04}{3,68} = 7,08, \quad F_B = \frac{S_B^2}{S_z^2} = \frac{187,04}{3,68} = 50,83,$$

$$F_C = \frac{S_C^2}{S_z^2} = \frac{117,04}{3,68} = 31,80, \quad F_{AB} = \frac{S_{AB}^2}{S_z^2} = \frac{0,38}{3,68} = 0,10,$$

$$F_{AC} = \frac{S_{AC}^2}{S_z^2} = \frac{0,38}{3,68} = 0,10, \quad F_{BC} = \frac{S_{BC}^2}{S_z^2} = \frac{0,38}{3,68} = 0,10,$$

$$F_{ABC} = \frac{S_{ABC}^2}{S_z^2} = \frac{4,80}{3,68} = 1,30$$

Аналіз результатів (стандартне значення критерію Фішера складає 4,41) дозволяє зробити висновок про незалежність дії факторів. Всі взаємодії факторів виявилися незначущими.

### **Дисперсійний аналіз комплексів з якісними ознаками**

У попередніх розділах розглядалися методи дисперсійного аналізу ознак, які можна оцінити кількісно (кількісних ознак). Але існує значна кількість ознак, що виражаються тільки якісно. Наприклад, суцвіття хризантем можуть бути різного кольору (жовті, червоні, білі, бузкові та ін.). Цю ознаку дуже важко оцінити кількісно. Якісні ознаки можуть мати декілька градацій, як у наведеному прикладі, або тільки дві градації. Прикладом двох градацій ознаки можуть слугувати хворі та здорові рослини на дослідній ділянці. Якщо ознака може мати лише дві градації, то ці градації зветься *альтернативними*. Одна з цих градацій, а саме та, яка цікавить дослідника, зветься *результативною*. Наприклад, якщо дослідника цікавить стійкість рослин до хвороб, то результативною ознакою буде кількість здорових рослин на дослідній ділянці. Практично всі якісні ознаки під час проведення дисперсійного аналізу можуть бути зведені до альтернативних. У наведеному прикладі у рослин хризантем можна за результативну ознаку прийняти жовтий колір суцвіть. Тоді всі інші рослини будуть мати альтернативну ознаку, тобто ці рослини не мають суцвіть жовтого кольору. Такі ознаки ще називають *нерезультативними* або *негативними*. Після виді-

лення результативної та негативної ознак, можна проводити дисперсійний аналіз.

Дисперсійний аналіз якісних ознак проводиться за тими ж самими схемами, що і дисперсійний аналіз кількісних ознак. Різниця полягає у тому, що кількісна ознака може бути охарактеризована середніми показниками – вибірковими середніми, показниками варіювання, генеральною середньою, тоді як якісні ознаки виражаються частками (відносними частотами) або відсотками від загальної кількості спостережень. Як і комплекси кількісних ознак, комплекси якісних ознак поділяються на ортогональні та неортогональні, однофакторні та багатфакторні.

### Дисперсійний аналіз комплексів з ознаками, що виражаються частками від цілого.

Спочатку наведемо формули для обчислення сум квадратів відхилень та кількостей ступенів вільності для загального, факторного та випадкового варіювання:

Таблиця 20

Формули для проведення дисперсійного аналізу комплексів якісних ознак, виражених частками

Виток варіювання	Суми квадратів відхилень	Ступені вільності
Загальне варіювання	$SS_y = \sum m - \frac{(\sum m)^2}{N}$	$N - 1$
Факторне варіювання	$SS_x = \sum m \cdot p - \frac{(\sum m)^2}{N}$	$k - 1$
Випадкове варіювання	$SS_z = \sum m - \sum m \cdot p$	$N - k$

де:  $m$  – кількість варіант з результативною ознакою;

$p$  – відносна частота або частка варіант з результативною ознакою ( $p=m/n$ );

$N$  – загальна кількість спостережень (об'єм комплексу);

$n$  – кількість спостережень в окремих градаціях фактора (об'єми вибірок).

Враховуючи, що

$$p = \frac{m}{n},$$

формулу для обчислення факторної суми квадратів відхилень можна записати так:

$$SS_x = \sum \frac{m^2}{n} - \frac{(\sum m)^2}{N}.$$

А формула для обчислення випадкової суми квадратів ще може мати такий вигляд:

$$SS_z = SS_y - SS_x$$

Формули для обчислень середніх квадратів відхилень та вірогідності впливу фактора такі самі, як і в кількісному дисперсійному аналізі:

$$S_x^2 = \frac{SS_x}{f_x}, \quad S_z^2 = \frac{SS_z}{f_z} \quad \text{та} \quad F_x = \frac{S_x^2}{S_z^2}.$$

Методику розрахунків краще показати на числовому прикладі. Вивчалася стійкість різних сортів жита до сажки. Результативною ознакою вважалася наявність ознак захворювання. Дані обстежень та деякі допоміжні величини наведені в табл. 21.

Далі розраховуються суми квадратів відхилень:

$$\text{загальна} - SS_y = \sum m - \frac{(\sum m)^2}{N} = 105 - \frac{105^2}{240} = 105 - 45,94 = 56,09,$$

$$\text{факторна} - SS_x = \sum \frac{m^2}{n} - \frac{(\sum m)^2}{N} = 67,85 - 45,94 = 21,94,$$

$$\text{випадкова} - SS_z = SS_y - SS_x = 56,09 - 21,94 = 37,15.$$

Таблиця 21

Дані обстежень захворюваності рослин жита сажкою

Показники	Сорти жита				
	1	2	3	4	5
Усього обстежено (n)	40	50	50	40	60
Уражених рослин (m)	15	39	5	34	12
Здорових рослин	35	11	45	6	48
m <sup>2</sup>	225	1521	25	1156	144
m <sup>2</sup> /n	5,63	30,42	0,50	28,90	2,40



Далі обчислюються кількості ступенів вільності, середні квадрати відхилень та вірогідність впливу:

$$\text{ступені вільності факторного варіювання} - f_x = k - 1 = 5 - 1 = 4 ,$$

$$\text{ступені вільності випадкового варіювання} - f_z = N - k = 240 - 5 = 235 ,$$

$$\text{факторний середній квадрат відхилень} - S_x^2 = \frac{SS_x}{f_x} = \frac{21,94}{4} = 5,49 ,$$

$$\text{випадковий середній квадрат відхилень} - S_z^2 = \frac{SS_z}{f_z} = \frac{37,15}{235} = 0,16 ,$$

$$\text{вірогідність впливу фактора} - F_x = \frac{S_x^2}{S_z^2} = \frac{5,49}{0,16} = 34,31 .$$

Обчислене значення критерію вірогідності виявилось досить високим і значно перевищує стандартні значення (для  $\alpha=0,05$  воно дорівнює 2,39). З цього можна зробити висновок, що вивчені сорти жита вірогідно розрізняються за стійкістю до сажки і для вирощування треба відбирати більш стійкі сорти (у прикладі це сорти 1, 3 та 5).

Дисперсійний аналіз комплексів з ознаками, вираженими у відсотках від загальної кількості спостережень.

Загальні формули для обчислення наведені в табл. 22.

Таблиця 22

Формули для проведення дисперсійного аналізу якісних ознак, виражених відсотками від загальної кількості спостережень

Виток варіювання	Суми квадратів відхилень	Ступені вільності
Загальне варіювання	$SS_y = \sum p\% - \frac{(\sum p\%)^2}{100 \cdot k}$	$100 \cdot k - 1$
Факторне варіювання	$SS_x = \sum \frac{p\%^2}{100} - \frac{(\sum p\%)^2}{100 \cdot k}$	$k - 1$
Випадкове варіювання	$SS_z = SS_y - SS_x$	$100 \cdot k - k$

де:  $N$  – загальна кількість спостережень (об'єм комплексу):

$k$  – кількість градацій фактора;

$p$  – частка варіант з результативною ознакою від загальної кількості спостережень ( $N$ ), виражена у відсотках.

Для прикладу обчислимо комплекс, який було наведено у попередньому розділі. За результативну ознаку буде використана кількість здорових рослин (табл. 23).

Таблиця 23

Кількість здорових рослин на дослідній ділянці

Показник	Сорт жита					Суми
	1	2	3	4	5	
Здорових рослин	35	11	45	6	48	145
p%	24,1	7,6	31,0	4,1	33,2	100
p% <sup>2</sup>	580,81	57,76	961,00	16,81	1102,24	
p% <sup>2</sup> /100	5,81	0,58	9,61	0,17	11,02	27,19

На основі наведених у таблиці даних проводяться обчислення сум квадратів відхилень:

$$SS_y = \sum p\% - \frac{(\sum p\%)^2}{100 \cdot k} = 100 - \frac{100^2}{100 \cdot 5} = 100 - 20 = 80,$$

$$SS_x = \sum \frac{p\%^2}{100} - \frac{(\sum p\%)^2}{100 \cdot k} = 27,19 - 20 = 7,19,$$

$$SS_z = SS_y - SS_x = 80 - 7,19 = 72,81.$$

Далі обчислюються кількості ступенів вільності, середні квадрати відхилень та вірогідність впливу:

$$f_x = k - 1 = 5 - 1 = 4, \quad f_z = 100 \cdot k - k = 100 \cdot 5 - 5 = 500 - 5 = 495,$$

$$S_x^2 = \frac{SS_x}{f_x} = \frac{7,19}{4} = 1,79, \quad S_z^2 = \frac{SS_z}{f_z} = \frac{72,81}{495} = 0,15,$$

$$F_x = \frac{S_x^2}{S_z^2} = \frac{1,79}{0,15} = 11,93.$$

Порівняння обчисленого показника вірогідності зі стандартним значенням ( $\alpha=0,05$   $F_{st}=2,38$ ) дозволяє зробити висновок про наявність різниці між стійкістю різних сортів жита до захворювання сажкою.

Дисперсійний аналіз ортогональних двофакторних комплексів.

Схема дисперсійного аналізу двофакторних комплексів якісних ознак аналогічна кількісному аналізу двофакторних комплексів. Формули для проведення обчислень сум квадратів відхилень у випадку, коли якісні ознаки виражені частками, такі:

$$SS_y = \sum m - \frac{(\sum m)^2}{N}, \quad SS_x = \sum \frac{m^2}{n} - \frac{(\sum m)^2}{N}, \quad SS_z = SS_y - SS_x,$$

$$SS_A = \sum \frac{(\sum m_A)^2}{n_A} - \frac{(\sum m)^2}{N}, \quad SS_B = \sum \frac{(\sum m_B)^2}{n_B} - \frac{(\sum m)^2}{N},$$

$$SS_{AB} = SS_x - (SS_A + SS_B)$$

Для обчислення кількостей ступенів вільності, середніх квадратів відхилень та вірогідностей впливу використовують такі формули:

$$f_x = k - 1, \quad k = g_A \cdot g_B, \quad f_z = N - k, \quad f_A = g_A - 1, \quad f_B = g_B - 1,$$

$$f_{AB} = f_A \cdot f_B,$$

$$S_x^2 = \frac{SS_x}{f_x}, \quad S_z^2 = \frac{SS_z}{f_z}, \quad S_A^2 = \frac{SS_A}{f_A}, \quad S_B^2 = \frac{SS_B}{f_B}, \quad S_{AB}^2 = \frac{SS_{AB}}{f_{AB}},$$

$$F_x = \frac{S_x^2}{S_z^2}, \quad F_A = \frac{S_A^2}{S_z^2}, \quad F_B = \frac{S_B^2}{S_z^2}, \quad F_{AB}^2 = \frac{S_{AB}^2}{S_z^2}.$$

Методику розрахунків розглянемо на прикладі. При вибірковому дослідженні учнів старших класів міських та сільських шкіл були виявлені аномалії зору, які розподілилися наступним чином (табл. 24).

Таблиця 24

Результати обстеження аномалій зору в учнів шкіл

Показники	A1 - хлопці		A2 - дівчата		Сума
	B1 - місто	B2 - село	B1 - місто	B2 - село	
Обстежено	25	25	25	25	100
Аномалії (m)	3	2	8	2	15
p=m/n	0,12	0,08	0,32	0,08	
m <sup>2</sup>	9	4	64	4	
m <sup>2</sup> /n	0,36	0,16	2,56	0,16	3,24

Добуваються суми квадратів відхилень для всього комплексу:

$$SS_y = \sum m - \frac{(\sum m)^2}{N} = 15 - \frac{15^2}{100} = 15 - 2,25 = 12,75$$

$$SS_x = \sum \frac{m^2}{n} - \frac{(\sum m)^2}{N} = 3,24 - 2,25 = 0,99$$

$$SS_z = SS_y - SS_x = 12,75 - 0,99 = 11,76$$

Для обчислення сум квадратів за градаціями окремих факторів потрібно обчислити допоміжні величини  $\sum \frac{(m_A)^2}{n_A}$  та  $\sum \frac{(m_B)^2}{n_B}$ . Для їх обчислення складається таблиця, до якої заносяться вихідні величини та суми за факторами (табл. 25).

Обчислюються суми квадратів відхилень за факторам  $A$ ,  $B$  та їх взаємодією:

$$SS_A = \sum \frac{(\sum m_A)^2}{n_A} - \frac{(\sum m)^2}{N} = 2,50 - 2,25 = 0,25$$

$$SS_B = \sum \frac{(\sum m_B)^2}{n_B} - \frac{(\sum m)^2}{N} = 2,77 - 2,25 = 0,52$$

$$SS_{AB} = SS_x - (SS_A + SS_B) = 0,99 - (0,25 + 0,52) = 0,22$$

Таблиця 25

Обчислення допоміжних величин двофакторного комплексу

Градації факторів	$n_i$	$m_i$	$(\sum m_i)^2$	$\frac{(\sum m_i)^2}{n_i}$
A <sub>1</sub>	50	5	25	0,5
A <sub>2</sub>	50	10	100	2,0
Сума по А	100			$\sum \frac{(\sum m_A)^2}{n_A} = 2,0 + 0,5 = 2,5$
B <sub>1</sub>	50	11	121	2,45
B <sub>2</sub>	50	4	16	0,32
Сума по В	100			$\sum \frac{(\sum m_B)^2}{n_B} = 2,45 + 0,32 = 2,77$

Визначаються кількості ступенів вільності, середні квадрати відхилень та вірогідності впливу факторів:

$$f_x = k - 1 = 4 - 1 = 3, \quad f_z = N - k = 100 - 4 = 96, \quad f_A = g_A - 1 = 2 - 1 = 1,$$

$$f_B = g_B - 1 = 2 - 1 = 1, \quad f_{AB} = f_A \cdot f_B = 1 \cdot 1 = 1$$

$$\begin{aligned}
S_x^2 &= \frac{SS_x}{f_x} = \frac{0,99}{3} = 0,33, & S_z^2 &= \frac{SS_z}{f_z} = \frac{11,76}{96} = 0,12, \\
S_A^2 &= \frac{SS_A}{f_A} = \frac{0,25}{1} = 0,25 \\
S_B^2 &= \frac{SS_B}{f_B} = \frac{0,52}{1} = 0,52, & S_{AB}^2 &= \frac{SS_{AB}}{f_{AB}} = \frac{0,22}{1} = 0,22, \\
F_x &= \frac{S_x^2}{S_z^2} = \frac{0,33}{0,12} = 2,8, & F_A &= \frac{S_A^2}{S_z^2} = \frac{0,25}{0,12} = 2,1, \\
F_B &= \frac{S_B^2}{S_z^2} = \frac{0,52}{0,12} = 0,43, & F_{AB} &= \frac{S_{AB}^2}{S_z^2} = \frac{0,22}{0,12} = 1,8
\end{aligned}$$

Порівняння обчислених значень вірогідності зі стандартними значеннями критерію Фішера дозволяють зробити висновок про вірогідність впливу тільки фактора В (місце проживання). Вплив всіх інших факторів виявився невірогідним.

#### 4. МЕТОДИ МНОЖИННИХ ПОРІВНЯНЬ

Після проведення дисперсійного аналізу логічно постає питання, які саме градації факторів вірогідно впливають на значення ознаки. Нагадаємо, що перевірка нульової гіпотези (дослідні варіанти не відрізняються за значеннями ознаки) проводиться за допомогою F-критерію із загальним рівнем значущості  $\alpha$ , який характеризує ризик відхилення від цієї гіпотези, коли вона вірогідна. Значення ризику звичайно вибирається для  $\alpha = 0,05$  або  $\alpha = 0,01$ . У процесі подальшого узагальнення даних і перевірки гіпотез про рівність середніх значень та інших, більш складних, порівнянь може бути проведено декілька перевірок вірогідності. Якщо кожна перевірка проводиться з рівнем значущості  $\alpha$ , ризик отримання однієї або більше хибно значущих різниць буде більшим ніж  $\alpha$ . Для  $k$  таких незалежних перевірок різниці (тестів значущості) ризик отримання хибно значущого результату, коли всі нульові гіпотези вірогідні, буде дорівнювати:

$$\alpha' = 1 - (1 - \alpha)^k$$

Наприклад, при застосуванні в дисперсійному комплексі, в якому порівнюються п'ять градацій фактора замість F-критерію серії t-критеріїв з рівнем значущості  $\alpha=0,05$  ( $k=5$ ), ймовірність помилкового передбачення збільшується з 5% до

$$\alpha' = 1 - (1 - 0,05)^5 = 0,23 \text{ (23\%)}$$

Для зменшення ризику отримання "надто великої" кількості значущих різниць необхідно, щоб загальний рівень значущості якоюсь мірою контролювався. Процедури перевірки гіпотез для отримання висновків одночасно про декілька порівнянь за допомогою методів, що контролюють рівень ризику помилкових висновків називаються **процедурами множинних порівнянь**.

Множинні порівняння проводяться після дисперсійного аналізу, коли F-критерій перевищує стандартне значення, тобто вплив визнаний вірогідним на обраному рівні значущості. Однак процедури множинних порівнянь дозволяють отримати цінну інформацію навіть у тих випадках, коли дисперсійний аналіз не дав вірогідних результатів. У дисперсійному аналізі обчислюється також випадковий середній квадрат відхилення  $S_z$  та кількість ступенів вільності для випадкового варіювання, які використовуються для наступних множинних порівнянь.

Процедури множинних порівнянь зводяться до обчислення допуску, з яким порівнюються різниці між середніми, що цікавлять дослідника. Різницю вважають вірогідною, якщо її значення перевищує допуск або дорівнює йому.

## Метод Даннета (порівняння з контролем)

Порівняння середніх різних дослідних варіантів з контролем є одним з найпоширеніших завдань статистичної обробки експериментальних даних. Контролем можуть слугувати, наприклад, варіанти з нульовим або зі стандартним (еталонним) рівнем фактора. У цьому випадку для вияву значущості різниці між варіантами дослідження застосовується тест Даннета.

Якщо середню арифметичну контрольної вибірки позначити через  $y_{kt}$ , а дослідні вибіркові середні через  $y_1, y_2$  та  $y_3$ , то можна записати завдання цього порівняння так:  $y_{kt} \rightarrow y_1, y_{kt} \rightarrow y_2$  та  $y_{kt} \rightarrow y_3$ . Для розрахунків використовуються випадковий середній квадрат відхилень комплексу ( $S_z$ ) та ступені вільності  $f_z=N-k$  та  $f_x=k-1$ . У таблиці стандартних значень Даннета знаходять значення коефіцієнта  $t_{\alpha, f_x, f_z}^D$ .

Треба мати на увазі, що значення коефіцієнта залежать від характеру зміни дослідних середніх відносно контрольного значення, а самі формули – від рівномірності комплексу. За характером розподілу дослідних середніх відносно контрольного середнього значення комплекси поділяються на одnobічні та двобічні. В одnobічних комплексах усі дослідні середні або перевищують за значенням, або менші, ніж контрольна середня, тобто виконується одна з двох умов:

$$y_{kt} \leq y_1 \leq y_2 \leq \dots \leq y_k \text{ або} \\ y_{kt} \geq y_1 \geq y_2 \geq \dots \geq y_k$$

Якщо частина дослідних середніх менша за своїм значенням, ніж контрольна середня, а інша частина перевищує її значення, то такі комплекси відносяться до двобічних. Залежно від типу комплексу застосовуються різні таблиці стандартних значень (табл. 15 та 16 додатків). Після визначення коефіцієнта Даннета проводиться обчислення допуску за відповідними формулами.

Для нерівномірних комплексів:

$$D = t_{\alpha, m}^D \cdot S \cdot \sqrt{\frac{1}{n_k} + \frac{1}{n_o}}$$

де  $S$  – обчислюється як квадратний корінь з випадкового квадрата відхилень дисперсійного комплексу.

Для рівномірних комплексів:

$$D = t_{\alpha, m}^D \cdot \sqrt{2} \cdot S_y \text{ де} \\ S_y = \sqrt{\frac{S_z^2}{n}}$$

**Приклад.**

Наведемо приклад застосування критерію Даннета, який показує застосування однобічного тесту для рівномірних комплексів.

Вивчався вплив різних доз мінеральних добрив на ріст рослин. Наприкінці досліду довжина рослин за варіантами досліду складала: контроль – 5,61 см, 1-ша доза добрив – 7,48, 2-га доза – 11,94, 3-тя доза – 22,09 та 4-та доза – 11,83 см. Усі вибірки мали однакову кількість спостережень ( $n = 7$ ). Значення випадкового квадрату відхилень, знайдене у дисперсійному аналізі, дорівнює 18,16, а кількість ступенів вільності для випадкового варіювання  $f_z = 35 - 5 = 30$ . Оскільки комплекс рівномірний ( $n_k = n_1 = n_2 = n_3$ ), однобічний (всі дослідні середні перевищують за значенням контрольне), застосовується формула для однобічного рівномірного комплексу. Спочатку обчислюється значення

$$S_y = \sqrt{\frac{18,16}{7}} = 1,61$$

Далі в табл. 15 додатків знаходиться значення  $t_{\alpha}^D$  для  $f_y = 5 - 1 = 4$  та  $f_z = 30$  (це значення дорівнює 2,25). Розраховується допуск

$$D = t_{\alpha,m}^D \cdot \sqrt{2} \cdot S_y = 2,25 \cdot 1,41 \cdot 1,61 = 5,11$$

Перевіряються різниці між контрольним та дослідними середніми (знак різниці не враховується, тому від більшого значення віднімається менше):

$$D = 7,48 - 5,61 = 1,47 < 5,11$$

$$D = 11,94 - 5,61 = 6,32 > 5,11$$

$$D = 22,09 - 5,61 = 16,48 > 5,11$$

$$D = 11,83 - 5,61 = 6,22 > 5,11$$

Усі різниці, крім першої, перевищують обчислене значення допуску. З цього можна зробити висновок, що друга, третя та четверта дози добрив мають вірогідний позитивний вплив на рослини, тоді як перша доза не впливає на їх ріст.

### **Процедура Шеффе для будь-яких типів порівнянь (S-метод)**

Для обчислення допуску під час використання процедури Шеффе для перевірки будь-якого контрасту (різниці між двома порівнюваними вибірковими середніми) застосовується така формула:

$$D^S = \sqrt{(k-1) \cdot F} \cdot \sqrt{\sum C_{gu}^2} \cdot S_y$$

де:  $k$  – кількість порівнюваних вибірок;

$F$  – стандартне значення критерію Фішера для рівня значущості  $\alpha$  і кількості ступенів вільності  $f_x = (k-1)$  та  $f_z = \sum n_k - k$ ;

$C$  – узагальнюючі контрасти;



$S_x$  – дисперсія середнього значення.

Розрахунок узагальнюючих контрастів можна показати на такому прикладі. Припустимо, що є п'ять вибірових середніх ( $y_1, y_2, y_3, y_4, y_5$ ). Треба порівняти перші два з них з трьома останніми. Контраст можна записати таким чином:  $(y_1 + y_2)/2 - (y_3 + y_4 + y_5)/3$ , тоді узагальнюючі контрасти можна записати як  $C_1 = C_2 = 1/2$  і  $C_3 = C_4 = C_5 = 1/3$ . Звідси  $C_{gu} = (1/2) + (1/2) + (1/3) + (1/3) + (1/3)$ . У попарних порівняннях цей показник буде дорівнювати одиниці, а формула за умови  $n_1 = n_2 = \dots = n_k$  набуває вигляду:

$$D^S = \sqrt{2 \cdot (k-1) \cdot F} \cdot S_{\bar{y}}$$

З нерівними об'ємами порівнюваних вибірок у парних порівняннях використовується формула:

$$D^S = \sqrt{(k-1) \cdot F} \cdot \sqrt{n_1 \cdot n_2 / (n_1 + n_2)} \cdot S_z$$

Треба враховувати, що цей метод у парних порівняннях дає найбільшу величину допуску відносно інших методів. Однак він дозволяє проводити не тільки парні порівняння, але і порівняння будь-якої кількості вибірок комплексу між собою. Інша перевага цього методу полягає в тому, що можна порівнювати рівномірні та нерівномірні комплекси без упорядкування середніх.

### **Процедура Т'юкі (N-метод)**

Ця процедура застосовується для аналізу рівномірних комплексів і, як і попередня, контролює рівень помилки, яка відноситься до ряду порівнюваних середніх. Вона особливо зручна, коли головний інтерес становить перевірка всіх різниць між середніми кількох варіантів, узятими попарно, тому що допуски для цих контрастів значно менші, ніж для S-методу.

Допуск для будь-якої різниці  $y_1 - y_k$  обчислюється як

$$D^T = q_{\alpha, k} \cdot S_{\bar{y}}$$

де:  $q$  – верхня точка розподілення ст'юдентизованого розмаху  $k$  середніх (табл. 17 додатків) з кількістю ступенів вільності  $f_z = k \cdot (n - k)$  ( $k$  – кількість вибірок у комплексі).

### **Методи множинних розмахів Н'юмана – Кейлса та Дункана**

Ці методи можна вважати модифікаціями процедури Т'юкі, хоча вони і були розроблені абсолютно незалежно. Вони застосовуються тоді, коли контрастами, які цікавлять дослідника, є тільки порівняння між середніми трьох або більше варіантів досліджень, взятими попарно; більш загальні контрасти не обчислюються.

У цих процедурах  $k$  середніх упорядковуються від меншого до більшого:

$$\bar{y}_1 \leq \bar{y}_2 \leq \dots \leq \bar{y}_k$$

Щоб перевірити різницю між двома середніми, наприклад  $y_1 \rightarrow y_2$ , у процедурі Т'юкі розраховується один для всього комплексу допуск. В методах множинних розмахів обчислюється декілька допусків, величина яких залежить від кількості середніх, які знаходяться між значеннями, які порівнюються. Для методу Н'юмана – Кейлса формула для розрахунку допуску має вигляд:

$$D^H = q_{\alpha, k'} \cdot S_{\bar{y}}$$

де  $k'$  – кількість середніх, які знаходяться в інтервалі між двома порівнюваними середніми, включаючи ці два (тобто  $k' = u - h + 1$ , де  $u$  та  $h$  являють собою порядкові номери двох варіантів в упорядкованому комплексі). Стандартні значення  $q_{\alpha, k'}$  знаходяться у тій самій таблиці, що і для процедури Т'юкі. Таким чином, допуск за методом Н'юмана – Кейлса менше, ніж допуск за методом Т'юкі, виключаючи порівняння між крайніми варіантами, коли допуски рівні. Треба враховувати, що якщо різниця між парою середніх ( $y_u, y_h$ ) перевірена і є невірогідною, ніяка інша різниця між парами, що знаходяться між ними, не може бути визнана значущою. У зв'язку з цим у першу чергу треба перевірити значущість різниці між двома крайніми варіантами (мінімальною та максимальною середніми). Якщо ця різниця незначуща, то подальші обчислення втрачають сенс.

За методом Дункана величина допуску визначається за тією ж формулою, що і в процедурі Н'юмана – Кейлса:

$$D^{DK} = q_{\alpha, k'} \cdot S_{\bar{y}}$$

де  $k'$ , як і в минулому способі, є числом середніх між двома варіантами, різниця між якими перевіряється, включаючи і ці два, але значення контрольованої помилки дещо більше і визначається за формулою  $\alpha = 1 - (1 - \alpha)$ , що призводить до зниження критичності тесту. Допуски в цьому тесті виходять менші, ніж в методах Т'юкі і Ньюмана – Кейлса. Тому метод Дункана знаходить широке застосування у дослідників. Він дозволяє виявити більше вірогідних різниць, хоча це пов'язано з певним ризиком завищити кількість вірогідних результатів. Стандартні значення вибираються з таблиці Дункана.

Процедура порівнянь середніх за цими методами проводиться кількома етапами.

1. Розташовують вибіркові середні в порядку збільшення.
2. Обчислюють помилку середнього для кожного варіанта, застосовуючи оцінку випадкового середнього квадрату відхилень ( $S_z$ ) з числом ступенів вільності  $f = N - k$ , отриману у дисперсійному аналізі

$$S_{\bar{y}} = \sqrt{\frac{S_z}{n}}$$

де  $n$  – число повторностей для кожного варіанта (об'єми вибірок).

3. З таблиць Т'юкі (табл. 17 додатків) або Дункана (табл. 18 додатків) критичних ст'юдентизованих розмахів для обраного рівня значущості  $\alpha$  і числа  $f$ , рівного числу ступенів вільності випадкового середнього квадрату відхилень  $f_z$  для  $k' = 2, \dots, k$ , записують  $k - 1$  стандартних значень розмахів. Помножуючи ці значення на  $S_{\bar{y}}$ , отримують  $(k' - 1)$  найменших значущих різниць – допусків.

4. Перевіряють значущість різниць між порівнюваними середніми, починаючи з крайніх: різниця мінімального і максимального середніх порівнюють з найменшою значущою різницею  $D$  при  $k' = k$ , далі різницю максимального середнього і першого, яке перевищує мінімальне. Аналогічне порівняння застосовують для другого за величиною середнього, яке порівнюють з найменшим і т.д., доки не будуть досліджені всі можливі пари середніх.

### **Метод Ст'юдента, як окремий випадок множинних порівнянь**

У біологічних дослідженнях часто застосовується для визначення вірогідності різниці між вибірковими середніми метод Ст'юдента. Цей тест проводиться за формулою:

$$t^S = \frac{|y_1 - y_2|}{\sqrt{(S_{y_1}^2 - S_{y_2}^2)}}$$

Далі отриманий показник  $t$  порівнюється з табличним значенням  $t_{\alpha, f}$  (де:  $f$  – кількість ступенів вільності – визначається за формулою  $f = n_1 + n_2 - 2$ , –  $n_1$  і  $n_2$  об'єми порівнюваних вибірок). Якщо обчислене значення перевищує стандартне або дорівнює йому, робиться висновок про вірогідність різниці з рівнем значущості  $\alpha$ . В іншому випадку різниця між середніми є невірогідною.

Цей метод дає задовільні результати при порівнянні комплексів, в яких міститься всього дві вибірки. У такому разі вірогідність безпомилкових прогнозів буде дорівнювати  $\alpha$ . Якщо комплекс складається з більшої кількості вибірок, ймовірність отримання хибно значущих різниць між середніми збільшується пропорційно ступеню кількості пар порівнюваних середніх. Таким чином застосування тесту Ст'юдента в комплексах з кількістю градацій факторів більше двох є грубою помилкою. Слід також враховувати, що цей метод в умовах порівняння вибірок з різними об'ємами може давати хибно значущі різниці.

Для порівняння двох вибірок з рівними об'ємами зручно користуватися формулою:

$$t = \frac{|\bar{y}_1 - \bar{y}_2|}{s_d} = \frac{|\bar{y}_1 - \bar{y}_2|}{S_{\bar{y}} \cdot \sqrt{2}}$$

де:  $s_d$  – стандартна помилка різниці.

Існує модифікація метода Ст'юдента для парних порівнянь в комплексах, які складаються з кількох вибірок ( $k > 2$ ). Цей метод дозволяє контролювати ступінь похибки на визначеному рівні. Обчислення ведуться за формулою:

$$D_g = t_{st} \cdot s_g = t_{st} \cdot \sqrt{\sum c_{gu}^2 \cdot S_{\bar{y}}}$$

де:  $D_g$  – найменша значуща різниця;

$t_{st}$  – стандартне значення критерію Ст'юдента для обраного рівня значущості;

$s_g$  – стандартна помилка контрасту;

$c_{gu}^2$  – узагальнюючі контрасти (обчислення див. у методі Шеффе);

$S_{\bar{y}}$  – дисперсія середнього значення комплексу –  $S_{\bar{y}} = \sqrt{\frac{S_z^2}{n}}$ .

Різниця між середніми визнається вірогідною, якщо вона перевищує найменшу значущу різницю або дорівнює їй.

## Приклади використання процедур парних порівнянь

У результаті первинної статистичної обробки даних та дисперсійного аналізу комплексу, що складається з п'яти вибірок (п'ять градацій фактора, кожна вибірка складалася з шести дат) були отримані такі результати:

- вибіркові середні – **9,53; 20,27; 22,75; 29,00; 15,52;**
- випадковий середній квадрат відхилень – **3,306;**
- кількість ступенів вільності для випадкового варіювання – **25;**
- стандартне значення критерію Фішера – **2,78.**

Оскільки комплекс рівномірний, то можна розрахувати дисперсію середнього значення:

$$S_{\bar{y}} = \sqrt{\frac{S_z^2}{n}} = \sqrt{\frac{3,306}{6}} = 0,742$$

Спочатку обчислимо вірогідність різниць за допомогою процедури Шеффе. Для цього обчислимо різниці між вибірковими середніми:

$$D_{12} = 9,53 - 20,27 = -10,74,$$

$$D_{13} = 9,53 - 22,75 = -13,22,$$

$$D_{14} = 9,53 - 29,00 = -19,47,$$

$$D_{15} = 9,53 - 15,52 = -5,22,$$

$$\begin{aligned}
D_{23} &= 20,27 - 22,75 = -2,48, \\
D_{24} &= 20,27 - 29,00 = -8,73, \\
D_{25} &= 20,27 - 15,52 = -4,75, \\
D_{34} &= 22,75 - 29,00 = -6,25, \\
D_{35} &= 22,75 - 15,52 = -7,23, \\
D_{45} &= 29,00 - 15,52 = 13,48.
\end{aligned}$$

Далі обчислимо допуск (найменшу значущу різницю) і порівняємо з нею отримані різниці між середніми:

$$D^S = \sqrt{2 \cdot (k-1) \cdot F} \cdot S_{\bar{y}} = \sqrt{2 \cdot (5-1) \cdot 2,78} \cdot 0,742 = 3,50$$

Аналіз порівняння різниць зі стандартним допуском показує, що всі різниці крім, різниці між середніми 2 та 3, є вірогідними. Різниця між середніми 2 та 3 незначуща, оскільки її значення менше найменшої значущої різниці.

Тепер обчислимо допуски за методами Т'юкі, Н'юмана – Кейлса та Дункана. Для цього спочатку упорядковуються середні значення за зростанням їх величин, тобто буде отримано ряд:

$$9,53; 15,52; 20,27; 22,75; 29,00.$$

Для методу Т'юкі це робити необов'язково, але якщо різниця між мінімальним та максимальним значенням невірогідна, це допоможе позбутися зайвої обчислювальної роботи.

Далі розраховуються різниці між середніми:

$$\begin{aligned}
D_{12} &= 9,53 - 15,52 = -5,99, \\
D_{13} &= 9,53 - 20,27 = -10,74, \\
D_{14} &= 9,53 - 22,75 = -13,22, \\
D_{15} &= 9,53 - 29,00 = -19,47, \\
D_{23} &= 15,52 - 20,27 = -4,75, \\
D_{24} &= 15,52 - 22,75 = -7,23, \\
D_{25} &= 15,52 - 29,00 = -13,48, \\
D_{34} &= 20,27 - 22,75 = -2,48, \\
D_{35} &= 20,27 - 29,00 = -8,73, \\
D_{45} &= 22,75 - 29,00 = -6,25.
\end{aligned}$$

Треба зазначити, що знак мінус у порівняннях не враховується. Він тільки вказує на напрям зміни середніх. Тепер обчислюються допуски:

$$\text{для методу Т'юкі} - D^T = q_{\alpha k} \cdot S_{\bar{y}} = 4,17 \cdot 0,742 = 3,09;$$

для методу Н'юмана – Кейлса (буде обчислюватися декілька допусків залежно від кількості порівнюваних середніх, які знаходяться між ними):

для пари середніх 1 та 5  $- D^H = q_{\alpha k} \cdot S_{\bar{y}} = 4,17 \cdot 0,742 = 3,09$  ,  
 для пар середніх 1–4 та 2–5  $- D^H = 3,90 \cdot 0,742 = 2,89$  ,  
 для пар середніх 1–3, 2–4 та 3–5  $- D^H = 3,53 \cdot 0,742 = 2,62$  ,  
 для пар середніх 1–2, 2–3, 3–4 та 4–5  $- D^H = 2,92 \cdot 0,742 = 2,17$  ;

**для методу Дункана** (також буде обчислюватися декілька допусків залежно кількості порівнюваних середніх, які знаходяться між ними):

для пари середніх 1 та 5  $- D^{DK} = q_{\alpha k} \cdot S_{\bar{y}} = 3,23 \cdot 0,742 = 2,40$  ,  
 для пар середніх 1–4 та 2–5  $- D^{DK} = 3,16 \cdot 0,742 = 2,34$  ,  
 для пар середніх 1–3, 2–4 та 3–5  $- D^{DK} = 3,07 \cdot 0,742 = 2,28$  ,  
 для пар середніх 1–2, 2–3, 3–4 та 4–5  $- D^{DK} = 2,92 \cdot 0,742 = 2,17$  ;

**для методу Ст'юдента** –

$$D_g = t_{st} \cdot \sqrt{\sum c_{gu}^2} \cdot S_{\bar{y}} = 2,57 \cdot \sqrt{2 \cdot \left(\frac{1}{2}\right)^2} \cdot 0,742 = 1,35 .$$

Як бачимо, при застосуванні процедур Т'юкі, Н'юмана – Кейлса, Дункана та Ст'юдента всі різниці виявилися вірогідними (крім різниці між середніми 3 та 4 у методі Т'юкі). Причому метод Ст'юдента дає найменшу з усіх різницю, що може призвести до появи хибно значущих різниць; тобто цей метод з малою вірогідністю контролює рівень похибки і ним краще не користуватись.

## 5. КОРЕЛЯЦІЙНИЙ АНАЛІЗ

Кореляційний аналіз застосовується з метою визначення близькості або ступеня супряженості між ознаками, які варіюють, та для визначення форми і спрямованості існуючого між ними зв'язку. За спрямованістю існує позитивна (або пряма) та негативна (або обернена) кореляція, а за формою – лінійна (прямолінійна) та нелінійна (криволінійна). Якщо кореляція позитивна, значення групових середніх однієї ознаки збільшуються із збільшенням значень іншої ознаки. При негативній кореляції збільшення значень однієї ознаки веде до зменшення значень іншої ознаки. Кореляція зветься лінійною, коли напрям зв'язку між ознаками  $X$  та  $Y$  графічно та аналітично виражається прямою лінією. Якщо кореляційна залежність між  $X$  і  $Y$  має іншу форму залежності, то вона називається нелінійною.

Під терміном *близькість зв'язку* розуміють ступінь супряженості ознак. Для позначення близькості зв'язку між змінними величинами  $X$  і  $Y$  у математиці застосовується поняття *функція*. За такого зв'язку певному значенню, що набуває змінна  $X$ , що називається аргументом, відповідає тільки одне значення змінної  $Y$ , яка називається функцією. Залежність, що зветься *функціональною*, можна записати як  $Y = f(X)$ . Прикладом функціональної залежності може слугувати залежність довжини кола від його радіуса. Однак такі однозначні залежності зустрічаються далеко не завжди. У біології частіше зустрічаються залежності, коли певному числовому значенню однієї ознаки відповідає ціла група значень іншої, супряженої з першою, ознаки (або ознак). Такого типу залежність між змінними випадковими величинами  $X$  та  $Y$ , за якої кожному значенню однієї з них відповідає не яке-небудь конкретне значення, а певне групове середнє, тобто  $y = f(X)$  або  $\bar{x} = f(Y)$ , називається *кореляцією*.

Важливим обмеженням кореляційного аналізу є біологічний аналіз ознак перед проведенням розрахунку показників кореляції. Якщо зв'язок між ознаками не має біологічного сенсу, то обчислення показників кореляції також не має ніякого сенсу, бо ці показники не будуть відбивати біологічні закономірності. Наприклад, два не пов'язані між собою процеси: біосинтез хлорофілу та синтез каротиноїдів. Досліджуючи вплив забруднення навколишнього середовища на рослини, можна часто виявити таку закономірність (особливо у стійких видів). Кількість хлорофілу зменшується, а кількість каротину в листках зростає. Якщо провести кореляційний аналіз, можна виявити майже функціональну залежність між цими показниками, хоча біохімічного зв'язку між ними нема. У такому випадку завдання поставлене некоректно. Коректним буде з'ясування таких питань:

1. Залежність швидкості руйнування хлорофілу від міцності його зв'язку з ліпо-протеїдним комплексом.
2. Залежність ступеня руйнування хлорофілу від активності ферменту хлорофілази.

3. Залежність ступеня руйнування хлорофілу від ступеня пошкодження рослин забруднювачами.

4. Залежність змін вмісту каротиноїдів від стійкості рослин.

Щоб визначити ступінь супряженості між ознаками  $X$  та  $Y$ , які варіюють, необхідно порівняти відповідним чином їх значення одне з одним. Якщо із збільшенням однієї ознаки пропорційно збільшуються середні значення іншої ознаки, це вказує на наявність позитивного зв'язку, і навпаки, якщо збільшення значень першої ознаки веде до зменшення середніх значень другої – є зворотний або негативний зв'язок між ними.

У зв'язку з тим, що за наявності кореляції доводиться мати справу не із зростанням функції, а з супряженою варіацією ознак, то виражати її треба у вигляді взаємозалежних відхилень від середніх значень, які характеризують супряжену варіацію двох ознак  $X$  та  $Y$ . Одним з таких показників є емпірична коваріація ( $Cov$ ):

$$Cov = \frac{1}{n} \cdot \sum (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{n} \cdot \sum (x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y})$$

Звичайно, в силу різної розмірності скорельованих величин  $X$  і  $Y$ , порівнюють не самі відхилення від середніх, а їх перетворені (неіменовані) значення у вигляді нормованих відхилень:

$$t_x = \frac{x_i - \bar{x}}{s_x} \quad \text{або} \quad t_y = \frac{y_i - \bar{y}}{s_y}.$$

Звідси витікає емпіричний коефіцієнт кореляції ( $r$ ):

$$r = \frac{Cov}{s_x \cdot s_y} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot s_x \cdot s_y}$$

Коефіцієнт кореляції – величина відносна і виражається у частках одиниці. Таким чином, коефіцієнт кореляції може набувати будь-яких значень в інтервалі від -1 до +1. Причому, значення  $r = 0$  вказує на повну відсутність кореляції, а значення +1 або -1 вказують на прямопропорційну або оберненопропорційну функціональну залежність відповідно.

Як і будь-який біометричний показник, коефіцієнт кореляції обчислюється з певною помилкою, яку можна розрахувати за формулою:

$$S_r = \frac{1 - r^2}{\sqrt{n}}$$

Якщо  $n \leq 100$ , то критерієм для перевірки нульової гіпотези є показник  $t$ :



$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}.$$

Нульова гіпотеза відхиляється якщо  $t \geq t_{st}$  для  $f=n-2$  і обраного рівня значущості  $\alpha$ . Це означає, що в генеральній сукупності  $r \neq 0$  і, отже, вибірковий коефіцієнт кореляції вірогідно відрізняється від нуля, між  $X$  і  $Y$  існує кореляційний зв'язок. Коли  $t < t_{st}$  нульова гіпотеза зберігається (кореляції не існує).

Недоліком коефіцієнта кореляції є обмеженість інтервалу варіювання (-1 .. +1). У зв'язку з цим за малих об'ємів вибірок його розподіл сильно відхиляється від нормального, що може призвести до помилок в оцінці кореляції між ознаками. Тому для малочисельних вибірок застосовується показник  $Z$ , варіювання якого може відбиватися у межах від  $-\infty$  до  $+\infty$  і розподіл якого практично не залежить від об'єму вибірок. Цей показник розраховується за формулою:

$$Z = \frac{1}{2} \cdot \ln \frac{1+r}{1-r} = 1,15129 \cdot \lg \frac{1+r}{1-r}$$

Значення показника  $Z$  можна знайти у спеціальній таблиці, складеній Фішером. У ній наводяться значення  $Z$ , відповідні різним величинам коефіцієнту кореляції ( $r$ ). Величину помилки показника  $Z$  обчислюють за формулою:

$$S_z^2 = \frac{1}{n-3},$$

а показник вірогідності – за формулою:

$$t_z = \frac{Z}{S_z} = Z \cdot \sqrt{n-3}.$$

### Обчислення показників парної кореляції.

Якщо порівнюється лише дві супряжених ознаки, то така кореляція називається *парною*. Під час дослідження парної кореляції у більшості випадків розраховується один з двох можливих коефіцієнтів кореляції:  $Y$  за  $X$  або  $X$  за  $Y$ , тобто як залежать зміни ознаки  $Y$  від змін ознаки  $X$  або навпаки, як змінюється ознака  $X$  із змінами значення ознаки  $Y$ .

Для вибірок невеликого об'єму коефіцієнт кореляції обчислюється без розподілу вибіркового матеріалу в варіаційні ряди та складання кореляційних таблиць. Розрахунки можна вести за формулами, які наведені вище. Однак зручніше користуватися дещо зміненими робочими формулами.

$$r = \frac{n \cdot \sum x \cdot y - \sum x \cdot \sum y}{\sqrt{n \cdot \sum x^2 - (\sum x)^2} \cdot \sqrt{n \cdot \sum y^2 - (\sum y)^2}} \text{ або}$$

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

де через  $x$  і  $y$  позначені парні варіанти супряжених ознак  $X$  та  $Y$ ,  $\bar{x}$  та  $\bar{y}$  – середні арифметичні значення кожної ознаки,  $n$  – кількість парних спостережень або об'єм вибірки.

Щоб використати одну з цих формул, необхідно попередньо знайти проміжні значення. Частина з них обчислюється в процесі первинної статистичної обробки експериментальних даних. Це такі показники, як сума вибірових дат, сума квадратів вибірових дат і вибірові середні арифметичні. Для обчислення суми добутків значень першої і другої ознак краще за все скласти таблицю, до якої заносяться пари ознак і їх добутки (табл. 26, виключаючи графи  $x$  і  $y$ ). Таблиця значень ознак складається і тоді, коли первинна статистична обробка даних не здійснюється. У цьому випадку зручніше заповнити всі графи в таблиці 26.

За підсумковими даними цієї таблиці, якщо необхідно, розраховуються середні значення для першої та другої ознак. Далі за допомогою однієї з наведених вище формул розраховується коефіцієнт кореляції. Наступним кроком аналізу є обчислення помилки коефіцієнта кореляції та його надійного інтервала. На останньому етапі, якщо вибірка складається менше ніж з 50 варіантів, коефіцієнт кореляції перетворюється на показник  $Z$  та розраховуються помилка і вірогідність показника  $Z$ .

Для розрахунку коефіцієнтів кореляції у великих комплексах ( $n > 100$ ) складаються спеціальні кореляційні таблиці. Однак розгляд методів обробки великих комплексів не є завданням цього посібника.

Таблиця 26

Таблиця значень парних ознак для обчислення коефіцієнту кореляції

$x$	$y$	$x \cdot y$	$x^2$	$y^2$
$x_1$	$y_1$	$x_1 \cdot y_1$	$x_1^2$	$y_1^2$
$x_2$	$y_2$	$x_2 \cdot y_2$	$x_2^2$	$y_2^2$
...	...	...	...	...
$x_n$	$y_n$	$x_n \cdot y_n$	$x_n^2$	$y_n^2$
$\sum x_i$	$\sum y_i$	$\sum x_i \cdot y_i$	$\sum x_i^2$	$\sum y_i^2$

Треба враховувати, що невірогідність коефіцієнта кореляції або показника  $Z$  не в усіх випадках вказує на відсутність зв'язку між ознаками, які вивчаються. У більшості випадків невірогідний коефіцієнт кореляції вказує тільки на відсутність прямолінійного зв'язку між супряженими

ознаками, тобто залежності, яка може бути описана рівнянням прямої лінії  $y=a+b \cdot x$ . Для того, щоб остаточно довести наявність або відсутність кореляції між ознаками необхідно додатково обчислити *кореляційне відношення*, яке використовується для вияву криволінійної кореляції.

### Кореляційне відношення.

Для визначення криволінійної залежності між змінними величинами  $X$  та  $Y$  коефіцієнт кореляції непридатний. У таких випадках застосовується інший показник, який був запропонований К. Пірсоном і названий *кореляційним відношенням*, який позначають літерою  $\eta$ . На відміну від коефіцієнта кореляції, який характеризує залежність між змінними  $X$  та  $Y$  з точки зору прямої пропорційності, кореляційне відношення описує її двобічно.

Для обчислення кореляційного відношення вибірки упорядковують спочатку за значеннями одного фактора. Наприклад, у дослідах отримані такі дані:

$X$	2	4	6	8	4	6	2	6
$Y$	4	8	8	7	4	10	6	12

Упорядковуємо цю сукупність за значеннями ознаки  $X$ :

$X$	2	2	4	4	6	6	6	8
$Y$	4	6	4	8	8	10	12	7

Видно, що деякі значення  $X$  повторюються, що дозволяє розподілити цю сукупність наступним чином:

$X$	2	4	6	8
$Y'$	5	6	10	7

де  $Y'$  – окремі групові середні, які отримані усередненням значень  $Y$ , які відповідають однаковим значенням величин  $X$ . Наприклад,  $(4+6)/2=5$ ,  $(10+8+12)/3=10$  і т.д.

Якщо впорядкувати цю сукупність за значеннями  $Y$ , ми отримаємо такий ряд:

$Y$	4	4	6	7	8	8	10	12
$X$	2	4	2	8	6	4	6	6

Цей ряд складається не з чотирьох, як у попередньому випадку, а з шести груп – 4; 6; 7; 8; 10; 12, яким відповідають окремі середні ( $X'$ ) –  $(2+4)/2=3$ ; 2; 8;  $(6+4)/2=5$ ; 6; 6, тобто виходить такий розподіл:

$Y$	4	6	7	8	10	12
$X'$	3	2	8	5	6	6

Таким чином, з'ясовується, що залежність між змінними  $X$  та  $Y$  виражається по різному залежно від того, за значеннями якої з них –  $X$  або  $Y$  – впорядковується сукупність. Цей приклад дозволяє зрозуміти, чому кореляційне відношення характеризує зв'язок між ознаками, що варіюють двобічно –  $Y$  за  $X$  та  $X$  за  $Y$ , а тому і виражається не одним, а двома показниками  $\eta_{x/y}$  та  $\eta_{y/x}$ . Вони обчислюються за аналогічними формулами:

$$\eta_{x/y} = \sqrt{\frac{s_{yx}^2}{s_y^2}} \text{ та } \eta_{y/x} = \sqrt{\frac{s_{xy}^2}{s_x^2}},$$

значення  $s_{yx}^2$  та  $s_{xy}^2$  визначаються як

$$s_{yx}^2 = \frac{\sum (\bar{y}_x - \bar{y})^2}{n} \text{ та } s_{xy}^2 = \frac{\sum (\bar{x}_y - \bar{x})^2}{n},$$

де:  $\bar{y}_x$  – групові середні значення ознаки  $Y$  у впорядкованому за значеннями  $X$  ряді;  $\bar{y}$  – вибіркова середня ознаки  $Y$ ;  $\bar{x}_y$  – групові середні ознаки  $X$  у сукупності, впорядкованій за значеннями ознаки  $Y$ ;  $\bar{x}$  – вибіркова середня ознаки  $X$ . Ці формули можна записати в іншому вигляді, в якому у більшості випадків вони застосовуються для розрахунків:

$$\eta_{y/x} = \sqrt{\frac{\sum (\bar{y}_x - \bar{y})^2}{\sum (y_i - \bar{y})^2}} \text{ та } \eta_{x/y} = \sqrt{\frac{\sum (\bar{x}_y - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Обчислення показників вірогідності кореляційного відношення проводиться за формулою:

$$t = \eta \cdot \sqrt{\frac{n-2}{1-\eta^2}}$$

критичні значення якого для визначеного рівня значущості ( $\alpha$ ) та кількості ступенів вільності ( $f=n-2$ ) знаходять у таблиці Ст'юдента. Нульова гіпотеза, тобто припущення про відсутність зв'язку між ознаками, відхиляється, якщо  $t > t_{st}$ .

Слід враховувати, що кореляційне відношення може приймати тільки позитивні значення (від 0 до +1).

Оскільки коефіцієнт кореляції характеризує тільки лінійний зв'язок, а кореляційне відношення – будь-яку форму залежності, то при точно лінійній залежності між змінними  $X$  та  $Y$  повинна виконуватися рівність між коефіцієнтами кореляційного відношення  $\eta_{y/x}=\eta_{x/y}$ , а також рівність між коефіцієнтом кореляції та кореляційним відношенням ( $\eta=r$ ). За наявності нелінійного зв'язку  $\eta_{y/x}\neq\eta_{x/y}$  та  $\eta\neq r$ . Звідси витікає, що з різниці між цими показниками можна робити висновок про форму кореляційної залежності між ознаками, які варіюють. Як показник лінійності зв'язку, який позначається грецькою літерою  $\gamma$ , використовується різниця між квадратами кореляційного відношення і коефіцієнта кореляції

$$\gamma = \eta^2 - r^2.$$

Вибіркова помилка цього показника визначається за такою приблизною формулою:

$$S_\gamma = 2 \cdot \sqrt{\frac{\gamma}{n}} \cdot \sqrt{(1-\eta^2) - (1-r^2)} + 1 \text{ або}$$

$$S_\gamma = \frac{\sqrt{\gamma - \gamma^2 \cdot (2 - \eta^2 - r^2)}}{\sqrt{n}}$$

У дослідженнях малих вибірок замість  $n$  використовується кількість ступенів вільності ( $f=n-2$ ). Критерієм вірогідності показника  $\gamma$  є його відношення до своєї помилки:

$$t = \frac{\gamma}{S_\gamma}.$$

Якщо  $t < 3$ , кореляція між ознаками оцінюється як прямолінійна. У більш відповідальних випадках оцінки за прямолінійну приймають кореляцію, якщо  $t < 2,5$ .

### Приклад проведення кореляційного аналізу

Вивчалася залежність лінійного росту міцелію гриба пеніцил (вибірка  $Y$ ) від вмісту сахарози в поживному середовищі (вибірка  $X$ ). Потрібно встановити існування залежності між цими показниками та її форму. Для цього потрібно обчислити коефіцієнт кореляції, кореляційне відношення та показник лінійності зв'язку. Вихідні дані та проміжні показники для обчислення коефіцієнта кореляції наведені в табл. 27. Користуючись сумами з таблиці за допомогою формул для обчислення коефіцієнта кореляції та його вірогідності, проводиться кореляційний аналіз.

Спочатку розраховується коефіцієнт кореляції, його помилка та вірогідність :

$$r = \frac{n \cdot \sum x \cdot y - \sum x \cdot \sum y}{\sqrt{n \cdot \sum x^2 - (\sum x)^2} \cdot \sqrt{n \cdot \sum y^2 - (\sum y)^2}} = \frac{10 \cdot 1898,3 - 123 \cdot 145}{57,29 \cdot 27,29} = 0,73,$$

$$S_r = \frac{1 - r^2}{\sqrt{n}} = \frac{1 - 0,533}{3,16} = 0,148, \quad t = \frac{r \cdot \sqrt{n - 2}}{\sqrt{1 - r^2}} = \frac{0,73 \cdot 2,83}{0,68} = 3,03.$$

Таблиця 27

Вихідні дані для проведення кореляційного аналізу

№	X	Y	X·Y	X <sup>2</sup>	Y <sup>2</sup>
1	23,3	19,0	442,7	542,89	361,00
2	6,9	9,0	62,1	47,61	81,00
3	10,4	14,0	145,6	108,16	196,00
4	13,1	12,0	157,2	136,24	144,00
5	10,4	16,0	166,4	108,16	256,00
6	8,7	14,0	121,8	75,69	196,00
7	3,1	14,0	43,4	9,61	196,00
8	22,3	17,0	379,1	497,29	289,00
9	14,4	17,0	244,8	207,36	289,00
10	10,4	13,0	135,2	108,16	169,00
<b>Суми:</b>	<b>123,0</b>	<b>145,0</b>	<b>1898,3</b>	<b>1841,17</b>	<b>2177,00</b>

Порівняння обчисленого значення вірогідності (**3,03**) зі стандартним для  $\alpha=0,05$  та  $f=8$  (**2,31**) дає змогу зробити висновок про існування кореляції між цими показниками. Але вибірки, між якими встановлюється кореляційна залежність мають невеликий об'єм, тому правильнішим буде використання показника  $Z$ . Його можна або обчислити, або знайти у таблиці (табл. 13 додатків). Скористаємося таблицями. Для  $r=0,73$  показник  $Z$  буде дорівнювати **0,929**. Тепер визначимо помилку та вірогідність показника:

$$S_z = \frac{1}{\sqrt{n - 3}} = \frac{1}{2,65} = 0,377, \quad t_z = Z \cdot \sqrt{n - 3} = 0,929 \cdot 2,65 = 2,46.$$

Порівняння обчисленого критерію зі стандартним вказує на вірогідність прямопропорційної кореляції. Для уточнення лінійності кореляції обчислимо кореляційне відношення та показник лінійності зв'язку, їх помилки та вірогідність. На початку аналізу розраховуються середні значення  $X$  та  $Y$ , використовуючи суми з попередньої таблиці:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{123}{10} = 12,3 \text{ та } \bar{y} = \frac{\sum y_i}{n} = \frac{145}{10} = 14,5.$$

Для розрахунків кореляційного відношення переписуються вибірки та впорядковуються за значеннями  $X$ .

$X$	3,1	6,9	8,7	10,4	10,4	10,4	13,1	14,4	22,3	23,3
$Y$	14,0	9,0	14,0	14,0	16,0	13,0	12,0	17,0	17,0	19,0
$(y_i - \bar{y})$	0,5	5,0	0,5	0,5	1,5	1,5	2,5	2,5	2,5	4,5
$(y_i - \bar{y})^2$	0,25	25,0	0,25	0,25	2,25	2,25	6,26	6,25	6,25	20,25

Знаходяться середні значення  $Y$  для однакових значень  $X$ :

$X$	3,1	6,9	8,7	10,4	13,1	14,4	22,3	23,3
$\bar{y}_x$	14,0	9,0	14,0	13,7	12,0	17,0	17,0	19,0
$(\bar{y}_x - \bar{y})$	0,5	5,0	0,5	0,8	2,5	2,5	2,5	4,5
$(\bar{y}_x - \bar{y})^2$	0,25	25,00	0,25	0,64	6,25	6,26	6,25	20,25

Тепер обчислюються суми квадратів відхилень та кореляційне відношення:

$$\sum (\bar{y}_x - \bar{y})^2 = 0,25 + 25,00 + \dots + 6,25 + 20,25 = 62,14,$$

$$\sum (y_i - \bar{y})^2 = 0,25 + 25,00 + 0,25 + \dots + 6,25 + 20,25 = 69,25,$$

$$\eta_{x/y} = \sqrt{\frac{\sum (\bar{y}_x - \bar{y})^2}{\sum (y_i - \bar{y})^2}} = \sqrt{\frac{62,14}{69,25}} = \sqrt{0,897} = 0,95,$$

$$t_\eta = \eta \cdot \sqrt{\frac{n-2}{1-\eta^2}} = 0,95 \cdot \sqrt{\frac{10-2}{1-0,897}} = 8,37.$$

Порівняння обчисленого критерію вірогідності зі стандартним дає змогу зробити висновок про вірогідність кореляційного відношення з ймовірністю 99% ( $t_{st}$  для  $\alpha=0,05$  дорівнює 2,31, а для  $\alpha=0,01$  – 3,41). Далі обчислюється показник лінійності зв'язку та його вірогідність:

$$\gamma = \eta^2 - r^2 = 0,897 - 0,533 = 0,364,$$

$$S_\gamma = \frac{2 \cdot \sqrt{\gamma - \gamma^2} \cdot (2 - \eta^2 - r^2)}{\sqrt{n}} = \frac{2 \cdot \sqrt{0,364 - 0,132} \cdot (2 - 0,897 - 0,533)}{\sqrt{10}} = 0,341,$$

$$t_\gamma = \frac{\gamma}{S_\gamma} = \frac{0,364}{0,341} = 1,07.$$

Оскільки отримане значення  $t_\gamma$  менше 2,5, кореляція визнається лінійною.

### Множинна кореляція.

Обчислюючи коефіцієнт кореляції між двома ознаками  $X$  і  $Y$ , ми не враховуємо їх залежності від інших ознак організму, які варіюють. Між тим відомо, що біологічний об'єкт – це складна система, частини і ознаки якої, складним чином взаємодіють між собою. Тому поряд з вивченням парних кореляцій перед дослідником часто виникає завдання обчислення множинних зв'язків між ознаками організму. Воно вирішується за допомогою коефіцієнтів множинної і окремих (парціальних) кореляцій. Сукупний коефіцієнт кореляції між ознаками  $X$ ,  $Y$  і  $Z$  обчислюється за формулою:

$$r_{xyz} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2 \cdot r_{xy} \cdot r_{xz} \cdot r_{yz}}{1 - r_{xy}^2}},$$

де:  $r_{xy}$ ,  $r_{xz}$  і  $r_{yz}$  парні коефіцієнти кореляції між ознаками  $X$  і  $Y$ ,  $X$  і  $Z$  та  $Y$  і  $Z$ . Сукупний коефіцієнт кореляції має ті самі властивості, що і парні коефіцієнти кореляції. Але, на відміну від них, цей показник завжди має позитивні значення (коливається в межах від 0 до +1).

Крім того, на практиці часто використовуються парціальні коефіцієнти кореляції, які дозволяють виявити зв'язок кожного з двох факторів, виключаючи вплив третього. Формули для їх обчислення такі (у дужках у формулах наводяться фактори, які виключаються):

$$r_{xy(z)} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2) \cdot (1 - r_{yz}^2)}},$$

$$r_{xz(y)} = \frac{r_{xz} - r_{xy} \cdot r_{yz}}{\sqrt{(1 - r_{xy}^2) \cdot (1 - r_{yz}^2)}},$$

$$r_{yz(x)} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{(1 - r_{xy}^2) \cdot (1 - r_{xz}^2)}}.$$

Послідовність розрахунків така.

1. Складається таблиця комплексу (табл. 28).
2. Розраховуються значення парних коефіцієнтів кореляції за формулами, наведеними у попередньому розділі.

Таблиця 28

Таблиця значень ознак для обчислення коефіцієнту множинної кореляції



$X$	$Y$	$Z$	$X^2$	$Y^2$	$Z^2$	$X \cdot Y$	$X \cdot Z$	$Y \cdot Z$
$x_1$	$y_1$	$z_1$	$x_1^2$	$y_1^2$	$z_1^2$	$x_1 \cdot y_1$	$x_1 \cdot z_1$	$y_1 \cdot z_1$
$x_2$	$y_2$	$z_2$	$x_2^2$	$y_2^2$	$z_2^2$	$x_2 \cdot y_2$	$x_2 \cdot z_2$	$y_2 \cdot z_2$
...	...	...	...	...	...	...	...	...
$x_n$	$y_n$	$z_n$	$x_n^2$	$y_n^2$	$z_n^2$	$x_n \cdot y_n$	$x_n \cdot z_n$	$y_n \cdot z_n$
$\sum x_i$	$\sum y_i$	$\sum z_i$	$\sum x_i^2$	$\sum y_i^2$	$\sum z_i^2$	$\sum x_i \cdot y_i$	$\sum x_i \cdot z_i$	$\sum y_i \cdot z_i$

3. Використовуючи парні коефіцієнти кореляції, обчислюють сукупний коефіцієнт кореляції, окремі коефіцієнти кореляції, їх помилки та вірогідність.

### Приклад обчислення показників множинної кореляції

Вивчалася залежність між довжиною колосу жита ( $X$ ), кількістю колосків ( $Y$ ) та кількістю зернин у кожному колосі ( $Z$ ). Отримані дані та проміжні значення наведені в табл. 29. Потрібно встановити зв'язок між цими показниками, тобто обчислити загальний та окремі коефіцієнти кореляції. На початку аналізу, використовуючи формули парних коефіцієнтів кореляції, обчислюється зв'язок між парами  $X \rightarrow Y$ ,  $X \rightarrow Z$ ,  $Y \rightarrow Z$ .

Таблиця 29

Вихідні дані для обчислення показників множинної кореляції

№	$X$	$Y$	$Z$	$X^2$	$Y^2$	$Z^2$	$X \cdot Y$	$Y \cdot Z$	$X \cdot Z$
1	70	18	36	4900	324	1296	1260	648	2520
2	60	17	29	3600	289	841	1020	493	1740
3	70	22	40	4900	484	1600	1540	880	2800
4	46	10	12	2116	100	144	460	120	552
5	58	16	31	3364	256	961	928	496	1798
6	69	18	32	4761	324	1024	1242	576	2208
7	32	9	13	1024	81	169	288	117	416
8	62	18	35	3844	324	1225	1116	630	2170
9	46	15	30	2116	225	900	690	450	1380
10	62	22	36	3844	484	1296	1364	792	2232
<b>Суми</b>	<b>575</b>	<b>165</b>	<b>294</b>	<b>34469</b>	<b>2891</b>	<b>9456</b>	<b>9908</b>	<b>5202</b>	<b>17816</b>

$$r_{xy} = \frac{n \cdot \sum x \cdot y - \sum x \cdot \sum y}{\sqrt{n \cdot \sum x^2 - (\sum x)^2} \cdot \sqrt{n \cdot \sum y^2 - (\sum y)^2}} = \frac{4205}{4868,53} = 0,865,$$

$$r_{xz} = \frac{n \cdot \sum x \cdot z - \sum x \cdot \sum z}{\sqrt{n \cdot \sum x^2 - (\sum x)^2} \cdot \sqrt{n \cdot \sum z^2 - (\sum z)^2}} = \frac{9110}{10748,72} = 0,852,$$

$$r_{yz} = \frac{n \cdot \sum y \cdot z - \sum y \cdot \sum z}{\sqrt{n \cdot \sum y^2 - (\sum y)^2} \cdot \sqrt{n \cdot \sum z^2 - (\sum z)^2}} = \frac{3510}{3720,36} = 0,941.$$

Далі обчислюються окремі та загальний коефіцієнти кореляції:

$$r_{xy(z)} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2) \cdot (1 - r_{yz}^2)}} = \frac{0,864 - 0,941 \cdot 0,852}{\sqrt{(1 - 0,941^2) \cdot (1 - 0,852^2)}} = 0,336,$$

$$r_{xz(y)} = \frac{r_{xz} - r_{xy} \cdot r_{yz}}{\sqrt{(1 - r_{xy}^2) \cdot (1 - r_{yz}^2)}} = \frac{0,852 - 0,864 \cdot 0,941}{\sqrt{(1 - 0,864^2) \cdot (1 - 0,941^2)}} = 0,201,$$

$$r_{yz(x)} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{(1 - r_{xy}^2) \cdot (1 - r_{xz}^2)}} = \frac{0,941 - 0,864 \cdot 0,852}{\sqrt{(1 - 0,864^2) \cdot (1 - 0,852^2)}} = 0,807,$$

$$r_{xyz} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2 \cdot r_{xy} \cdot r_{xz} \cdot r_{yz}}{1 - r_{xy}^2}} = \sqrt{\frac{0,864^2 + 0,852^2 - 2 \cdot 0,864 \cdot 0,852 \cdot 0,941}{1 - 0,864^2}} = \sqrt{\frac{0,0752}{0,0994}} = 0,870,$$

Останнім кроком є обчислення вірогідності коефіцієнтів кореляції, яке проводиться за формулою:

$$t_r = \frac{r_{окр} \cdot \sqrt{n - \nu}}{\sqrt{1 - r_{окр}^2}},$$

де:  $r_{окр}$  – окремий коефіцієнт кореляції;

$n$  – кількість супряжених даних;

$\nu$  – кількість враховуваних ознак.

$$t_{xy(z)} = \frac{0,336 \cdot \sqrt{10 - 3}}{\sqrt{1 - 0,336^2}} = 0,94,$$

$$t_{xz(y)} = \frac{0,201 \cdot \sqrt{10 - 3}}{\sqrt{1 - 0,201^2}} = 0,54,$$

$$t_{yz(x)} = \frac{0,807 \cdot \sqrt{10 - 3}}{\sqrt{1 - 0,807^2}} = 3,61,$$

Порівняння цих величин зі стандартними вказує на вірогідний зв'язок лише між ознаками  $Y$  та  $Z$ , всі інші окремі коефіцієнти кореляції виявилися невірогідними. Разом з тим, загальний коефіцієнт кореляції є досить високим, що може свідчити про взаємовплив між ознаками, які вивчалися.

## 6. РЕГРЕСІЙНИЙ АНАЛІЗ

### Основні положення та застосування в статистичній обробці експериментальних даних.

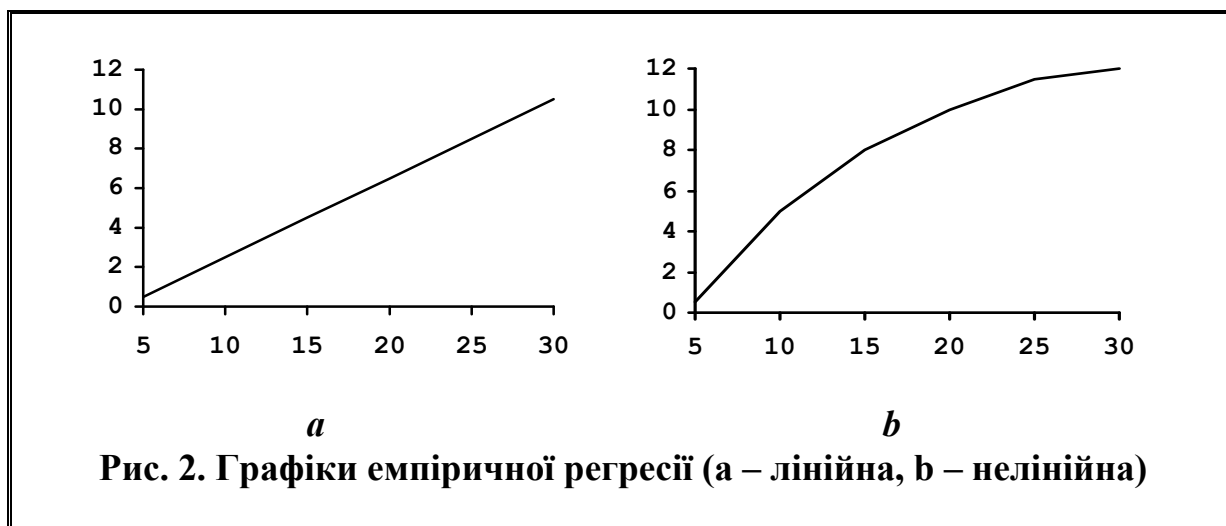
Проведення дисперсійного та кореляційного аналізу дозволяє встановити існування впливу факторів на результативну ознаку або з'ясувати близькість супряженості двох або більше ознак. При проведенні кореляційного аналізу можна також встановити напрям (прямопропорційний або оберненопропорційний) та форму (прямолінійна або криволінійна) залежності. Але ці методи статистичної обробки не дозволяють встановити ступінь впливу фактора або взаємозалежності ознак. За результатами обчислень дисперсійного або кореляційного аналізу не можна встановити на скільки одиниць в середньому зміниться значення результативної ознаки при зміні на одиницю виміру фактора або іншої, пов'язаної з першою, ознаки. Між тим такі залежності становлять для дослідника не абиякий інтерес і займають важливе місце у статистичному аналізі групових властивостей. Такі завдання дозволяє вирішувати *регресійний аналіз*. Цей метод статистичної обробки проводиться з метою встановлення математичного очікування величини результативної ознаки від змін значень діючої ознаки або фактора. Функція, яка дозволяє за величиною однієї (незалежної) ознаки ( $X$ ) знаходити очікувані середні значення іншої ознаки ( $Y$ ), пов'язаної із значеннями незалежної змінної, називається *регресією* або *рівнянням регресії*. У результаті регресійного аналізу розраховуються коефіцієнти рівняння теоретичної залежності між ознаками, які вивчаються.

У завдання регресійного аналізу входять два практично самостійних завдання: *вирівнювання емпіричних рядів регресії та встановлення зв'язку між результатами експерименту (розрахунок коефіцієнтів рівняння регресії)*.

#### Вирівнювання емпіричних рядів регресії

Для вирівнювання емпіричних рядів регресії можуть використовуватися графічний та аналітичний методи.

Першим етапом графічного методу є побудова графіка залежності результативної ознаки від незалежної змінної (рис. 2). Аналіз графічного зображення регресії може дати уяву про форму залежності та її напрям. Якщо в завдання експерименту входить елімінація випадкових коливань емпіричної лінії регресії і встановлення загального напрямку залежності, то найлегшим методом є графічне вирівнювання ряду.



Графічний метод вирівнювання зручніше використовувати для лінійних залежностей. Складнішим є вирівнювання криволінійних графіків. У будь-якому випадку спочатку будується графік емпіричної залежності. А далі між крайніми значеннями ламаної лінії проводиться пряма або плавна крива таким чином, щоб сума відстаней від точок теоретичної (вирівняної) лінії регресії до точок емпіричної лінії регресії була найменшою (рис. 3).

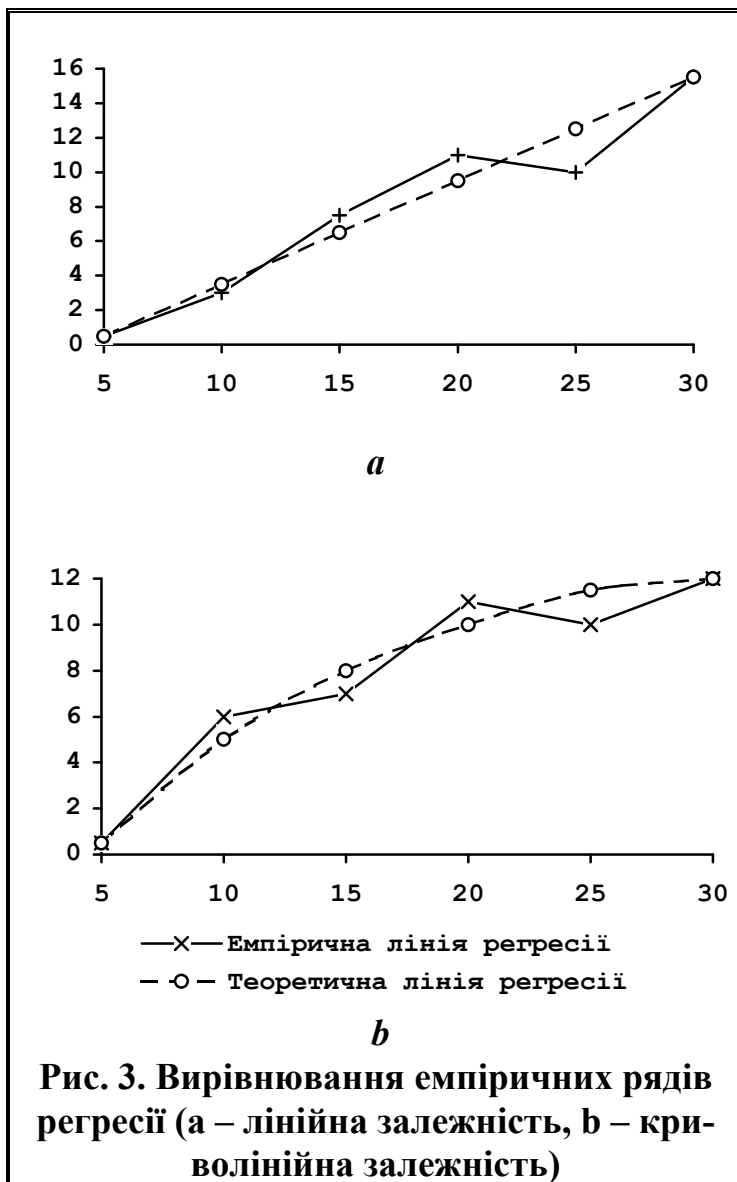
Слід враховувати, що графічним методом можна лише приблизно вирівнювати емпіричні ряди. Якщо потрібне точніше вирівнювання, застосовуються аналітичні методи. Найчастіше використовуються методи ковзкої середньої та зваженої ковзкої середньої. Ці методи можна застосовувати і в тих випадках, коли залежність між параметрами невідома.

#### Метод ковзкої середньої

Метод ковзкої середньої полягає в тому, що отримані емпіричним шляхом значення  $y_i$  розташовані за фіксованими значеннями  $x_i$ , замінюють новими, отриманими усередненням трьох або п'яти розташованих поруч значень  $y_i$  (суму перших в ряді трьох або п'яти дат ділять відповідно на 3 або 5). Для отримання наступного значення  $y'$  беруть нові три або п'ять значень  $y_i$ , зсунутих на одиницю, що можна записати у вигляді формул:

$$y'_1 = \frac{y_1 + y_2 + y_3}{3}; \quad y'_2 = \frac{y_2 + y_3 + y_4}{3} \text{ і т.д., або}$$

$$y'_1 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}; \quad y'_2 = \frac{y_2 + y_3 + y_4 + y_5 + y_6}{5} \text{ і т.д.}$$



**Рис. 3. Вирівнювання емпіричних рядів регресії (а – лінійна залежність, б – криволінійна залежність)**

Отриману величину ділять на 2. Це буде перше додаткове значення ( $y_{-1}$  або  $y_{+1}$ ). Для обчислення другого додаткового значення використовують перше додаткове значення. Це може бути записано такими формулами:

$$y_{-1} = \frac{2 \cdot y_1 + y_2 - y_4}{2}, \quad y_{+1} = \frac{2 \cdot y_n + y_{n-1} - y_{n-3}}{2},$$

$$y_{-2} = \frac{2 \cdot y_{-1} + y_1 - y_3}{2}, \quad y_{+2} = \frac{2y_{+1} + y_n - y_{n-2}}{2}$$

Якщо емпіричні ряди короткі, то для обчислення краще використовувати формули:

$$y_{-1} = \frac{4 \cdot y_1 + y_2 - 2 \cdot y_3}{3}, \quad y_{+1} = \frac{4 \cdot y_n + y_{n-1} - 2 \cdot y_{n-2}}{3},$$

Недолік цього методу полягає в тому, що втрачаються значення крайніх варіант. Якщо об'єми вибірок великі, то така втрата не позначається суттєво на подальшій роботі. Коли об'єми вибірок малі (4–10 дат), тоді втрата хоча б однієї з них суттєво впливає на подальшу роботу.

Точніші і не пов'язані з втратою крайніх значень результати можна отримати, використовуючи метод зваженої ковзкої середньої. У цьому методі з обох кінців ряду додаються по два нових члени ряду. Вони визначаються наступним чином. Перше з початку або кінця ряду помножується на 2, до отриманого добутку додається наступне (попереднє) в ряді значення, третє значення пропускається, а від отриманої суми віднімають четверте значення.

$$y_{+2} = \frac{4 \cdot y_{-1} + y_1 - 2 \cdot y_2}{3}, \quad y_{+2} = \frac{4 \cdot y_{+1} + y_n - 2 \cdot y_{n-1}}{3}$$

Вирівняні значення отримують шляхом обчислення зваженої середньої з п'яти сусідніх емпіричних значень, взятих з відповідними коефіцієнтами 1; 2; 4; 2; 1. Отримане значення ділять на 10. Наприклад  $y'_1$  можна отримати за формулою:

$$y'_1 = \frac{y_{-2} + 2 \cdot y_{-1} + 4 \cdot y_1 + 2 \cdot y_2 + y_3}{10}$$

Інші значення отримують аналогічно, зсуваючи кожного разу емпіричний ряд на одиницю, тобто для обчислення  $y'_2$  беруть значення  $y_{-1}$ ,  $y_1$ ,  $y_2$ ,  $y_3$  та  $y_4$ , і т.д. Цей метод є найточнішим з усіх методів вирівнювання і не призводить до втрати даних.

### Приклад застосування методу ковзкої середньої

Потрібно провести вирівнювання емпіричного ряду з 6 дат:

**12,3; 18,5; 18,9; 24,5; 30,6; 30,7.**

Середнє арифметичне ряду дорівнює **22,58**. Оскільки ряд малий і втрата хоча б однієї дати може призвести до значної похибки, використовується метод зваженої ковзкої середньої. Спочатку обчислюються додаткові дані:

$$y_{-1} = \frac{2 \cdot y_1 + y_2 - y_4}{2} = \frac{2 \cdot 12,3 + 18,5 - 24,5}{2} = 9,3,$$

$$y_{-2} = \frac{2 \cdot y_{-1} + y_1 - y_3}{2} = \frac{2 \cdot 9,3 + 12,3 - 18,9}{2} = 6,0,$$

$$y_{+1} = \frac{2 \cdot y_n + y_{n-1} - y_{n-3}}{2} = \frac{2 \cdot 30,7 + 30,6 - 18,9}{2} = 36,55,$$

$$y_{+2} = \frac{2 \cdot y_{+1} + y_n - y_{n-2}}{2} = \frac{2 \cdot 36,55 + 30,7 - 24,5}{2} = 39,65.$$

Після обчислень ряд буде виглядати так:

**6,0; 9,3; 12,3; 18,5; 18,9; 24,5; 30,6; 30,7; 36,55; 39,65.**

Далі, використовуючи додані дані, обчислюються виправлені значення ряду  $y'_i$ . Для цього застосовується усереднення п'яти сусідніх даних з відповідними ваговими коефіцієнтами:

$$y'_1 = \frac{6,0 + 2 \cdot 9,3 + 4 \cdot 12,3 + 2 \cdot 18,5 + 18,9}{10} = 12,97,$$

$$y'_2 = \frac{9,3 + 2 \cdot 12,3 + 4 \cdot 18,5 + 2 \cdot 18,9 + 24,5}{10} = 17,02,$$

$$y'_3 = \frac{12,3 + 2 \cdot 18,5 + 4 \cdot 18,9 + 2 \cdot 24,5 + 30,6}{10} = 20,45,$$

$$y'_4 = \frac{18,5 + 2 \cdot 18,9 + 4 \cdot 24,5 + 2 \cdot 30,6 + 30,7}{10} = 24,62,$$

$$y'_5 = \frac{18,9 + 2 \cdot 24,5 + 4 \cdot 30,6 + 2 \cdot 30,7 + 36,55}{10} = 28,83,$$

$$y'_6 = \frac{24,5 + 2 \cdot 30,6 + 4 \cdot 30,7 + 2 \cdot 36,55 + 39,65}{10} = 32,13.$$

У результаті буде отримано вирівняний ряд регресії, який складається з таких варіант:

**12,97; 17,02; 20,45; 24,62; 28,83; 32,13.**

Середня арифметична цього ряду становить **22,67**, тобто досить близька до середньої арифметичної похідного ряду.

### **Обчислення коефіцієнтів рівняння лінійної регресії**

Якщо потрібно встановити закономірності змін значень результативної ознаки від значень незалежної змінної, потрібно обчислити коефіцієнти рівняння регресії. Таке рівняння дозволяє обчислювати очікувані (теоретичні) значення результативної ознаки для визначеного значення незалежної змінної. Закономірності зміни значень результативної ознаки можуть бути різними. Тому рівняння регресії можуть бути лінійними, гіперболічними, логарифмічними та іншими. Це завдання вирішується тільки аналітичним методом.

Найпростішим є рівняння прямолінійної регресії, яке виражається формулою

$$\tilde{y}_i = b_0 + b_1 \cdot x_i,$$

де:  $y_i$  – очікуване (теоретичне) значення результативної ознаки;

$b_0$  – вільний член рівняння регресії;

$b_1$  – коефіцієнт пропорційності;

$x_i$  – незалежна змінна.

Найчастіше для знаходження коефіцієнтів рівняння регресії застосовується *метод найменших квадратів*, який полягає в тому, що сума квадратів відхилень теоретичних значень від емпіричних значень буде най-



меншою. Цей метод застосовується для вирішення рівнянь лінійної та криволінійної залежності в однофакторних та багатфакторних експериментах. Обчислення коефіцієнтів проводяться за такою схемою (незалежно від ступеню рівняння):

1. визначення загального вигляду рівняння регресії на основі попереднього аналізу даних (можна побудувати графік залежності);
2. складання системи нормальних рівнянь. Для цього спочатку залежна змінна переноситься у праву частину рівняння. Далі всі члени рівняння помножуються на величини, які знаходяться поруч з коефіцієнтом, визначається  $-1$ ,  $x$ ,  $x^2$  і т.д. Перед змінними ставиться знак суми.

Для рівняння лінійної залежності цю послідовність можна записати у вигляді наступних формул:

1. Вихідне рівняння –  $y = b_0 + b_1 \cdot x$ ;
2. Перетворене рівняння –  $b_0 + b_1 \cdot x = y$ ;
3. Всі члени рівняння помножуються на  $1$  (величина для  $b_0$ ) і ставиться знак суми перед змінними –  $k \cdot b_0 + b_1 \cdot \sum x = \sum y$ ;
4. Всі члени рівняння помножуються на  $x$  (величина для  $b_1$ ) і ставиться знак суми перед змінними –  $b_0 \cdot \sum x + b_1 \cdot \sum x^2 = \sum x \cdot y$ .

Тепер можна записати систему нормальних рівнянь для визначення коефіцієнтів:

$$\begin{cases} k \cdot b_0 + b_1 \cdot \sum x = \sum y \\ b_0 \cdot \sum x + b_1 \cdot \sum x^2 = \sum x \cdot y. \end{cases}$$

Вирішення цієї системи рівнянь в алгебраїчному вигляді дозволяє отримати робочі формули, за якими і проводиться обчислення коефіцієнтів рівняння в практиці статистичного аналізу:

$$b_1 = \frac{k \cdot \sum x \cdot y - \sum x \cdot \sum y}{k \cdot \sum x^2 - (\sum x)^2}$$

$$b_0 = \frac{\sum y - b_1 \cdot \sum x}{k}$$

Як і будь-який біометричний показник, коефіцієнт пропорційності обчислюється з певною похибкою, яку можна обчислити за формулою:

$$S_b = \sqrt{\frac{\sum (y_i - \bar{y})^2 - \frac{[\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})]^2}{\sum (x_i - \bar{x})^2}}{(k-2) \cdot \sum (x_i - \bar{x})^2}}$$

Критерієм вірогідності коефіцієнта пропорційності є відношення його значення до його помилки:

$$t = \frac{b}{S_b}$$

Якщо обчислене значення дорівнює або перевищує стандартне значення критерію Ст'юдента для кількості ступенів вільності  $f=k-1$  ( $k$  – кількість пар даних) та обраного рівня значущості  $\alpha$ , коефіцієнт пропорційності вірогідно відмінний від нуля для обраного рівня значущості.



Слід враховувати, що точки емпіричної лінії регресії практично ніколи не будуть знаходитися на прямій лінії. Тому теоретична лінія регресії є такою прямою, яка проходить найближче до всіх точок емпіричного ряду. Вибіркове значення коефіцієнту регресії є оцінкою відповідного генерального показника і відрізняється від нього в середньому на величину  $S_b$ . Це означає, що "дійсна" лінія регресії знаходиться (при великих об'ємах вибірки) в межах пари вертикальних кутів, утворених у точці перетину (точка  $\bar{x}, \bar{y}$ ) двох

прямих  $PP'$  та  $QQ'$  (рис. 4) з нахилами  $b_0 - S_b$  та  $b_0 + S_b$ . Тому для кожного значення  $x$  "дійсне" значення розрахункового (теоретичного середнього) буде знаходитися в інтервалі  $y_x = \bar{y} \pm S_b \cdot (x - \bar{x})$ . Оскільки інтервал буде збільшуватися з віддаленням значень  $x$  від середнього значення  $\bar{x}$ , то теоретичне передбачення буде мати меншу ймовірність на краях діапазону змін незалежної змінної. Інтервал, в якому буде знаходитися генеральна середня (з ймовірністю 67%), позначений на малюнку відрізком  $AB$ .

прямих  $PP'$  та  $QQ'$  (рис. 4) з нахилами  $b_0 - S_b$  та  $b_0 + S_b$ . Тому для кожного значення  $x$  "дійсне" значення розрахункового (теоретичного середнього) буде знаходитися в інтервалі  $y_x = \bar{y} \pm S_b \cdot (x - \bar{x})$ . Оскільки інтервал буде збільшуватися з віддаленням значень  $x$  від середнього значення  $\bar{x}$ , то теоретичне передбачення буде мати меншу ймовірність на краях діапазону змін незалежної змінної. Інтервал, в якому буде знаходитися генеральна середня (з ймовірністю 67%), позначений на малюнку відрізком  $AB$ .

## Обчислення коефіцієнтів рівнянь регресії при нелінійних залежностях

У багатьох випадках залежності зміни результативної ознаки від зміни значень незалежної змінної мають нелінійний характер. Найчастіше в біології зустрічаються гіперболічна, логарифмічна, степенева, показникова, параболічна та логістична залежності. Рідше зустрічаються періодичні залежності, які описуються формулами тригонометричних функцій. Далі будуть охарактеризовані основні нелінійні залежності.

### Гіперболічна залежність

*Гіперболічна* залежність досить часто зустрічається в практиці біологічних досліджень. Наприклад, залежність росту паростків рослин на забрудненому ґрунті від концентрації забруднювачів у більшості випадків описується саме гіперболічною залежністю. Гіперболічна залежність може бути виражена наступними рівняннями:

$$y = b_0 + \frac{b_1}{x} \quad \text{або} \quad y = b_0 + b_1 \cdot x + \frac{b_2}{x}.$$

Для обчислення коефіцієнтів рівняння складається система нормальних рівнянь, яка виглядає так:

$$\begin{cases} b_0 \cdot k + b_1 \cdot \sum \frac{1}{x} = \sum y \\ b_0 \cdot \sum \frac{1}{x} + b_1 \cdot \sum \frac{1}{x^2} = \sum \frac{y}{x} \end{cases}$$

Алгебраїчне вирішення цієї системи рівнянь дає наступні робочі формули для визначення коефіцієнтів першого рівняння:

$$b_0 = \frac{\sum y \cdot \sum \frac{1}{x^2} - \sum \frac{y}{x} \cdot \sum \frac{1}{x}}{k \cdot \sum \frac{1}{x^2} - \left( \sum \frac{1}{x} \right)^2}$$
$$b_1 = \frac{k \cdot \sum \frac{y}{x} - \sum y \cdot \sum \frac{1}{x}}{k \cdot \sum \frac{1}{x^2} - \left( \sum \frac{1}{x} \right)^2}$$

Таким чином, для розрахунку коефіцієнтів рівняння гіперболічної регресії потрібно обчислити декілька допоміжних сум ( $\sum y$ ,  $\sum y/x$ ,  $\sum 1/x$  та  $\sum 1/x^2$ ). Пі-

Для обчислення коефіцієнтів рівняння регресії визначають помилку коефіцієнта та адекватність рівняння.

Показникова (експоненціальна) функція

У тих випадках, коли основна тенденція зміни значень емпіричного ряду регресії близька до геометричної прогресії, вона задовільно описується рівняннями *показникової* або *експоненціальної* функції:

$$y = b_0 \cdot b_1^x \text{ або } y = b_0 \cdot e^{b_1 \cdot x}$$

Обчислення коефіцієнтів рівнянь такого типу починається з того, що функції приводяться до вигляду лінійних, які можуть графічно бути виражені прямою лінією. Для цього ліву і праву частину рівнянь логарифмують, що веде до наступного перетворення вихідних рівнянь:

$$y = b_0 \cdot b_1^x \text{ перетворюється на } \lg y = \lg b_0 + x \cdot \lg b_1, \text{ а}$$

$$y = b_0 \cdot e^{b_1 \cdot x} \text{ – на } \ln y = \ln b_0 + x \cdot b_1$$

Для першого рівняння коефіцієнти знаходять за такими робочими формулами:

$$\lg b_0 = \frac{\sum \lg y \cdot \sum x^2 - \sum (x \cdot \lg y) \cdot \sum x}{k \cdot \sum x^2 - (\sum x)^2},$$

$$\lg b_1 = \frac{k \cdot \sum (x \cdot \lg y) - \sum x \cdot \sum \lg y}{k \cdot \sum x^2 - (\sum x)^2}.$$

Таким чином, для обчислення коефіцієнтів рівняння потрібно обчислити суми  $x_i$ ,  $x_i^2$ ,  $\lg y_i$  та  $x \cdot \lg y_i$ .

Степенева функція

У деяких випадках залежність між біологічними процесами задовільно описується рівнянням степеневі функції:

$$y = b_0 \cdot x^{b_1},$$

яке в результаті логарифмування перетворюється на рівняння прямої лінії:

$$\lg y = \lg b_0 + b_1 \cdot \lg x.$$

Після складання системи нормальних рівнянь та їх вирішення в алгебраїчному вигляді отримуються робочі формули для обчислення коефіцієнтів рівняння степеневі регресії:

$$\lg b_0 = \frac{\sum \lg y \cdot \sum (\lg x)^2 - \sum (\lg x \cdot \lg y) \cdot \sum \lg x}{k \cdot \sum (\lg x)^2 - (\sum \lg x)^2},$$

$$b_1 = \frac{n \cdot \sum (\lg x \cdot \lg y) - \sum \lg x \cdot \sum \lg y}{k \cdot \sum (\lg x)^2 - (\sum \lg x)^2}.$$

Після обчислення коефіцієнтів рівнянь показникової, експоненціальної та степеневі залежності, потрібно зробити зворотні перетворення коефіцієнтів та значень результативної ознаки.

#### Параболічні залежності

Параболічні залежності виражаються квадратичними, кубічними та рівняннями більш високого ступеня. Найбільш простим методом вирішення таких рівнянь є використання поліномів Чебишева. Але для застосування цього методу необхідно дуже чітко дотримуватися певних правил планування експерименту та варіювання незалежної змінної, що є не завжди можливим в умовах біологічного експерименту. Однак і у випадку параболічних залежностей можна використовувати метод найменших квадратів. Але обсяг обчислювальної роботи при цьому набагато збільшується порівняно з методом Чебишева. Разом з тим у біології дуже рідко зустрічаються залежності, які виражаються поліномами високого ступеня. Тому ми розглянемо тільки метод вирішення рівняння *другого ступеню* (або *квадратичної залежності*), яке виражається формулою:

$$y = b_0 + b_1 \cdot x + b_2 \cdot x^2.$$

Така залежність досить задовільно характеризує процеси, які мають один максимум або один мінімум. Наприклад, вплив вмісту деяких речовин в ґрунті на ріст рослин. Так, сірка та мікроелементи при малих концентраціях зумовлюють покращення ростових процесів рослин. Більш високі концентрації, які перевищують оптимальні, викликають пригнічення росту аж до загибелі рослин, тобто ці елементи стають забруднювачами.

Як бачимо, в рівнянні три параметри, які необхідно знайти для його вирішення, тому система нормальних рівнянь буде складатися з трьох рівнянь:

$$\begin{cases} b_0 \cdot k + b_1 \cdot \sum x + b_2 \cdot \sum x^2 = \sum y \\ b_0 \cdot \sum x + b_1 \cdot \sum x^2 + b_2 \cdot \sum x^3 = \sum x \cdot y \\ b_0 \cdot \sum x^2 + b_1 \cdot \sum x^3 + b_2 \cdot \sum x^4 = \sum x^2 \cdot y \end{cases}$$

Робочі формули для обчислення коефіцієнтів рівняння параболічної залежності мають такий вигляд:

$$b_0 = \frac{\sum y \cdot \sum x^4 - \sum x^2 \cdot \sum x^2 \cdot y}{k \cdot \sum x^4 - (\sum x^2)^2},$$

$$b_1 = \frac{\sum x \cdot y}{\sum x^2}, \quad b_2 = \frac{k \cdot \sum x^2 \cdot y - \sum x^2 \cdot \sum y}{k \cdot \sum x^4 - (\sum x^2)^2}.$$

Таким чином, для обчислення параметрів рівняння необхідно обчислити суми  $x$ ,  $y$ ,  $x \cdot y$ ,  $x^2$ ,  $x^3$ ,  $x^2 \cdot y$  та  $x^4$ .

#### Логістична залежність

При статистичній обробці мікробіологічних досліджень, наприклад, вивченні швидкості росту мікроорганізмів, дуже часто спостерігаються так звані логістичні залежності, які характеризуються похилою початковою частиною, далі функція швидко зростає і наприкінці виходить на плато. Така сама залежність зустрічається при вивченні процесів росту, збільшення чисельності популяцій та деяких інших процесів. Ці процеси характеризуються диференціальним рівнянням типу:

$$\frac{dy}{dx} = KY,$$

де:  $K$  – середня швидкість процесу;  
 $x$  – час, протягом якого відбувається процес.

Вирішення цього рівняння приводить до формули експоненціальної залежності

$$Y = b_0 \cdot e^{K \cdot x},$$

де  $b_0$  – початкове значення параметра, коли  $X=0$ .

Це рівняння задовільно описує наведені залежності (наприклад, ріст популяції), якщо процеси відбуваються в умовах необмеженого часу та простору, і є еквівалентним геометричній прогресії.

Якщо процеси відбуваються в умовах обмеження (наприклад, експериментальні умови), то вони описуються рівнянням логістичної функції:

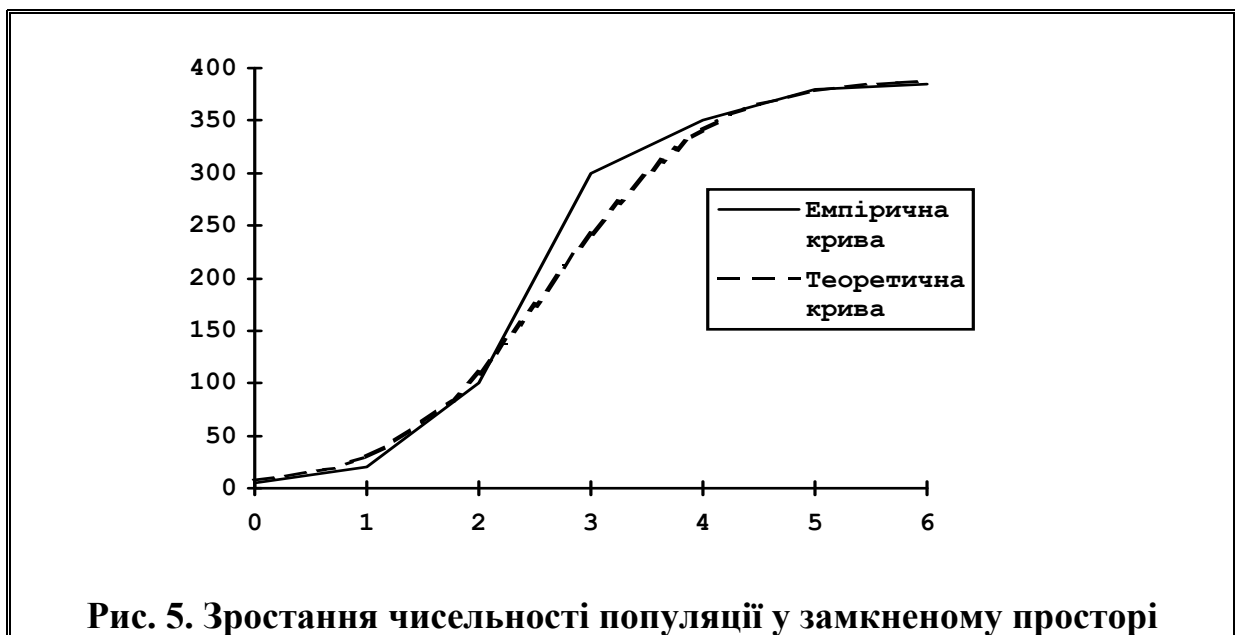
$$\frac{dy}{dx} = K \cdot Y \cdot (A - Y), \text{ або після інтегрування } x = \frac{A}{1 + e^{-A \cdot K \cdot (-x)}},$$

де:  $x$  – час, протягом якого досягається половина кінцевої величини ознаки ( $A$ ), тобто верхньої асимптоти кривої. Цю залежність можна показати на наступному графіку (рис. 5).

Зручнішою для розрахунків є формула Ферхюльста:

$$Y = \frac{A}{1 + 10^{a+b \cdot x}} + C,$$

де:  $Y$  – значення ознаки, яка вивчається;  
 $x$  – час, який пройшов з початку до закінчення процесу;  
 $C$  – початкове значення ознаки, з якого розпочато вимірювання;  
 $a$  та  $b$  – параметри рівняння, які потрібно визначити.



**Рис. 5. Зростання чисельності популяції у замкненому просторі**

Для проведення обчислень рівняння логарифмують:

$$\lg\left(\frac{A}{Y - C} - 1\right) = a + b \cdot x.$$

Якщо позначити ліву частину рівняння через  $\lg z$ , отримаємо параболу першого порядку  $\lg z = a + b \cdot x$ . Для визначення параметрів цього рівняння використовується система нормальних рівнянь:

$$\begin{cases} \sum \lg z = k \cdot a + b \cdot \sum x \\ \sum x \cdot \lg z = a \cdot \sum x + b \cdot \sum x^2 \end{cases}$$

Розв'язання цієї системи дозволяє визначити параметри рівняння. Для цього потрібно обчислити суми  $x$ ,  $x^2$ ,  $\lg z$ ,  $x \cdot \lg z$ .

## Обчислення коефіцієнтів рівнянь множинної регресії.

Інколи буває необхідним дослідити залежність між кількома ознаками, які варіюють одночасно. У цьому випадку застосовують методи множинного регресійного аналізу. Принципово цей метод не відрізняється від методів однофакторного регресійного аналізу. У найпростішому випадку лінійної залежності рівняння регресії для трьох змінних має такий вигляд:

$$y = b_0 + b_1 \cdot x + b_2 \cdot z$$

Розв'язання цього рівняння, тобто обчислення його параметрів, пов'язано зі складанням системи нормальних рівнянь, до якої входить три рівняння:

$$\begin{cases} k \cdot b_0 + b_1 \cdot \sum x + b_2 \cdot \sum z = \sum y \\ b_0 \cdot \sum x + b_1 \cdot \sum x^2 + b_2 \cdot \sum x \cdot z = \sum x \cdot y \\ b_0 \cdot \sum z + b_1 \cdot \sum x \cdot z + b_2 \cdot \sum z^2 = \sum z \cdot y \end{cases}$$

Сумісне вирішення цієї системи нормальних рівнянь дозволяє визначити коефіцієнти рівняння регресії. Якщо залежності між змінними нелінійні, то попередньо потрібно провести перетворення даних аналогічним тим, що описані вище.

### Адекватність рівняння регресії. Інтерполяція та екстраполяція.

У попередніх розділах зверталася увага на те, що коефіцієнт пропорційності рівняння регресії обчислюється з певною статистичною помилкою, що зумовлює певний інтервал, в якому знаходяться теоретичні (обчислені за рівнянням регресії) значення результативної ознаки. Іншою важливою властивістю рівняння регресії є його *адекватність*. Адекватність рівняння регресії є параметром, який дозволяє встановити наскільки точно теоретична залежність (рівняння регресії) описує емпіричну залежність (результати експерименту). За цим показником рівняння регресії може бути *адекватним*, тобто таким, що задовільно характеризує емпіричну залежність, або *неадекватним*, якщо теоретичні значення значно відхиляються від відповідних показників емпіричного ряду. Для встановлення адекватності рівняння найчастіше використовується порівняння за Фішером *дисперсії адекватності* та *дисперсії відтворення*.

Дисперсія адекватності обчислюється за формулою:

$$S_{ad}^2 = \frac{(\tilde{y}_i - \bar{y}_i)^2}{k - u},$$



де:  $S_{ad}^2$  – дисперсія адекватності;  
 $\tilde{y}_i$  – обчислені (теоретичні) значення результативної ознаки;  
 $\bar{y}_i$  – експериментальні середні значення результативної ознаки;  
 $k$  – кількість пар даних, за якою проводилось обчислення параметрів рівняння;  
 $u$  – кількість коефіцієнтів (параметрів) рівняння регресії.

З формули обчислення дисперсії адекватності витікає важлива особливість, яку потрібно постійно враховувати досліднику при плануванні експерименту для проведення регресійного аналізу – кількість дослідів (рівнів фактора або пар даних) повинна хоча б на одиницю перевищувати кількість коефіцієнтів рівняння. В іншому випадку знаменник дроби ( $k-u$ ) буде дорівнювати нулю і обчислення дисперсії адекватності стає неможливим. Для нелінійних залежностей у будь-якому випадку кількість дослідів не повинна бути менше трьох.

Дисперсію відтворення (в окремих підручниках наводиться термін "дисперсія середнього значення") знаходять шляхом усереднення суми всіх вибірових середніх квадратів відхилень за формулою:

$$S_y^2 = \frac{\sum s_i^2}{k \cdot n},$$

де:  $S_y^2$  – дисперсія відтворення;  
 $s_i^2$  – вибірові середні квадрати відхилень;  
 $k$  – кількість пар даних (кількість дослідів);  
 $n$  – об'єм вибірок.

Слід особливо підкреслити, що в регресійному аналізі завжди використовуються рівномірні комплекси, тобто комплекси з однаковою кількістю повторювань у всіх дослідях.

**Критерієм для порівняння дисперсій адекватності та відтворення слугує їх відношення:**

$$F_{ad} = \frac{S_{ad}^2}{S_y^2}.$$

Якщо обчислене значення  $F$  перевищує стандартне значення для порівнянь дисперсій за Фішером для обраного рівня значущості ( $\alpha$ ), кількості ступенів вільності дисперсії адекватності ( $k-u$ ) та кількості ступенів вільності для дисперсії відтворення ( $k-n$ ), то різниця між дисперсіями визнається вірогідною, а рівняння регресії – неадекватним. Якщо різниця між дисперсією адекватності та дисперсією відтворення невірогідна, рівняння регресії визнається адекватним, а математична модель задовільно описує емпіричну залежність.

Остання проблема, яка постає перед дослідником під час проведення регресійного аналізу, стосується меж поширення результатів обчислень на інтервал змін незалежної змінної. Існує два способи застосування математичної моделі – *інтерполяція* або використання математичної моделі в межах експериментальних значень незалежної змінної, та *екстраполяція* або спроба використати математичну модель для прогнозування поведінки результативної ознаки за межами експериментальних змін незалежної змінної.

Для вирішення цієї проблеми необхідно мати уяву про поведінку об'єкта досліджень у широкому інтервалі змін умов експерименту та враховувати властивості математичних моделей, про які йшлося в раніше. Якщо повернутися до графіка, наведеного на рис. 4, то можна побачити, що границі варіювання теоретичного ряду найближче підходять до емпіричного ряду при значенні незалежної змінної, яке дорівнює середній арифметичній із її рівнів. На межах діапазону змін аргументу похибка значно збільшується, а при виході за межі цього діапазону, форма залежності взагалі може змінитися. Тому математичні моделі найчастіше застосовуються тільки для інтерполяції поведінки об'єктів досліджень. Екстраполяція можлива тільки в тих випадках, коли дослідник впевнений, що форма відгуку об'єкта не зміниться при збільшенні або зменшенні значень аргумента за межами дослідних значень. Для отримання якісної математичної моделі поведінки об'єкта слід обирати широкі інтервали варіювання значень аргумента. Бажано, щоб вони повністю включали в себе потрібні значення для передбачення поведінки результативної ознаки. Якщо немає впевненості в лінійності функції відклику, то кількість градацій незалежної змінної також повинна бути максимально можливою.

### **Приклад послідовності обчислення коефіцієнтів рівняння регресії**

Для пояснення методики проведення регресійного аналізу скористаємось лінійною моделлю. Вивчалася залежність швидкості росту міцелію

гливи (мм за добу) від температури в інтервалі від 18°C до 30°C. Кількість повторювань дослідів ( $n$ ) при кожній становила 3. Отримані дані наведені в табл. 30. Для обчислень коефіцієнтів лінійної моделі необхідно розрахувати суми  $x$ ,  $x^2$ ,  $y$  та  $x \cdot y$ . Ці показники також подані в таблиці.

Таблиця 30

Вихідні дані для обчислення рівняння лінійної регресії

№	Температура, °C (x)	Швидкість росту (y)		$x^2$	x·y
		Середні	$s_i^2$		
1	18	5	2,0	324	90
2	20	9	1,2	400	180
3	22	11	1,6	484	242
4	24	12	1,0	576	288
5	26	14	2,0	676	364
6	28	18	1,3	784	504
7	30	22	2,0	900	660
<b>Суми</b>	<b>168</b>	<b>91</b>		<b>4144</b>	<b>2328</b>

Спочатку розраховуються коефіцієнти рівняння регресії для лінійної моделі:

$$b_1 = \frac{k \cdot \sum x \cdot y - \sum x \cdot \sum y}{k \cdot \sum x^2 - (\sum x)^2} = \frac{7 \cdot 2328 - 168 \cdot 91}{7 \cdot 4144 - 168^2} = \frac{1008}{784} = 1,286,$$

$$b_0 = \frac{\sum y - b_1 \cdot \sum x}{k} = \frac{91 - 1,286 \cdot 168}{7} = -17,864.$$

Наступним кроком є обчислення помилки коефіцієнту пропорційності ( $b_1$ ) та перевірка його вірогідності:

$$S_b = \sqrt{\frac{\sum (y_i - \bar{y})^2 - \frac{[\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})]^2}{\sum (x_i - \bar{x})^2}}{(k-2) \cdot \sum (x_i - \bar{x})^2}} = \sqrt{\frac{192 - \frac{20736}{112}}{5 \cdot 112}} = 0,111,$$

$$t_b = \frac{b_1}{S_b} = \frac{1,286}{0,111} = 11,58$$

Як бачимо, вірогідність коефіцієнту досить висока. Обчислене значення перевищує стандартне значення Ст'юдента як для  $\alpha=0,05$  (2,57), так

і для  $\alpha=0,01$  (4,03). Таким чином, коефіцієнт пропорційності вірогідно відрізняється від 0 і повинен бути включений до рівняння регресії. У результаті проведених розрахунків отримано рівняння регресії, яке виглядає так:

$$\bar{y} = 1,286 \cdot x - 17,864$$

Тепер потрібно перевірити адекватність рівняння регресії. Для цього спочатку обчислюються очікувані значення росту міцелію при всіх досліджених температурах:

$$\begin{aligned} \bar{y}_1 &= 1,286 \cdot 18 - 17,864 = 5,28, & \bar{y}_2 &= 1,286 \cdot 20 - 17,864 = 7,86, \\ \bar{y}_3 &= 1,286 \cdot 22 - 17,864 = 10,43, & \bar{y}_4 &= 1,286 \cdot 24 - 17,864 = 13,00, \\ \bar{y}_5 &= 1,286 \cdot 26 - 17,864 = 15,57, & \bar{y}_6 &= 1,286 \cdot 28 - 17,864 = 18,14, \\ \bar{y}_7 &= 1,286 \cdot 20 - 17,864 = 20,72. \end{aligned}$$

Отримані дані, а також експериментальні середні і вибіркові квадрати відхилень заносяться до таблиці (табл. 31).

Таблиця 31

Експериментальні та очікувані середні

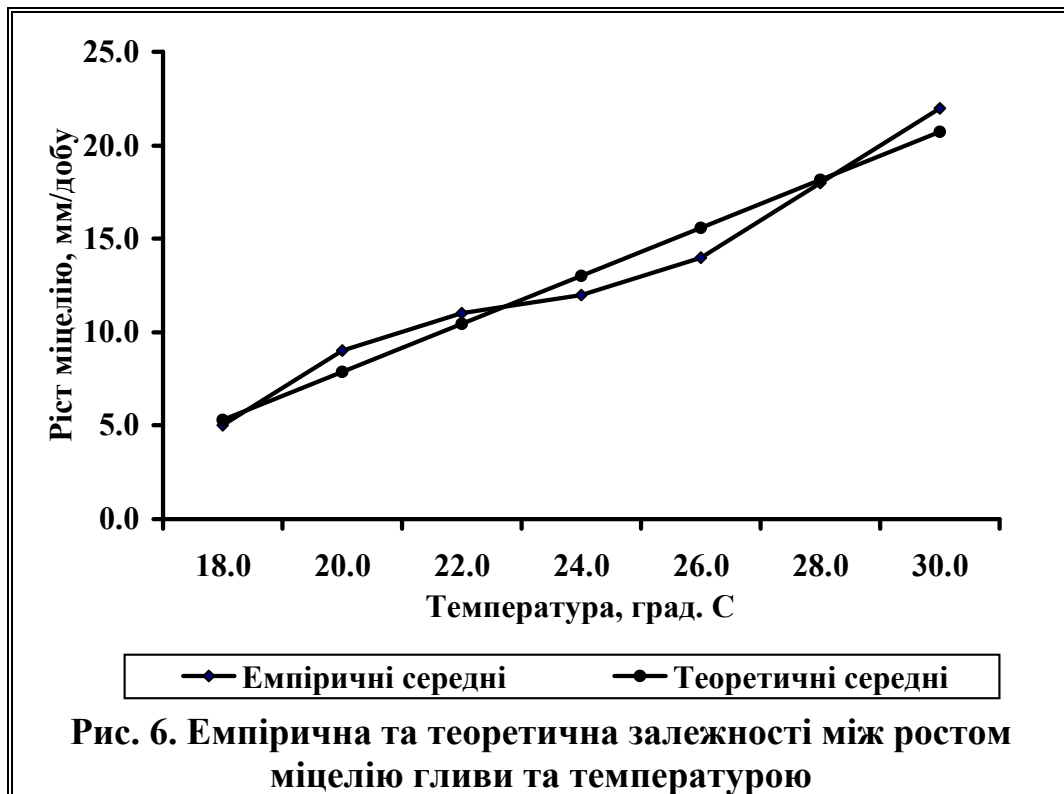
№	$\bar{y}_i$	$s_i^2$	$y_i$	$y_i - \bar{y}_i$	$(y_i - \bar{y}_i)^2$
1	5	2,0	5,28	-0,28	0,0784
2	9	1,2	7,86	1,14	1,2990
3	11	1,6	10,43	0,57	0,3250
4	12	1,0	13,00	-1,00	1,0000
5	14	2,0	15,57	-1,57	2,4650
6	18	1,3	18,14	-0,14	0,0196
7	22	2,0	20,72	1,28	1,6380
<b>Суми</b>		<b>11,1</b>		<b>0</b>	<b>6,8250</b>

Обчислюються дисперсії адекватності та відтворення:

$$\begin{aligned} S_{ad}^2 &= \frac{\sum (\bar{y}_i - \bar{y}_i)^2}{k - u} = \frac{6,825}{7 - 2} = 1,365, & S_{\bar{y}}^2 &= \frac{\sum s_i^2}{k \cdot n} = \frac{11,1}{7 \cdot 3} = 0,529, \\ F_{ad} &= \frac{S_{ad}^2}{S_{\bar{y}}^2} = \frac{1,365}{0,529} = 2,58 \end{aligned}$$

Отримане значення менше за стандартне значення критерію Фішера для  $f_{ad}=5$  та  $f_y=21$  (2,7), тому отримане рівняння регресії потрібно визнати

адекватним. Для ілюстрації отриманої залежності наведемо емпіричний та теоретичний графіки залежності росту міцелію від температури в дослідних межах (рис. 6).



## 7. НЕПАРАМЕТРИЧНІ МЕТОДИ СТАТИСТИЧНОЇ ОБРОБКИ

Критерії Ст'юдента ( $t$ ) та Фішера ( $F$ ) пов'язані з певними формами нормального розподілу вибірових дат: їх застосування передбачає оцінку розходжень між генеральними параметрами на основі вивчення вибірових показників порівнюваних сукупностей. Їх статистична потужність реалізується повністю лише за умови, що розподіл сукупностей підпорядковується закону нормального розподілу або не дуже відхиляється від нього.

У практичній роботі зустрічаються не тільки нормально розподілені, але і сукупності, які мають інший тип розподілу, про який досліднику мало що відомо. У таких випадках застосування параметричних методів не гарантує відсутності похибок. Тому для оцінки розходжень між вибірками, поруч з параметричними критеріями, застосовуються й інші, засновані на порівнянні не самих варіант, а їх порядкових номерів у впорядкованому ряді. Критерії, засновані на цьому принципі, називаються порядковими або ранговими, або непараметричними. Використання непараметричних критеріїв не пов'язано з визначенням форми розподілу і розрахунком середніх значень, середніх квадратів відхилень та інших параметрів. Саме тому такі методи і називаються непараметричними.

Непараметричні або порядкові критерії прості за своєю конструкцією, вони не потребують складної обчислювальної роботи, що вигідно відрізняє їх від параметричних критеріїв. Але, не зважаючи на зручність рангових методів, перевага залишається на боці параметричних критеріїв, які мають більшу статистичну потужність. Тому у всіх випадках, коли розподіл практично не відхиляється від закону нормального розподілу, рекомендується застосовувати параметричні критерії. Там, де виникають сумніви стосовно точності висновків, які робляться на основі використання параметричних методів, або коли форма розподілу залишається невідомою, або недостатньо зрозумілою, доречно застосовувати непараметричні критерії різниці. Всебічний контроль вірності висновків ніколи не стане на заваді. Якщо при обробці вибірового матеріалу різними засобами отримані однакові статистичні результати, це тільки збільшує гарантії правильності висновків.

У цьому розділі будуть наведені деякі з непараметричних методів статистичної обробки результатів експериментів.

При поясненні методів застосування всіх критеріїв буде використовуватися нульова гіпотеза про відсутність різниці між градаціями, тобто припущення, що  $k$  градацій являють собою об'єднану вибірку із загальної генеральної сукупності і різниці між ними не існує. Результати обчислень повинні або підтвердити, або відхилити цю гіпотезу на визначеному рівні значущості  $\alpha$ .

## Встановлення вірогідності різниці між двома вибірками (Т-критерій Уайта)

Одним з критеріїв, які застосовуються для встановлення різниці при порівнянні двох незалежних розподілів, є непараметричний критерій Т-Уайта (в літературі можна також зустріти назви “критерій Вілксона для не-супряжених” ознак або “критерій Манна-Уїтні”), який рівною мірою застосовується як для вибірок рівновеликих, так і для вибірок неоднакового об’єму. Сутність методу полягає в тому, що всі варіанти порівнюваних вибірок впорядковуються за зростанням значень в один загальний ряд та знаходяться ранги кожної варіанти. Далі ранги сумують окремо з кожної вибірки. Якщо вибірки, які порівнюються між собою, зовсім не відрізняються одна від одної, то і суми рангів повинні бути рівними між собою. У протилежному разі такої рівності не спостерігається. Чим більшими будуть розходження між вибірками, тим більше будуть відрізнятися суми їх рангів. У зв’язку з тим, що ця різниця може бути випадковою, вона оцінюється за допомогою критерію Уайта. Стандартні значення цього критерію для різного об’єму вибірок та двох рівнів значущості –  $\alpha = 0,95$  та  $\alpha = 0,99$  наведені в табл. 19 додатків.

Для оцінювання критерію  $T$  завжди застосовується *менша* сума, яка і порівнюється із стандартним (табличним) значенням цього критерію для  $n_1$  та  $n_2$ , тобто об’ємів порівнюваних сукупностей і визначеного критичного значення ймовірності. Якщо  $T_{st} > T_f$ , спостерігається вірогідна різниця між вибірками, нульова гіпотеза відкидається. Якщо стандартне значення критерію Уайта ( $T_{st}$ ) менше або дорівнює фактичному значенню ( $T_f$ ), нульова гіпотеза зберігається, а різниця між вибірками визнається статистично невірогідною. Ранги – це числа натурального ряду, якими позначаються члени впорядкованих сукупностей. Однаковим значенням цих сукупностей відповідає однаковий загальний середній ранг. Загальна сума рангів знаходиться за формулою:

$$N = 1 + 2 + 3 + \dots + n = \frac{n \cdot (n + 1)}{2},$$

де  $N$  загальна кількість спостережень.

Застосування критерію Т-Уайта можна показати на наступному прикладі. Було встановлено, що довжина тіла личинок ковалика в посівах озимого жита та проса, виражена в міліметрах, варіює так:

Місце збору	Значення дат, мм							Середні ( $y$ )
У посівах жита	7	10	14	15	12	16	12	12,3
У посівах проса	11	12	16	13	18	15		14,2

При аналізі цих значень складається враження, що личинки на посівах проса мають більші розміри. Перевіримо це припущення за допомогою критерію Т-Уайта. Впорядковуємо сукупність спостережень:

Довжина личинок	7	10	11	12	12	12	13	14	15	15	16	16	18
Номери дат	1	2	3	4	5	6	7	8	9	10	11	12	13
Ранги	1	2	3	5	5	5	7	8	9,5	9,5	11,5	11,5	13

Обчислюємо суми рангів:

Для 1-ої вибірки (жито):  $1+2+8+9,5+5+11,5+5=42$

Для 2-ої вибірки (просо):  $3+5+11,5+7+13+9,5=49$

---

Всього 91

Видно, що суми рангів для першої та другої вибірок відрізняються. Щоб визначити вірогідність різниці, меншу суму ( $T_f=42$ ) порівнюють зі стандартним значенням критерію для  $n_1=7$  і  $n_2=6$ , яке дорівнює 27. Оскільки  $T_{st}=27 < T_f=42$ , треба зробити висновок про відсутність різниці між розмірами личинок ковалика, який мешкає в посівах озимого жита і проса, тобто ця різниця має випадковий характер і не є статистично вірогідною.

#### Х-критерій Ван дер Вардена

Порівняно з Т-критерієм Уайта цей метод має більшу статистичну потужність. Однак Х-критерій Ван дер Вардена досить чутливий до функції розподілу дат у вибірках. Найбільшу статистичну спроможність він має при розподілі, який наближається до нормального.

Процедура його використання виглядає так. Перш за все, впорядковуються дати в обох вибірках в один загальний ряд за зростанням їх значень, як і у методі Уайта. Наприклад, ряди

$X_i$	3,8	3,6	4,6	3,9	
$Y_j$	4,4	6,1	6,3	4,2	5,7

запишемо у вигляді ( $v_k$  – номер дати в упорядкованому ряді)

$x_i$	3,6	3,8	3,9		4,6				
$y_j$				4,2		5,4	5,7	6,1	6,3
$v_k$	1	2	3	4	5	6	7	8	9



В упорядкованому ряді значення  $x_i$  мають номери  $v_k = 1; 2; 3; 5$ . Далі треба обчислити суму

$$X = \sum_{k=1}^{n_x} \Psi\left(\frac{v_k}{n+1}\right)$$

де  $n=n_x+n_y$  – загальна кількість варіант у вибірках (у нашому випадку  $n=4+5=9$ ), а  $\Psi$  функція, обернена інтегралу ймовірностей  $\Phi(u)$  (її значення наведені в табл. 20 додатків). У нашому прикладі значення відношень

$\frac{v_k}{n+1}$  дорівнюють:

$$\frac{1}{10} = 0,10; \quad \frac{2}{10} = 0,20; \quad \frac{3}{10} = 0,30; \quad \frac{5}{10} = 0,50.$$

Цим значенням  $\frac{v_k}{n+1}$  відповідають значення  $\Psi\left(\frac{v_k}{n+1}\right)$ : -1,28; -0,84; -0,52; 0,00. Тому

$$X = (-1,28) + (-0,84) + (-0,52) + (0,00) = -2,64$$

Для перевірки вірогідності різниці між вибірками використовують абсолютне значення  $X$  (знак мінус, якщо отримано негативне значення, відкидають), яке порівнюється зі стандартним значенням (табл. 21 додатків). Нульова гіпотеза відхиляється (вибірки різняться між собою), якщо отримане  $X$  буде більше, ніж відповідне критичне значення для даної кількості варіант та прийнятого рівня значущості. Для пошуку критичних значень треба визначити, крім загальної кількості спостережень, різницю між кількостями варіант у вибірках (на знак різниці не звертають уваги).

У нашому прикладі  $n=n_x+n_y=4+5=9$ ,  $n_y-n_x=5-4=1$ . Критичні значення, знайдені у таблиці, дорівнюють 2,48 для  $\alpha=5\%$  та 7 для  $\alpha=1\%$ . У зв'язку з тим, що отримане значення  $X = 2,64$  більше, ніж 2,38 (для  $\alpha = 5\%$ ), але менше, ніж 7 (для  $\alpha=1\%$ ), можливість відхилити нульову гіпотезу досить сумнівна і залежить від мети дослідження. Якщо 5-ти відсотковий рівень значущості задовольняє, то можна вважати різницю між вибірками вірогідною, в іншому випадку треба визнати різницю між вибірковими значеннями незначущою.

Ще раз підкреслимо, що переваги  $X$ -критерію Ван дер Вардена можуть виявитися лише за умови близького до нормального розподілу сукупностей.

### Серійний критерій

Цей критерій дозволяє виявити різницю між двома сукупностями не тільки за центральною тенденцією, а і за іншими властивостями.

Як і у інших випадках, нульова гіпотеза  $H_0$  полягає у припущенні, що ряди  $x_i$  та  $y_j$  є двома вибірками з однієї генеральної сукупності. Якщо це так, то окремі ранги або вибіркові значення з обох рядів повинні чергуватися при об'єднанні їх в один загальний ряд. Наприклад, ми маємо два ряди:  $x_1 x_2 x_3 x_4 x_5$  та  $y_1 y_2 y_3 y_4 y_5 y_6$ ; тоді розташування за величинами значень:

$$x_1 x_2 x_3 x_4 x_5 y_1 y_2 y_3 y_4 y_5 y_6$$

явно суперечить нульовій гіпотезі. Більше відповідає  $H_0$  розташування, яке має вигляд

$$x_1 x_2 y_1 x_3 x_4 y_2 y_3 y_4 y_5 x_5 y_6$$

Кількісним показником, за яким можна відрізнити ці розташування одне від одного, може слугувати кількість серій  $S$ , кожна з яких є безперервною послідовністю варіант, що належать до одного з двох рядів. Наприклад, перше розташування складається з двох серій:

1)  $x_1 x_2 x_3 x_4 x_5$ ; 2)  $y_1 y_2 y_3 y_4 y_5 y_6$ ,

а друге розташування – з шести серій:

1)  $x_1 x_2$ ; 2)  $y_1$ ; 3)  $x_3 x_4$ ; 4)  $y_2 y_3 y_4 y_5$ ; 5)  $x_5$ ; 6)  $y_6$ .

Серійний критерій різниці між двома сукупностями полягає в тому, що нульова гіпотеза відхиляється, якщо кількість серій занадто мала (звідси зрозуміло, що у такому випадку повинен використовуватися однобічний критерій). У табл. 22 додатків подані граничні значення кількості серій для  $\alpha=0,05$ ; приблизно можна вважати, що за  $n_x, n_y < 20$   $S_{01}=S_{05-2}$ .

Нульова гіпотеза приймається, якщо  $S > S_{st}$ , і відхиляється, якщо  $S < S_{st}$ .

Наприклад, треба порівняти два ряди

$x_i$	11,5	26,0	29,1	19,7	2,3	22,6	30,9	10,8	23,2	38,8	21,5
$y_j$	18,4	15,5	25,2	16,9	24,0	13,3	17,9	13,2			

Розташовуємо всі значення в один ряд із зростаючими значеннями, розміщуючи для зручності варіанти  $x_i$  та  $y_j$  у різних рядках:

$x_i$	2,3	10,8	11,5								19,7
$y_j$				13,2	13,3	15,5	16,9	17,9	18,4		

$x_i$	21,5	22,6	23,2			26,0	29,1	30,9	38,8	
$y_j$				24,0	25,2					

Як бачимо, ряд розділюється на п'ять серій. У таблиці стандартних значень знаходимо  $S_{0,5(11,8)}=6$ . Оскільки  $S$  менше, ніж  $S_{0,5}$ , але не менше, ніж  $S_{0,5-2}$ , питання про справедливість нульової гіпотези (відсутності різниці між вибірками) залишається нез'ясованим.

Якщо кількість варіант в одному з рядів перевищує 20, можна скористатися тим, що за умови вірності нульової гіпотези значення  $S$  мають майже нормальний розподіл із середнім значенням  $S$  та середнім квадратом відхилення  $s_S$ , які дорівнюють:

$$\bar{S} = \frac{2 \cdot n_x \cdot n_y}{n_x + n_y}; \quad s_S^2 = \frac{2 \cdot n_x \cdot n_y \cdot (2 \cdot n_x \cdot n_y - n_x - n_y)}{(n_x + n_y)^2 \cdot (n_x + n_y - 1)}$$

Звідси, за допомогою формули

$$U_S = \frac{\bar{S} - S - 0,5}{s_S}$$

обчислюємо значення критерію вірогідності. Нульова гіпотеза приймається, якщо  $U_S < 1,96$  та відхиляється за умови, що  $U_S > 2,58$ ; величина 0,5 є поправкою на дискретність.

Наприклад, при порівнянні двох рядів чисельністю  $n_x=14$ ,  $n_y=26$  в об'єднаному ряді виявилось  $S=11$  серій. Маємо:

$$\bar{S} = \frac{2 \cdot 14 \cdot 26}{14 + 26} + 1 = \frac{278}{40} + 1 = 19,2$$

$$s_S = \frac{2 \cdot 14 \cdot 26 \cdot (2 \cdot 14 \cdot 26 - 14 - 26)}{(14 + 26)^2 \cdot (14 + 26 - 1)} = \frac{278 \cdot (278 - 40)}{40^2 \cdot (40 - 1)} = 8,03$$

так що

$$U_S = \frac{19,2 - 11 - 0,5}{\sqrt{8,03}} = \frac{7,7}{2,84} = 2,72$$

Оскільки  $U_S=2,72 > 2,58$ , нульова гіпотеза відхиляється.

### Критерій Колмогорова–Смирнова

Коли об'єми порівнюваних сукупностей малі, серійний критерій виявляється недостатньо чутливим. У тих випадках, коли  $S$  дорівнює  $S_{st}$  або відрізняється від  $S_{st}$  не більше ніж, на одиницю, слід перевіряти результат за допомогою більш чутливого критерію Колмогорова-Смирнова.

Цей критерій має за основу порівняння рядів нагромаджених частот обох сукупностей. Нехай  $z_{i\{x\}}$  та  $z_{i\{y\}}$  – нагромаджені частоти рядів, впорядкованих за зростанням їх значень. Обчисливши різниці  $h_i = z_{i\{x\}} - z_{i\{y\}}$ , знаходимо найбільшу за абсолютною величиною різницю

$$D = \max |z_{i\{x\}} - z_{i\{y\}}|$$

Різниця вважається вірогідною, якщо вона перевищує стандартне значення, яке розраховується за формулою:

$$D_\alpha = \sqrt{\frac{1}{2} \cdot \ln \frac{2}{\alpha} \cdot \left( \frac{1}{n_x} + \frac{1}{n_y} \right)}$$

Якщо вираз  $\sqrt{\frac{1}{2} \cdot \ln \frac{2}{\alpha}}$  позначити через  $\lambda_\alpha$  і провести перетворення другого множника, можна записати

$$D_\alpha = \lambda_\alpha \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} = \lambda_\alpha \cdot \sqrt{\frac{n_x + n_y}{n_x \cdot n_y}}$$

Тепер, замість того щоб порівнювати емпіричне значення  $D$  з  $D_\alpha$ , можна обчислити значення  $\lambda_\alpha$ , яка виходить з формули:

$$\lambda^2 = D^2 \cdot \frac{n_x + n_y}{n_x \cdot n_y}$$

Ці значення порівнюють зі стандартними величинами  $\lambda^2$ , які дорівнюють:

$$\lambda_{0,05}^2 = \frac{1}{2} \cdot \ln \frac{2}{0,05} = \frac{1}{2} \cdot \ln 40 = 1,84$$

$$\lambda_{0,01}^2 = \frac{1}{2} \cdot \ln \frac{2}{0,01} = \frac{1}{2} \cdot \ln 200 = 2,65$$

Якщо  $\lambda^2 > \lambda_{0,05}^2$ , то нульова гіпотеза відкидається, а якщо  $\lambda^2 < \lambda_{0,05}^2$ , то вона приймається.

Наведемо приклад проведення розрахунків. Для цього використаємо значення, подані у першому прикладі застосування серійного критерію. Для проведення обчислень потрібно дані занести до таблиці (табл. 32). У стовпчику 1 записані всі значення у порядку зростання. У стовпчиках 2 та 3 подані частоти обох рядів. Значення ряду  $X$  будуть мати у стовпчику 2 частоти, які відповідають кількості варіант з цим значенням (у нашому ви-

падку 1), а в стовпчику 3 їх частота буде дорівнювати 0. Навпаки, значення ряду  $Y$  будуть мати ненульову частоту в стовпчику 3 і нульову – у стовпчику 2. Далі складаються ряди нагромаджених частот для значень  $X$  та  $Y$  (стовпчики 4 та 5). Поділивши ці значення на об'єм відповідного ряду ( $n_x=11$ ,  $n_y=8$ ), отримуємо значення  $z_{i\{x\}}$  та  $z_{i\{y\}}$  (стовпчики 6 та 7 відповідно). В останньому стовпчику подані значення різниці  $|\eta_i| = |z_{i\{x\}} - z_{i\{y\}}|$  (знак різниці не враховується).

Як можна бачити, найбільшою різницею є значення  $|\eta_i|=0,477=D$ . Це значення використовується для розрахунку величини  $\lambda^2$  за формулою:

$$\lambda^2 = D^2 \cdot \frac{n_x \cdot n_y}{n_x + n_y} = 0,477^2 \cdot \frac{11 \cdot 8}{11 + 8} = 1,06.$$

Оскільки отримане значення менше, ніж  $\lambda_{0,05}=1,84$ , то нульова гіпотеза зберігається.

Якщо обидві порівнювані вибірки мають однаковий об'єм ( $n_x=n_y=n$ ), розрахунки значно спрощуються. Перш за все, формула перетворюється на вираз;

$$\lambda^2 = D^2 \cdot \frac{n}{2}$$

також виявляється, що непотрібно обчислювати значення  $z_{i\{x\}}$  та  $z_{i\{y\}}$ . Достатньо віднайти максимальне абсолютне значення з різниць між нагромадженими частотами ( $\max|s_{i\{x\}} - s_{i\{y\}}|$ ), яка позначається літерою  $\Delta$ ). Тоді  $\Delta=D/n$ . Звідси, остаточна формула для обчислення  $\lambda^2$  набуває вигляд:

$$\lambda^2 = \frac{\Delta^2}{2 \cdot n}$$

Таблиця 32

Значення даних для обчислення критерію Колмогорова-Смирнова

$x_i, y_i$	$n_{i\{x\}}$	$n_{i\{y\}}$	$s_{i\{x\}}$	$s_{i\{y\}}$	$z_{i\{x\}}$	$z_{i\{y\}}$	$ \eta_i $
2,3	1	0	1	0	0,091	0,000	0,091
10,8	1	0	2	0	0,182	0,000	0,182

11,5	1	0	3	0	0,273	0,000	0,273
13,2	0	1	3	1	0,273	0,125	0,148
13,3	0	1	3	2	0,273	0,250	0,023
15,5	0	1	3	3	0,273	0,375	0,102
16,9	0	1	3	4	0,273	0,500	0,227
17,9	0	1	3	5	0,273	0,650	0,352
18,4	0	1	3	6	0,273	0,750	<b>0,477</b>
19,7	1	0	4	6	0,364	0,750	0,386
21,5	1	0	5	6	0,455	0,750	0,295
22,6	1	0	6	6	0,546	0,750	0,204
23,2	1	0	7	6	0,636	0,750	0,114
24,0	0	1	7	7	0,636	0,875	0,239
25,2	0	1	7	8	0,636	1,000	0,364
26,1	1	0	8	8	0,727	1,000	0,273
29,1	1	0	9	8	0,818	1,000	0,182
30,9	1	0	10	8	0,910	1,000	0,090
38,8	1	0	11	8	1,000	1,000	0,000
<b>Сума</b>	<b>11</b>	<b>8</b>					

Наведені методи перевірки застосовуються для порівняння вибірок з ознаками, які варіюють незалежно. Якщо значення двох порівнюваних ознак супряжені між собою, застосування цих методів може призвести до значної помилки. Далі подані декілька способів порівняння вибірових розподілів із супряженими ознаками.

### Критерій знаків

У випадках застосування непараметричних критеріїв порівняння вибірок із супряженими ознаками порівнюються не ранги впорядкованих варіант, а ранги впорядкованих різниць між супряженими ознаками.

Найпростішим і, разом з тим, найменш потужним з цих методів є *критерій знаків*. Застосування методу пояснимо на такому прикладі. Вивчалася дія внесення гербіциду на врожайність жита. Досліди проводилися на 10 дослідних та 10 контрольних ділянках (гербіцид не вносився). У зв'язку з тим, що поля, на яких проводилися дослідження, відрізнялися ґрунтовими умовами, ділянки були попарно розподілені між всіма полями. Таким чином, отримано 10 супряжених пар результатів:

<b>Номер пари</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
Контроль	30,2	28,7	31,4	24,6	26,8	21,2	34,1	30,8	25,7	32,7
Дослід	32,3	29,1	28,5	22,5	23,1	26,3	30,9	36,7	24,1	27,2
Відношення	1,07	1,01	0,91	0,91	0,86	1,24	0,91	1,19	0,94	0,84
<b>Знаки</b>	<b>+</b>	<b>+</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>+</b>	<b>-</b>	<b>+</b>	<b>-</b>	<b>-</b>

Знайдемо відношення між врожайностями жита на дослідних і контрольних ділянках (четвертий рядок). У чотирьох випадках (1, 2, 6 та 8 варіанти) врожайність жита при внесенні гербіциду дещо підвищується, а в інших варіантах, навпаки, зменшується. Оскільки у більшості дослідних варіантів врожайність зменшується, складається враження про негативний вплив гербіциду на рослини. Але такі коливання врожайності можуть мати суто випадковий характер. Для перевірки вірогідності різниці за відношеннями закріплюються знаки: якщо відношення менше 1 – знак мінус, якщо більше – знак плюс. Якщо відношення дорівнює 1, то воно не враховується; також зменшується відповідно і кількість спостережень ( $n$ ). Знаки різниць для нашого прикладу наведені у п'ятому рядку запису результатів. Тепер окремо підраховується кількість позитивних та негативних знаків (у нашому випадку  $\Sigma+=4$ ,  $\Sigma-=6$ ) і менша сума порівнюється з критичними значеннями, поданими в табл. 23 додатків,  $Z$ . У прикладі  $n=10$ ,  $Z=4$ ,  $Z_{05,10}=2$ ,  $Z_{01,10}=1$ . У зв'язку з тим, що  $Z > Z_{05,10}$  нульова гіпотеза зберігається (вплив гербіциду виявився невірогідним). Нульова гіпотеза відхиляється за умови, що  $Z < Z_{01,n}$ . У випадку, коли  $Z_{01,n} < Z < Z_{05,n}$ , питання про вірогідність різниці залишається відкритим і потребує перевірки за допомогою методу з більшою статистичною потужністю.

Основним недоліком цього методу є те, що не враховуються значення різниць між величинами супряжених ознак. Більш надійним критерієм є критерій Вілксона.

### **Критерій Вілксона ( $Z$ ) для супряжених пар ознак**

$Z$ -критерій Вілксона має достатню потужність при аналізі вірогідності різниці між супряженими ознаками. Методика його використання зводиться до наступного. Спочатку знаходять різниці між парними варіантами супряжених рядів: при цьому враховуються знаки різниць. Далі ці різниці впорядковують за абсолютними значеннями у зростаючий ряд та визначають їх ранги. Ранги сумують окремо для позитивних та негативних різниць. Якщо різниця між парними варіантами дорівнює нулю, вона не враховується, і кількість спостережень ( $n$ ) відповідно зменшується.

Менша сума рангів, незалежно від знаку, порівнюється зі стандартними значеннями, наведеними в табл. 24 додатків, для обраного рівня значущості та кількості парних спостережень, яка повинно бути більше 6. Якщо табличне (стандартне) значення критерію  $Z$  перевищує його фактичне значення, тобто меншу суму рангів, це вказує на вірогідність різниці, яка спостерігається між вибірками. У протилежному випадку нульова гіпотеза зберігається і різниця, яка спостерігається між рядами супряжених варіант, визнається статистично невірогідною.

Наприклад, методом серійних випробувань вивчався ступінь зараженості яблук двома різновидностями гнилизни **A** і **B**, які викликають загнивання плодів. Результати десяти випробувань подані в табл. 33.

## Результати випробувань патогенності різновидностей гриба

Різновидність гриба	Кількість тканини, яка загниває за серіями дослідів, г									
	1	2	3	4	5	6	7	8	9	10
А	5,9	4,9	3,6	8,8	7,4	5,7	6,0	6,3	9,4	9,1
В	4,7	5,1	3,7	7,4	6,1	4,2	5,0	7,1	6,3	8,0
Різниця	1,2	-0,2	-0,1	1,4	1,3	1,5	1,0	-0,8	3,1	1,1

Впорядковуються різниці за їх абсолютними значеннями і обчислюються ранги:

Різниця	-0,1	-0,2	-0,8	1,0	1,1	1,2	1,3	1,4	1,5	3,1
Ранги	1	2	3	4	5	6	7	8	9	10

Сумуються ранги:

з негативними знаками:  $1+2+3=6$

з позитивними знаками:  $4+5+6+7+8+9+10=49$

---

Всього 55

Менша сума рангів ( $Z=6$ ) порівнюється з табличним значенням критерію для  $n=10$  і рівня значущості  $\alpha=0,05$ , яке дорівнює 8. У зв'язку з тим, що обчислене значення менше стандартного ( $6 < 8$ ), можна зробити висновок, що різновидність гриба А вірогідно має більшу патогенність, ніж різновидність В.

### Кореляція рангів

Поряд з параметричними показниками кореляційного зв'язку існують і непараметричні або порядкові показники. Ці критерії застосовуються у тих випадках, коли об'єкти, що вивчаються, впорядковуються за значеннями ознак. Найбільш відомим показником сили зв'язку є *ранговий коефіцієнт кореляції Спірмена*, який обчислюється за формулою:

$$r_{\rho} = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)}$$

де:  $d$  – різниця між рангами супряжених значень ознак  $X$  і  $Y$ , тобто  $d=x_{\rho}-y_{\rho}$ ;  
 $n$  – об'єм вибірки або кількість супряжених пар.

Щоб з'ясувати, чи існує зв'язок між ознаками  $X$  і  $Y$ , чи його немає, треба впорядкувати їх значення і встановити, як вони розташовані відносно один одного. Якщо зростаючим значенням однієї ознаки ( $X$ ) відповіда-



ють зростаючі значення другої ознаки ( $Y$ ), то існує позитивний зв'язок. Якщо із зростанням першої ознаки значення другої зменшуються, то має місце негативний зв'язок. Коли зв'язок між ознаками відсутній, впорядкованим значенням однієї з них будуть відповідати неупорядковані значення іншої ознаки.

Позначивши впорядковані значення ознак порядковими номерами натурального ряду, можна обчислити ранги цих значень і за різницею рангів зробити висновок про ступінь супряженості між ознаками. При повному (функціональному) зв'язку ранги впорядкованих ознак повністю збігаються між собою і різниця між ними буде дорівнювати нулю. У таких випадках  $r_\rho=1$ . Якщо ознаки  $X$  і  $Y$  варіюють незалежно один від одного, то

$$\frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)} = 1$$

і ранговий коефіцієнт буде дорівнювати нулю. Ранговий коефіцієнт кореляції, як і параметричний показник, виражається частками одиниці і може варіювати від -1 до +1.

Техніку розрахунків зручніше показати на числовому прикладі. Вивчалася залежність між вмістом кальцію та ступенем пошкодження листків деревних рослин промисловими викидами (табл. 34). Потрібно перевірити близькість зв'язку між цими показниками, тобто з'ясувати, чи впливає вміст кальцію на стійкість рослин до забруднювачів повітря.

Для зручності упорядкування значень ознаки  $Y$  їх показники заносяться до окремого стовпця, а далі ранги переносяться до графі  $y_\rho$ , яка розташована поряд з графою  $x_\rho$ . Впорядкування комплексу і обчислення різниць є найвідповідальнішою роботою. Як і в Т-критерії Уайта, однаковим значенням ознаки відповідають ранги, які визначаються як середні арифметичні значення їх номерів. Після впорядкування, обчислення різниць рангів та їх квадратів, отриману суму  $d^2$  підставляють у формулу

$$r_\rho = 1 - \frac{6 \cdot 2,5}{11 \cdot (11^2 - 1)} = 1 - \frac{15}{1320} = 0,989$$

Таблиця 34

Результати визначення вмісту кальцію у різних за стійкістю до забруднення рослин

№	X	Y	Ранги рядів		$X_\rho - Y_\rho = d^2$	$d^2$	Розрахунок рангів Y		
			$X_\rho$	$Y_\rho$			№	Y	$Y_\rho$

1	1,6	60,0	1	9,5	-1	1	1	1,0	2
2	6,7	80,0	2	11	0	0	2	1,0	2
3	8,0	50,0	3	8	1	1	3	1,0	2
4	9,5	60,0	4	9,5	0	0	4	10,0	4
5	10,0	1,0	5	2	0	0	5	12,0	5
6	10,8	30,0	6	7	0	0	6	15,0	6
7	13,9	15,0	7	6	0	0	7	30,0	7
8	14,5	10,0	8	4	0	0	8	50,0	8
9	14,6	12,0	9	5	-0,5	0,25	9	60,0	9,5
10	16,7	1,0	10	2	0,5	0,25	10	60,0	9,5
11	16,2	1,0	11	2	0	0	11	80,0	11
Σ						2,5			

Критерієм вірогідності показника рангової кореляції є таке відношення:

$$t_r = r_\rho \cdot \sqrt{\frac{n-2}{1-r_\rho^2}}$$

Нульова гіпотеза відкидається, якщо  $t_r$  перевищує критичне значення  $t_{st}$  (або буде дорівнювати йому), яке знаходиться в таблиці Ст'юдента для заданого рівня значущості ( $\alpha$ ) і кількості ступенів вільності  $f=n-2$ .

Обчислимо вірогідність коефіцієнту кореляції для нашого прикладу:

$$t_r = 0,989 \cdot \sqrt{\frac{11-2}{1-0,989^2}} = 0,989 \cdot \sqrt{46,39} = 6,735$$

Стандартне значення критерію Ст'юдента для  $f=11-2=9$  та  $\alpha=0,05$  дорівнює 2,26, а для  $\alpha=0,01$  – 3,25. Як бачимо, ці значення значно менші за обчислене – 6,735. Таким чином, зв'язок між вмістом кальцію та стійкістю рослин визнається вірогідним з ймовірністю 99%.

Цей показник дає надійну оцінку при кількості пар не менше 10. Якщо об'єми вибірок малі, то треба використовувати такі формули для розрахунку критичних значень:

$$\text{для } \alpha=0,05 - t_r = \frac{1,96}{\sqrt{n-1}} \cdot \left(1 - \frac{0,16}{n-1}\right)$$

$$\text{для } \alpha=0,01 - t_r = \frac{2,58}{\sqrt{n-1}} \cdot \left(1 - \frac{0,69}{n-1}\right)$$

Ранговий коефіцієнт кореляції слід використовувати лише у тих випадках, коли з тієї або іншої причини неможливе застосування параметричних показників. До таких випадків належать об'єкти, розподіл яких дуже

відрізняється від нормального закону, або коли закон їх розподілу залишається нез'ясованим, а також у тих випадках, коли варіюючі ознаки оцінюються балами або іншими умовними одиницями виміру, і зв'язок між ними визначається лише загальним напрямом мінливості, що і дозволяє впорядковувати сукупність спостережень за значеннями рангів. У всіх таких випадках показник рангової кореляції може і повинен знайти широке використання. Але він не може повністю замінити параметричні показники кореляційного зв'язку.

### Рангова оцінка сили впливу

Непараметричний показник сили впливу є аналогом параметричного дисперсійного аналізу. Цей показник використовується під час вивчення порядкових ознак. Ми познайомимось з обчисленням непараметричного показника сили впливу на прикладі однофакторного комплексу.

Для розрахунку непараметричної сили впливу об'єкти повинні бути упорядковані без повторних рангів від  $R=1$  для найслабшого до  $R=N$  для найсильнішого вираження ознаки. Тоді група дат буде групою натуральних чисел 1, 2, 3, ..., N, для якої всі групові характеристики можна отримати на основі тільки одного показника – об'єму групи.

$$\text{Сума рангів} - \sum R = \frac{N \cdot (N + 1)}{2};$$

$$\text{середня арифметична} - M_R = \frac{N + 1}{2};$$

$$\text{середній квадрат відхилень загальний} - SS_y = \frac{(N - 1) \cdot N \cdot (N + 1)}{12};$$

$$\text{загальна поправка} - H = \left[ \frac{N \cdot (N + 1)}{2} \right]^2 \div N = \frac{N \cdot (N + 1)^2}{4};$$

$$\text{окремі поправки за градаціями} - H_i = \frac{(\sum R)^2}{n};$$

$$\text{середній квадрат відхилень факторний} - SS_x = \sum H_i - H.$$

Подальші обчислення ведуться звичайним методом:

$$\eta_x^2 = \frac{SS_x}{SS_y}, \quad S_{\eta_x^2} = (1 - \eta_x^2) \cdot \frac{k - 1}{N - k}, \quad F_x = \frac{\eta_x^2}{S_{\eta_x^2}}$$

Використання цих формул для рівномірного розподілення дат (але не для нормального!) може дати не завжди правильні результати.

Тому у відповідальних дослідженнях треба перевірити вірогідність впливу за критерієм  $\chi$ -квадрат, який у цьому випадку дорівнює:

$$\chi^2 = (N - 1) \cdot \eta_x^2$$

Під час серійних робіт, коли потрібно визначити тільки вірогідність впливів, критерій  $\chi^2$  можна обчислити за формулою:

$$\chi^2 = \frac{12}{N \cdot (N + 1)} \cdot \sum H_i - 3 \cdot (N + 1)$$

Для пояснення послідовності обчислень використаємо такий приклад. У селекціонерів почали з'являтися тварини з особливим забарвленням хутра. Дослідники припустили, що поява забарвлення передається за батьківською лінією. Були проаналізовані нащадки п'яти батьків. З аномаліями виявилось 30 тварин. У зв'язку з тим, що ця ознака не може бути оцінена ні точно, ані приблизно, всі 30 нащадків були впорядковані без повторних рангів. В якості дати кожного нащадка встановлювався ранг у загальному ряді. Дані наведені в табл. 35.

Спочатку проводиться розрахунок допоміжних величин, середніх квадратів відхилень, показника сили впливу та його вірогідності:

$$H = \frac{N \cdot (N + 1)^2}{4} = 7207,5,$$

$$SS_y = \frac{(N - 1) \cdot N \cdot (N + 1)}{12} = \frac{29 \cdot 30 \cdot 31}{12} = 2247,5,$$

$$SS_x = \sum H_i - H = 8525,7 - 7207,5 = 1318,2,$$

$$\eta_x^2 = \frac{SS_x}{SS_y} = \frac{2247,5}{1318,2} = 0,5865,$$

$$\chi^2 = (N - 1) \cdot \eta_x^2 = 29 \cdot 0,5865 = 17,0,$$

$$\chi_{st}^2 = \{9,5 - 13,3 - 18,5\}.$$

Таблиця 35

Непараметричний показник спадкоємності в однофакторному комплексі

Батьки	Ранги дітей	n	$\sum R$	$H_i = \frac{(\sum R)^2}{n}$	$M_R = \frac{\sum R}{n}$
<b>A</b>	2, 17, 20, 23, 25, 28	7	136	2642,3	19,4
<b>B</b>	1, 3, 8, 12	4	24	144,0	6,0
<b>C</b>	16, 19, 22, 24, 30	5	111	2464,2	22,2
<b>D</b>	4, 5, 6, 7, 9, 10, 11, 14	8	66	544,5	8,2
<b>E</b>	13, 15, 18, 26, 27, 29	6	128	2730,7	21,3
<b>Суми</b>		30	465	8525,7	

Вірогідність впливу фактора (батько) дорівнює 99%, що може бути причиною спонтанної мутації, яка передається у спадщину за батьківською лінією.

### Множинні порівняння, засновані на критерії Краскела – Уолліса

Порівняння всіх вибірок

Якщо об'єми вибірок рівні порівняння комплексу за градаціями ведуть за такою схемою:

1. Впорядковують усі вибірки і знаходять суми рангів для кожної градації фактора.
2. Обчислюють  $k \cdot V \cdot (k-1)/2$  абсолютних значень різниць  $|R_u - R_h|$ , де  $u$  і  $h$  – номери вибірок від 1 до  $k$ . Величина  $k \cdot V \cdot (k-1)/2$  визначає кількість пар, порівнюваних між собою.
3. Різниця між ефектами значуща, якщо  $|R_u - R_h| \geq Y_{a,k,n}$  де  $Y_{a,k,n}$  стандартне значення критерію для рівня значущості  $\alpha$ , кількості градацій фактора  $k$  і об'єму вибірок  $n$ .

Наприклад, на чотирьох стадіях розвитку рослин вивчалась активність фермента. Дані наведені в табл. 36.

Таблиця 36

#### Активність фермента на різних стадіях розвитку

Активність фермента на різних стадіях розвитку							
1		2		3		4	
Значення	Ранг	Значення	Ранг	Значення	Ранг	Значення	Ранг
7	1,5	9	3	11	5	15	8
10	4	7	1,5	16	9	18	11
12	6	14	7	20	12	17	10

Методи впорядкування комплексів і знаходження середніх рангів для варіант, значення яких повторюються були описані раніше. Обчислюємо абсолютні значення різниць між ранговими сумами  $R_j$ :

$$\begin{aligned}
 R_1 &= 1,5 + 4 + 6 = 11,5; & R_2 &= 3 + 1,5 + 7 = 11,5 \\
 R_3 &= 5 + 9 + 12 = 26; & R_4 &= 8 + 11 + 10 = 29 \\
 |R_1 - R_2| &= 11,5 - 11,5 = 0; & |R_1 - R_3| &= |11,5 - 26| = 14,5; \\
 |R_1 - R_4| &= |11,5 - 29| = 17,5; & |R_2 - R_3| &= |11,5 - 26| = 14,5; \\
 |R_2 - R_4| &= 11,5 - 29 = 17,5; & |R_3 - R_4| &= |26 - 29| = 3.
 \end{aligned}$$

Отримані різниці порівнюються зі стандартним значенням, яке дорівнює 22. Всі різниці менші від стандартного. Це означає, що стадія розвитку рослини не впливає на активність фермента.

Якщо об'єми порівнюваних вибірок, не дорівнюють один одному, то критерієм вірогідності відмінності між різницями рангів є така умова:

$$|R_u - R_v| \geq \sqrt{h'_{(a,k(n_1, \dots, n_k))} \cdot \frac{N \cdot (N-1)}{12} \cdot \left( \frac{1}{n_u} + \frac{1}{n_v} \right)}$$

де:  $h'_{(\alpha, k(n_1, \dots, n_k))}$  – стандартне значення критерію Краскела – Уолліса.

Порівняння дослідних вибірок з контрольною

Вирізняють однобічну і двобічну перевірки різниці дослідних та контрольної градацій. Послідовність операцій така.

1. Упорядковується комплекс і знаходяться суми рангів для кожної з  $k$  вибірок комплексу.
2. Обчислюються  $k-1$  різниць, взятих за абсолютною величиною  $|R_u - R_k|$ .
3. Різниця між дослідним і контрольним варіантами вірогідна при однобічній перевірці, якщо  $|R_u - R_{kt}| \geq Y^*_{a,k-1,n}$ , при двобічній перевірці, якщо  $|R_u - R_{kt}| \geq Y^{**}_{a,k-1,n}$ , де  $Y^*_{a,k-1,n}$  і  $Y^{**}_{a,k-1,n}$  стандартні значення критерію вірогідності різниці для однобічної або двобічної перевірок відповідно.

Якщо об'єми вибірок нерівні між собою, то застосовується наближення Данна:

– при однобічній перевірці:

$$|R_u - R_{kt}| \geq \sqrt{z_{(\alpha, k-1)} \cdot \frac{N \cdot (N+1)}{12} \cdot \left( \frac{1}{n_{kt}} + \frac{1}{n} \right)}$$

– при двобічній перевірці:

$$|R_u - R_{kt}| \geq \sqrt{z_{(\alpha, 2 \cdot (k-1))} \cdot \frac{N \cdot (N+1)}{12} \cdot \left( \frac{1}{n_{kt}} + \frac{1}{n} \right)}$$

де:  $z_{(a, k-1)}$  та  $z_{(a, 2 \cdot (k-1))}$  критичні значення однобічного та двобічного критерію Данна відповідно.

## РОЗДІЛ 2. ПЛАНУВАННЯ ТА ОБРОБКА АКТИВНИХ ЕКСПЕРИМЕНТІВ

### 8. ОСНОВНІ ПОЛОЖЕННЯ ТА ЗАВДАННЯ ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ

#### **Основи математичного планування експериментів**

До цього часу розглядалися так звані *пасивні* експерименти, коли дослідник використовує математичний апарат тільки наприкінці експерименту, тобто при обчисленні його результатів. Проблеми пасивного експерименту зводяться до правильного збору матеріалу та добору найліпших методів статистичної обробки. Для математичної обробки пасивних експериментів використовуються методи кореляційного, дисперсійного, регресійного (у тому числі і багатофакторного) аналізів. За допомогою регресійного аналізу можна отримати досить якісні математичні моделі поведінки об'єкта досліджень залежно від факторів середовища. Але в пасивному експерименті багатофакторні моделі будуються в результаті великої дослідної і обчислювальної роботи. На противагу пасивним експериментам, для багатофакторних комплексів у біологічних дослідженнях все частіше застосовуються *активні* експерименти, в яких дослідники використовують математичний апарат ще під час підготовки досліджень. Часто в практиці досліджень такі експерименти називають *спланованими*, а підготовку дослідів – *математичним плануванням експерименту*. Такі експерименти виконуються за спеціально розробленими планами, що дозволяє отримати значно більше статистичної інформації за мінімального об'єму дослідів. Вибір планів залежить від мети експерименту та властивостей об'єкта досліджень. Активні експерименти мають значні переваги порівняно з пасивними:

1. Плани експериментів складаються заздалегідь, до початку дослідів, що дозволяє максимально спростити наступну математичну обробку для обчислення математичних моделей.
2. Оптимальне використання факторного простору в активному експерименті дозволяє з мінімальною кількістю експериментів отримати максимум інформації про явища, які вивчаються; однак слід пам'ятати, що використання факторного простору є одним з найскладніших питань планування експериментів.
3. Сплановані експерименти дозволяють вирішувати завдання побудови емпіричної залежності (апроксимація функції відгуку), пошуку екстремумів (максимуму або мінімуму) процесів та оптимізації управління біологічними процесами.

4. Методи планування експериментів дозволяють дослідним шляхом встановити ступінь впливу різних факторів на реакцію відповіді біологічних об'єктів.
5. Сплановані експерименти дозволяють отримати математичні моделі таких процесів, моделювання яких у пасивних експериментах значно ускладнено (наприклад, моделі типу склад–властивості), та формалізувати методами спеціального дисперсійного аналізу вивчення явищ, які залежать від якісних факторів.
6. Планування експерименту дозволяє вивчати та математично описувати процеси та явища з неповним знанням їх механізмів.

Слід враховувати, що регресійний аналіз, який проводиться за спеціальними планами, чутливий до наступних вихідних передумов:

- *результати спостережень (функція відгуку) являють собою незалежні, нормально розподілені величини;*
- *вибіркові середні квадрати відхилень однорідні;*
- *значення незалежних змінних вимірюються із зневажливо малою похибкою порівняно з помилкою визначення функції відгуку.*

Наведемо основні визначення, які будуть використовуватися у наступних розділах:

1. *Планування експерименту* – процедура вибору кількості та умов проведення дослідів, необхідних і достатніх для вирішення визначених завдань з певною точністю:
2. *Фактор* (незалежна змінна, вхідний параметр) – один з можливих способів впливу на об'єкт досліджень, який може приймати значення, визначені дослідником (контрольований фактор).
3. *Факторний простір* – простір, в якому будується графічна поверхня відгуку. Він задається координатними осями, на яких подаються значення факторів та функції відгуку. Розмірність факторного простору залежить від кількості факторів. У багатофакторному експерименті сприймається ускладнено через багатовимірності простору.
4. *Відгук* (результативна ознака) – значення реакції відповіді біологічної системи на дію факторів.
5. *Параметр оптимізації* (тільки для екстремальних експериментів) – кількісно визначена характеристика мети експерименту. Останнє може полягати в пошуку мінімального або максимального значення відгуку системи.
6. *Область визначення* – множина значень, які можуть приймати відгук, параметр оптимізації або фактор.
7. *Математична модель* – опис залежності змін функції відгуку або параметру оптимізації від варіювання факторів, яка виражена математичною формулою.



8. *Функція відгуку* (статистична модель об'єкта) – залежність, яку необхідно визначити за результатами експерименту. У загальному вигляді –  $y=f(X_1, X_2, \dots, X_k)$ .
9. *Поверхня відгуку* – графічне зображення змін функції відгуку від змін рівнів факторів.
10. *Рівняння регресії* – формула, яка відбиває з певною ймовірністю залежність функції відгуку ( $Y$ ) від змін факторів ( $X$ ).
11. *Ап'юріорна інформація* – інформація про поведінку біологічної системи, яка відома до побудови плану експерименту. Може бути отримана в результаті попередніх експериментів або з літератури.

#### Теоретичні основи планування експериментів

У статистиці для пояснення планів експерименту дуже часто використовують кібернетичну модель, яку називають ще "чорним ящиком". Поняття "чорного ящика" відповідає функціональній системі, внутрішня структура якої невідома, а відомі лише параметри її входу та виходу (рис. 7). Стрілки праворуч відбивають вихідні параметри (вони позначені літерою  $Y$ ), які характеризують результат дії факторів і називаються результативними ознаками. Це може бути врожай, вихід клітин, фізіологічні показники та інше.

Для проведення експерименту необхідно мати можливість впливати на поведінку "чорного ящика". Всі способи такого впливу називаються факторами і позначаються літерами  $X$  (стрілки ліворуч). Їх також називають вхідними параметрами. Це може бути дія фізичних, хімічних, біологічних (температура, вологість, живлення, доза та вид добрив) та інших способів впливу на об'єкт досліджень, різний стан (сорт, стать, вік) біологічних об'єктів досліджень.

Під час вирішення завдання використовуються математичні моделі об'єкта досліджень. Під математичною моделлю мають на увазі рівняння, яке зв'язує відгук системи з факторами. У загальному вигляді це рівняння записується так:

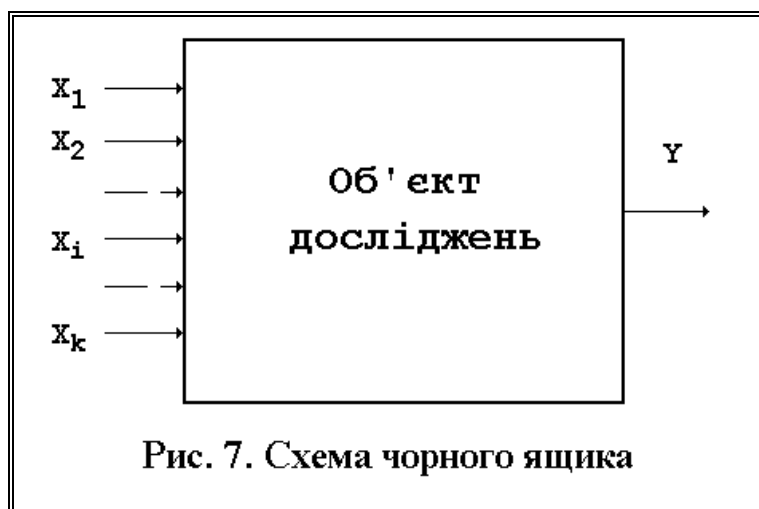
$$Y = f(x_1, x_2, \dots, x_k),$$

де  $f()$  замінює слова "функція від".

Кожний фактор може приймати в досліді одно або декілька значень, які називаються *рівнями фактору*. Фактори можуть мати нескінченну множину значень (*безперервний фактор*) або декілька фіксованих значень (*дискретний фактор*). Разом з тим, значення навіть безперервних факторів фіксуються і встановлюються з певною точністю, що визначає їх певну дискретність при проведенні експерименту. Тому можна вважати, що будь-який фактор має певну кількість дискретних значень. Це припущення сут-

тево полегшує побудову "чорного ящика" і плану експерименту та спрощує оцінку їх складності.

Фіксований набір рівнів факторів визначає один з можливих станів



"чорного ящика". Одночасно цей набір є умовами одного з експериментів плану. Якщо перебрати всі можливі стани факторів, можна отримати множину різних станів дослідного об'єкта. Водночас це є кількістю дослідів, необхідних для проведення експерименту. Кількість різних станів об'єкта визначає складність даної системи. Якщо відома складність

системи, легко визначити кількість дослідів, які потрібно включити до експерименту для вирішення завдань дослідження.

### **Характеристика основних завдань запланованого експерименту**

Виходячи з властивостей планованих експериментів, можна зробити висновок, що планування експерименту дозволяє на суто науковій основі поставити завдання, які в пасивному експерименті вирішуються інтуїтивними методами, дати відповіді на такі запитання:

- Як використовувати апріорну інформацію;
- Скільки і яких дослідів потрібно провести, і як їх статистично обробити;
- Як вирішити поставлені завдання найефективніше.

Однак слід зазначити, що відсутній єдиний метод планування експерименту. Планування експерименту являє собою певну методологічну концепцію, яка поєднує декілька методів. Використання того або іншого з них зумовлено поставленим завданням та можливостями, які знаходяться у розпорядженні експериментатора.

Перелік біологічних завдань, які вирішуються методами планування експерименту, достатньо широкий. Наприклад, за допомогою планування експерименту можуть бути вирішені проблеми знаходження оптимальних умов культивування мікроорганізмів та клітинних систем; оптимізація процесів біотехнології; з'ясування основних факторів, які визначають поведінку біологічних об'єктів та виключення факторів, що заважають; пошук біологічно активних препаратів та умов їх синтезу; визначення синер-

гізму дії факторів; оцінка параметрів математичних моделей біологічних процесів та інші завдання.

Експерименти, що виконуються за планами, можна розділити на ряд категорій. Залежно від того, передбачається або не передбачається поведінка об'єкта у часовому просторі, моделі поділяють на динамічні та статичні. В обох випадках модель може бути використана для передбачення в межах області експериментальних значень змінних (*інтерполяційна модель*) або за межами цих значень (*екстраполяційна модель*). Математичні моделі також можуть бути використані для вибору визначених (наприклад, оптимальних) умов функціонування об'єктів досліджень.

Інший бік використання моделей пов'язаний з перевіркою гіпотез. Модель можна спеціально створити для перевірки певної гіпотези або множини гіпотез. Однак якщо вона створювалась і для іншої мети, ця модель також повинна задовольняти вимогам існуючих теорій та гіпотез щодо об'єкта досліджень. Перевірка такої відповідності зветься *інтерпретацією*. Нарешті, можливо, що отримана модель буде сприяти появі нових гіпотез, які потребують експериментального підтвердження. Придатність моделі для висунення нових гіпотез називається *евристичністю*.

Природно, що така класифікація, як і будь яка інша, є умовною, тому не існує чіткої межі між різними призначеннями моделей. Всю різноманітність перелічених завдань можна умовно віднести до таких основних груп.

1. Завдання *порівняльного* та *елімінуючого* експерименту – порівняння дослідних варіантів з максимальною точністю та усунення факторів, які “заважають” (джерел неоднорідностей), на результати цих порівнянь.
2. Завдання *математичного опису багатофакторного об'єкта* – отримання адекватних моделей, які описують поведінку об'єкта в області режимів, що цікавлять дослідника, включно із завданнями з обмеженням на незалежні змінні типу  $x=I$  у так званих експериментах на симплексі. Використовують здебільшого для інтерполяції поведінки об'єкта, оскільки екстраполяція пов'язана із зменшенням точності передбачення.
3. Завдання *екстремального* експерименту – пошук оптимальних режимів функціонування багатофакторного об'єкта. Це можуть бути максимальні (наприклад, вихід продукту) або мінімальні (наприклад, вартість продукту або процесу) значення.
4. Завдання *відсіваючого* експерименту – з'ясування найбільш суттєвих факторів і відсів несуттєвих.
5. Методи планування експериментів для спеціальних випадків (наприклад, вивчення діаграм склад→властивості та інші).

Основні типи експериментів з рекомендаціями щодо використання методів планування наведені в табл. 37.

Типи планованих експериментів та плани, які застосовуються для їх реалізації

Тип експерименту	Призначення	Плани, що застосовуються
Елімінуючий	Видалення впливу неоднорідностей (зовнішніх, неконтрольованих факторів), зменшення випадкової дисперсії комплексу	Латинські квадрати, куби та ін., спеціальні плани типу $2^k$ для видалення безперервного дрейфу.
Порівняльний (у тому числі експертні оцінки)	Побудова ряду переважності для деяких факторів, що вивчаються. Порівняння дослідних варіантів з максимальною точністю	Латинські плани (квадрати, куби, гіпер-куби), блок-схеми, неповні та повні факторні експерименти, неповні блочні плани (для експертного оцінювання). Результати подаються у вигляді рангів. Обробка здійснюється непараметричними методами.
Відсіваючий	Виділення групи суттєвих факторів та відсівання несуттєвих	Плани, побудовані на базі [гіпер]-греко-латинських конфігурацій, комбіновані плани, які дозволяють скоротити кількість варіантів
Екстремальний	Пошук оптимальних умов функціонування багатофакторних об'єктів	Комбінаторні плани типу $2^k$ (ПФЕ, ДФЕ) з прийняттям умов лінійності поверхні відгуку в межах змін рівнів факторів
Побудова математичної моделі об'єкта з кількісними змінними	Отримання інтерполяційної (рідше екстраполяційної) моделі поведінки об'єкта залежно від змін рівнів кількісних факторів	Плани типу $r^k$ (повний факторний експеримент), де $r$ – кількість рівнів фактора (2, 3, 5)
Отримання математичної моделі об'єкта з кількісними та якісними змінними	Отримання математичних моделей в умовах, коли область дії факторів має дискретно-безперервний характер	Плани і методи коваріаційного аналізу
Спеціальні	Вирішення спеціальних завдань ("склад $\rightarrow$ властивості")	Плани досліджень поверхні відгуку, комбіновані плани.

Вибір плану експерименту значною мірою залежить від мети досліджень та поведінки функції відгуку. Якщо є необхідною математична модель для інтерполяції поведінки системи, то необхідно використовувати плани з трьома або п'ятьма рівнями факторів та широкими інтервалами варіювання факторів. Під час пошуку оптимальних умов проведення різних процесів для вибору діапазону варіювання рівнів факторів, дуже важливою є апріорна інформація. На кількість рівнів факторів сильний вплив чинить лінійність змін функції відгуку. За лінійних змін достатньо двох рівнів фактора, тоді як при значній нелінійності змін функції відгуку потрібно використовувати досліди з трьома або п'ятьма рівнями факторів.

На початку дослідження будь-якого об'єкта дуже важливим є аналіз наявної апріорної інформації. Ця інформація може характеризуватися повнотою та надійністю. Якщо ще до проведення досліджень можна відповісти на будь-яке питання про об'єкт досліджень, яке нас цікавить, то рівень апріорної інформації дуже високий і завдання, яке поставлено дослідником, вже вирішене. Протилежний випадок – це повна відсутність апріорної інформації. У такому випадку нічого не залишається, крім спроб висунення гіпотез за аналогією до інших схожих об'єктів досліджень та їх експериментальної перевірки. Найчастіше в практиці досліджень зустрічаються різноманітні проміжні випадки. Зрозуміло, що чим більшою інформацією про об'єкт володіє дослідник, тим легше вирішувати нові завдання і тим легше планувати нові дослідження.

Не все, що відомо про об'єкт досліджень, відомо з однаковою надійністю. Більш того, дуже часто не відомо наскільки вірогідне те або інше твердження. Тому постійно виникає необхідність приймати рішення, засновані на апріорних відомостях, вірогідність яких невідома. Такі рішення приймаються на основі досвіду, знань, інтуїції та індивідуальності експериментатора, на протиположні рішення, які приймаються на основі формалізованої математичної процедури. Найчастіше планування експериментів у біології засновано на складному сполученні формальних процедур та інтуїтивних рішеннях.

У будь-якому випадку для методології активного експерименту характерним є урахування всієї апріорної інформації при обранні плану і організації дослідів. Загальними принципами, які властиві для цієї методології і забезпечують точність аналізу, є:

- проведення повторних дослідів;
- створення однорідних умов та утворення блоків;
- використання методів рандомізації для усунення впливу факторів, які заважають.

**Проведення повторних дослідів** є необхідним перш за все для обчислення помилки експерименту, значення якої використовується при аналізі результатів. Під час аналізу важливо не тільки визначити результати

різних методів впливів, але і визначити якою мірою існуюча варіація дослідного матеріалу впливає на порівняння результатів різних варіантів. По-друге, використання великої кількості об'єктів на кожний варіант підвищує точність результатів порівняння, оскільки дисперсія середнього значення, пропорції або будь якої іншої величини, яка застосовується для порівняння, оберненопропорційна до кількості спостережень.

Повторність необхідна також для збільшення області використання результатів. Оскільки будь-який експеримент допускає тільки обмежене узагальнення, бажано створювати різні умови для проведення експерименту, використовувати різні сорти рослин, різні методи та інше, щоб, використавши принцип повторності, відокремити ефект, який цікавить дослідника, на фоні шуму, який утворюється за рахунок наведеного різноманіття. Якими повинні бути чисельності вибірок залежить від потрібної точності порівнянь та варіабельності об'єктів досліджень. Якщо дослідник знає ці параметри, легко обчислити необхідні об'єми вибірок. Формули для таких розрахунків наводилися у першому розділі.

**Створення однорідних умов і утворення блоків.** Для зменшення природної варіації необхідно враховувати всю апіорну інформацію під час вибору плану та організації експерименту для створення однорідних умов при порівнянні різних варіантів. Наприклад, бажано проводити порівняння різних ліків на одній групі хворих, щоб позбутися впливу на результати індивідуальних особливостей хворих; використовувати у дослідках на тваринах особин однієї статі, забезпечити їм однакові умови живлення, утримання, введення різних препаратів. Щоб підвищити чутливість досліду і створити основу для широкого використання результатів, дослідний матеріал поміщають в однорідні блоки. Наприклад, можна дослідити 4 варіанту лікування, розподіливши тварин на чотири блоки за приплодами і випробувавши всі варіанти на кожному припліді. Кожний приплід у такому плані типу “випадкових блоків” являє собою маленький експеримент, і на порівняння варіантів буде впливати тільки випадкова варіація у припліді; вплив додаткової варіації між приплодами (блоками) виключається. При бажанні ефекти цих блоків також можуть бути враховані.

**Рандомізація.** Введення до експерименту елементів випадковості є необхідним для того, щоб позбутися впливу неконтрольованих факторів та будь-якого зміщення результатів з боку експериментатора. Останні завжди наявні у будь-якому експерименті навіть при самому ретельному плануванні і створенні однорідних умов для порівняння. Тому розподіл об'єктів досліджень у групи та дослідних груп за характером випробувань необхідно здійснювати у випадковому порядку. Рандомізація гарантує, що в середньому групи не дуже будуть відрізнятися між собою через дію на них неконтрольованих факторів.

Відповідність між моделлю та об'єктом досягається вибором методу дослідження та його реалізацією. Іноді трапляються випадки, коли вибір методу однозначний або коли будь-який метод дозволяє вирішити проблему. Як правило, вибір методу ускладнений. Тоді потрібно обирати метод з детерміністичних, тобто методів безпосереднього визначення функції відгуку, та статистичних. При ускладненні об'єктів досліджень зростає значення статистичних методів. Тому є необхідним сполучення обох типів методів. На початковому етапі найчастіше обираються максимально прості моделі, наприклад, лінійні. Якщо вони не задовольняють, тоді переходять до вивчення більш складних залежностей. Для перевірки придатності моделі (адекватності) існують спеціальні статистичні методи.

Однак потрібно мати на увазі, що під час планування експерименту, по-перше враховуються властивості об'єкта досліджень, по-друге враховуються вимоги до факторів.

#### Вимоги до об'єкта досліджень

Перш за все, необхідно, щоб функція відгуку була випадковою нормально розподіленою кількісною величиною, яку можна легко виміряти або обчислити з допомогою простих методів. Якщо неможливо дати безпосередню кількісну оцінку об'єкта досліджень (якісна ознака), то за певним значеннями цієї ознаки закріплюються відповідні ранги. Прикладами кількісних ознак об'єктів досліджень є інтенсивність фотосинтезу, вміст пігментів у листках, лінійний ріст та накопичення маси міцелієм грибів, рослинами та ін. Ці показники можуть бути або безпосередньо виміряні протягом аналізу (ріст, маса рослин) або обчислені за простими формулами після вимірювання (вміст пігментів, інтенсивність фотосинтезу). Прикладом якісних ознак може слугувати колір квітів. Наприклад, квіти червоного кольору можуть бути рожевими, блідо-червоними, яскраво-червоними, малиновими і т.д. Для проведення порівняння відтінки кольору або його прояв можна позначити балами: рожевий – 1 бал, блідо-червоний – 2, яскраво-червоний – 3 і т.д. Таким чином якісна ознака перетворюється в кількісну.

Основною характеристикою об'єкта досліджень є його складність. Вона визначається кількістю різноманіття або, як прийнято в кібернетиці, кількістю різних станів, у кожному з яких може перебувати об'єкт. Характеризуючи складність системи, можна казати про прості об'єкти, складні об'єкти та великі системи. Слід мати на увазі, що чіткої межі між всіма цими видами складності не існує і що кожний об'єкт може розглядатися і як простий, і як складний залежно від мети досліджень. Розглядаючи методи планування експерименту, ми будемо розглядати об'єкти як прості або складні, оскільки теорія дослідження великих систем ще недостатньо розроблена. Крім того, у деяких випадках доцільно вважати великою системою сам процес дослідження (системний підхід).

Другою важливою характеристикою об'єкта досліджень є його керованість. Керованим вважається такий об'єкт, який дослідник може за своїм бажанням перевести в будь-який стан і підтримувати його в цьому стані з визначеною точністю протягом визначеного часу. Об'єкти, до яких ця вимога не виконується, називаються некерованими. На практиці більшість біологічних об'єктів досліджень є частково керованими.

Ще однією важливою характеристикою, необхідною для проведення експерименту, є відтворюваність результатів. Для перевірки відтворюваності результатів потрібно вибрати один який-небудь або декілька рівнів фактора і провести експеримент у цих умовах. Далі експерименти повторюють декілька разів через нерівні проміжки часу. Неоднорідність результатів експериментів буде характеризувати відтворюваність результатів. Якщо ця неоднорідність задовольняє вимоги (обраний рівень похибки), то об'єкт задовольняє вимоги відтворюваності, якщо перевищує – то не задовольняє. Вирішити проблему відтворюваності інколи можна завдяки збільшенню кількості повторювань досліджу. Якщо збільшення повторювань не допомагає, то можливо на об'єкт досліджень впливає який-небудь неконтрольований фактор, що постійно дрейфує. У цьому випадку потрібно використовувати спеціальні плани експериментів для врахування дрейфуючого фактору.

#### Вимоги до факторів

Не менш важливим для правильного складання плану експерименту є дотримання певних вимог до факторів.

У планованих експериментах найчастіше випробовується дія декількох факторів на об'єкт досліджень. Тому першою вимогою до факторів є їх сумісність. Сумісність означає, що всі комбінації факторів здійсненні і небезпечні. Це дуже важлива вимога, оскільки її недотримання може призвести до такого поєднання факторів, за якого може статися, наприклад, вибух або інші небажані явища. Несумісність факторів може спостерігатися на границях їх областей визначення. Позбутися цього можна скороченням областей варіювання. Ситуація ускладнюється, якщо несумісність спостерігається всередині областей визначення. Одним з можливих вирішень цієї проблеми є поділ областей визначення на підобласті і рішення двох окремих завдань.

Під час планування експерименту важливою є незалежність факторів, тобто можливість встановлення фактора на будь-якому рівні незалежно від рівнів інших факторів, інакше кажучи, фактори не повинні бути скорельовані. Якщо ця умова не може бути виконана, то планування експерименту неможливе. Для прикладу можна навести дослідження певної термодинамічної системи, в якій передбачалося включити три фактори:  $X_1$  – тиск, атм.,  $X_2$  – об'єм, л та  $X_3$  – температуру, К. У цьому випадку один з трьох факторів не може змінюватися незалежно від інших, оскільки за фо-



рмулою Менделєєва – Клапейрона ( $PV=nRT$ ) при встановлені значень двох з трьох факторів останній буде автоматично змінюватися. Тому для проведення експериментів можна використати тільки одну з комбінацій двох факторів: 1 – тиск та об'єм, 2 – тиск та температура, 3 – об'єм та температура.

Наступною важливою характеристикою фактора є можливість операційного визначення його рівня, тобто потрібно мати можливість вимірювання рівнів факторів за допомогою приладів. Також повинен бути наведений спосіб регуляції та підтримання будь-якого значення фактора. Наприклад, часто кажуть, що процес визначається дифузією. Однак дифузія є фактором, який не можна точно виміряти операційно і сама залежить від концентрацій речовин, температури та інших умов. Тому такий фактор у плануванні експерименту використовувати не можна.

Кожний фактор має область визначення. Для планування експерименту важливо, щоб область визначення фактора була дискретною, тобто могла приймати тільки кінцеві значення. Однак область визначення може бути і безперервною величиною, але використання приладів, що мають кінцеву точність та чутливість, дозволяють говорити про дискретність таких безперервних факторів. Наприклад, температура може змінюватися за безперервним законом. Разом з тим при використанні спиртового термометра ми не зможемо відчувати зміни менше  $0,25^{\circ}\text{C}$  (половина ціни ділення шкали), ртутного –  $0,1^{\circ}\text{C}$ , електронного –  $0,01^{\circ}\text{C}$ . Цей факт і визначає дискретність області визначення температури як фактора.

Крім того, області визначення факторів завжди обмежені. Обмеження можуть бути принциповими та технічними. Прикладом принципового обмеження може слугувати абсолютний нуль температури (нижчої температури неможливо отримати навіть теоретично). Технічні обмеження пов'язані з властивостями об'єкта досліджень, апаратури, обладнання. Наприклад, не можна під час культивування живих об'єктів (грибів, мікроорганізмів, рослин) встановлювати температуру вище порогу коагуляції білка (для більшості організмів –  $60\text{--}70^{\circ}\text{C}$ ). Під час проведення технічних експериментів прикладом технічних обмежень може слугувати температура плавлення матеріалу експериментальної установки.

До точності фіксації факторів не ставлять особливих вимог. Важливо, щоб вона була відома. Слід враховувати, що чим менша точність фіксації рівнів факторів, тим менша кількість можливих дискретних значень, і, відповідно, менша кількість рівнів фактора. Чим більша точність, тим швидше будуть отримані результати з визначеною вірогідністю. Тому точність фіксування факторів повинна бути якомога вищою. Ще одне важливе зауваження – помилка визначення значення фактору повинна бути набагато меншою за помилку визначення функції відгуку. Ступінь точності визначається діапазоном вимірів значень факторів.

## Види факторів

У розділі, присвяченому дисперсійному аналізу, вже розглядалися види факторів, які можуть впливати на об'єкт досліджень. Як і в дисперсійних комплексах, у багатофакторному регресійному аналізі всі фактори поділяються на кількісні та якісні. До кількісних факторів відносяться такі, значення яких встановлюється певними числами (рН, температура, освітлення та інші). Їх можна кількісно оцінити інструментальними методами та точно дозувати під час підготовки та проведення досліду. Прикладом використання якісних факторів може слугувати вивчення кількості мікроорганізмів на полях, засіяних сільськогосподарськими культурами, або в природних екосистемах. У цьому випадку неможливо кількісно оцінити рівні факторів. Тому кожному місцю помешкання присвоюється відповідний



ранг, який є номером відповідного місця помешкання (наприклад, пшеничному полю дають номер 1, кукурудзяному – 2 і т.д.). Інколи буває вигідно замінити кількісну оцінку фактора якісною ранговою оцінкою. Наприклад, потрібно визначити вплив місця розташування колб з культурою гриба на стелажі з освітленням на ріст культури гриба. Місце розташування можна описати кількісно координатами відносно меж полиці

(мал. 8 а, координати X та Y). Однак можна зробити інакше – розбити полицю на квадрати і закріпити за кожним квадратом номер (мал. 8 б), використовуючи ці номери як градації фактора. Правда, у такому випадку змінюється вигляд рівняння для оцінювання впливу. За умов кількісних ознак використовується рівняння, яке має вигляд:

$$y = \bar{y} + \beta_i \cdot x_i + \varepsilon.$$

Якісні фактори вимагають проведення дисперсійного аналізу зі встановленням вірогідності впливу та його закономірностей за формулою:

$$y = \bar{y} + \gamma_j \cdot z_j + \varepsilon.$$

У випадку сполучення кількісних та якісних факторів використовується коваріаційний аналіз з рівнянням:

$$y = \bar{y} + \beta_i \cdot x_i + \gamma_j \cdot z_j + \varepsilon,$$

де  $y$  – очікуване значення функції відгуку;  
 $\bar{y}$  – середнє значення функції відгуку у комплексі;  
 $\beta_i$  – коефіцієнт при  $i$ -тому кількісному факторі;  
 $x_i$  – значення  $i$ -того кількісного фактора;  
 $\gamma_j$  – коефіцієнт при  $j$ -тому якісному факторі;  
 $z_j$  – значення  $j$ -того якісного фактора;  
 $\varepsilon$  – випадкова дисперсія.

Крім загальних принципів експериментування (чітке формулювання завдання досліджень, ретельна підготовка плану експерименту з урахуванням усієї апріорної інформації), для планування експериментів характерні такі особливості.

1. Використовується комплексний підхід до вивчення об'єктів, який передбачає одночасну дію багатьох факторів з метою оцінки їх впливу та впливу їх взаємодій. Одночасне варіювання факторів за спеціальною програмою забезпечує вивчення кожного з цих факторів у різних умовах, створених змінами інших факторів.

2. У процесі експериментування більшість дій дослідника формалізовано. Досліди виконуються невеликими серіями за певними алгоритмами. Після кожної серії дослідів проводиться аналіз та інтерпретація результатів і приймаються обґрунтовані рішення про подальшу роботу. В результаті формується оптимальна послідовна стратегія експериментування.

3. Результати дослідів подаються у вигляді математичної моделі – рівняння регресії, яке зв'язує між собою цільовий показник з факторами, що варіюють.

4. Отримані моделі мають оптимальні статистичні властивості і забезпечують компактне представлення результатів у зручній для друку та зберігання формі.

## 9. ФАКТОРНІ ПЛАНИ $2^k$

Факторні плани  $2^k$  використовуються для отримання лінійної моделі поведінки об'єкта досліджень, які можуть бути використані безпосередньо як модель (моделюючий експеримент) або для вирішення проблем знаходження оптимальних умов протікання процесу (екстремальний експеримент). Під час складання плану експерименту постає одна з найважливіших проблем – вибір основних рівнів факторів та інтервалів їх варіювання.

Вибираючи області експерименту перш, за все потрібно оцінити границі визначення факторів. При цьому повинні враховуватися обмеження на варіювання факторів (принципові та технічні обмеження, сумісність та незалежність факторів). У деяких екстремальних експериментах важливе значення мають також техніко-економічні та економічні обмеження.

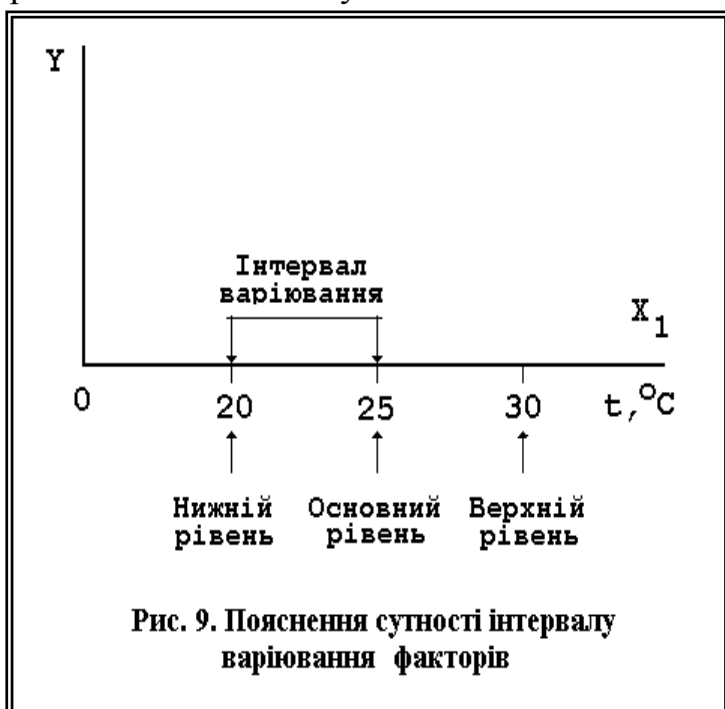
### **Вибір основного рівня та інтервалів варіювання факторів**

Визначаючи експериментальну область необхідно враховувати всю апріорну інформацію, наявну у дослідника. Перш за все визначається мета експерименту. По-друге, потрібно визначити лінійність функції відгуку. По-третє, визначається точність, з якою можна підтримувати значення факторів. Від цих параметрів значною мірою залежить вибір інтервалів варіювання факторів. Під час проведення експериментів з метою отримання математичної моделі поведінки об'єкта, обирають максимально можливі інтервали варіювання факторів (основний рівень повинен знаходитися в центрі області визначення або в центрі області передбачення), якщо поведінка об'єкта лінійна. У випадку нелінійних змін функції відгуку або обирають послідовність вузьких інтервалів варіювання, або використовують плани з більшою кількістю рівнів фактора (плани  $3^k$  або  $5^k$ ).

При проведенні екстремальних експериментів важливо правильно обрати *основний рівень (центр експерименту)*. У цьому може значно допомогти апріорна інформація. Може статися, що з літератури відомі умови найкращого протікання процесу, яким відповідають одна або декілька комбінацій рівнів факторів. Ці комбінації можна використовувати як основні рівні факторів для побудови плану експерименту. Якщо відомі межі варіювання факторів і найліпші умови знаходяться в цих межах, то основний рівень фактора обирається в центрі визначеної області змін фактору. Завдання ускладнюється, якщо точка оптимуму знаходиться на межі або близько до границі області визначення. У такому випадку основний рівень обирають з деяким зсувом від найліпших умов. Може статися, що координати найкращої точки невідомі, але відома підобласть, в якій процес проходить достатньо добре. Тоді основний рівень обирається або в центрі підобласті, або у випадковій її точці. І наприкінці, можливий випадок, коли є декілька еквівалентних точок, координати яких відмінні. Якщо відсутні

додаткові дані (технологічного, економічного характеру), вибір основного рівня експерименту довільний.

Наступним етапом підготовки планованого експерименту є вибір інтервалу варіювання факторів. *Інтервалом варіювання факторів називають певну чисельну величину, додавання якої до основного рівня дає значення верхнього (максимального) рівня, а віднімання – значення нижнього (мінімального) рівня.* Для пояснення сутності інтервалу варіювання можна скористатися рис. 8. На числовій осі відбито основний рівень фактора ( $25^{\circ}\text{C}$ ). Інтервал варіювання обрано рівним  $5^{\circ}\text{C}$ . Тоді максимальний рівень буде отримано в результаті додавання до значення основного рівня інтервалу варіювання ( $25^{\circ}\text{C}+5^{\circ}\text{C}=30^{\circ}\text{C}$ ), а нижній рівень – віднімання ( $25^{\circ}\text{C}-5^{\circ}\text{C}=20^{\circ}\text{C}$ ). Для якісних ознак поняття верхнього, нижнього та основного рівнів не мають сенсу.



На вибір інтервалів варіювання накладаються обмеження зверху та знизу. Інтервал варіювання не може бути меншим ніж помилка, з якою дослідник фіксує та підтримує фактор. Це обмеження знизу. З іншого боку, інтервал варіювання не може бути таким, щоб верхній або нижній рівні виходили за межі області визначення.

Потрібно звернути увагу на те, що при вирішенні завдання оптимізації для першої серії дослідів по-

трібно обирати такі межі варіювання факторів, які давали б можливість для подальшого руху до оптимальних значень.

Вибір інтервалів варіювання – це, мабуть, найважча процедура планування експерименту, оскільки вона пов'язана з неформалізованим етапом планування. У цьому процесі може допомогти наявна апріорна інформація, така як відомості про точність фіксування факторів, кривизну та діапазон змін функції відгуку. Звичайно, ця інформація є орієнтовною, а інколи може виявитися навіть помилковою. Але це єдина розумна підстава, на якій можна починати планування експерименту. В ході експериментів її досить часто доводиться корегувати.

Точність фіксування факторів визначається точністю приладів і стабільністю рівнів протягом досліду. У плануванні експерименту можна

прийняти такі градації точності фіксування факторів: висока точність – помилка фіксування менше 1%; середня точність – помилка фіксування від 1 до 5%; низька точність – помилка фіксування від 5 до 10%.

Джерелом відомостей про кривизну поверхні відгуку можуть слугувати графіки однофакторних залежностей та теоретичні міркування. Певне уявлення про кривизну дає аналіз табличних даних, оскільки наявності кривизни відповідають непропорційні зміни відгуку при рівномірних змінах факторів. У посібнику будуть розглядатися три випадку: **функція відгуку лінійна; функція відгуку суттєво нелінійна; інформація про кривизну відсутня.**

Останньою характеристикою, яка може допомогти у виборі інтервалів варіювання факторів, є інтервал змін параметра оптимізації. Якщо дослідник має в своєму розпорядженні певну множину дослідів, то завжди можна знайти мінімальне та максимальне значення функції відгуку. Різниця між цими значеннями і є *діапазоном змін відгуку (параметра оптимізації)* для цієї множини дослідів. Вирізняють вузький та широкий діапазони. **Діапазон змін відгуку може вважатися вузьким, якщо він несуттєво відрізняється від варіювання відгуку в повторних дослідах (природної варіабельності). В іншому випадку діапазон змін параметра оптимізації вважається широким. Третій випадок пов'язаний з відсутністю інформації про діапазон змін відгуку.**

Для певної формалізації процедури вибору можна користатися даними, поданими в табл. 38. В ній наведені 27 сполучень умов експерименту та рекомендації щодо вибору рівнів факторів.

Таблиця 38

Рекомендації щодо вибору інтервалу варіювання факторів залежно від сполучення умов

Точність фіксування фактора	Характеристика поверхні функції відгуку	Діапазон змін відгуку	Інтервал варіювання фактора	Додаткові відомості
1	2	3	4	4
Низька	Лінійна	Широкий	Широкий	
		Вузький	Широкий	
		Невідомо	Широкий	

Продовження табл. 38

1	2	3	4	4
	Нелінійна	Широкий	Рішення не-однозначне	1. Збільшення точності підтримання факторів
		Вузький	Рішення не-однозначне	2. Збільшення кількості повторних дослідів
		Невідомо	Рішення не-однозначне	3. Інтуїтивне рішення
	Невідомо	Широкий	Середній	
		Вузький	Широкий, рідко середній	
		Невідомо	Широкий	
Середня	Лінійна	Широкий	Середній	
		Вузький	Широкий	
		Невідомо	Широкий	
	Нелінійна	Широкий	Вузький	
		Вузький	Середній	
		Невідомо	Середній	
	Невідомо	Широкий	Середній	
		Вузький	Середній, рідко широкий	
		Невідомо	Середній	
Висока	Лінійна	Широкий	Середній	
		Вузький	Широкий	
		Невідомо	Середній	
	Нелінійна	Широкий	Вузький	
		Вузький	Вузький	
		Невідомо	Вузький	
	Невідомий	Широкий	Середній	
		Вузький	Широкий	
		Невідомо	Середній	

З наведених у таблиці даних витікає, що перші дев'ять ситуацій, які відповідають низькій точності фіксування факторів, визначають широкий інтервал варіювання факторів. Разом з тим, за умови нелінійної залежності

функції відгуку виникає невизначена ситуація, коли неможливо прийняти однозначне рішення. У цьому випадку широкий інтервал варіювання факторів вступає в конфлікт з нелінійністю зміни функції відгуку, а вузький неможливо використати через недостатню точність фіксування факторів. У цьому випадку можливі декілька рішень. Перше з них полягає у підвищенні точності фіксування та підтримання рівнів фактора. Для цього можна використовувати точніші прилади для виміру параметрів, застосовувати автоматичне підтримання рівня та інші прийоми, які дозволяють позбутися низької точності фіксування факторів. Крім того, можна збільшити повторність дослідів, щоб зменшити похибку, або прийняти інше (інтуїтивне) рішення. При високій точності фіксування факторів інтервал їх варіювання цілком залежить від поведінки функції відгуку.

Декілька слів про поняття вузького, середнього та широкого інтервалу варіювання факторів. Чіткої межі між цими поняттями не існує. Все залежить від області визначення та обмежень на варіювання факторів. Але можна обумовити широкий інтервал варіювання приблизно 30% і більше області визначення (тобто при переході від низького до високого рівня фактору охоплюється біля 60% області визначення фактора). Середній інтервал варіювання становить від 10% до 30% області визначення фактора і вузький інтервал відповідає охопленню менше 10% області визначення фактора. У кожному конкретному випадку можливі й інші рішення.

Наступним етапом для побудови плану є перехід від реальних змінних до кодованих. Ця операція необхідна для побудови матриць планування, які мають певні властивості. У кодованих змінних рівні факторів в експерименті з двома рівнями факторів мають лише три значення: верхній рівень дорівнює +1, нижній – -1, а основний рівень – 0. Для факторів з безперервною областю визначення використовується така формула для перетворення реальних змінних у кодовані:

$$x_j = \frac{\bar{x}_j - \bar{x}_{j(0)}}{I_j},$$

де:  $x_j$  – кодоване значення фактора;  
 $\bar{x}_j$  – натуральне значення фактора на нижньому або верхньому рівні;  
 $\bar{x}_{j(0)}$  – натуральне значення фактора на основному рівні;  
 $I_j$  – інтервал варіювання;  
 $j$  – номер фактора.





Таке перетворення можна проілюструвати наступним графіком (рис. 10). Припустимо, що обрано основний рівень фактора 25°C, а інтервал варіювання – 5°C. Тоді в реальних змінних мінімальний рівень фактора приймає значення 20°C, а високий – 30°C.

Для перетворення точку початку координат переносимо у точку, яка відповідає основному рівню (точка 0), а інтервал варіювання обираємо рівним 1. Тоді нижній рівень буде дорівнювати  $0-1=-1$ , а верхній –  $0+1=+1$ .

Для якісних ознак, які варіюють на двох рівнях, один рівень позначається як +1, а інший – -1. Порядок рівнів значення не має.

#### Побудова планів експериментів $2^k$

Після визначення основного, верхнього та нижнього рівнів факторів можна скласти план експерименту. В експериментах  $2^k$  фактори варіюють на двох рівнях. У зв'язку з цим можливе отримання лінійної моделі. Якщо відома кількість факторів, включених до експерименту, відразу можна підрахувати кількість дослідів за формулою:

$$N = 2^k,$$

де:  $N$  – кількість необхідних дослідів;

$2$  – кількість рівнів фактора;

$k$  – кількість факторів, включених до експерименту.

Так, для двох факторів кількість необхідних дослідів буде дорівнювати:

$$N = 2^2 = 4.$$

У плані експерименту необхідно врахувати всі можливі сполучення рівнів факторів. Нагадаємо, що в плануванні факторного експерименту рівні факторів позначаються +1 (верхній рівень) та -1 (нижній рівень). При побудові планів у матрицях одиниця звичайно пропускається. Таким чином у матрицях планів верхній рівень фактора позначається знаком «+», а нижній рівень – знаком «-». Умови експерименту записуються у вигляді таблиць, рядки яких відповідають дослідом, а стовпці – факторам. Крім того, до таблиці включають стовпці для середніх значень функції відгуку.

Саме ці таблиці і називають матрицями експерименту. Для двофакторного експерименту матрицю експерименту наведено в табл. 39.

Таблиця 39

Матриця планування експерименту  $2^2$

Номер досліду	X1	X2	$Y_i$
1	-1	-1	$Y_1$
2	+1	-1	$Y_2$
3	-1	+1	$Y_3$
4	+1	+1	$Y_4$

Стовпці в таблиці, які відповідають факторам та значенням функції відгуку, називаються *вектор-стовпцями*, а рядки – *вектор-рядками*. Таким чином, у матриці експерименту  $2^2$  два вектор-стовпця незалежних змінних та один вектор-стовпець відгуку, і чотири вектор-рядки. У таблиці в алгебраїчному вигляді записані умови чотирьох дослідів, які потрібно

виконати для отримання лінійної моделі залежності функції відгуку від змін двох факторів.

Геометрично ці умови можна подати за допомогою системи координат (рис. 11). Для цього на осях, які відповідають факторам знаходять основні рівні (0 точки на осях  $\tilde{x}_1, \tilde{x}_2$ ) і проводять через них нові осі ( $x_1, x_2$ ). Точка їх перетину дає нульовий рівень експериментів у кодіваних одиницях. Далі

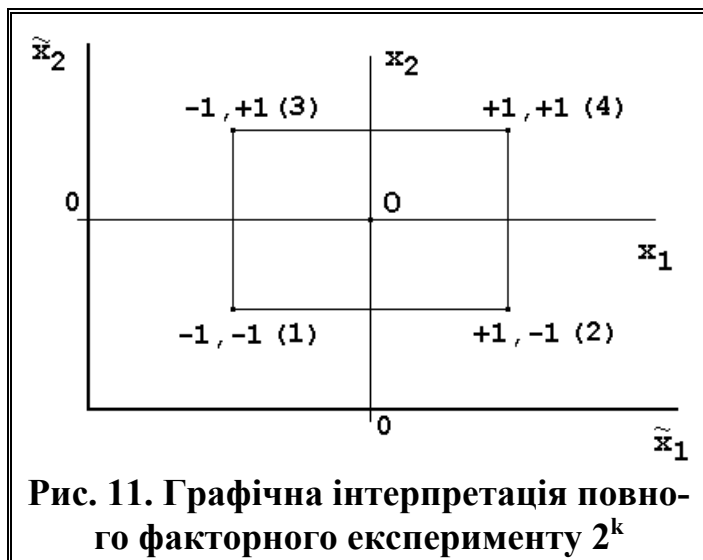


Рис. 11. Графічна інтерпретація повного факторного експерименту  $2^k$

на кодіваних осях проводять відрізки, які відповідають кодіваним значенням інтервалу варіювання (1). Точки перетину прямих, проведених через кодівані значення максимального та мінімального значень факторів, дають умови експериментів. На рисунку вони подані в дужках і відповідають номерам дослідів в табл. 39.

План експерименту з двома факторами є найпростішим планом. Але у практиці експериментування досить часто зустрічаються складніші плани

з трьома, чотирма і більшою кількістю факторів. Якщо для двофакторного експерименту всі можливі сполучення факторів встановити досить легко, а можна і просто їх запам'ятати, то при переході до складніших планів виникає проблема побудови матриць. Найчастіше для створення багатофакторних матриць використовують три методи. Найпростіший з них полягає в тому, що при введенні нового фактора кожна комбінація рівнів попереднього плану повторюється двічі: перший раз у сполученні з нижнім рівнем нового фактора, а другий раз – з верхнім рівнем нового фактора. Наприклад, для трьох факторів матриця, побудована таким способом, наведена в табл. 40.

Таблиця 40

Побудова матриці експерименту з трьома факторами на матриці двофакторного експерименту

<i>Область експерименту <math>2^2</math></i>				
Номер дослід	$X_1$	$X_2$	$X_3$	$Y_i$
1	-1	-1	-1	$Y_1$
2	+1	-1	-1	$Y_2$
3	-1	+1	-1	$Y_3$
4	+1	+1	-1	$Y_4$
5	-1	-1	+1	$Y_5$
6	+1	-1	+1	$Y_6$
7	-1	+1	+1	$Y_7$
8	+1	+1	+1	$Y_8$

Цей метод може бути поширений на матриці будь-якого розміру.

Другий метод ґрунтується на правилі чергування знаків. У першому стовпці знаки чергуються через 1, у другому – через 2, в третьому – через 4 і так далі за ступенями двійки. Ступінь являє собою номер фактора мінус один ( $2^0, 2^1, 2^2, 2^3$  і т.д.).

Третій спосіб полягає в перемножуванні стовпців матриці. Перемножування рядків матриці з одноіменними знаками дає +1, з різнойменними – -1. Користуючись цим правилом, можна отримати таку схему для варіювання третього фактора (перші чотири дослід):

$$1 \text{ дослід} - -1 \cdot -1 = +1; \quad 2 \text{ дослід} - +1 \cdot -1 = -1;$$

3 дослід  $- +1 \cdot -1 = -1$ ; 4 дослід  $- +1 \cdot +1 = +1$

Далі для третього фактору знаки міняються на зворотні. Це буде друга половина матриці (табл. 41):

Таблиця 41

Матриця експерименту  $2^3$ , отримана методом перемножування

Номер дос- ліду	X1	X2	X3	$y_i$
1	-1	-1	+1	$y_1$
2	+1	-1	-1	$y_2$
3	-1	+1	-1	$y_3$
4	+1	+1	+1	$y_4$
5	-1	-1	-1	$y_5$
6	+1	-1	+1	$y_6$
7	-1	+1	+1	$y_7$
8	+1	+1	-1	$y_8$

У перших двох випадках матриці планування мають однаковий вигляд (табл. 40), а третій метод дає матрицю, відмінну від перших. Статистична обробка матриць планування, отриманих всіма методами, здійснюється однаково, але інтерпретація результатів буде дещо відрізнятись.

#### Властивості повного факторного експерименту

Для побудови моделі важливо знати, які властивості має матриця експерименту незалежно від кількості факторів. Від цих властивостей у певній залежності знаходяться і властивості математичної моделі (оптимальність, точність передбачення, тощо). Під час проведення розрахунків за результатами дослідів будуть отримані коефіцієнти моделі, які повинні бути найліпшими. Точність передбачення не повинна залежати від напрямку в факторному просторі.

Дві властивості витікають безпосередньо з методів побудови матриці. Перша властивість – *симетричність відносно центру експерименту*. Ця властивість формулюється наступним чином: **алгебраїчна сума елементів вектор-стовпця дорівнює нулю**, або у вигляді формули –

$$\sum x_{ji} = 0,$$

де:  $j$  – номер фактора;

$N$  – кількість дослідів;  
 $i$  – номер дослідів.

Друга властивість – *правило нормування* – сума квадратів елементів кожного вектор–стовпця дорівнює кількості дослідів, або:

$$\sum x_{ji}^2 = N.$$

Це витікає з того, що значення факторів дорівнюють +1 або –1.

Третя властивість – *сума добутоків будь-яких двох вектор–стовпців матриці дорівнює нулю*. Ця властивість називається *ортогональністю матриці планування* і виражається формулою:

$$\sum x_{ji} \cdot x_{ui} = 0 \text{ де } j \neq u.$$

Остання, четверта властивість зветься *ротатабельністю*, тобто точки в матриці планування підбираються так, що точність передбачення значень відгуку однакова на рівних відстанях від центру експерименту і не залежить від напрямку.

Наприклад, є дві матриці ( $a$  та  $b$ ). Потрібно встановити, яка з них відповідає усім вимогам планів.

Для перевірки обчислюються всі параметри:

Для матриці  $a$ :  $\sum x_{ji} = 0$ ,  $\sum x_{ji}^2 = 4$ ,  $\sum x_{ji} \cdot x_{ui} = 0$

Для матриці  $b$ :  $\sum x_{ji} = 0$ ,  $\sum x_{ji}^2 = 4$ ,  $\sum x_{ji} \cdot x_{ui} = -4 (\neq 0)$ .

a		b	
$X_1$	$X_2$	$X_1$	$X_2$
–1	–1	–1	+1
+1	–1	+1	–1
–1	+1	–1	+1
+1	+1	+1	–1

Друга матриця ( $b$ ) не відповідає вимогам, оскільки не виконується третя властивість.

### Статистична обробка результатів експериментів типу $2^k$

Як вже повідомлялося, метою багатофакторного планування є отримання математичної моделі поведінки об'єкта досліджень залежно від зміни рівнів факторів, яка відбивається рівнянням лінійної залежності. Способи статистичної обробки таких експериментів зводяться до обчислення коефіцієнтів рівняння лінійної регресії типу

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2.$$

У розділі 6 ми вже зупинялися на проблемі статистичної оцінки коефіцієнтів рівняння регресії. В експерименті, який проводиться за планом, це питання також є актуальним. Можна стверджувати, що експеримент здійснюється з метою перевірки гіпотези, що лінійна модель

$$\bar{\mu} = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2$$

адекватна. Тобто в генеральній сукупності (на що вказують літери грецького алфавіту) з певною ймовірністю залежність змін функції відгуку буде відбуватися саме за такою формулою. Оскільки під час експерименту можна отримати лише вибіркові параметри, їх точність і надійність цілком залежать від властивостей вибірки і потребують статистичної перевірки. Тому рекомендується така послідовність статистичної обробки результатів експериментів.

1. Первинна статистична обробка результатів з перевітками усіх вибірок на нормальність розподілу та однорідність.
2. Перевірка вибірових середніх квадратів відхилень.
3. Оцінка значущості коефіцієнтів рівняння регресії.
4. Обчислення коефіцієнтів рівняння регресії.
5. Перевірка адекватності рівняння регресії.

Проведення первинної статистичної обробки, перевірка вибірок на однорідність і нормальність розподілу та перевірка однорідності середніх квадратів відхилень наводилися у відповідних розділах частини 1 посібника.

Оцінка значущості коефіцієнтів регресії

Обчислення значущості коефіцієнтів рівняння регресії здійснюється для визначення значущих коефіцієнтів, тобто таких коефіцієнтів, які вірогідно відрізняються від нуля. Нульова гіпотеза (коефіцієнт рівняння дорівнює 0) зберігається, якщо коефіцієнт менший за певне розрахункове значення, і відкидається, якщо він дорівнює або перевищує це значення. Для експериментів з двома рівнями факторів розрахункове значення значущості коефіцієнтів обчислюється наступним чином. На першому етапі знаходиться загальна сума всіх вибірових середніх квадратів відхилень:

$$s_{заг}^2 = \sum s_i^2 .$$

Далі визначається дисперсія одиничного значення:

$$s_{\{y\}}^2 = \frac{s_{заг}^2}{N \cdot (n - 1)},$$

де:  $N$  – кількість дослідів у плані (кількість вибірок);

$n$  – кількість дат у вибірці (об'єм вибірки).

Обчислюється дисперсія середнього значення, отриманого з  $n$  повторностей у кожному досліді:

$$s_{\{\bar{y}\}}^2 = \frac{s_{\{y\}}^2}{n},$$

і дисперсія коефіцієнтів рівняння регресії:

$$s_{\{b\}}^2 = \frac{s_{\{\bar{y}\}}^2}{\sum x_i^2} = \frac{s_{\{\bar{y}\}}^2}{N}.$$

Остання частина формули, де замість суми квадратів значень незалежної змінної використовується кількість дослідів, витікає з другої властивості матриць експериментів. Подальші обчислення можна вести двома шляхами. Обидва вони пов'язані з використанням критерію Ст'юдента (табл. 1 додатків). У таблиці знаходиться стандартне значення критерію ( $t_{st}$ ) для кількості ступенів вільності  $f=N \cdot (n-1)$ . Далі у першому випадку обчислюється за формулою

$$\{b\} = \sqrt{s_{\{b\}}^2} \cdot t_{st}$$

значущість коефіцієнтів регресії, з якою порівнюються всі коефіцієнти регресії, взяті за абсолютною величиною (знак при коефіцієнті не враховується). Якщо коефіцієнт рівняння регресії дорівнює або перевищує обчислену значущість, він включається до рівняння. В іншому випадку вважають, що коефіцієнт дорівнює 0 і разом з незалежною змінною при ньому викидають з рівняння.

Другий спосіб полягає в обчисленні для кожного коефіцієнта критерію значущості за формулою

$$t_i = \frac{|b|}{\sqrt{s_{\{b\}}^2}}$$

і порівнянні цього показника зі стандартним значенням критерію Ст'юдента. Критерієм відмінності коефіцієнту рівняння регресії від нуля є рівність або перевищення обчисленого значення над стандартним.

Величина  $\{b\}$  одночасно є надійним інтервалом для оцінки відповідності вибіркового коефіцієнта рівняння регресії відповідному генеральному показнику і позначається ще як  $\Delta_{b_i}$ . Тобто можна записати, що в генеральній сукупності з ймовірністю  $\alpha$

$$\beta_i = b_i \pm \Delta_{b_i}.$$

Обчислення коефіцієнтів рівняння регресії.

Загальним методом для обчислення коефіцієнтів рівняння багатофакторної регресії є метод найменших квадратів, який вже розглядався у розділі, присвяченому регресійному аналізу. Але властивості матриць планування дозволяють значно спростити процедури обчислень. Так, загальною формулою для обчислення будь-якого коефіцієнта рівняння багатофакторної регресії є наступна формула:

$$b_i = \frac{\sum x_{iu} \cdot \bar{y}_u}{\sum x_{iu}^2},$$

або, враховуючи властивості матриці планування:

$$b_i = \frac{\sum x_{iu} \cdot \bar{y}_u}{N}.$$

Враховуючи, що значення незалежних змінних можуть мати тільки два значення ( $-1$  або  $+1$ ) в чисельнику буде алгебраїчна сума значень відгуку із знаками, відповідними до знаків незалежних змінних у вектор-стовпцях. Наприклад, для обчислення коефіцієнтів  $b_1$  та  $b_2$  відповідні формули будуть мати наступний вигляд:

$$b_1 = \frac{(-1) \cdot \bar{y}_1 + (+1) \cdot \bar{y}_2 + (-1) \cdot \bar{y}_3 + (+1) \cdot \bar{y}_4}{4} = \frac{-\bar{y}_1 + \bar{y}_2 - \bar{y}_3 + \bar{y}_4}{4},$$

$$b_2 = \frac{(-1) \cdot \bar{y}_1 + (-1) \cdot \bar{y}_2 + (+1) \cdot \bar{y}_3 + (+1) \cdot \bar{y}_4}{4} = \frac{-\bar{y}_1 - \bar{y}_2 + \bar{y}_3 + \bar{y}_4}{4}.$$

Таким чином обчислені два коефіцієнти в рівнянні регресії. Тепер постає завдання обчислення ще одного коефіцієнта лінійної моделі –  $b_0$ . Це можна зробити, якщо ввести до матриці планування ще один стовпчик з фіктивною змінною  $X_0$  (табл. 42).

З таблиці видно, що фіктивна змінна у всіх дослідах має значення тільки  $+1$ . Тому формула для обчислення коефіцієнта  $b_0$  буде такою:

$$b_0 = \frac{(+1) \cdot \bar{y}_1 + (+1) \cdot \bar{y}_2 + (+1) \cdot \bar{y}_3 + (+1) \cdot \bar{y}_4}{4} = \frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4}{4}.$$

Тобто значення коефіцієнта  $b_0$  являє собою загальну середню комплексу.

Таблиця 42

Матриця планування з введеними до неї стовпцями фіктивної змінної та взаємодії факторів



Номер досліду	$X_0$	$X_1$	$X_2$	$X_1 \cdot X_2$	$y_i$
1	+1	-1	-1	+1	$\bar{y}_1$
2	+1	+1	-1	-1	$\bar{y}_2$
3	+1	-1	+1	-1	$\bar{y}_3$
4	+1	+1	+1	+1	$\bar{y}_4$

У результаті проведених розрахунків обчислено три коефіцієнту рівняння за результатами чотирьох дослідів і отримано рівняння лінійної моделі

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2.$$

Але ще залишився четвертий дослід, який можна використати двома способами. Основна мета цього дослідів – це з'ясування взаємодії між факторами. Сутність взаємодії вже розглядалася в розділі, присвяченому дисперсійному аналізу. Для обчислення коефіцієнта рівняння регресії при взаємодії потрібно додати ще один стовпчик до матриці планування (п'ятий стовпчик у табл. 42), який отримують перемноженням значень незалежних змінних  $X_1$  та  $X_2$ . Після цього підставляють значення незалежних змінних у формулу обчислення коефіцієнтів:

$$b_{12} = \frac{(+1) \cdot \bar{y}_1 + (-1) \cdot \bar{y}_2 + (-1) \cdot \bar{y}_3 + (+1) \cdot \bar{y}_4}{4} = \frac{\bar{y}_1 - \bar{y}_2 - \bar{y}_3 + \bar{y}_4}{4}.$$

Такі вектор–стовпці можна отримати для будь-якої кількості взаємодіючих факторів. В результаті обчислення останнього коефіцієнта буде отримана неповна квадратична модель, яка має вигляд

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_{12} \cdot x_1 \cdot x_2.$$

У дослідях, коли метою досліджень є інтерполяція, врахування взаємодії обов'язкове, якщо коефіцієнт  $b_{12}$  вірогідно відрізняється від 0. Під час проведення екстремальних експериментів інколи навіть у випадках вірогідності цього коефіцієнта його не враховують, а четвертий дослід використовують для введення додаткового фактора. Про проведення таких дослідів йтиметься у наступному розділі.

Другим способом обчислення коефіцієнтів рівняння регресії є *алгоритм Йейтса*, який дозволяє суттєво спростити розрахунки. Цей метод складається з послідовних кроків попарного сумування та віднімання середніх значень функції відгуку. Кількість таких кроків дорівнює кількості факторів, і в результаті обчислень отримуються *контрасти*. Ділення конт-

растів на кількість дослідів ( $N$ ) дає коефіцієнти рівняння регресії. Послідовність цих операцій можна проілюструвати таблицею 43.

На першому кроці використовуються емпіричні середні значення. Спочатку до першого середнього додається друге, а до третього – четверте. У результаті отримують першу пару проміжних значень ( $G_1$  та  $G_2$ ). Далі від другого середнього віднімають перше, а від четвертого – третє і отримують наступні два проміжні значення ( $G_3$  та  $G_4$ ). Зверніть увагу на те, що віднімати завжди потрібно від значення, яке має більший номер (але не величину), значення з меншим номером. На другому кроці аналогічна операція проводиться з проміжними величинами, які були отримані в попередньому кроці. В шостому стовпчику наведені комбінації незалежних змінних, які відносяться до відповідних коефіцієнтів. Із збільшенням кількості факторів до трьох, кількість кроків також збільшується до трьох. Чотири фактори вимагають чотирьох кроків і так далі.

Таблиця 43

Послідовність розрахунків в алгоритмі Йейтса

Номер дослідів	$\bar{y}_i$	1-й крок	2-й крок (контрасти)	Коефіцієнти	Змінна
1	$\bar{y}_1$	$G_1 = \bar{y}_1 + \bar{y}_2$	$C_1 = G_1 + G_3$	$b_0 = C_1/n$	–
2	$\bar{y}_2$	$G_2 = \bar{y}_3 + \bar{y}_4$	$C_2 = G_3 + G_4$	$b_1 = C_2/N$	$X_1$
3	$\bar{y}_3$	$G_3 = \bar{y}_2 - \bar{y}_1$	$C_3 = G_2 - G_1$	$b_2 = C_3/N$	$X_2$
4	$\bar{y}_4$	$G_4 = \bar{y}_4 - \bar{y}_3$	$C_4 = G_4 - G_3$	$b_{12} = C_4/N$	$X_1 \cdot X_2$

Після обчислення коефіцієнтів рівняння регресії з кодованими значеннями змінних і визначення значущих коефіцієнтів можна провести переведення кодованих змінних у реальні. Для цього використовується наведена раніше формула для визначення кодованих рівнів факторів. Якщо тепер замість кодованого значення фактора до рівняння підставити ці формули, розкрити дужки та привести подібні члени, буде отримано рівняння регресії в реальних змінних. Для двох факторів формула такого перетворення має такий вигляд:

$$\bar{y} = b_0 + b_1 \cdot \left( \frac{\bar{x}_1 - \bar{x}_{1(0)}}{I_1} \right) + b_2 \cdot \left( \frac{\bar{x}_2 - \bar{x}_{2(0)}}{I_2} \right) + b_{12} \cdot \left( \frac{\bar{x}_1 - \bar{x}_{1(0)}}{I_1} \right) \cdot \left( \frac{\bar{x}_2 - \bar{x}_{2(0)}}{I_2} \right).$$

Це перетворення потрібно завжди робити при обчисленні моделі для інтерполяції. Якщо досліді проводяться для пошуку оптимальних умов, то таке перетворення необов'язкове, а інколи і небажане.

## Встановлення адекватності рівняння регресії

Адекватність рівняння регресії – важливий параметр, який дозволяє встановити якість математичної моделі (рівняння регресії), тобто як точно багатofакторна модель описує емпіричну залежність (результати експерименту). Як і у випадку регресійного аналізу для встановлення адекватності рівняння використовується порівняння за Фішером *дисперсії адекватності* та *дисперсії відтворення*.

Дисперсія адекватності обчислюється за наступною формулою:

$$S_{ad}^2 = \frac{(\tilde{y}_i - \bar{y}_i)^2}{N - u},$$

де:  $S_{ad}^2$  – дисперсія адекватності;

$\tilde{y}_i$  – обчислені (теоретичні) значення результативної ознаки;

$\bar{y}_i$  – експериментальні середні значення результативної ознаки;

$N$  – кількість дослідів у матриці планування;

$u$  – кількість коефіцієнтів (параметрів) рівняння регресії.

Якщо для обчислення коефіцієнтів рівняння використані всі дослідів і всі коефіцієнти значущі, виникає ситуація, коли знаменник дроби буде дорівнювати нулю. Тобто не буде ступенів вільності для обчислення дисперсії адекватності. У такому випадку план зветься *насиченим* і адекватність рівняння не обчислюється. Треба сказати, що у випадку насиченого плану, найчастіше розрахункові значення дорівнюють або дуже близькі до емпіричних середніх функції відгуку. Тому і чисельник дроби буде також дорівнювати 0 або дуже близьким до нього.

Методика обчислення дисперсії відтворення для багатofакторних експериментів наведена вище. Це дисперсія середнього значення, яка розраховується під час визначення значущості коефіцієнтів рівняння регресії ( $s_{\{y\}}^2$ ).

Критерієм для порівняння дисперсій адекватності та відтворення слугує їх відношення:

$$F_{ad} = \frac{S_{ad}^2}{S_{\bar{y}}^2}.$$

Якщо обчислене значення  $F$  менше стандартного значення для порівняння дисперсій за Фішером для обраного рівня значущості ( $\alpha$ ), кількості ступенів вільності дисперсії адекватності ( $N-u$ ) та кількості ступенів вільності для дисперсії відтворення ( $N-n$ ), тоді рівняння регресії визнається адекватним. В іншому випадку рівняння регресії неадекватне, а математична модель незадовільно описує емпіричну залежність. У випадку лінійного рів-

няння це може означати, що дійсна залежність має суттєву нелінійність і потрібно використовувати плани з більшою кількістю рівнів факторів, щоб можна було обчислити рівняння вищих ступенів, або звужити діапазон варіювання факторів.

### Приклад статистичної обробки експерименту $2^k$

Методику проведення статистичної обробки експериментів  $2^k$  можна показати на прикладі двофакторного експерименту. Наприклад, вивчалася дія температури та рН середовища на ріст міцелію. Матриця експерименту, середні значення функції відгуку та вибіркові квадрати відхилень наведені в табл. 44.

Таблиця 44

Результати експерименту, виконаного за планом  $2^2$

Номер досліджу	X1	X2	$y_i$	$s_i^2$
1	-1	-1	10	5,0
2	+1	-1	20	10,0
3	-1	+1	15	7,0
4	+1	+1	10	4,0

Спочатку обчислюється значущість коефіцієнтів рівняння регресії:

$$s_{\{y\}}^2 = \frac{s_{заг}^2}{N \cdot (n - 1)} = \frac{26,0}{4 \cdot (3 - 1)} = 3,25,$$

$$s_{\{y\}}^2 = \frac{s_{\{y\}}^2}{n} = \frac{3,25}{3} = 1,08,$$

$$s_{\{b\}}^2 = \frac{s_{\{y\}}^2}{N} = \frac{1,08}{4} = 0,27.$$

У таблиці 1 додатків знаходимо значення критерію Ст'юдента (2,16) і обчислюємо показник значущості коефіцієнтів рівняння:

$$\{b\} = \sqrt{s_{\{b\}}^2} \cdot t_{st} = \sqrt{0,27} \cdot 2,16 = 1,1$$

Таким чином, усі коефіцієнти рівняння, які за абсолютною величиною будуть дорівнювати або перевищувати 1,1, повинні бути внесені до рівняння регресії.

Для обчислення коефіцієнтів рівняння регресії використаємо обидва методи. Спершу застосуємо метод найменших квадратів:

$$b_0 = \frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4}{4} = \frac{10 + 20 + 15 + 10}{4} = 13,75,$$

$$b_1 = \frac{-\bar{y}_1 + \bar{y}_2 - \bar{y}_3 + \bar{y}_4}{4} = \frac{-10 + 20 - 15 + 10}{4} = 1,25,$$

$$b_2 = \frac{-\bar{y}_1 - \bar{y}_2 + \bar{y}_3 + \bar{y}_4}{4} = \frac{-10 - 20 + 15 + 10}{4} = -1,25,$$

$$b_{12} = \frac{\bar{y}_1 - \bar{y}_2 - \bar{y}_3 + \bar{y}_4}{4} = \frac{10 - 20 - 15 + 10}{4} = -3,75.$$

Як бачимо, всі коефіцієнти перевищують значення критерію вірогідності і повинні увійти до рівняння. Таким чином, рівняння регресії у кодованих змінних буде таким:

$$\hat{y} = 13,75 + 1,25 \cdot x_1 - 1,25 \cdot x_2 - 3,75 \cdot x_1 \cdot x_2$$

Знаки біля коефіцієнтів рівняння показують напрям впливу (знак «+» – позитивний вплив, знак «-» – негативний). Виходячи з цього, збільшення температури (у межах досліджених температур) веде до збільшення росту, а підвищення рН, навпаки, пригнічує ріст. Сумісна дія обох факторів (взаємодія) має негативний знак, тобто підвищення значень одного з факторів викликає зниження впливу іншого. Тепер проведемо обчислення коефіцієнтів за алгоритмом Йейтса (табл. 45).

## Обчислення коефіцієнтів рівняння регресії за алгоритмом Йейтса

Номер досліджу	Вибіркові середні	1-й крок	2-й крок	Коефіцієнти	Незалежні змінні
1	10	10+20=30	30+25=55	13,75	–
2	20	15+10=25	10+(-5)=5	1,25	$X_1$
3	15	20-10=10	25-30=-5	-1,25	$X_2$
4	10	10-15=-5	-5-(-10)=-15	-3,75	$X_1 \cdot X_2$

Як бачимо, обома способами отримані однакові результати. Тепер потрібно перевірити адекватність рівняння. Але в цьому прикладі перевірка адекватності неможлива через насиченість плану. Останнім кроком є переведення кодованих змінних у реальні. Для цього до отриманого рівняння регресії підставляються формули для переходу від реальних змінних в кодовані. Припустимо, що низький рівень фактору  $X_1$  становить 2, а високий – 10°C, для фактору  $X_2$  ці показники становлять відповідно 1 та 5 од. Основний рівень для першого фактора становить  $x_{1(0)} = \frac{2+10}{2} = 6$ , а інтервал варіювання –  $I_1 = 10 - 6 = 6 - 2 = 4$ .

Для другого фактора ці показники відповідно становлять 3 і 2. Тоді:

$$\begin{aligned}
 y &= b_0 + b_1 \cdot \left( \frac{\bar{x}_1 - \bar{x}_{1(0)}}{I_1} \right) + b_2 \cdot \left( \frac{\bar{x}_2 - \bar{x}_{2(0)}}{I_2} \right) + b_{12} \cdot \left( \frac{\bar{x}_1 - \bar{x}_{1(0)}}{I_1} \right) \cdot \left( \frac{\bar{x}_2 - \bar{x}_{2(0)}}{I_2} \right) = \\
 &= 13,75 + 1,25 \cdot \left( \frac{x_1 - 6}{4} \right) - 1,25 \cdot \left( \frac{x_2 - 3}{2} \right) - 3,75 \cdot \left( \frac{x_1 - 6}{4} \right) \cdot \left( \frac{x_2 - 3}{2} \right) = \\
 &= 13,75 + 0,3125 \cdot x_1 - 1,875 - 0,625 \cdot x_2 + 1,875 - 0,625 \cdot x_1 \cdot x_2 + 1,875 \cdot x_1 + \\
 &+ 3,75 \cdot x_2 - 11,25 = 2,5 + 2,19 \cdot x_1 + 3,125 \cdot x_2 - 0,635 \cdot x_1 \cdot x_2
 \end{aligned}$$

## 10. НЕПОВНИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ

### Складання планів дослідів неповного факторного експерименту

Кількість дослідів у повному факторному експерименті набагато перевищує кількість коефіцієнтів, які визначаються для побудови лінійної моделі. Наприклад, у двофакторному експерименті обчислюється три коефіцієнти, а кількість дослідів становить чотири; в експериментах з трьома факторами кількість коефіцієнтів лінійного рівняння дорівнює чотирьом, а проводиться вісім дослідів. Іншими словами, повний факторний експеримент, якщо не враховуються взаємодії факторів, має значну надмірність. Надмірність швидко зростає із збільшенням кількості факторів. Якщо дослідника цікавлять тільки лінійні ефекти або коефіцієнти при взаємодіях факторів не відрізняються вірогідно від нуля, таку надмірність можна використати для введення в плани додаткових факторів. Однак слід зазначити, що коректне використання такого введення можливе тільки за відсутності вірогідної взаємодії, оскільки ефекти нових факторів прирівнюються до ефектів взаємодії. Інша особливість цього прийому полягає в тому, що він використовується в переважно тільки для екстремальних експериментів. Експерименти, в яких використовуються плани, коли новий фактор прирівнюється до вектор-стовпця взаємодій, називають *неповними факторними експериментами*, а самі плани – *неповними репліками*.

Для ознайомлення з методами побудови планів неповних реплік ще раз розглянемо матрицю планування експерименту  $2^2$  (табл. 45).

Таблиця 45

Матриця експерименту  $2^2$

Номер дослідів	$X_1$	$X_2$	$X_1 \cdot X_2$	$Y_i$
1	-1	-1	+1	$Y_1$
2	+1	-1	-1	$Y_2$
3	-1	+1	-1	$Y_3$
4	+1	+1	+1	$Y_4$

За результатами такого експерименту можна обчислити чотири коефіцієнти і подати результати у вигляді неповного квадратичного рівняння (рівняння другого ступеня), як було показано у попередньому розділі. Якщо ж є підстави вважати, що в обраних інтервалах варіювання процес може бути описаний виключно лінійною моделлю, то достатньо обчислити всього три коефіцієнти. Залишається один ступінь вільності, який можна використати для зменшення кількості дослідів. Якщо прийняти лінійне наближення, то  $b_{12} \rightarrow 0$  і вектор-стовпець  $X_1 X_2$  можна використати для нового

фактора  $X_3$  (табл. 46). У табл. 46 також включені вектор-стовпці взаємодії факторів.

Таблиця 46

Матриця експерименту  $2^{3-1}$

Номер досліджу	$X_1$	$X_2$	$X_3$	$X_1X_2$	$X_1X_3$	$X_2X_3$	$y_i$
1	-1	-1	+1	+1	-1	-1	$\bar{y}_1$
2	+1	-1	-1	-1	-1	+1	$\bar{y}_2$
3	-1	+1	-1	-1	+1	-1	$\bar{y}_3$
4	+1	+1	+1	+1	+1	+1	$\bar{y}_4$

Запланований в такий спосіб експеримент зберігає всі властивості матриць планування (симетричність відносно центра експерименту, ортогональність та інші). Однак втрачається дуже важлива властивість – незалежність оцінки коефіцієнтів моделі. Якщо уважно розглянути вектор-стовпці таблиці, то можна звернути увагу, що стовпці всіх окремих факторів повторюють який-небудь стовпець взаємодії. Так, вектор-стовпець  $X_1$  повторює вектор-стовпець  $X_2X_3$ , вектор-стовпець  $X_2$  повторює вектор-стовпець  $X_1X_3$ , а вектор-стовпець  $X_3$  повторює вектор-стовпець  $X_1X_2$ . В такому випадку кажуть, що ефекти факторів змішані з ефектами взаємодії. Це можна також проілюструвати такими формулами:

$$b_1 = \beta_1 + \beta_{23}, \quad b_2 = \beta_2 + \beta_{13}, \quad b_3 = \beta_3 + \beta_{12},$$

де:  $b_i$  – коефіцієнти рівняння регресії, обчислені за результатами експерименту;

$\beta_i$  – відповідні показники генеральної сукупності.

Однак при плануванні неповних експериментів передбачається лінійна модель. Тому всі взаємодії факторів приймаються рівними нулю і в результаті обчислень будуть отримані чисті оцінки ефектів факторів. Головним при побудові таких планів є зменшення кількості дослідів і можливість визначення напряму руху до оптимуму.

Склавши план з чотирьох дослідів для оцінки впливу трьох факторів, ми отримали половину повного факторного експерименту  $2^3$  або так звану напіврепліку трьохфакторного експерименту. Крім того, існує можливість прирівняти  $X_3$  до  $-X_1X_2$ . Для цього потрібно змінити знаки у вектор-стовпці на протилежні. У такому випадку

$$b_1 = \beta_1 - \beta_{23}, \quad b_2 = \beta_2 - \beta_{13}, \quad b_3 = \beta_3 - \beta_{12}.$$



Після реалізації обох напівреплік можна отримати окремі оцінки лінійних ефектів і ефектів взаємодії, як в повному факторному експерименті  $2^3$ . Власне кажучи, виконання обох напівреплік  $2^{3-1}$  і є проведення повного факторного експерименту.

**Вибір напівреплік. Генеруюче відношення та визначальні контрасти**

Якщо використовуються напіврепліки  $2^{3-1}$ , існує всього дві можливості: прирівняти  $X_3$  до  $+X_1X_2$  або до  $-X_1X_2$ . Тому існує всього дві напіврепліки експерименту  $2^{3-1}$  (табл. 47).

Таблиця 47

Можливі напіврепліки експерименту  $2^{3-1}$

Номер досліджу	1-ша напіврепліка				2-га напіврепліка			
	$X_1$	$X_2$	$X_3$	$X_1X_2X_3$	$X_1$	$X_2$	$X_3$	$X_1X_2X_3$
1	-1	-1	+1	+1	-1	-1	-1	-1
2	+1	-1	-1	+1	+1	-1	+1	-1
3	-1	+1	-1	+1	-1	+1	+1	-1
4	+1	+1	+1	+1	+1	+1	-1	-1

Для добутоків трьох стовпців ( $X_1, X_2, X_3$ ) матриці 1 виконується співвідношення  $+I=X_1 \cdot X_2 \cdot X_3$ , а для матриці 2:  $-I=X_1 \cdot X_2 \cdot X_3$ . З таблиці видно, що всі добутки однакові і в першому випадку дорівнюють плюс одиниці, а в другому – мінус одиниці.

Символічне позначення добутоків стовпців, рівних  $+1$  або  $-1$ , називається визначальним контрастом. Визначальний контраст допомагає визначати змішані ефекти. Для того, щоб визначити, який ефект змішаний з певним, потрібно обидві частини визначального контрасту помножити на стовпець, який відповідає певному ефекту. Так, якщо  $I=X_1X_2X_3$  то для  $X_1$ :

$$X_1 = X_1^2 \cdot X_2 \cdot X_3 = X_2 \cdot X_3$$

оскільки  $X_1^2 = 1$ .

Відповідні показники для  $X_2$  та  $X_3$  будуть такими:

$$X_2 = X_1 \cdot X_2^2 \cdot X_3 = X_1 \cdot X_3 \text{ та } X_3 = X_1 \cdot X_2 \cdot X_3^2 = X_1 \cdot X_2.$$

Це означає, що коефіцієнти лінійного рівняння будуть оцінками:

$$b_1 \rightarrow \beta_1 + \beta_{23}, \quad b_2 \rightarrow \beta_2 + \beta_{13}, \quad b_3 \rightarrow \beta_3 + \beta_{12}.$$

Співвідношення, що показує з яким з ефектів змішаний даний ефект, називається генеруючим відношенням.

Напіврепліки, в яких основні ефекти змішані з двофакторними взаємодіями, називають планами з розв'язувальною спроможністю III (по найбільшій кількості у визначальному контрасті). Такі плани прийнято позначати як  $2_{III}^{3-1}$ .

При виборі напіврепліки  $2^{4-1}$  можливі вісім рішень:

- |                    |                    |                       |
|--------------------|--------------------|-----------------------|
| 1) $X_4 = X_1X_2$  | 4) $X_4 = -X_2X_3$ | 7) $X_4 = X_1X_2X_3$  |
| 2) $X_4 = -X_1X_2$ | 5) $X_4 = X_1X_3$  | 8) $X_4 = -X_1X_2X_3$ |
| 3) $X_4 = X_2X_3$  | 6) $X_4 = -X_1X_3$ |                       |

Розв'язувальна здатність цих напівреплік різна. Так, репліки 1 – 6 мають по три фактори у визначальному контрасті, а 7 та 8 – по чотири. Напіврепліки 7 та 8 мають максимальну розв'язувальну здатність і називаються головними. Розв'язувальна здатність визначається системою змішування певної репліки. Вона буде максимальною, якщо лінійні ефекти змішані із взаємодіями найбільшого можливого порядку.

За відсутності апріорної інформації про ефекти взаємодій дослідник повинен прагнути обрати репліку з максимальною розв'язувальною здатністю, оскільки потрібні взаємодії менш ймовірні і менш важливі, ніж парні. Якщо існує інформація про ефекти взаємодії факторів, то вона обов'язково повинна враховуватися під час вибору напівреплік.

Репліки, в яких немає жодного головного ефекту змішаного з іншим головним ефектом або парним ефектом, а всі парні взаємодії змішані один з одним, називаються планами з розв'язувальною здатністю IV (за найбільшою кількістю факторів у визначальному контрасті) і позначаються як плани  $2_{IV}^{4-1}$ .

Наприклад, обрані напіврепліки, які задані визначальними контрастами  $I = X_1X_2X_3X_4$  та  $I = -X_1X_2X_3X_4$ . Сумісні оцінки у цьому випадку визначаються співвідношеннями:

$X_1 = X_2X_3X_4$	$X_1 = -X_2X_3X_4$
$X_2 = X_1X_3X_4$	$X_2 = -X_1X_3X_4$
$X_3 = X_1X_2X_4$	$X_3 = -X_1X_2X_4$
$X_4 = X_1X_2X_3$	$X_4 = -X_1X_2X_3$
$X_1X_2 = X_3X_4$	$X_1X_2 = -X_3X_4$
$X_1X_3 = X_2X_4$	$X_1X_3 = -X_2X_4$
$X_1X_4 = X_2X_3$	$X_1X_4 = -X_2X_3$

Такий тип змішування дає можливість оцінювати лінійні ефекти сумісно з ефектами взаємодії другого порядку (потрійними), а взаємодії першого порядку (парні) – сумісно один з одним.

Якщо напіврепліки задані генеруючим відношенням  $X_4 = X_1X_2$  або  $X_4 = -X_1X_2$ , то у цьому випадку визначальними контрастами є  $I = X_1X_2X_4$  та  $I = -$

$X_1X_2X_4$ . Розв'язувальна здатність такого плану становить III і деякі основні ефекти змішуються з парними взаємодіями:

$X_1 = X_2X_4$	$X_1 = -X_2X_4$
$X_2 = X_1X_4$	$X_2 = -X_1X_4$
$X_3 = X_1X_2X_3X_4$	$X_3 = -X_1X_2X_3X_4$
$X_4 = X_1X_2$	$X_4 = -X_1X_2$
$X_1X_3 = X_2X_3X_4$	$X_1X_3 = -X_2X_3X_4$
$X_2X_3 = X_1X_3X_4$	$X_2X_3 = -X_1X_3X_4$
$X_2X_4 = X_1X_2X_3$	$X_2X_4 = -X_1X_2X_3$

Розв'язувальна здатність таких напівреплік менша, ніж у планів з розв'язувальною здатністю IV. Такі напіврепліки має сенс обирати, якщо дослідник володіє апріорною інформацією про наявність значущої потрійної взаємодії або про незначущість однієї з трьох парних взаємодій ( $X_1X_2$ ,  $X_1X_3$  або  $X_2X_3$ ).

Таким чином, використання неповного планування потребує точного знання системи змішування і повного уявлення, яка інформація втрачається. У загальному випадку найменш значущими в біологічних системах є взаємодії більш високих порядків (потрійні, почетверні і т.д.). Тому саме ці взаємодії рекомендується використовувати для прирівнювання нового фактора при створенні напівреплік. **Статистична обробка результатів експериментів, виконаних за планами напівреплік нічим не відрізняється від статистичної обробки повних факторних планів.** Тільки слід враховувати, що відповідні коефіцієнти у неповних планах відбивають не взаємодію факторів, а ефекти нових факторів або змішані ефекти взаємодії і нового фактора.

#### Метод послідовної реалізації та метод перевалу

Використовуючи плани типу  $2^k$ , іноді доцільно спочатку виконати невеликі групи дослідів – мінімально можливі плани типу  $2^{p-k}$ , щоб швидко отримати очікувані ефекти. Якщо в результаті проведених дослідів з'ясується їх недостатність, потрібно реалізувати наступну групу дослідів (наступний блок  $2^{k-p}$ ) з метою уточнення коефіцієнтів поліному моделі та встановлення її нових членів. У деяких випадках до такого способу реалізації неповного факторного експерименту спонукає встановлення значущого ефекту взаємодії факторів, який був прирівняний до нуля. В інших випадках така реалізація є єдиним можливим способом реалізації повних факторних планів. Такі ситуації трапляються за постійного дрейфу факторів і розглядаються в спеціальних посібниках. У зв'язку з тим, що цей метод передбачає послідовну реалізацію окремих блоків, він отримав назву *методу послідовної реалізації*.

Для ілюстрації цього методу ще раз повернемося до табл. 47. У напіврепліці **a** фактор  $X_3$  прирівняний до взаємодії  $X_1X_2$ , а у напіврепліці **b** – до  $-X_1X_2$ . Припустимо, що виконані експерименти першої напіврепліки і

після їх статистичної обробки з'ясована недостатність матеріалу для побудови якісної математичної моделі (рівняння виявилось неадекватним і може бути припущена значущість взаємодії факторів). У такому випадку рекомендується реалізувати і другу напіврепліку. Після виконання другої серії експериментів її результати об'єднуються з першими і проводиться сумісна статистична обробка. При цьому будуть отримані незмішані оцінки лінійних ефектів факторів і виокремлені ефекти взаємодій. Ефект виокремлення лінійних ефектів можна показати, якщо просумувати коефіцієнти першої та другої матриць

$$\begin{array}{lll}
 b_1 = \beta_1 + \beta_{23} & b_2 = \beta_2 + \beta_{13} & b_3 = \beta_3 + \beta_{12} \\
 + b_1 = \beta_1 - \beta_{23} & + b_2 = \beta_2 - \beta_{13} & + b_3 = \beta_3 - \beta_{12} \\
 \hline
 b_1 = \beta_1 & b_2 = \beta_2 & b_3 = \beta_3
 \end{array}$$

З наведених формул витікає, що лінійні ефекти звільнилися від ефектів взаємодії. Самі ефекти взаємодії також можуть бути встановлені після проведення експериментів за планами обох напівреплік. Практично це витікає з того, що дві напіврепліки  $2^{k-1}$  складають повну репліку  $2^k$ .

Іншим способом позбутися змішування лінійних ефектів з парними взаємодіями є *метод перевалу*. Цей метод полягає в тому, що знаки при всіх незалежних змінних змінюються на протилежні у всіх вектор-стовпцях. Метод перевалу є окремим випадком методу послідовної реалізації і часто використовується в матрицях неповних планів для вивільнення лінійних ефектів від ефектів взаємодій.

#### Репліки значної неповноти. Узагальнюючий визначальний контраст

У плані повного факторного експерименту  $2^3$  крім трьох вектор-стовпців лінійних ефектів, є ще чотири вектор-стовпці взаємодій (табл. 48). Ці вектор-стовпці можна використати для прирівнювання до них нових факторів, якщо буде встановлено незначущість парних та потрійних взаємодій. Максимальна кількість факторів, яка може бути введена до матриці плану  $2^3$  дорівнює семи.

Таблиця 48

Матриця експерименту  $2^3$  з вектор-стовпцями взаємодій

Номер досліджу	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>1</sub> X <sub>2</sub>	X <sub>1</sub> X <sub>3</sub>	X <sub>2</sub> X <sub>3</sub>	X <sub>1</sub> X <sub>2</sub> X <sub>3</sub>
1	-1	-1	-1	+1	+1	+1	-1
2	+1	-1	-1	-1	-1	+	+1
3	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	-1	+1	-1	-1	-1
5	-1	-1	+1	+1	-1	-1	+
6	+1	-1	+1	-1	+1	-1	-1
7	-1	+1	+1	-1	-1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1

Якщо до плану вводиться новий фактор, то кожного разу збільшується неповнота планів. Так, у результаті введення одного додаткового фактора буде отримана напіврепліка, двох – 1/4 репліки (чверть-репліка), трьох – 1/8 репліки і так далі.

Разом з тим під час планування такого експерименту значно ускладнюється вибір генеруючих відношень та визначальних контрастів для побудови матриці. Наприклад, при аналізі п'яти факторів (побудові 1/4 репліки) можливі дванадцять варіантів прирівнювання факторів тільки у тому разі, якщо один фактор прирівнюється до потрійної взаємодії, а другий – до парної:

- |     |               |                 |     |               |                  |
|-----|---------------|-----------------|-----|---------------|------------------|
| 1)  | $X_4=X_1X_2$  | $X_5=X_1X_2X_3$ | 2)  | $X_4=X_1X_2$  | $X_5=-X_1X_2X_3$ |
| 3)  | $X_4=-X_1X_2$ | $X_5=X_1X_2X_3$ | 4)  | $X_4=-X_1X_2$ | $X_5=-X_1X_2X_3$ |
| 5)  | $X_4=X_1X_3$  | $X_5=X_1X_2X_3$ | 6)  | $X_4=X_1X_3$  | $X_5=-X_1X_2X_3$ |
| 7)  | $X_4=-X_1X_3$ | $X_5=X_1X_2X_3$ | 8)  | $X_4=-X_1X_3$ | $X_5=-X_1X_2X_3$ |
| 9)  | $X_4=X_2X_3$  | $X_5=X_1X_2X_3$ | 10) | $X_4=X_2X_3$  | $X_5=-X_1X_2X_3$ |
| 11) | $X_4=-X_2X_3$ | $X_5=X_1X_2X_3$ | 12) | $X_4=-X_2X_3$ | $X_5=-X_1X_2X_3$ |

Вибір варіантів залежить від того, які взаємодії вважаються незначущими. Припустимо, що обраний п'ятий варіант ( $X_4=X_1X_3$   $X_5=X_1X_2X_3$ ). Тоді визначальними контрастами будуть  $I=X_1X_3X_4$  та  $I=X_1X_2X_3X_5$ . Якщо перемножити ці визначальні контрасти, то буде отримане третє відношення, яке визначає елементи стовпця  $I=X_2X_4X_5$ . Щоб повністю охарактеризувати розв'язувальну спроможність, необхідно записати узагальнюючий визначальний контраст  $I=X_1X_3X_4=X_2X_4X_5=X_1X_2X_3X_4X_5$ . Система змішування визначається множенням узагальнюючого контрасту послідовно на  $X_1$ ,  $X_2$ ,  $X_3$  і так далі.

$$\begin{aligned}
 X_1 &= X_3X_4 = X_1X_2X_4X_5 = X_2X_3X_5, \\
 X_2 &= X_1X_2X_3X_4 = X_4X_5 = X_1X_3X_5, \\
 X_3 &= X_1X_4 = X_2X_3X_4X_5 = X_1X_2X_5, \\
 X_4 &= X_1X_3 = X_2X_5 = X_1X_2X_3X_4X_5, \\
 X_5 &= X_1X_3X_4X_5 = X_2X_4 = X_1X_2X_3, \\
 X_1X_2 &= X_2X_3X_4 = X_1X_4X_5 = X_3X_5, \\
 X_1X_5 &= X_3X_4X_5 = X_1X_2X_4 = X_2X_3.
 \end{aligned}$$

Отримана досить складна система змішування лінійних ефектів з ефектами взаємодії першого, другого, третього і четвертого порядків. Якщо під час проведення такого експерименту з'ясується вірогідність хоча б однієї з парних взаємодій, то виникають сумніви, по-перше, у незначущості всіх інших взаємодій, а по-друге, виникає неоднозначність в оцінці лінійних ефектів. В такому випадку слід зробити другу серію дослідів, скориставшись методом послідовної реалізації або методом перевалу. Використання методу перевалу приводить до зміни знаків добутоків потрійних

ефектів на протилежні, порівняно з першою чверть–реплікою. Усереднення результатів двох дослідів дає можливість звільнити лінійні ефекти від парних взаємодій.

Під час вибору реплік значної неповноти (1/8, 1/16 і т.д.) аналіз генеруючих відношень і узагальнюючих контрастів значно ускладнюється. Так, до узагальнюючого контрасту 1/8 репліки входять сім рівностей, а кожна оцінка коефіцієнтів при лінійних факторах змішана з трьома взаємодіями (як парними, так і потрійними).

Дуже важливим при складанні матриць неповних планів є незначущість практично всіх взаємодій між факторами, які змішані з лінійними ефектами, а не тільки тих, що змішані з новими факторами. Оскільки під час планування біологічних експериментів ніколи не можна бути впевненим у відсутності вірогідної взаємодії між факторами, вважається, що використання планів значної неповноти на базі повного факторного експерименту  $2^3$ , тобто з кількістю дослідів  $N=8$  є малоефективним. Тому для вирішення більшості завдань неповні плани рекомендується будувати на планах типу  $2^4$  або  $2^5$ .

## 11. ПЛАНИ ЕКСПЕРИМЕНТІВ ДЛЯ ВСТАНОВЛЕННЯ НЕЛІНІЙНИХ ЗАЛЕЖНОСТЕЙ

Після реалізації експерименту типу  $2^k$  може бути отримана лише лінійна модель або неповна квадратична, якщо враховані взаємодії факторів. Усі квадратичні ефекти та ефекти вищих ступенів входять в оцінку вільного члена рівняння ( $b_0$ ). Тому у випадку значної нелінійності функції відгуку вільний член рівняння регресії приймає необґрунтовано велике, а іноді і беззмістовне, значення. У багатьох випадках нелінійних залежностей лінійні моделі можуть виявитися неадекватними. У такому разі постає питання про подальші дії дослідника. Про одне з можливих рішень вже йшлося у попередніх розділах – це використання менших інтервалів варіювання. Але таких захід можливий тільки в екстремальних експериментах. Коли метою досліджень є отримання математичної моделі поведінки біологічного об'єкта, такий прийом неприпустимий. Тоді для з'ясування типу нелінійної залежності застосовуються плани з більшою кількістю рівнів факторів, хоча це призводить до значного збільшення обсягу експерименту. Найчастіше використовуються експерименти з трьома або з п'ятьма рівнями факторів.

### Експерименти типу $3^k$

Знайомство з експериментами з трьома рівнями факторів краще за все почати з однофакторного експерименту. Припустимо, вже обраний основний рівень та інтервал варіювання фактора. В експерименті з двома рівнями факторів досліди виконувалися у двох точках (високий та низький рівень). З трьома рівнями потрібна ще третя точка. Вибір її однозначний – це центр експерименту (основний рівень). План такого експерименту наведено у табл. 49.

Таблиця 49

Матриця експерименту  $3^1$

Номер дослід	X1	$\bar{y}_i$
1	-1	$\bar{y}_1$
2	0	$\bar{y}_2$
3	+1	$\bar{y}_3$

Побудувавши такий план після виконання дослідів і проведення обчислень, можна отримати повну квадратичну модель поведінки об'єкта:

$$\bar{y} = b_0 + b_1 \cdot x_1 + b_{11} \cdot x_1^2$$

Тепер спробуємо розібратися у сутності коефіцієнта  $b_{11}$ . З попередніх розділів відомо, що коефіцієнт  $b_0$  дорівнює загальній середній комплексу і для однофакторного комплексу може бути визначений за формулою:

$$b_0 = \frac{\bar{y}_1 + \bar{y}_3}{2}.$$

Якщо лінійна модель адекватна, то в результаті виконання третього досліду в центрі експерименту повинне бути отримане значення, яке дорівнює загальній середній (у межах помилки досліду), тобто  $b_0$ . Це витікає з властивостей матриць планування (симетричність результатів відносно центру експерименту). Тому за умови лінійності моделі повинна виконуватися рівність:

$$b_0 - \bar{y}_2 = 0.$$

Якщо ця умова не виконується, то це свідчить про наявність квадратичного ефекту факторів.

Обчислення коефіцієнтів квадратичного рівняння вже розглядалося в розділі, присвяченому регресійному аналізу. Нагадаємо, що для знаходження коефіцієнтів рівняння другого ступеня потрібно вирішити систему з трьох рівнянь. При використанні кодованих змінних ( $-1$ ,  $0$  та  $+1$ ) формули для обчислень значно спрощуються:

$$b_0 = \bar{y}_2, \quad b_1 = \frac{\bar{y}_3 - \bar{y}_1}{2}, \quad b_{11} = \frac{\bar{y}_1 - 2 \cdot \bar{y}_2 + \bar{y}_3}{2}.$$

Тепер можна перейти до експериментів з більшою кількістю факторів. Так само, як і у експерименті з двома факторами, для введення до матриці планування нового фактора необхідно взяти вихідну матрицю і продублювати її з усіма рівнями нового фактора (табл. 50).

У таблиці у стовпці «взаємодії» показано до яких ефектів відносяться відповідні рядки матриці. У експерименті з трьома рівнями фактора можна визначити лінійні та квадратичні ефекти факторів, а також їх взаємодії першого та другого ступеня. Для полегшення визначення ступеня ефекту в таблицю введені перетворені кодовані значення ( $-1=0$ ,  $0=1$  та  $+1=2$ , стовпці 4 та 5). Враховуючи, що будь яке число в ступені 0 завжди дорівнює одиниці, якщо в рядку зустрічається 0, то фактор не входить у цю взаємодію. 1 відповідає першому ступеню фактора, а 2 – другому ступеню.



Матриця експерименту  $3^2$  і перехід до нових кодованих значень

Номер досліджу	Кодовані позначення		Нові кодовані значення		$\bar{y}_i$	Взаємодія
	$X_1$	$X_2$	$X_1$	$X_2$		
1	-1	-1	0	0	$\bar{y}_1$	-
2	0	-1	1	0	$\bar{y}_2$	$X_1$
3	+1	-1	2	0	$\bar{y}_3$	$X_1^2$
4	-1	0	0	1	$\bar{y}_4$	$X_2$
5	0	0	1	1	$\bar{y}_5$	$X_1 \cdot X_2$
6	+1	0	2	1	$\bar{y}_6$	$X_1^2 \cdot X_2$
7	-1	+1	0	2	$\bar{y}_7$	$X_2^2$
8	0	+1	1	2	$\bar{y}_8$	$X_1 \cdot X_2^2$
9	+1	+1	2	2	$\bar{y}_9$	$X_1^2 \cdot X_2^2$

Статистична обробка результатів експериментів  $3^k$  проводиться за тією ж схемою, що і експериментів  $2^k$ , тобто спочатку виконується первинна статистична обробка та перевірка вибірок на нормальність розподілу і однорідність, перевірка комплексу на однорідність дисперсій, обчислення коефіцієнтів рівняння регресії, перевірка значущості коефіцієнтів і перевірка адекватності рівняння регресії.

#### Обчислення та перевірка значущості коефіцієнтів рівняння регресії

Обчислення коефіцієнтів рівняння регресії для експериментів з трьома рівнями факторів можна вести з допомогою методу найменших квадратів. Але формули для цього методу в планах  $3^k$  набагато складніші, ніж у планах  $2^k$ . Тому краще використовувати алгоритм Йейтса, який дещо складніший від відповідного способу для обробки двофакторних експериментів. Схема алгоритму Йейтса для експерименту з трьома рівнями фактора та двома факторами наведена в табл. 51.

Коефіцієнти рівняння регресії отримують діленням контрастів на відповідні дільники, наведені в останньому стовпці таблиці. Якщо проводиться експеримент з більшою кількістю факторів, то ці коефіцієнти можна легко обчислити, виходячи з дільників для однофакторного комплексу. Ці дільники становлять **3**, **2** та **6**. Для двофакторного комплексу їх отримують множенням кожного вихідного множника на всі інші. Так, для отримання перших трьох дільників, необхідно перший дільник (**3**) помножити послідовно на **3**, **2** та **6**, друга трійка отримується множенням другого дільника на **3**, **2** та **6**, і остання трійка – множенням **6** на **3**, **2** та **6**. Якщо тепер значення

Алгоритм Йейтса для експерименту  $3^k$ 

Номер досліджу	$\bar{y}_i$	1-й крок	2-й крок (контрасти), $C_i^2$	Дільник, $d_i$
1	$\bar{y}_1$	$G_1 = \bar{y}_1 + \bar{y}_2 + \bar{y}_3$	$C_1 = G_1 + G_2 + G_3$	9
2	$\bar{y}_2$	$G_2 = \bar{y}_4 + \bar{y}_5 + \bar{y}_6$	$C_2 = G_4 + G_5 + G_6$	6
3	$\bar{y}_3$	$G_3 = \bar{y}_7 + \bar{y}_8 + \bar{y}_9$	$C_3 = G_7 + G_8 + G_9$	18
4	$\bar{y}$	$G_4 = \bar{y}_3 - \bar{y}_1$	$C_4 = G_3 - G_1$	6
5	$\bar{y}_5$	$G_5 = \bar{y}_6 - \bar{y}_4$	$C_5 = G_6 - G_4$	4
6	$\bar{y}_6$	$G_6 = \bar{y}_9 - \bar{y}_7$	$C_6 = G_9 - G_7$	12
7	$\bar{y}_7$	$G_7 = \bar{y}_1 - 2 \cdot \bar{y}_2 + \bar{y}_3$	$C_7 = G_1 - 2 \cdot G_2 + G_3$	18
8	$\bar{y}_8$	$G_8 = \bar{y}_4 - 2 \cdot \bar{y}_5 + \bar{y}_6$	$C_8 = G_4 - 2 \cdot G_5 + G_6$	12
9	$\bar{y}_9$	$G_9 = \bar{y}_7 - 2 \cdot \bar{y}_8 + \bar{y}_9$	$C_9 = G_7 - 2 \cdot G_8 + G_9$	36

дільників для експерименту з двома факторами знову послідовно помножити на 3, 2 та 6 будуть отримані 27 дільників для експерименту з трьома факторами і так далі. Ще одна особливість плану з трьома рівнями полягає в тому, що коефіцієнти розраховуються за перетвореними кодованими значенням (0, 1, 2). Тому рівняння регресії має такий вигляд:

$$\bar{y} = B_0 + B_1 \cdot x_1 + B_{11} \cdot P_1 + B_2 \cdot x_2 + B_{12} \cdot x_1 \cdot x_2 + B_{112} \cdot P_1 \cdot x_2 + B_{22} \cdot P_2 + B_{221} \cdot x_1 \cdot P_2 + B_{1122} \cdot P_1 \cdot P_2$$

в якому замість значення  $x_i^2$  стоїть певна величина  $P_i$ , зв'язана з основним кодованим значенням змінної наступним виразом:

$$P_i = 3 \cdot x_i^2 - 2$$

Для визначення значущості коефіцієнтів рівняння регресії з трьома та п'ятьма рівнях факторів застосовується інша процедура, ніж у експерименті з двома рівнями факторів. Застосовуючи наведені в розділі 9 формули, спочатку обчислюється дисперсія відтворення (дисперсія середнього значення)  $s_{\{\bar{y}\}}^2$ . Подальша оцінка значущості коефіцієнтів проводиться за допомогою критерію Фішера для кожного коефіцієнта окремо. Для цього використовується формула

$$F = \frac{C_i^2}{d_i \cdot s_{\{\bar{y}\}}^2}$$

де:  $C_i^2$  – контрасти обчислені за алгоритмом Йейтса;  
 $d_i$  – дільники, які використовуються в алгоритмі.

Знайдене значення порівнюється зі стандартним значенням критерію Фішера для обраного рівня значущості ( $\alpha$ ), кількості ступенів вільності  $I$  для кожного з контрастів і  $N$  для дисперсії відтворення. Якщо обчислене значення дорівнює або перевищує стандартне, коефіцієнт вважається значущим. Після перевірки значущості здійснюється перехід до основних кодованих змінних ( $-1, 0, +1$ ), для чого в рівняння замість значень  $P_i$  потрібно підставити наведену вище формулу, розкрити дужки та привести подібні. Останнім кроком розрахунків, який проводиться в разі потреби, є перетворення основних кодованих значень змінних у реальні з використанням формул, наведених у розділі 9.

### Плани експериментів типу $5^k$

Як і всі повні факторні експерименти, плани  $5^k$  будуються таким чином, що в них послідовно вивчаються всі рівні одного з факторів на фоні кожного з рівнів інших факторів, так що в цілому випробовуються всі можливі комбінації п'яти рівнів кожного фактора. Потрібно підкреслити, що в експериментах з п'ятьма рівнями факторів кількість необхідних дослідів дуже швидко збільшується з збільшенням кількості факторів. Тому плани з кількістю факторів більше трьох використовуються дуже рідко. Навіть дослідження трьох факторів, які варіюють на п'яти рівнях, потребують проведення 125 дослідів, що є досить трудомістким.

Для виконання експерименту з п'ятьма рівнями факторів обирається основний рівень (центр експерименту) та інтервал варіювання для кожного фактора. Досліди виконуються в точках центру експерименту та на відстанях, які дорівнюють одному ( $-1$  і  $+1$ ) та двом ( $-2$  і  $+2$ ) інтервалам варіювання факторів (табл. 52).

Обробка результатів такого експерименту проводиться за алгоритмом Йейтса. Для цього матриця розбивається на блоки з п'ятьма рядками в кожному і далі обчислюються суми:

Таблиця 52

Матриця експерименту  $5^2$

Номер дослідів	Кодовані значення		Нові кодовані значення		$\bar{y}_i$	Взаємодія
	X1	X2	X1	X2		
1	-2	-2	0	0	$\bar{y}_1$	–
2	-1	-2	1	0	$\bar{y}_2$	$X_1$
3	0	-2	2	0	$\bar{y}_3$	$X_1^2$
4	+1	-2	3	0	$\bar{y}_4$	$X_1^3$
5	+2	-2	4	0	$\bar{y}_5$	$X_1^4$

6	-2	-1	0	1	$\bar{y}_6$	$X_2$
7	-1	-1	1	1	$\bar{y}_7$	$X_1X_2$
8	0	-1	2	1	$\bar{y}_8$	$X_1^2X_2$
9	+1	-1	3	1	$\bar{y}_9$	$X_1^3X_2$
10	+2	-1	4	1	$\bar{y}_{10}$	$X_1^4X_2$
11	-2	0	0	2	$\bar{y}_{11}$	$X_2^2$
12	-1	0	1	2	$\bar{y}_{12}$	$X_1X_2^2$
13	0	0	2	2	$\bar{y}_{13}$	$X_1^2X_2^2$
14	+1	0	3	2	$\bar{y}_{14}$	$X_1^3X_2^2$
15	+2	0	4	2	$\bar{y}_{15}$	$X_1^4X_2^2$
16	-2	+1	0	3	$\bar{y}_{16}$	$X_2^3$
17	-1	+1	1	3	$\bar{y}_{17}$	$X_1X_2^3$
18	0	+1	2	3	$\bar{y}_{18}$	$X_1^2X_2^3$
19	+1	+1	3	3	$\bar{y}_{19}$	$X_1^3X_2^3$
20	+2	+1	4	3	$\bar{y}_{20}$	$X_1^4X_2^3$
21	-2	+2	0	4	$\bar{y}_{21}$	$X_2^4$
22	-1	+2	1	4	$\bar{y}_{22}$	$X_1X_2^4$
23	0	+2	2	4	$\bar{y}_{23}$	$X_1^2X_2^4$
24	+1	+2	3	4	$\bar{y}_{24}$	$X_1^3X_2^4$
25	+2	+2	4	4	$\bar{y}_{25}$	$X_1^4X_2^4$

$$G_{11} = 2 \cdot \bar{y}_1 - \bar{y}_2 - 2 \cdot \bar{y}_3 - \bar{y}_4 + 2 \cdot \bar{y}_5, \quad G_{12} = 2 \cdot \bar{y}_6 - \bar{y}_7 - 2 \cdot \bar{y}_8 - \bar{y}_9 + 2 \cdot \bar{y}_{10},$$

$$G_{13} = 2 \cdot \bar{y}_{11} - \bar{y}_{12} - 2 \cdot \bar{y}_{13} - \bar{y}_{14} + 2 \cdot \bar{y}_{15},$$

$$G_{14} = 2 \cdot \bar{y}_{16} - \bar{y}_{17} - 2 \cdot \bar{y}_{18} - \bar{y}_{19} + 2 \cdot \bar{y}_{20},$$

$$G_{15} = 2 \cdot \bar{y}_{21} - \bar{y}_{22} - 2 \cdot \bar{y}_{23} - \bar{y}_{24} + 2 \cdot \bar{y}_{25},$$

$$G_{16} = -\bar{y}_1 + 2 \cdot \bar{y}_2 - 2 \cdot \bar{y}_4 + \bar{y}_5, \quad G_{17} = -\bar{y}_6 + 2 \cdot \bar{y}_7 - 2 \cdot \bar{y}_9 + \bar{y}_{10},$$

$$G_{18} = -\bar{y}_{11} + 2 \cdot \bar{y}_{12} - 2 \cdot \bar{y}_{14} + \bar{y}_{15}, \quad G_{19} = -\bar{y}_{16} + 2 \cdot \bar{y}_{17} - 2 \cdot \bar{y}_{19} + \bar{y}_{20},$$

$$G_{20} = -\bar{y}_{21} + 2 \cdot \bar{y}_{22} - 2 \cdot \bar{y}_{24} + \bar{y}_{25}, \quad G_{21} = \bar{y}_1 - 4 \cdot \bar{y}_2 + 6 \cdot \bar{y}_3 - 4 \cdot \bar{y}_4 + \bar{y}_5,$$

$$G_{22} = \bar{y}_6 - 4 \cdot \bar{y}_7 + 6 \cdot \bar{y}_8 - 4 \cdot \bar{y}_9 + \bar{y}_{10},$$

$$G_{23} = \bar{y}_{11} - 4 \cdot \bar{y}_{12} + 6 \cdot \bar{y}_{13} - 4 \cdot \bar{y}_{14} + \bar{y}_{15},$$

$$G_{24} = \bar{y}_{16} - 4 \cdot \bar{y}_{17} + 6 \cdot \bar{y}_{18} - 4 \cdot \bar{y}_{19} + \bar{y}_{20},$$

$$G_{25} = \bar{y}_{21} - 4 \cdot \bar{y}_{22} + 6 \cdot \bar{y}_{23} - 4 \cdot \bar{y}_{24} + \bar{y}_{25}.$$

Якщо експеримент проводиться з двома або трьома факторами, то кількість кроків таких обчислень дорівнює кількості факторів. У кожному наступному кроці використовуються результати, отримані у попередньому. Результати останнього кроку є контрастами, з яких шляхом ділення на дільники отримують коефіцієнти рівняння регресії. Дільники для експерименту з п'ятьма рівнями фактора для однофакторного експерименту становлять **5, 10, 14, 10, 70**. Для більшої кількості факторів дільники отримують послідовним перемноженням один на одного, як в експерименті з трьома рівнями факторів. Для двофакторного експерименту ці дільники

становлять: 25, 50, 70, 50, 350, 50, 100, 140, 100, 700, 70, 140, 196, 140, 980, 50, 100, 140, 100, 700, 350, 700, 980, 700, 4900.

Далі перевіряється значущість коефіцієнтів рівняння регресії за формулами, які наведені для аналогічної перевірки експериментів з трьома рівнями факторів, і складається рівняння регресії. Для двофакторного експерименту це рівняння має такий вигляд:

$$\begin{aligned}
 y = & B_0 + B_1 \cdot X_1 + B_2 \cdot P_1^2 + B_3 \cdot P_1^3 + B_4 \cdot P_1^4 + B_5 \cdot X_2 + B_6 \cdot X_1 \cdot X_2 + \\
 & + B_7 \cdot P_1^2 \cdot X_2 + B_8 \cdot P_1^3 \cdot X_2 + B_9 \cdot P_1^4 \cdot X_2 + B_{10} \cdot P_2^2 + B_{11} \cdot X_1 \cdot P_2^2 + \\
 & + B_{12} \cdot P_1^2 \cdot P_2^2 + B_{13} \cdot P_1^3 \cdot P_2^2 + B_{14} \cdot P_1^4 \cdot P_2^2 + B_{15} \cdot P_2^3 + B_{16} \cdot X_1 \cdot P_2^3 + \\
 & + B_{17} \cdot P_1^2 \cdot P_2^3 + B_{18} \cdot P_1^3 \cdot P_2^3 + B_{19} \cdot P_1^4 \cdot P_2^3 + B_{20} \cdot P_2^4 + B_{21} \cdot X_1 \cdot P_2^4 + \\
 & + B_{22} \cdot P_1^2 \cdot P_2^4 + B_{23} \cdot P_1^3 \cdot P_2^4 + B_{24} \cdot P_1^4 \cdot P_2^4
 \end{aligned}$$

В отриманому рівнянні здійснюється заміна незалежних змінних, які мають другий, третій і четвертий ступінь ( $P_i^2$ ,  $P_i^3$ ,  $P_i^4$ ) на відповідні вирази:

$$P_i^2 = x_i^2 - 1, \quad P_i^3 = \frac{1}{3} \cdot (5 \cdot x_i^3 - 17 \cdot x_i), \quad P_i^4 = \frac{1}{12} (35 \cdot x_i^4 - 155 \cdot x_i^2 + 72)$$

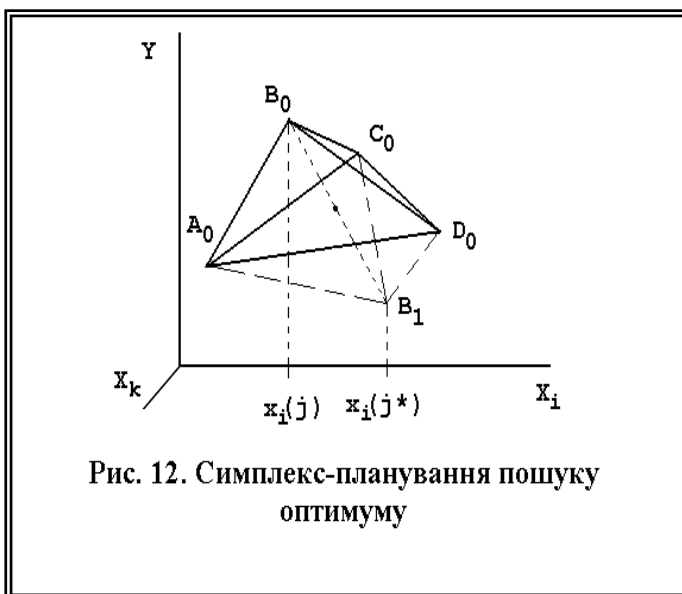
Останнім кроком статистичної обробки, якщо це необхідно, є заміна кодованих значень факторів на реальні, яка здійснюється за допомогою формул кодування змінних (розділ 9).

## 12. СИМПЛЕКС–ПЛАНУВАННЯ. ПОБУДОВА ПЛАНІВ. ПРИЙНЯТТЯ РІШЕНЬ ПРО НАПРЯМ РУХУ ДЛЯ ПОШУКУ ОПТИМУМУ

*Симплекс–планування* застосовується для пошуку оптимальних умов протікання біологічних процесів (екстремальні експерименти) і на початкових етапах потребує значно меншої експериментальної та обчислювальної роботи, ніж інші плани. *Симплексом* називають найпростішу, тобто з мінімальною кількістю елементів, геометричну фігуру, яка не містить криволінійних елементів. У просторі з двох вимірів (на площині) такою фігурою є правильний трикутник, а у просторі трьох вимірів – правильний тетраедр (трикутна піраміда, яка має всього чотири вершини, тоді як у куба їх шість).

Симплекс–планування починається з серії дослідів, кожний з яких проводиться з комбінацією значень факторів, які відповідають одній з вершин правильного симплексу (правильного трикутника, правильного тетраедру і т.д.). Значення факторів повинні вимірюватися не в натуральних одиницях, а в кодованих змінних, які не мають розмірності. В  $k$ -вимірному просторі симплекс має  $k+1$  вершин, тобто кількість дослідів буде дорівнювати кількості факторів ( $k$ ) плюс один.

Після реалізації  $k+1$  дослідів у вершинах симплекса, вибирають ту з вершин, дослід в якій має найгірший результат, і проводять новий дослід в точці, яка є дзеркальним відбиттям найгіршої точки відносно протилежної грані. Приклад такого плану для симплекса в просторі трьох вимірів наведено на рис. 12.



Припустимо, що в симплексі  $A_0B_0C_0D_0$  найгіршою виявилася точка  $B_0$ . Тоді потрібно провести новий дослід в точці  $B_1$ , яка разом з точками  $A_0C_0D_0$  утворить новий симплекс  $A_0B_1C_0D_0$ . Далі найгіршу точку обирають з нового симплекса (наприклад,  $D_0$ ) і проводять новий дослід з координатами  $D_1$ , який утворює симплекс  $A_0B_1C_0D_1$ . Проведення таких дослідів продовжується до тих пір, поки останній дослід дасть результати, які будуть гіршими за всі попередні.

Оскільки на кожному кроці симплекс–планування виключається найгірша точка, весь симплекс поступово переміщується в область оптимальних значень функції відгуку. Якщо нова точка виявилася гіршою в новому симплексі, то по-

трібно повернутися до попереднього симплекса і відбити другу за якістю вершину.

Чим ближче симплекс перекочується до оптимуму, тим частіше доведеться повертатися до попередніх вершин. Показником досягнення оптимуму є найгірші результати всіх останніх дослідів порівняно з попередніми. Для більшої точності визначення оптимуму всі останні досліді потрібно робити в декількох повторюваннях.

Таблиця 53

### Матриця симплекс-планування

№ вершини симплексу (№ дослідів, j)	№ координати вершини симплекса, який приймає значення від 1 до k (номер фактору, i)						
	1	2	3	4	...	k-1	k
1	$r_1$	$r_2$	$r_3$	$r_4$	...	$r_{k-1}$	$r_k$
2	$-R_1$	$r_2$	$r_3$	$r_4$	...	$r_{k-1}$	$r_k$
3	0	$-R_2$	$r_3$	$r_4$	...	$r_{k-1}$	$r_k$
4	0	0	$-R_3$	$r_4$	...	$r_{k-1}$	$r_k$
...	...	...	...	...	...	...	...
k	0	0	0	0	...	$-R_{k-1}$	$r_k$
k+1	0	0	0	0	...	0	$-R_k$

Вибір умов проведення дослідів

Якщо помістити початок координат у центрі рівнобічного трикутника (точка 0 на рис. 13.), а довжину ребра обрати рівною 1, то координати вершин симплекса будуть визначатися матрицею, наведеною в табл. 53.

Величини  $r_i$   $R_i$  відповідно дорівнюють:

$$r_i = \sqrt{\frac{1}{2 \cdot i \cdot (i+1)}} \text{ та}$$

$$R_i = \sqrt{\frac{1}{2 \cdot (i+1)}}$$

$i$  є, відповідно, радіусами



Рис. 13. Двомірний правильний симплекс

вписаної та описаної  $i$ -мірних сфер (рис. 13).

Під час планування експерименту зручніше користуватися для знаходження вершин симплекса змінними, вираженими в натуральних одиницях. Формула переходу від кодованих змінних до натуральних майже така сама як, і у факторних планах:

$$x_i = \frac{\tilde{x}_i - \tilde{x}_{i(0)}}{\lambda_{\tilde{x}_i}}$$

де:  $x_i$  – кодоване значення фактора;  
 $\tilde{x}_i$  – натуральне (реальне) значення фактора;  
 $\tilde{x}_{i0}$  – натуральне (реальне) значення фактора в центрі експерименту;  
 $\lambda_{\tilde{x}_i}$  – інтервал варіювання, який при симплекс-плануванні дорівнює:

$$\lambda_{\tilde{x}_i} = \tilde{x}_{i(\max)} - \tilde{x}_{i(\min)}.$$

З цих формул витікає, що:

$$x_i = \frac{\tilde{x}_i - \tilde{x}_{i(0)}}{\tilde{x}_{i(\max)} - \tilde{x}_{i(\min)}}$$

Виходячи з попередніх формул та матриці планування, можна обчислити значення координат вершин для максимального ( $r_i$ ) та мінімального ( $-R_i$ ) рівнів кожного фактора:

$$\frac{\tilde{x}_{i(\max)} - \tilde{x}_{i(0)}}{\tilde{x}_{i(\max)} - \tilde{x}_{i(\min)}} = r_i \quad \text{та} \quad \frac{\tilde{x}_{i(\min)} - \tilde{x}_{i(0)}}{\tilde{x}_{i(\max)} - \tilde{x}_{i(\min)}} = -R_i$$

Звідки нульове значення фактору дорівнює

$$\tilde{x}_{i(0)} = \frac{i \cdot \tilde{x}_{i(\max)} + \tilde{x}_{i(\min)}}{i + 1}$$

Відповідно, вершини початкового симплекса мають натуральні координати, які визначаються матрицею, поданою в табл. 54.

Для наочності наведемо приклад обчислення умов проведення симплекс-планування. Припустимо, що потрібно знайти найліпші умови росту грибів при різних температурах, рН середовища та концентрації цукру в середовищі. Визначимо межі змін цих показників (табл. 55).



Таблиця 54

## Матриця симплекс–планування в натуральних змінних

№ вершини симплексу (№ досліду, j)	№ координати вершини симплекса, який набуває значення від 1 до k (номер фактору, i)						
	1	2	3	4	...	k-1	k
1	X <sub>1(max)</sub>	X <sub>2(max)</sub>	X <sub>3(max)</sub>	X <sub>4(max)</sub>	...	X <sub>k-1(max)</sub>	X <sub>k(max)</sub>
2	X <sub>1(min)</sub>	X <sub>2(max)</sub>	X <sub>3(max)</sub>	X <sub>4(max)</sub>	...	X <sub>k-1(max)</sub>	X <sub>k(max)</sub>
3	X <sub>1(0)</sub>	X <sub>2(min)</sub>	X <sub>3(max)</sub>	X <sub>4(max)</sub>	...	X <sub>k-1(max)</sub>	X <sub>k(max)</sub>
4	X <sub>1(0)</sub>	X <sub>2(0)</sub>	X <sub>3(min)</sub>	X <sub>4(max)</sub>	...	X <sub>k-1(max)</sub>	X <sub>k(max)</sub>
...	...	...	...	...	...	...	...
k	X <sub>1(0)</sub>	X <sub>2(0)</sub>	X <sub>3(0)</sub>	X <sub>4(0)</sub>	...	X <sub>k-1(min)</sub>	X <sub>k(max)</sub>
k+1	X <sub>1(0)</sub>	X <sub>2(0)</sub>	X <sub>3(0)</sub>	X <sub>4(0)</sub>	...	X <sub>k-1(0)</sub>	X <sub>k(min)</sub>

Таблиця 55

## Граничні значення факторів для оптимізації росту гриба

Фактор	Нижня межа	Верхня межа
t°C (X <sub>1</sub> )	27,0	33,0
pH (X <sub>2</sub> )	6,8	7,2
C цукру (X <sub>3</sub> )	44,0	52,0

Скориставшись формулою, обчислюються нульові рівні для X<sub>1</sub>, X<sub>2</sub> (третій фактор у перших трьох дослідах буде приймати максимальне значення, а в четвертому – мінімальне, тому нульовий рівень обчислювати не потрібно):

$$x_{1(0)} = \frac{1 \cdot 33,0 + 27,0}{1 + 1} = 30,0$$

$$x_{2(0)} = \frac{2 \cdot 7,2 + 6,8}{2 + 1} = 7,06 \approx 7,1$$

Координати вершин визначаються з допомогою табл. 54. Ці координати вершин початкового симплекса будуть такими:

Вершини (№ дослід)	Координати вершин		
	t°C	pH	C цукру
1 (A <sub>0</sub> )	33,0	7,2	52,0
2 (B <sub>0</sub> )	27,0	7,2	52,0
3 (C <sub>0</sub> )	30,0	6,8	52,0
4 (D <sub>0</sub> )	30,0	7,1	44,0

Це будуть умови перших чотирьох дослідів симплекса. Перед кожним новим дослідом потрібно обчислити умови цього дослідів ( $j^*$ ), симетричні найгіршій вершині ( $j$ ) симплекса відносно протилежної грані. Для цього використовується наступна формула:

$$x_{i(j^*)} = \frac{2}{k} \cdot \sum_{l \neq j} x_{i(l)} - x_{i(j)}$$

Припустимо, що "найгіршою" точкою виявилася C<sub>0</sub>. Тоді потрібно поставити дослід в точці C<sub>1</sub>, симетричній точці C<sub>0</sub> відносно грані A<sub>0</sub>B<sub>0</sub>D<sub>0</sub>. Користуючись формулою, обчислимо координати точки C<sub>1</sub>:

$$t^{\circ}(C_1) = \frac{2}{3} \cdot (33,0 + 27,0 + 30,0) - 30,0 = 30,0$$

$$pH(C_1) = \frac{2}{3} \cdot (7,2 + 7,2 + 7,1) - 6,8 = 7,5$$

$$C(C_1) = \frac{2}{3} \cdot (52,0 + 52,0 + 44,0) - 52,0 = 46,7$$

Таким чином, новий дослід потрібно поставити із такими значеннями факторів: температура – 30°C, pH – 7,5 та концентрація цукру – 46,7 мг/г.

Наприкінці потрібно зазначити, що симплекс планування має ряд переваг під час пошуку оптимуму порівняно з факторним експериментом.

1. Числові значення параметра оптимізації не входять до будь яких обчислень, тому симплекс–планування можна застосовувати для оптимізації якісних характеристик об'єкта.
2. Пошук області оптимуму в симплекс–плануванні не пов'язаний з будь-якою поверхнею відгуку, що веде до економії дослідів порівняно з методом крутого сходження (повний факторний експеримент), особливо в межах близьких до оптимуму з нелінійними залежностями.
3. Правила симплекс–планування легко піддаються повній формалізації, що дозволяє автоматизувати пошук оптимуму за допомогою обчислювальної техніки.

Разом з тим симплекс–планування має і певні недоліки, які обмежують його використання.

1. Якщо не враховувати першої серії дослідів, всі інші досліди можна виконувати тільки послідовно після того, як отримано результати попереднього. Це визначає підвищені вимоги щодо дотримання умов проведення експериментів.
2. У симплекс–плануванні для отримання коефіцієнтів лінійної моделі необхідно було б робити додаткові досить складні розрахунки.
3. Хоча в симплекс–плануванні не враховується безпосередньо поверхня відгуку, це не означає, що метод зовсім не чутливий до функції відгуку. Експеримент краще проходить при великій крутизні та симетричності поверхні відгуку та мало ефективний за умови пологої кривої поверхні відгуку. Але форма залежності підпорядкована інтервалу варіювання факторів. Тому до їх вибору у симплекс–плануванні ставляться підвищені вимоги. Якщо у повному факторному експерименті їх можна виправити за результатами аналізу рівняння регресії, то в симплекс–плануванні неправильно обрані початкові рівні факторів ведуть до необхідності переробляти весь експеримент.

### 13. ЛАТИНСЬКІ КВАДРАТИ. ПОБУДОВА ПЛАНІВ. СТАТИСТИЧНА ОБРОБКА РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТУ.

Латинські квадрати є планами експериментів, за якими проводиться дисперсійний аналіз. Вони призначені для усунення неоднорідностей, які спотворюють вплив фактору на поведінку об'єкта досліджень.

Латинські квадрати розміром  $p \times p$  являють собою квадратну таблицю, у  $p^2$  клітинок якої записані  $p$  рівнів дослідної змінної так, щоб кожний рівень зустрічався у кожному рядку та у кожному стовпці тільки один раз. Ці плани призначені для усунення впливу двох факторів, що заважають, і відповідають обмеженням за рядками та стовпцями. Приклади латинських квадратів для  $p=2, 3, 4$  наведені в табл. 56.

Таблиця 56

Латинські квадрати для  $p=2, 3, 4$

p=2		p=3			p=4			
<i>A</i>	<i>B</i>	<i>C</i>	<i>B</i>	<i>A</i>	<i>D</i>	<i>A</i>	<i>C</i>	<i>B</i>
<i>B</i>	<i>A</i>	<i>A</i>	<i>C</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>D</i>	<i>C</i>
		<i>B</i>	<i>A</i>	<i>C</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>D</i>
					<i>C</i>	<i>D</i>	<i>B</i>	<i>A</i>

З поданих планів витікає, що в латинських квадратах поділ на однорідні блоки проводиться за стовпцями та рядками. Наприклад, латинський квадрат  $4 \times 4$  можна використовувати в експерименті з дослідження чотирьох рівнів фактора (доз лікарських препаратів, добрив і таке інше) на чотирьох різних об'єктах (тварини, рослини). Стовпці такого плану можуть відповідати різним об'єктам досліджень, які відрізняються між собою внаслідок своєї природної мінливості, а рядки – відповідному терміну проведення досліду. Так, якщо вивчалася дія інсуліну на вміст цукру в крові тварин, то літерами *A, B, C, D* позначаються дози препарату. Кожний стовпчик відповідає тварині, яка залучена до експерименту і може відрізнитися від іншої початковим вмістом цукру в крові; а кожний рядок відбиває день проведення досліду, що також може якимось чином вплинути на тварин. Останні два фактори (об'єкт досліджень та термін проведення експерименту) заважають встановленню необхідної дози лікарського препарату, оскільки у звичайному дисперсійному аналізі будуть підвищувати випадкову дисперсію.

Для забезпечення рандомізації експерименту квадрат необхідного розміру обирають випадковим способом зі спеціальних таблиць. Кількість

латинських квадратів залежить від розмірів квадрата і для  $p > 3$  є досить великою. Так, для квадрату  $4 \times 4$  існує 576 матриць латинських квадратів.

Під час плануванні експерименту рядки та стовпці використовуються для позначення рівнів двох факторів. Ці два фактори утворюють факторний експеримент  $p^2$  для джерел неоднорідностей (табл. 57, фактори *A* та *B*), на який накладено латинський квадрат  $p \times p$ , утворений рівнями основного фактора, дія якого вивчається (фактор *C*). Латинський квадрат є частиною плану, однак у плануванні експерименту весь план прийнято називати латинським квадратом.

Таблиця 57

### Латинський квадрат $4 \times 4$

Фактор А	Фактор В			
	<i>b</i> <sub>1</sub>	<i>b</i> <sub>2</sub>	<i>b</i> <sub>3</sub>	<i>b</i> <sub>4</sub>
<i>a</i> <sub>1</sub>	<i>c</i> <sub>1</sub>	<i>c</i> <sub>2</sub>	<i>c</i> <sub>3</sub>	<i>c</i> <sub>4</sub>
<i>a</i> <sub>2</sub>	<i>c</i> <sub>2</sub>	<i>c</i> <sub>3</sub>	<i>c</i> <sub>4</sub>	<i>c</i> <sub>1</sub>
<i>a</i> <sub>3</sub>	<i>c</i> <sub>3</sub>	<i>c</i> <sub>4</sub>	<i>c</i> <sub>1</sub>	<i>c</i> <sub>2</sub>
<i>a</i> <sub>4</sub>	<i>c</i> <sub>4</sub>	<i>c</i> <sub>1</sub>	<i>c</i> <sub>2</sub>	<i>c</i> <sub>3</sub>

### Статистична обробка результатів експерименту

Використовуючи латинські квадрати, статистичний аналіз базують на припущенні про адитивність впливу досліджуваних факторів і може бути помилковим, якщо між ними в дійсності існує взаємодія. Розмір квадрата також відіграє суттєве значення у точності висновків. Тому не рекомендується в досліджах користуватися квадратами з  $p < 4$ . У дисперсійному аналізі латинського квадрата також велике значення має випадковий вибір квадрата з множини його трансформацій. Якщо квадрат обрано випадково, то можливі відхилення в оцінках моделі можуть розглядатися як одна із складових випадкової дисперсії. Слід враховувати, що латинський квадрат не є звичайною моделлю з припущенням нормальності, для якої помилки незалежні, а їх дисперсії рівні.

Схеми дисперсійного аналізу залежать від наявності або відсутності повторювань у межах кожного рівня фактора у кожній клітині квадрата. Ми розглянемо тільки спосіб статистичної обробки безповторних експериментів. Для проведення аналізів потрібно знайти суми результатів експериментів окремо за рядками, стовпцями та літерами і загальні суми дат та їх квадратів. Формули для розрахунків такі:

$$\text{Сума квадратів дат} - \sum_{ijk} y_{ijk}^2 = y_{a_1 b_1 c_1}^2 + y_{a_1 b_2 c_2}^2 + \dots;$$

сума дат для всього комплексу –  $T = \sum y_{ijk}$  ;  
сума дат за рядками –  $T_A = \sum y_{ijk}$  ;  
сума дат за стовпцями –  $T_B = \sum y_{jik}$  ;  
сума дат за літерами (рівням фактора) –  $T_C = \sum y_{kij}$  .

Далі отримані суми використовують для визначення сум квадратів відхилень:

загальна –  $SS_y = \sum_{ijk} y_{ijk}^2 - \frac{T^2}{N}$ , де  $N$  – кількість дослідів в комплексі;

за фактором  $A$  (рядок) –  $SS_A = \frac{\sum T_A^2}{n} - \frac{T^2}{N}$  ( $n$  – кількість дослідів в рядку);

за фактором  $B$  (стовпчик) –  $SS_B = \frac{\sum T_B^2}{n} - \frac{T^2}{N}$  ( $n$  – кількість дослідів у стовпчику);

за фактором  $C$  (літери) –  $SS_C = \frac{\sum T_C^2}{n} - \frac{T^2}{N}$  ( $n$  – кількість дослідів для кожної літери);

випадкову –  $SS_z = SS_y - SS_A - SS_B - SS_C$  .

Для всіх формул значення  $N$  та  $n$  залежить від розміру квадрата. Так, для квадрата  $4 \times 4$   $N=16$ , а  $n=4$ . У загальному вигляді ці показники можна визначити за формулами:

$$N = p^2 \text{ та } n = p .$$

Далі обчислюються ступені вільності:

$$f_y = N - 1, \quad f_A = f_B = f_C = p - 1, \quad f_z = (p - 1) \cdot (p - 2)$$

і середні квадрати відхилень:

$$S_A^2 = \frac{SS_A}{p - 1}, \quad S_B^2 = \frac{SS_B}{p - 1}, \quad S_C^2 = \frac{SS_C}{p - 1}, \quad S_z^2 = \frac{SS_z}{(p - 1) \cdot (p - 2)} .$$

Останнім етапом є розрахунок критерію Фішера і визначення вірогідності впливу факторів. Ці розрахунки ведуться за формулою:

$$F_u = \frac{S_u^2}{S_z^2}, \text{ де } u \text{ може приймати значення } A, B, \text{ або } C .$$

Умовою вірогідності впливу є перевищення обчисленого значення  $F$  над стандартною величиною критерію Фішера.

Таблиця 58

**Результати експерименту щодо внесення фосфорного добрива.**

Час вне- сення	Номер рослини				Т <sub>В</sub>	Доза	Т <sub>С</sub>
	1	2	3	4			
1	В - 24	С - 46	Д - 34	А - 48	152	А	227
2	Д - 33	А - 58	В - 57	С - 60	208	В	160
3	А - 57	Д - 26	С - 60	В - 45	188	С	212
4	С - 46	В - 34	А - 64	Д - 47	191	Д	140
Т <sub>А</sub>	160	164	215	200	Т=739		

$$SS_y = \sum y_{ijk}^2 - \frac{T^2}{N} = 24^2 + \dots + 47^2 - \frac{739^2}{16} = 2448,44$$

$$SS_A = \frac{\sum T_A^2}{n} - \frac{T^2}{N} = \frac{160^2 + \dots + 200^2}{4} - \frac{739^2}{16} = 547,69$$

$$SS_B = \frac{\sum T_B^2}{n} - \frac{T^2}{N} = \frac{152^2 + \dots + 191^2}{4} - \frac{739^2}{16} = 415,69$$

$$SS_C = \frac{\sum T_C^2}{n} - \frac{T^2}{N} = \frac{227^2 + \dots + 140^2}{4} - \frac{739^2}{16} = 1285,69$$

$$SS_z = SS_y - SS_A - SS_B - SS_C = 2448,44 - 547,69 - 415,69 - 1285,69 = 199,37$$

Наприклад, на чотирьох рослинах одного виду вивчалися різні дози фосфорного добрива, яке вносили з періодичністю у два тижні. Як результативну ознаку використовували приріст рослин, який вимірювали перед наступним внесенням добрива. Ґрунт у всіх дослідах був однаковий, тому до факторів, які заважають аналізу, відносилися тільки індивідуальні особливості рослин та час внесення добрив. Результати дослідів та статистична обробка подані в табл. 58, 59.

Результати дисперсійного аналізу даних табл. 57

Джерело варіювання	SS <sub>i</sub>	f	S <sub>i</sub> <sup>2</sup>	S <sub>i</sub> <sup>2</sup> / S <sub>z</sub> <sup>2</sup>
Загальне	2448,44	15	163,22	–
Рядки (час внесення)	547,69	3	182,56	5,49
Стовпці (рослина)	415,69	3	138,56	4,17
Літери (доза добрив)	1285,69	3	428,56	12,90
Випадкове	199,37	6	33,22	–

Латинські квадрати призначені для випробування декількох рівнів одного фактора. Для дослідження двох або більшої кількості використовуються греко–латинські квадрати та гіпер–греко–латинські квадрати. У греко–латинських квадратах для позначення рівнів факторів використовуються літери латинського (перший фактор) та грецького (другий фактор) алфавітів, а у гіпер–греко–латинських квадратах для позначення рівнів третього фактора використовують цифри. У цих випадках потрібно стежити, щоб у рядках і стовпцях не повторювалися не тільки самі літери або цифри, а і їх сполучення (табл. 60).

Приклад греко–латинського квадрата 4x4

Aα	Bδ	Cβ	Dγ
Bγ	Aβ	Dδ	Cα
Cδ	Dα	Aγ	Bβ
Dβ	Cγ	Bα	Aδ



**ДОДАТКИ**  
**(таблиці стандартних значень)**

Таблиця 1

Стандартні значення критерію  $t$  Ст'юдента

Кількість ступенів вільності $(f)$	Рівні значущості ( $\alpha$ ) (двобічне обмеження)			Кількість ступенів вільності $(f)$	Рівні значущості ( $\alpha$ ) (двобічне обмеження)		
	$0,05$	$0,01$	$0,001$		$0,05$	$0,01$	$0,001$
<i>1</i>	12,71	63,66	636,62	<i>18</i>	2,10	2,88	3,92
<i>2</i>	4,30	9,92	31,60	<i>19</i>	2,09	2,86	3,88
<i>3</i>	3,18	5,84	12,94	<i>20</i>	2,09	2,85	3,85
<i>4</i>	2,78	4,60	8,61	<i>21</i>	2,08	2,83	3,82
<i>5</i>	2,57	4,03	6,86	<i>22</i>	2,07	2,82	3,79
<i>6</i>	2,45	3,71	5,96	<i>23</i>	2,07	2,81	3,77
<i>7</i>	2,36	3,50	5,40	<i>24</i>	2,06	2,80	3,74
<i>8</i>	2,31	3,36	5,04	<i>25</i>	2,06	2,79	3,72
<i>9</i>	2,26	3,25	4,78	<i>26</i>	2,06	2,78	3,71
<i>10</i>	2,23	3,17	4,59	<i>27</i>	2,05	2,77	3,69
<i>11</i>	2,20	3,11	4,49	<i>28</i>	2,05	2,76	3,66
<i>12</i>	2,18	3,05	4,32	<i>29</i>	2,05	2,76	3,66
<i>13</i>	2,16	3,01	4,22	<i>30</i>	2,04	2,75	3,65
<i>14</i>	2,14	2,98	4,14	<i>40</i>	2,02	2,70	3,55
<i>15</i>	2,13	2,95	4,07	<i>60</i>	2,00	2,66	3,46
<i>16</i>	2,12	2,92	4,01	<i>120</i>	1,98	2,62	3,37
<i>17</i>	2,11	2,90	3,96	$\infty$	1,96	2,58	3,29
	<i>0,025</i>	<i>0,005</i>	<i>0,0005</i>		<i>0,025</i>	<i>0,005</i>	<i>0,0005</i>
	Рівні значущості ( $\alpha$ ) (однобічне обмеження)				Рівні значущості ( $\alpha$ ) (однобічне обмеження)		

Таблиця 2 а

Критичні значення відношення R/S для оцінки нормальності розподілу  
(для ймовірностей 0,025 – 0,10)

Об'єм вибірки n	Нижні границі			Верхні границі		
	Ймовірності помилки					
	0,025	0,050	0,100	0,100	0,050	0,025
3	1,745	1,758	1,782	1,997	1,999	2,000
4	1,930	1,980	2,040	2,409	2,429	2,439
5	2,090	2,150	2,220	2,712	2,753	2,782
6	2,220	2,280	2,370	2,949	3,012	3,056
7	2,330	2,400	2,490	3,143	3,222	3,282
8	2,430	2,500	2,590	3,308	3,399	3,741
9	2,510	2,590	2,680	3,449	3,552	3,634
10	2,590	2,670	2,760	3,570	3,685	3,777
11	2,660	2,740	2,840	3,680	3,800	3,903
12	2,720	2,800	2,900	3,780	3,910	4,020
13	2,780	2,860	2,960	3,870	4,000	4,120
14	2,830	2,920	3,020	3,950	4,090	4,210
15	2,880	2,970	3,070	4,020	4,170	4,290
16	2,930	3,010	3,120	4,090	2,240	4,370
17	2,970	3,060	3,170	4,150	4,310	4,440
18	3,010	3,100	3,210	4,210	4,370	4,510
19	3,050	3,140	3,250	4,270	4,430	4,570
20	3,090	3,180	3,290	4,320	4,490	4,630
25	3,240	3,340	3,450	4,530	4,710	4,870
30	3,370	3,470	3,590	4,700	4,890	5,060
35	3,480	3,580	3,700	4,840	5,040	5,210
40	3,570	3,670	3,790	4,960	5,160	5,340
45	3,660	3,750	3,880	5,060	5,260	5,450
50	3,730	3,830	3,950	5,140	5,350	5,540
55	3,800	3,900	4,020	5,220	5,430	5,630
60	3,860	3,960	4,080	5,290	5,510	5,700
65	3,910	4,010	4,140	5,350	5,570	5,770
70	3,960	4,060	4,190	5,410	5,630	5,830
75	4,010	4,110	4,240	5,460	5,680	5,880
80	4,050	4,160	4,280	5,510	5,730	5,930
85	4,090	4,200	4,330	5,560	5,780	5,980
90	4,130	4,240	4,360	5,600	5,820	6,030
95	4,170	4,270	4,400	5,640	5,860	6,070
100	4,210	4,310	4,440	5,680	5,900	6,110
150	4,480	4,590	4,720	5,960	6,180	6,390
200	4,680	4,780	4,900	6,150	6,390	6,600
500	5,250	5,370	5,490	6,720	6,940	7,150
1000	5,680	5,790	5,920	7,110	7,330	7,540

Таблиця 2 б

Критичні значення відношення R/S для оцінки нормальності розподілу  
(для ймовірностей 0,000 – 0,010)

Об'єм вибірки n	Нижні границі			Верхні границі		
	Ймовірності помилки					
	0,000	0,005	0,010	0,010	0,005	0,000
3	1,732	1,735	1,737	2,000	2,000	2,000
4	1,732	1,830	1,870	2,445	2,447	2,449
5	1,826	1,980	2,020	2,803	2,813	2,828
6	1,826	2,110	2,150	3,095	3,115	3,162
7	1,871	2,220	2,260	3,338	3,369	3,465
8	1,871	2,310	2,350	3,543	3,585	3,742
9	1,897	2,390	2,440	3,720	3,772	3,935
10	1,897	2,460	2,510	3,875	3,935	4,000
11	1,915	2,530	2,580	4,012	4,079	4,472
12	1,915	2,590	2,640	4,134	4,208	4,690
13	1,927	2,640	2,700	4,244	4,325	4,899
14	1,927	2,700	2,750	4,340	4,431	5,099
15	1,936	2,740	2,800	4,440	4,530	5,292
16	1,936	2,790	2,840	4,520	4,620	5,477
17	1,944	2,830	2,880	4,600	4,700	5,657
18	1,944	2,870	2,920	4,670	4,780	5,831
19	1,949	2,900	2,960	4,740	4,850	6,000
20	1,949	2,940	2,990	4,800	4,910	6,164
25	1,961	3,090	3,150	5,060	5,190	6,930
30	1,966	3,210	3,270	5,260	5,400	7,620
35	1,972	3,320	3,380	5,420	5,570	8,250
40	1,975	3,410	3,470	5,560	5,710	8,830
45	1,978	3,490	3,550	5,670	5,830	9,380
50	1,980	3,560	3,620	5,770	5,930	9,900
55	1,982	3,620	3,690	5,860	6,020	10,390
60	1,983	3,680	3,750	5,940	6,100	10,860
65	1,985	3,740	3,800	6,010	6,170	11,310
70	1,986	3,790	3,850	6,070	6,240	11,750
75	1,987	3,830	3,900	6,130	6,300	12,170
80	1,987	3,880	3,940	6,180	6,350	12,570
85	1,988	3,920	3,990	6,230	6,400	12,960
90	1,989	3,960	4,020	6,270	6,450	13,340
95	1,990	3,990	4,060	6,320	6,490	13,710
100	1,990	4,030	4,100	6,360	6,530	14,070
150	1,993	4,320	4,380	6,640	6,820	17,260
200	1,995	4,530	4,590	6,840	7,010	19,950
500	1,998	5,060	5,130	7,420	7,600	31,590
1000	1,999	5,500	5,570	7,800	7,990	44,700

Таблиця 3

Коефіцієнти  $a_{n-i+1}$  для перевірки нормальності розподілу за методом  $W$ -критерію

Номер пари ( $i$ )	Об'єм вибірки ( $n$ )					
	3	4	5	6	7	8
1	0,7071	0,6872	0,6646	0,6431	0,6233	0,6052
2		0,1677	0,2413	0,2806	0,3031	0,3164
3				0,0875	0,1401	0,1743
4						0,0561

Номер пари ( $i$ )	Об'єм вибірки ( $n$ )					
	9	10	11	12	13	17
1	0,5888	0,5739	0,5601	0,5475	0,5359	0,5251
2	0,3244	0,3291	0,3315	0,3325	0,3325	0,3318
3	0,1976	0,2141	0,2260	0,2347	0,2412	0,2460
4	0,0947	0,1224	0,1429	0,1586	0,1707	0,1802
5		0,0399	0,0695	0,0922	0,1099	0,1240
6				0,0303	0,0539	0,0727
7						0,0240

Таблиця 4

Мінімальні значення критерію Вілсона ( $W$ ), які відповідають ймовірностям нормального розподілу ( $P$ )

Об'єм вибірки ( $n$ )	Ймовірності нормального розподілу ( $P$ )				
	1%	2%	5%	10%	50%
	Рівень значущості ( $\alpha$ )				
	0,01	0,02	0,05	0,1	0,5
3	0,753	0,756	0,767	0,789	0,959
4	0,687	0,707	0,748	0,792	0,936
5	0,686	0,715	0,762	0,806	0,927
6	0,713	0,743	0,788	0,826	0,927
7	0,730	0,760	0,803	0,838	0,928
8	0,749	0,778	0,818	0,851	0,932
9	0,764	0,791	0,829	0,59	0,935
10	0,781	0,806	0,842	0,869	0,938
11	0,792	0,817	0,850	0,876	0,940
12	0,805	0,828	0,859	0,883	0,943
13	0,814	0,837	0,866	0,889	0,945
14	0,825	0,846	0,874	0,895	0,947
15	0,835	0,855	0,881	0,901	0,950

Таблиця 5

Стандартні значення критерію  $t$  для перевірки вибірки на однорідність

Об'єм вибірки (n)	Рівень значущості (α)		Об'єм вибірки (n)	Рівень значущості (α)	
	0,05	0,01		0,05	0,01
5	3,04	5,04	20	2,145	2,932
6	2,78	4,36	25	2,105	2,852
7	2,62	3,96	30	2,079	2,802
8	2,51	3,71	35	2,061	2,768
9	2,43	3,54	40	2,048	2,742
10	2,37	3,41	45	2,038	2,722
11	2,33	3,31	50	2,030	2,707
12	2,29	3,23	60	2,018	2,683
13	2,26	3,17	70	2,009	2,667
14	2,24	3,12	80	2,003	2,655
15	2,22	3,08	90	1,998	2,646
16	2,20	3,04	100	1,994	2,639
17	2,18	3,01	∞	1,960	2,576
18	2,17	2,98			

Примітка: для обчислення критичних значень  $t$ , які не подані в таблиці використовується формула:  $t = t_{(>100)} + 100 \cdot \frac{t_{(100)} - t_{(>100)}}{n}$ .

Таблиця 6

Критерій для відкидання крайніх варіант ряду ( $\tau_\alpha$ )

n	Рівень значущості		n	Рівень значущості		n	Рівень значущості	
	0,05	0,01		0,05	0,01		0,05	0,01
5	1,92	1,97	21	2,80	3,11	80	3,33	3,70
6	2,07	2,16	22	2,82	3,13	90	3,37	3,74
7	2,18	2,31	23	2,84	3,16	100	3,40	3,77
8	2,27	2,43	24	2,86	3,18	120	3,46	3,83
9	2,35	2,53	25	2,88	3,20	150	3,53	3,90
10	2,41	2,62	26	2,90	3,22	200	3,61	3,98
11	2,47	2,69	27	2,91	3,24	300	3,73	4,09
12	2,52	2,75	28	2,93	3,26	400	3,80	4,17
13	2,56	2,81	29	2,94	3,28	500	3,87	4,24
14	2,60	2,86	30	2,96	3,29	600	3,92	4,28
15	2,64	2,90	35	3,02	3,36	700	3,96	4,32
16	2,67	2,94	40	3,03	3,42	800	3,99	4,35
17	2,70	2,98	45	3,12	3,48	900	4,02	4,38
18	2,73	3,02	50	3,16	3,52	1000	4,05	4,41
19	2,75	3,05	60	3,22	3,58	1500	4,14	4,50
20	2,78	3,08	70	3,28	3,64	2000	4,21	4,56

Таблиця 7

$$\text{Критичні значення } \tau' = \frac{y_{(n)} - y_{(n-1)}}{y_{(n)} - y_{(2)}} \text{ та } \tau' = \frac{y_{(2)} - y_{(1)}}{y_{(n-1)} - y_{(1)}}$$

n	Рівні значущості		n	Рівні значущості		n	Рівні значущості	
	0,05	0,01		0,05	0,01		0,05	0,01
4	0,991	0,955	13	0,520	0,410	22	0,414	0,320
5	0,916	0,807	14	0,502	0,395	23	0,407	0,314
6	0,805	0,689	15	0,486	0,381	24	0,400	0,309
6	0,740	0,610	16	0,472	0,369	25	0,394	0,304
8	0,683	0,554	17	0,462	0,359	26	0,389	0,299
9	0,635	0,512	18	0,449	0,349	27	0,383	0,295
10	0,597	0,477	19	0,439	0,341	28	0,378	0,291
11	0,566	0,450	20	0,430	0,334	29	0,374	0,287
12	0,541	0,428	21	0,421	0,327	30	0,369	0,283

Таблиця 8

Значення  $\rho^{(n)}=S/R^{(n)}$  для оцінки середнього квадрату відхилень за середнім розмахом відхилень  $R^{(n)}$

n`	$\rho^{(n)}$	n`	$\rho^{(n)}$	n`	$\rho^{(n)}$
2	0,886	6	0,395	10	0,325
3	0,591	7	0,370	11	0,315
4	0,486	8	0,351	12	0,307
5	0,430	9	0,337	13	0,294

Таблиця 9

Критичні значення критерію Кохрена перевірки однорідності дисперсій для  $\alpha=0,05$

Ступені вільності (f)	Кількість дисперсій, що порівнюються (k)								
	2	3	4	5	6	7	8	9	10
1	0,999	0,967	0,907	0,841	0,781	0,727	0,680	0,639	0,602
2	0,975	0,871	0,768	0,684	0,616	0,561	0,516	0,478	0,445
3	0,939	0,798	0,684	0,598	0,532	0,480	0,438	0,403	0,373
4	0,906	0,746	0,629	0,544	0,480	0,431	0,391	0,358	0,331
5	0,877	0,707	0,589	0,506	0,445	0,397	0,359	0,329	0,303
6	0,853	0,677	0,560	0,478	0,418	0,373	0,336	0,307	0,283
7	0,833	0,653	0,537	0,456	0,398	0,354	0,319	0,290	0,267
8	0,816	0,633	0,518	0,439	0,382	0,338	0,304	0,277	0,254
9	0,801	0,617	0,502	0,424	0,368	0,326	0,293	0,266	0,244
10	0,788	0,603	0,488	0,412	0,357	0,315	0,283	0,257	0,235

Таблиця 10

Критичні значення критерію Хартлі ( $F$ ) для перевірки дисперсій на однорідність ( $\alpha=0,05$ )

Ступені вільності ( $f$ )	Кількість порівнюваних дисперсій ( $k$ )										
	2	3	4	5	6	7	8	9	10	11	12
2	39,00	87,50	142,00	202,00	266,00	333,00	403,00	475,00	550,00	626,00	704,00
3	15,40	27,80	39,20	50,70	62,00	72,90	83,50	93,90	104,00	114,00	124,00
4	9,60	15,50	20,60	25,20	29,50	33,60	37,50	41,10	44,60	48,00	51,40
5	7,15	10,80	13,70	16,30	18,70	20,80	22,90	24,70	26,50	28,20	29,90
6	5,82	8,38	10,40	12,10	13,70	15,00	16,30	17,50	18,60	19,70	20,70
7	4,99	6,94	8,44	9,70	10,80	11,80	12,70	13,50	14,30	15,10	15,80
8	4,43	6,00	7,18	8,12	9,03	9,78	10,50	11,10	11,70	12,20	12,70
9	4,03	5,34	6,31	7,11	7,80	8,41	8,95	9,45	9,91	10,30	10,70
10	3,72	4,85	5,67	6,34	6,92	7,42	7,86	8,28	8,66	9,01	9,34
12	3,28	4,16	4,79	5,30	5,72	6,09	6,42	6,72	7,00	7,25	7,48

Таблиця 11

Критичні значення відношення  $s^2_{max}/s^2_{min}$  для перевірки дисперсій за методом Леслі та Брауну ( $\alpha=0,05$ )

Ступені вільності ( $f$ )	Кількість порівнюваних дисперсій ( $k$ )										
	2	3	4	5	6	7	8	9	10	11	12
3	6,267	9,392	11,990	14,300	16,400	18,350	20,190	21,930	23,590	25,180	26,710
4	3,971	5,335	6,371	7,237	7,992	8,669	9,285	9,854	10,380	10,880	11,350
5	3,157	4,018	4,643	5,149	5,580	5,958	6,298	6,607	6,892	7,156	7,404
6	2,744	3,381	3,831	4,187	4,487	4,747	4,978	5,187	5,377	5,553	5,717
7	2,494	3,007	3,362	3,640	3,871	4,070	4,245	4,403	4,546	4,678	4,800
8	2,325	2,760	3,056	3,286	3,476	3,638	3,781	3,909	4,024	4,130	4,227
9	2,203	2,584	2,841	3,039	3,201	3,340	3,461	3,568	3,666	3,755	3,837
10	2,110	2,452	2,681	2,855	2,998	3,120	3,226	3,320	3,404	3,482	3,553
11	2,037	2,349	2,556	2,714	2,842	2,951	3,046	3,129	3,205	3,274	3,337
13	1,928	2,198	2,375	2,508	2,617	2,708	2,787	2,857	2,920	2,977	3,029
16	1,820	2,049	2,198	2,309	2,399	2,475	2,540	2,597	2,648	2,695	2,737

Критичні значення  $\chi^2$  для  $\alpha=0,05$ 

<i>f</i>	$\chi^2$	<i>f</i>	$\chi^2$	<i>f</i>	$\chi^2$
<i>1</i>	3,8415	<i>6</i>	12,5920	<i>11</i>	19,6750
<i>2</i>	5,9915	<i>7</i>	14,0670	<i>12</i>	21,0260
<i>3</i>	7,8147	<i>8</i>	15,5070	<i>13</i>	22,3620
<i>4</i>	9,4877	<i>9</i>	16,9190	<i>14</i>	23,6850
<i>5</i>	11,0700	<i>10</i>	18,3070	<i>15</i>	24,9960

Таблиця 13

Значення *Z*, які відповідають величинам коефіцієнта кореляції *r*

<i>r</i>	Соті частки коефіцієнта кореляції <i>r</i>									
	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
<i>0,0</i>	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
<i>0,1</i>	0,100	0,111	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
<i>0,2</i>	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
<i>0,3</i>	0,310	0,321	0,332	0,343	0,354	0,365	0,377	0,388	0,400	0,412
<i>0,4</i>	0,424	0,436	0,448	0,460	0,472	0,485	0,498	0,510	0,523	0,536
<i>0,5</i>	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,663	0,678
<i>0,6</i>	0,693	0,709	0,725	0,741	0,758	0,776	0,793	0,811	0,829	0,848
<i>0,7</i>	0,867	0,887	0,908	0,929	0,951	0,973	0,996	1,020	1,045	1,071
<i>0,8</i>	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
<i>0,9</i>	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647
<i>0,99</i>	2,647	2,700	2,759	2,826	2,903	2,995	3,106	3,250	3,453	3,800



Критичні значення  $F$  (критерію Фішера) для  $\alpha=0,05$ 

$f_2$	$f_1$											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161,00	200,00	216,00	225,00	230,00	234,00	237,00	239,00	241,00	242,00	243,00	244,00
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31

Продовження табл. 14

	1	2	3	4	5	6	7	8	9	10	11	12
<b>20</b>	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28
<b>22</b>	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23
<b>24</b>	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18
<b>26</b>	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15
<b>28</b>	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,15	2,12
<b>30</b>	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09
<b>34</b>	4,13	3,28	2,88	2,65	2,49	2,38	2,30	2,23	2,17	2,12	2,08	2,05
<b>38</b>	4,10	3,25	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02
<b>42</b>	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,02	1,99
<b>46</b>	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,14	2,09	2,04	2,00	1,97
<b>50</b>	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95
<b>60</b>	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,93
<b>70</b>	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,01	1,97	1,93	1,89
<b>80</b>	3,96	3,11	2,72	2,48	2,33	2,21	2,12	2,05	1,99	1,95	1,91	1,89
<b>100</b>	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85
<b>150</b>	3,91	3,06	2,67	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82
<b>200</b>	3,89	3,04	2,65	2,41	2,26	2,14	2,05	1,98	1,92	1,87	1,83	1,80
<b>400</b>	3,86	3,02	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,78
<b>1000</b>	3,85	3,00	2,61	2,38	2,22	2,10	2,02	1,95	1,89	1,84	1,80	1,76
$\infty$	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,79	1,75

Продовження табл. 14

<i>f<sub>1</sub></i>	<i>f<sub>2</sub></i>											
	<i>14</i>	<i>16</i>	<i>20</i>	<i>24</i>	<i>30</i>	<i>40</i>	<i>50</i>	<i>75</i>	<i>100</i>	<i>200</i>	<i>500</i>	$\infty$
<i>1</i>	245,00	246,00	248,00	249,00	250,00	251,00	252,00	253,00	253,00	254,00	254,00	254,00
<i>2</i>	19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,50	19,50
<i>3</i>	8,71	8,69	8,66	8,64	8,62	8,60	8,58	8,57	8,56	8,54	8,54	8,53
<i>4</i>	5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,66	5,65	5,64	5,63
<i>5</i>	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36
<i>6</i>	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67
<i>7</i>	3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23
<i>8</i>	3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93
<i>9</i>	3,02	2,98	2,93	2,90	2,86	2,82	2,80	2,77	2,76	2,73	2,72	2,71
<i>10</i>	2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54
<i>11</i>	2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40
<i>12</i>	2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,36	2,35	2,32	2,31	2,30
<i>13</i>	2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,26	2,24	2,22	2,21
<i>14</i>	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
<i>15</i>	2,43	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,12	2,10	2,08	2,07
<i>16</i>	2,37	2,33	2,28	2,24	2,20	2,16	2,13	2,09	2,07	2,04	2,02	2,01
<i>17</i>	2,33	2,29	2,23	2,19	2,15	2,11	2,08	2,04	2,02	1,99	1,97	1,96
<i>18</i>	2,29	2,25	2,19	2,15	2,11	2,07	2,04	2,00	1,98	1,95	1,93	1,92
<i>19</i>	2,26	2,21	2,15	2,11	2,07	2,02	2,00	1,96	1,94	1,91	1,90	1,88
<i>20</i>	2,23	2,18	2,12	2,08	2,04	1,99	1,96	1,92	1,90	1,87	1,85	1,84
<i>22</i>	2,18	2,13	2,07	2,03	1,98	1,93	1,91	1,87	1,84	1,81	1,80	1,78
<i>24</i>	2,13	2,09	2,02	1,98	1,94	1,89	1,86	1,82	1,80	1,76	1,74	1,73

Закінчення табл. 14

$f_1$	$f_2$											
	14	16	20	24	30	40	50	75	100	200	500	$\infty$
26	2,09	2,05	1,99	1,95	1,90	1,85	1,82	1,78	1,76	1,72	1,70	1,69
28	2,07	2,02	1,96	1,91	1,87	1,81	1,78	1,75	1,72	1,69	1,67	1,65
30	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,69	1,66	1,64	1,62
34	2,00	1,95	1,89	1,84	1,80	1,74	1,71	1,67	1,64	1,61	1,59	1,57
38	1,96	1,92	1,85	1,80	1,76	1,71	1,67	1,63	1,60	1,57	1,54	1,53
42	1,94	1,89	1,82	1,78	1,73	1,68	1,64	1,60	1,57	1,54	1,51	1,49
46	1,91	1,87	1,80	1,75	1,71	1,65	1,62	1,57	1,54	1,51	1,48	1,46
50	1,90	1,85	1,78	1,74	1,69	1,63	1,60	1,55	1,52	1,48	1,46	1,44
60	1,86	1,81	1,75	1,70	1,65	1,59	1,56	1,50	1,48	1,44	1,41	1,39
70	1,84	1,79	1,72	1,67	1,62	1,56	1,53	1,47	1,45	1,40	1,37	1,35
80	1,82	1,77	1,70	1,65	1,60	1,54	1,51	1,45	1,42	1,38	1,35	1,32
100	1,79	1,75	1,68	1,63	1,57	1,51	1,48	1,42	1,39	1,34	1,30	1,28
150	1,76	1,71	1,64	1,59	1,54	1,47	1,44	1,37	1,34	1,29	1,25	1,22
200	1,74	1,69	1,62	1,57	1,52	1,45	1,42	1,35	1,32	1,26	1,22	1,19
400	1,72	1,67	1,60	1,54	1,49	1,42	1,38	1,32	1,28	1,22	1,16	1,13
1000	1,70	1,65	1,58	1,53	1,47	1,41	1,36	1,30	1,26	1,19	1,13	1,08
$\infty$	1,69	1,64	1,57	1,52	1,46	1,40	1,35	1,28	1,24	1,17	1,11	1,00

*Примітка:  $f_1$  – кількість ступенів вільності для більшої дисперсії (у дисперсійному аналізі – для факторіального середнього квадрата відхилень);  $f_2$  – кількість ступенів вільності для меншої дисперсії (у дисперсійному аналізі – для випадкового середнього квадрата відхилень).*

Таблиця 15

Стандартні значення однобічного критерію Даннета для  $\alpha=0,05$ 

f	Кількість середніх, виключаючи контроль								
	1	2	3	4	5	6	7	8	9
5	2,02	2,44	2,68	2,85	2,98	3,08	3,16	3,24	3,30
6	1,94	2,34	2,56	2,71	2,83	2,92	3,00	3,07	3,12
7	1,89	2,27	2,48	2,62	2,73	2,82	2,89	2,95	3,01
8	1,86	2,22	2,42	2,55	2,66	2,74	2,81	2,87	2,92
9	1,83	2,18	2,37	2,50	2,60	2,68	2,75	2,81	2,86
10	1,81	2,15	2,34	2,47	2,56	2,64	2,70	2,76	2,81
11	1,80	2,13	2,31	2,44	2,53	2,60	2,67	2,72	2,77
12	1,78	2,11	2,29	2,41	2,50	2,58	2,64	2,69	2,74
13	1,77	2,09	2,27	2,39	2,48	2,55	2,61	2,66	2,71
14	1,76	2,08	2,25	2,37	2,46	2,53	2,59	2,64	2,69
15	1,75	2,07	2,24	2,36	2,44	2,51	2,57	2,62	2,67
16	1,75	2,06	2,23	2,34	2,43	2,50	2,56	2,61	2,65
17	1,74	2,05	2,22	2,33	2,42	2,49	2,54	2,59	2,64
18	1,73	2,04	2,21	2,32	2,41	2,48	2,53	2,58	2,62
19	1,73	2,03	2,20	2,31	2,40	2,47	2,52	2,57	2,61
20	1,72	2,03	2,19	2,30	2,39	2,46	2,51	2,56	2,60
24	1,71	2,01	2,17	2,28	2,36	2,43	2,48	2,53	2,57
30	1,70	1,99	2,15	2,25	2,33	2,40	2,45	2,50	2,54
40	1,68	1,97	2,13	2,23	2,31	2,37	2,42	2,47	2,51
60	1,67	1,95	2,10	2,21	2,28	2,35	2,39	2,44	2,48
120	1,66	1,93	2,08	2,18	2,26	2,32	2,37	2,41	2,45
$\infty$	1,64	1,92	2,06	2,16	2,23	2,29	2,34	2,38	2,42

Таблиця 16

Стандартні значення двобічного критерію Даннета для  $\alpha=0,05$ 

f	Кількість середніх, виключаючи контроль									
	1	2	3	4	5	6	7	8	9	10
5	2,57	3,03	3,29	3,48	3,62	3,73	3,82	3,90	3,97	4,03
6	2,45	2,86	3,10	3,26	3,39	3,49	3,57	3,64	3,71	3,76
7	2,36	2,75	2,97	3,12	3,24	3,33	3,41	3,47	3,53	3,58
8	2,31	2,67	2,88	3,02	3,13	3,22	3,29	3,35	3,41	3,46
9	2,26	2,61	2,81	2,95	3,05	3,14	3,20	3,26	3,32	3,36
10	2,23	2,57	2,76	2,89	2,99	3,07	3,14	3,19	3,24	3,29
11	2,20	2,53	2,72	2,84	2,94	3,02	3,08	3,14	3,19	3,23
12	2,18	2,50	2,68	2,81	2,90	2,98	3,04	3,09	3,14	3,18
13	2,16	2,48	2,65	2,78	2,87	2,94	3,00	3,06	3,10	3,14
14	2,14	2,46	2,63	2,75	2,84	2,91	2,97	3,02	3,07	3,11
15	2,13	2,44	2,61	2,73	2,82	2,89	2,95	3,00	3,04	3,08
16	2,12	2,42	2,59	2,71	2,80	2,87	2,92	2,97	3,02	3,06
17	2,11	2,41	2,58	2,69	2,78	2,85	2,90	2,95	3,00	3,03
18	2,10	2,40	2,56	2,68	2,76	2,83	2,89	2,94	2,98	3,10
19	2,09	2,39	2,55	2,66	2,75	2,81	2,87	2,92	2,96	3,00
20	2,09	2,38	2,54	2,65	2,73	2,80	2,86	2,90	2,95	2,98
24	2,06	2,35	2,51	2,61	2,70	2,76	2,81	2,86	2,90	2,94
30	2,04	2,32	2,47	2,58	2,66	2,72	2,77	2,82	2,86	2,89
40	2,02	2,29	2,44	2,54	2,62	2,68	2,73	2,77	2,81	2,85
60	2,00	2,27	2,41	2,51	2,58	2,64	2,69	2,73	2,77	2,80
120	1,98	2,24	2,38	2,47	2,55	2,60	2,65	2,69	2,73	2,76
$\infty$	1,96	2,21	2,35	2,44	2,51	2,57	2,61	2,65	2,69	2,72

Верхня границя ст'юдентизованого розмаху для методів Т'юкі та Н'юмана-Кейлса

f	Кількість порівнюваних середніх										
	2	3	4	5	6	7	8	9	10	11	12
1	18,00	27,00	32,80	37,10	40,40	43,10	45,40	47,40	49,10	50,60	52,00
2	6,08	8,33	9,80	10,90	11,70	12,40	13,00	13,50	14,00	14,40	14,70
3	4,50	5,91	6,82	7,50	8,04	8,48	8,85	9,18	9,46	9,72	9,95
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83	8,03	8,21
5	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	6,99	7,17	7,32
6	3,46	4,34	4,90	5,30	5,63	5,90	6,12	6,32	6,49	6,65	6,79
7	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,30	6,43
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,05	6,18
9	3,20	3,95	4,41	4,76	5,02	5,24	5,43	5,59	5,74	5,87	5,98
10	3,15	3,88	4,33	4,65	4,91	5,12	5,30	5,46	5,60	5,72	5,83
11	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49	5,61	5,71
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,39	5,51	5,61
13	3,06	3,73	4,15	4,45	4,69	4,88	5,05	5,19	5,32	5,43	5,53
14	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,36	5,46
15	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,20	5,31	5,40
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15	5,26	5,35
17	2,98	3,63	4,02	4,30	4,52	4,70	4,86	4,99	5,11	5,21	5,31
18	2,97	3,61	4,00	4,28	4,49	4,67	4,82	4,96	5,07	5,17	5,27
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14	5,23
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01	5,11	5,20
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92	5,01	5,10
30	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	4,92	5,00
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,73	4,82	4,90
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81
120	2,80	3,36	3,68	3,92	4,10	4,24	4,36	4,47	4,56	4,64	4,71
$\infty$	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62

Таблиця 18

Стандартні значення ст'юдентизованих розмахів для методу Дункана

<b>f</b>	<b>Кількість порівнюваних середніх,</b>								
	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
<b>1</b>	17,97	17,97	17,97	17,97	17,97	17,97	17,97	17,97	17,97
<b>2</b>	6,08	6,08	6,08	6,08	6,08	6,08	6,08	6,08	6,08
<b>3</b>	4,50	4,52	4,52	4,52	4,52	4,52	4,52	4,52	4,52
<b>4</b>	3,93	4,01	4,03	4,03	4,03	4,03	4,03	4,03	4,03
<b>5</b>	3,64	3,75	3,80	3,81	3,81	3,81	3,81	3,81	3,81
<b>6</b>	3,46	3,59	3,65	3,68	3,69	3,70	3,70	3,70	3,70
<b>7</b>	3,34	3,48	3,55	3,59	3,61	3,62	3,63	3,63	3,63
<b>8</b>	3,26	3,40	3,48	3,52	3,55	3,57	3,58	3,58	3,58
<b>9</b>	3,20	3,34	3,42	3,47	3,50	3,52	3,54	3,54	3,54
<b>10</b>	3,15	3,29	3,38	3,43	3,46	3,49	3,50	3,52	3,52
<b>11</b>	3,11	3,26	3,34	3,40	3,44	3,46	3,48	3,49	3,50
<b>12</b>	3,08	3,22	3,31	3,37	3,41	3,44	3,46	3,47	3,48
<b>13</b>	3,06	3,20	3,29	3,35	3,39	3,42	3,44	3,46	3,47
<b>14</b>	3,03	3,18	3,27	3,33	3,37	3,40	3,43	3,44	3,46
<b>15</b>	3,01	3,16	3,25	3,31	3,36	3,39	3,41	3,43	3,45
<b>16</b>	3,00	3,14	3,24	3,30	3,34	3,38	3,40	3,42	3,44
<b>17</b>	2,98	3,13	3,22	3,28	3,33	3,37	3,39	3,41	3,43
<b>18</b>	2,97	3,12	3,21	3,27	3,32	3,36	3,38	3,40	3,42
<b>19</b>	2,96	3,11	3,20	3,26	3,31	3,35	3,38	3,40	3,42
<b>20</b>	2,95	3,10	3,19	3,26	3,30	3,34	3,37	3,39	3,41
<b>24</b>	2,92	3,07	3,16	3,23	3,28	3,32	3,34	3,37	3,39
<b>30</b>	2,89	3,04	3,13	3,20	3,25	3,29	3,32	3,35	3,37
<b>40</b>	2,86	3,01	3,10	3,17	3,22	3,27	3,30	3,33	3,35
<b>60</b>	2,83	2,98	3,07	3,14	3,20	3,24	3,28	3,31	3,33
<b>120</b>	2,80	2,95	3,04	3,12	3,17	3,22	3,25	3,29	3,31
<b>∞</b>	2,77	2,92	3,02	3,09	3,15	3,19	3,23	3,26	3,29



Значення критерію  $T$  Уайта для  $\alpha=0,95$ 

Більша кількість спостережень	Менша кількість спостережень													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4			11											
5		6	11	17										
6		7	12	18	26									
7		7	13	20	27	36								
8	3	8	14	21	29	38	49							
9	3	8	15	22	31	40	51	63						
10	3	9	15	23	32	42	53	65	78					
11	4	9	16	24	34	44	55	68	81	96				
12	4	10	17	26	35	46	58	71	85	99	115			
13	4	10	18	27	37	48	60	73	88	103	119	137		
14	4	11	19	28	38	50	63	76	91	106	123	141	160	
15	4	11	20	29	40	52	65	79	94	110	127	145	164	185
16	4	12	21	31	42	54	67	82	97	114	151	150	169	
17	5	12	21	32	43	56	70	84	100	117	135	154		
18	5	13	22	33	45	58	72	87	103	121	139			
19	5	13	23	34	46	60	74	90	107	124				
20	5	14	24	35	48	62	77	93	110					
21	6	14	25	37	50	64	79	95						
22	6	15	26	38	51	66	82							
23	6	15	27	39	53	68								
24	6	16	28	40	55									
25	6	16	28	42										
26	7	17	29											
27	7	17												



Таблиця 20

Значення  $\Psi(p)=u(\Phi)$  – функції оберненої до інтегралу ймовірностей

$P \rightarrow$ ↓	0	1	2	3	4	5	6	7	8	9
0,00	$-\infty$	-3,09	-2,88	-2,75	-2,65	-2,58	-2,51	-2,46	-2,41	-2,37
0,01	-2,33	-2,29	-2,26	-2,23	-2,20	-2,17	-2,14	-2,12	-2,10	-2,07
0,02	-2,05	-2,03	-2,01	-2,00	-1,98	-1,96	-1,94	-1,93	-1,91	-1,90
0,03	-1,88	-1,87	-1,85	-1,84	-1,83	-1,81	-1,80	-1,79	-1,77	-1,76
0,04	-1,75	-1,74	-1,73	-1,72	-1,71	-1,70	-1,68	-1,67	-1,66	-1,65
0,05	-1,64	-1,64	-1,63	-1,62	-1,61	-1,60	-1,59	-1,58	-1,57	-1,56
0,06	-1,55	-1,55	-1,54	-1,53	-1,52	-1,51	-1,51	-1,50	-1,49	-1,48
0,07	-1,48	-1,47	-1,46	-1,45	-1,45	-1,44	-1,43	-1,43	-1,42	-1,41
0,08	-1,41	-1,40	-1,39	-1,39	-1,38	-1,37	-1,37	-1,36	-1,35	-1,35
0,09	-1,34	-1,33	-1,33	-1,32	-1,32	-1,31	-1,30	-1,30	-1,29	-1,29
0,10	-1,28	-1,28	-1,27	-1,26	-1,26	-1,25	-1,25	-1,24	-1,24	-1,23
0,11	-1,23	-1,22	-1,22	-1,21	-1,21	-1,20	-1,20	-1,19	-1,19	-1,18
0,12	-1,18	-1,17	-1,17	-1,16	-1,16	-1,15	-1,15	-1,14	-1,14	-1,13
0,13	-1,13	-1,12	-1,12	-1,11	-1,11	-1,10	-1,10	-1,09	-1,09	-1,09
0,14	-1,08	-1,08	-1,07	-1,07	-1,06	-1,06	-1,05	-1,05	-1,05	-1,04
0,15	-1,04	-1,03	-1,03	-1,02	-1,02	-1,02	-1,01	-1,01	-1,00	-1,00
0,16	-0,99	-0,99	-0,99	-0,98	-0,98	-0,97	-0,97	-0,97	-0,96	-0,96
0,17	-0,95	-0,95	-0,95	-0,94	-0,94	-0,93	-0,93	-0,93	-0,92	-0,92
0,18	-0,92	-0,91	-0,91	-0,90	-0,90	-0,90	-0,89	-0,89	-0,89	-0,88
0,19	-0,88	-0,87	-0,87	-0,87	-0,86	-0,86	-0,86	-0,85	-0,85	-0,85
0,20	-0,84	-0,84	-0,83	-0,83	-0,83	-0,82	-0,82	-0,82	-0,81	-0,81
0,21	-0,81	-0,80	-0,80	-0,80	-0,79	-0,79	-0,79	-0,78	-0,78	-0,78
0,22	-0,77	-0,77	-0,77	-0,76	-0,76	-0,76	-0,75	-0,75	-0,75	-0,74
0,23	-0,74	-0,74	-0,73	-0,73	-0,74	-0,72	-0,72	-0,72	-0,71	0,71
0,24	-0,71	-0,70	-0,70	-0,70	-0,69	-0,69	-0,69	-0,68	-0,68	-0,68
0,25	-0,67	-0,67	-0,67	-0,67	-0,66	-0,66	-0,66	-0,65	-0,65	-0,65
0,26	-0,64	-0,64	-0,64	-0,63	-0,63	-0,63	-0,63	-0,62	-0,62	-0,62
0,27	-0,61	-0,61	-0,61	-0,60	-0,60	-0,60	-0,59	-0,59	-0,59	-0,59
0,28	-0,58	-0,58	-0,58	-0,57	-0,57	-0,57	-0,57	-0,56	-0,56	-0,56
0,29	-0,55	-0,55	-0,55	-0,54	-0,54	-0,54	-0,54	-0,53	-0,53	-0,53
0,30	-0,52	-0,52	-0,52	-0,52	-0,51	-0,51	-0,51	-0,50	-0,50	0,50
0,31	-0,50	-0,49	-0,49	-0,49	-0,48	-0,48	-0,48	-0,48	-0,47	-0,47
0,32	-0,47	-0,46	-0,46	-0,46	-0,46	-0,45	-0,45	-0,45	-0,45	-0,44
0,33	-0,44	-0,44	-0,43	-0,43	-0,43	-0,43	-0,42	-0,42	-0,42	-0,42

Продовження табл. 20

<b>P→</b> ↓	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>0,34</b>	-0,41	-0,41	0,41	-0,40	-0,40	-0,40	-0,40	-0,39	-0,39	-0,39
<b>0,35</b>	-0,39	-0,38	-0,38	-0,38	-0,37	-0,37	-0,37	-0,37	-0,36	-0,36
<b>0,36</b>	-0,36	-0,36	-0,35	-0,35	-0,35	-0,35	-0,34	-0,34	-0,34	-0,33
<b>0,37</b>	-0,33	-0,33	-0,33	-0,32	-0,32	-0,32	-0,32	-0,31	-0,31	-0,31
<b>0,38</b>	-0,31	-0,30	-0,30	-0,30	-0,30	-0,29	-0,29	-0,29	-0,28	-0,28
<b>0,39</b>	-0,28	-0,28	-0,27	-0,27	-0,27	-0,27	-0,26	-0,26	-0,26	-0,26
<b>0,40</b>	-0,25	-0,25	-0,25	-0,25	-0,24	-0,24	-0,24	-0,24	-0,23	-0,23
<b>0,41</b>	-0,23	-0,23	-0,22	-0,22	-0,22	-0,21	-0,21	-0,21	-0,21	-0,20
<b>0,42</b>	-0,20	-0,20	-0,20	-0,19	-0,19	-0,19	-0,19	-0,19	-0,18	-0,18
<b>0,43</b>	-0,18	-0,17	-0,17	-0,17	-0,17	-0,16	-0,16	-0,16	-0,16	-0,15
<b>0,44</b>	-0,15	-0,15	-0,15	-0,14	-0,14	-0,14	-0,14	-0,13	-0,13	-0,13
<b>0,45</b>	-0,13	-0,12	-0,12	-0,12	-0,12	-0,11	-0,11	-0,11	-0,11	-0,10
<b>0,46</b>	-0,10	-0,10	-0,10	-0,09	-0,09	-0,09	-0,09	-0,08	-0,08	-0,08
<b>0,47</b>	-0,08	-0,07	-0,07	-0,07	-0,07	-0,06	-0,06	-0,06	-0,06	-0,05
<b>0,48</b>	-0,05	-0,05	-0,05	-0,04	-0,04	-0,04	-0,04	-0,03	-0,03	-0,03
<b>0,49</b>	-0,03	-0,02	-0,02	-0,02	-0,02	-0,01	-0,01	-0,01	-0,01	0,00
<b>0,50</b>	0,00	0,00	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,02
<b>0,51</b>	0,03	0,03	0,03	0,03	0,04	0,04	0,04	0,04	0,05	0,05
<b>0,52</b>	0,05	0,05	0,06	0,06	0,06	0,06	0,07	0,07	0,07	0,07
<b>0,53</b>	0,08	0,08	0,08	0,08	0,09	0,09	0,09	0,09	0,10	0,10
<b>0,54</b>	0,10	0,10	0,11	0,11	0,11	0,11	0,12	0,12	0,12	0,12
<b>0,55</b>	0,13	0,13	0,13	0,13	0,14	0,14	0,14	0,14	0,15	0,15
<b>0,56</b>	0,15	0,15	0,16	0,16	0,16	0,16	0,17	0,17	0,17	0,17
<b>0,57</b>	0,18	0,18	0,18	0,18	0,19	0,19	0,19	0,19	0,20	0,20
<b>0,58</b>	0,20	0,20	0,21	0,21	0,21	0,21	0,22	0,22	0,22	0,23
<b>0,59</b>	0,23	0,23	0,23	0,24	0,24	0,24	0,24	0,25	0,25	0,25
<b>0,60</b>	0,25	0,26	0,26	0,26	0,26	0,27	0,27	0,27	0,27	0,28
<b>0,61</b>	0,28	0,28	0,28	0,29	0,29	0,29	0,30	0,30	0,30	0,30
<b>0,62</b>	0,31	0,31	0,31	0,31	0,32	0,32	0,32	0,32	0,33	0,33
<b>0,63</b>	0,33	0,33	0,34	0,34	0,34	0,35	0,35	0,35	0,35	0,36
<b>0,64</b>	0,36	0,36	0,36	0,37	0,37	0,37	0,37	0,38	0,38	0,38
<b>0,65</b>	0,39	0,39	0,39	0,39	0,40	0,40	0,40	0,40	0,41	0,41
<b>0,66</b>	0,41	0,42	0,42	0,42	0,42	0,43	0,43	0,43	0,43	0,44
<b>0,67</b>	0,44	0,44	0,45	0,45	0,45	0,45	0,46	0,46	0,46	0,46
<b>0,68</b>	0,47	0,47	0,47	0,48	0,48	0,48	0,48	0,49	0,49	0,49
<b>0,69</b>	0,50	0,50	0,50	0,50	0,51	0,51	0,51	0,52	0,52	0,52

Закінчення табл. 20

<b>P→</b> ↓	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>0,70</b>	0,52	0,53	0,53	0,53	0,54	0,54	0,54	0,54	0,55	0,55
<b>0,71</b>	0,55	0,56	0,56	0,56	0,57	0,57	0,57	0,57	0,58	0,58
<b>0,72</b>	0,58	0,59	0,59	0,59	0,59	0,60	0,60	0,60	0,61	0,61
<b>0,73</b>	0,61	0,62	0,62	0,62	0,63	0,63	0,63	0,63	0,64	0,64
<b>0,74</b>	0,64	0,65	0,65	0,65	0,66	0,66	0,66	0,67	0,67	0,67
<b>0,75</b>	0,67	0,68	0,68	0,68	0,69	0,69	0,69	0,70	0,70	0,70
<b>0,76</b>	0,71	0,71	0,71	0,72	0,72	0,72	0,73	0,73	0,73	0,74
<b>0,77</b>	0,74	0,74	0,75	0,75	0,75	0,76	0,76	0,76	0,77	0,77
<b>0,78</b>	0,77	0,78	0,78	0,78	0,79	0,79	0,79	0,80	0,80	0,80
<b>0,79</b>	0,81	0,81	0,81	0,82	0,82	0,82	0,83	0,83	0,83	0,84
<b>0,80</b>	0,84	0,85	0,85	0,85	0,86	0,86	0,86	0,87	0,87	0,87
<b>0,81</b>	0,88	0,88	0,89	0,89	0,89	0,90	0,90	0,90	0,91	0,91
<b>0,82</b>	0,92	0,92	0,92	0,93	0,93	0,93	0,94	0,94	0,95	0,95
<b>0,83</b>	0,95	0,96	0,96	0,97	0,97	0,97	0,98	0,98	0,99	0,99
<b>0,84</b>	0,99	1,00	1,00	1,01	1,01	1,02	1,02	1,02	1,03	1,03
<b>0,85</b>	1,04	1,04	1,05	1,05	1,05	1,06	1,06	1,07	1,07	1,08
<b>0,86</b>	1,08	1,09	1,09	1,09	1,10	1,10	1,11	1,11	1,12	1,12
<b>0,87</b>	1,13	1,13	1,14	1,14	1,15	1,15	1,16	1,16	1,17	1,17
<b>0,88</b>	1,18	1,18	1,19	1,19	1,20	1,20	1,21	1,21	1,22	1,22
<b>0,89</b>	1,23	1,23	1,24	1,24	1,25	1,25	1,26	1,26	1,27	1,28
<b>0,90</b>	1,28	1,29	1,29	1,30	1,30	1,31	1,32	1,32	1,33	1,33
<b>0,91</b>	1,34	1,35	1,35	1,36	1,37	1,37	1,38	1,39	1,39	1,40
<b>0,92</b>	1,41	1,41	1,42	1,43	1,43	1,44	1,45	1,45	1,46	1,47
<b>0,93</b>	1,48	1,48	1,49	1,50	1,51	1,51	1,52	1,53	1,54	1,55
<b>0,94</b>	1,55	1,56	1,57	1,58	1,59	1,60	1,61	1,62	1,63	1,64
<b>0,95</b>	1,64	1,65	1,66	1,67	1,68	1,70	1,71	1,72	1,73	1,74
<b>0,96</b>	1,75	1,76	1,77	1,79	1,8-	1,81	1,83	1,84	1,85	1,87
<b>0,97</b>	1,88	1,90	1,91	1,93	1,94	1,96	1,98	2,00	2,01	2,03
<b>0,98</b>	2,05	2,07	2,10	2,12	2,14	2,17	2,20	2,23	2,26	2,29
<b>0,99</b>	2,33	2,37	2,41	2,46	2,51	2,58	2,65	2,75	2,88	3,09

Таблиця 21

Критичні значення  $X_\alpha$  (критерію Вен дер Вардена) для  $\alpha=0,05$

<b>n</b> <b>(n<sub>x</sub>+n<sub>y</sub>)</b>	<b>n<sub>x</sub>-n<sub>y</sub></b>			<b>n</b> <b>(n<sub>x</sub>+n<sub>y</sub>)</b>	<b>n<sub>x</sub>-n<sub>y</sub></b>		
	<b>0-1</b>	<b>2-3</b>	<b>4-5</b>		<b>0-1</b>	<b>2-3</b>	<b>4-5</b>
<b>6</b>	∞	∞	∞	<b>29</b>	4,78	4,76	4,72
<b>7</b>	∞	∞	∞	<b>30</b>	4,88	4,87	4,84
<b>8</b>	2,40	2,30	∞	<b>31</b>	4,97	4,95	4,91
<b>9</b>	2,38	2,20	∞	<b>32</b>	5,07	5,06	5,03
<b>10</b>	2,60	2,49	2,30	<b>33</b>	5,15	5,13	5,10
<b>11</b>	2,72	2,58	2,40	<b>34</b>	5,25	5,24	5,21
<b>12</b>	2,86	2,79	2,68	<b>35</b>	5,33	5,31	5,28
<b>13</b>	2,96	2,91	2,78	<b>36</b>	5,42	5,41	5,38
<b>14</b>	3,11	3,06	3,00	<b>37</b>	5,50	5,48	5,45
<b>15</b>	3,24	3,19	3,06	<b>38</b>	5,59	5,58	5,55
<b>16</b>	3,39	3,36	3,28	<b>39</b>	5,67	5,65	5,62
<b>17</b>	3,49	3,44	3,36	<b>40</b>	5,75	5,74	5,72
<b>18</b>	3,63	3,60	3,53	<b>41</b>	5,83	5,81	5,79
<b>19</b>	3,73	3,69	3,61	<b>42</b>	5,91	5,90	5,88
<b>20</b>	3,86	3,84	3,78	<b>43</b>	5,99	5,97	5,95
<b>21</b>	3,96	3,92	3,85	<b>44</b>	6,08	6,06	6,04
<b>22</b>	4,08	4,06	4,01	<b>45</b>	6,14	6,12	6,10
<b>23</b>	4,18	4,15	4,08	<b>46</b>	6,23	6,21	6,19
<b>24</b>	4,29	4,27	4,23	<b>47</b>	6,29	6,27	6,25
<b>25</b>	4,39	4,36	4,30	<b>48</b>	6,36	6,35	6,34
<b>26</b>	4,50	4,48	4,44	<b>49</b>	6,43	6,42	6,39
<b>27</b>	4,59	4,56	4,51	<b>50</b>	6,50	6,50	6,48
<b>28</b>	4,69	4,68	4,64				

Таблиця 22

Критичні значення кількості серій  $S_{05}$ 

$n_x$	$n_y$																
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
6	3	3	3														
7	3	3	4	4													
8	3	3	4	4	5												
9	3	4	4	5	5	6											
10	3	4	5	5	6	6	6										
11	3	4	5	5	6	6	7	7									
12	4	4	5	6	6	7	7	8	8								
13	4	4	5	6	6	7	8	8	9	9							
14	4	5	5	6	7	7	8	8	9	9	10						
15	4	5	6	6	7	8	8	9	9	10	10	11					
16	4	5	6	6	7	8	8	9	10	10	11	11	11				
17	4	5	6	7	7	8	9	9	10	10	11	11	12	12			
18	4	5	6	7	8	8	9	10	10	11	11	12	12	13	13		
19	4	5	6	7	8	8	9	10	10	11	12	12	13	13	14	14	
20	4	5	6	7	8	9	9	10	11	11	12	12	13	13	14	14	15

Таблиця 23

Критичні значення кількості знаків  $Z$  (що менше зустрічаються)

$n$	$\alpha$		$n$	$\alpha$		$n$	$\alpha$		$n$	$\alpha$	
	0,05	0,01		0,05	0,01		0,05	0,01		0,05	0,01
8	1	1	20	6	4	32	10	9	44	16	14
9	2	1	21	6	5	33	11	9	45	16	14
10	2	1	22	6	5	34	11	10	46	16	14
11	2	1	23	7	5	35	12	10	47	17	15
12	3	2	24	7	6	36	12	10	48	17	15
13	3	2	25	8	6	37	13	11	49	18	16
14	3	2	26	8	7	38	13	11	50	18	16
15	4	3	27	8	7	39	13	12	51	19	16
16	4	3	28	9	7	40	14	12	52	19	17
17	5	3	29	9	8	41	14	12	53	19	17
18	5	4	30	10	8	42	15	13	54	20	18
19	5	4	31	10	8	43	15	13	55	20	18

Продовження табл. 23

n	$\alpha$		n	$\alpha$		n	$\alpha$		n	$\alpha$	
	0,05	0,01		0,05	0,01		0,05	0,01		0,05	0,01
56	21	18	68	26	23	80	31	29	92	37	34
57	21	19	69	26	24	81	32	29	93	37	34
58	22	19	70	27	24	82	32	29	94	38	35
59	22	20	71	27	25	83	33	30	95	38	35
60	22	20	72	28	25	84	33	30	96	38	35
61	23	21	73	28	26	85	33	31	97	39	36
62	23	21	74	29	26	86	34	31	98	39	36
63	24	21	75	29	26	87	34	32	99	40	37
64	24	22	76	29	27	88	35	32	100	40	37
65	25	22	77	30	27	89	35	32			
66	25	23	78	30	28	90	36	33			
67	26	23	79	31	28	91	36	33			

Таблиця 24

Критичні значення  $Z$  Вілсона (для супряжених ознак)

Кіль- кість па- рних спосте- режень	Рівень значущості		Кіль- кість па- рних спосте- режень	Рівень значущості		Кіль- кість па- рних спосте- режень	Рівень значущості	
	0,05	0,01		0,05	0,01		0,05	0,01
6	1	–	13	18	11	20	53	39
7	3	–	14	22	14	21	60	44
8	5	1	15	26	17	22	67	50
9	7	3	16	31	21	23	74	56
10	9	4	17	36	24	24	82	62
11	12	6	18	41	29	25	90	69
12	15	8	19	47	33			

Таблиця 25



Стандартні значення критерію Краскела–Уолліса ( $h_\alpha$ ) для  $\alpha=0,05$

Об'єми вибірок				$h_\alpha$	Об'єми вибірок					$h_\alpha$
$n_1$	$n_2$	$n_3$	$h_\alpha$		$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	
3	2	2	4,714	6	5	3			5,602	
3	3	1	5,143	6	5	4			5,661	
3	3	2	5,361	6	5	5			5,729	
3	3	3	5,600	6	6	2			5,410	
4	2	2	5,333	6	6	3			5,625	
4	3	1	5,208	6	6	4			5,724	
4	3	2	5,444	6	6	5			5,765	
4	3	3	5,791	6	6	6			5,801	
4	4	1	4,967	7	7	7			5,819	
4	4	2	5,455	8	8	8			5,805	
4	4	3	5,598							
4	4	4	5,692	2	2	2	2		6,167	
5	2	1	5,000	3	2	2	2		6,333	
5	2	2	5,160	3	3	2	2		6,527	
5	3	1	4,960	3	3	3	2		6,727	
5	3	2	5,251	3	3	3	3		7,000	
5	3	3	5,648	4	2	2	2		6,545	
5	4	1	4,986	4	3	2	2		6,621	
5	4	2	5,273	4	3	3	2		6,795	
5	4	3	5,656	4	3	3	3		6,984	
5	4	4	5,657	4	4	2	2		6,731	
5	5	1	5,127	4	4	3	2		6,874	
5	5	2	5,338	4	4	3	3		7,038	
5	5	3	5,705	4	4	4	2		6,957	
5	5	4	5,666	4	4	4	3		7,142	
5	5	5	5,780	4	4	4	4		7,235	
6	2	2	5,345							
6	3	2	5,348	2	2	2	2	2	7,418	
6	3	3	5,615	3	2	2	2	2	7,682	
6	4	2	5,340	3	3	2	2	2	7,910	
5	4	3	5,610	3	3	3	2	2	8,044	
6	4	4	5,681	3	3	3	3	2	8,200	
6	5	2	5,338	3	3	3	3	3	8,333	

Таблиця 26

Критичні значення для множинних порівнянь всіх  $k$  вибірок з рівними об'ємами ( $y_{(\alpha,k,n)}$ ) для  $\alpha=0,05$

n	k												
	3	4	5	6	7	8	9	10	11	12	13	14	15
2	–	12	15	19	23	26	30	33	37	41	45	48	51
3	16	22	29	36	43	50	–	–	–	–	–	–	–
4	24	34	45	–	–	–	–	–	–	–	–	–	–
5	33	–	–	–	–	–	–	–	–	–	–	–	–
6	43	–	–	–	–	–	–	–	–	–	–	–	–

Таблиця 27

Критичні значення ( $q$ ) розмаху  $k$  випадкових незалежних величин

k	Рівень значущості		k	Рівень значущості		k	Рівень значущості	
	0,05	0,01		0,05	0,01		0,05	0,01
2	3,643	2,772	9	5,078	4,387	16	5,493	4,845
3	4,120	3,314	10	5,157	4,474	17	5,535	4,891
4	4,403	3,633	11	5,227	4,552	18	5,574	4,934
5	4,603	3,858	12	5,290	4,622	19	5,611	4,974
6	4,575	4,030	13	5,348	4,685	20	5,645	5,012
7	4,882	4,170	13	5,400	4,743	22	5,709	5,081
8	4,987	4,286	15	5,448	4,796	24	5,766	5,144

Таблиця 28

Критичні значення ( $y^*$ ) односторонніх множинних порівнянь з контролем для  $k=3, n_1=n_2=n_3=n$

Рівень значущості	Об'єм вибірок (n)				
	2	3	4	5	6
0,05	8	14	20	28	36
0,01	–	16	26	36	47

Таблиця 29

Критичні значення ( $y^{**}$ ) двобічних множинних порівнянь з контролем для  
 $k=3, n_1=n_2=n_3=n$

Рівень значущості	Об'єм вибірок (n)				
	2	3	4	5	6
0,05	8	15	23	32	41
0,01	–	17	27	38	50

Таблиця 30

Критичні значення ( $m$ ) для максимуму абсолютного значення  $k-1$  випадкових величин із загальним коефіцієнтом кореляції  $\rho=1/2$

k-1	Рівень значущості		k-1	Рівень значущості		k-1	Рівень значущості	
	0,01	0,05		0,01	0,05		0,01	0,05
1	2,58	1,96	6	3,11	2,57	11	3,27	2,74
2	2,79	2,21	7	3,15	2,61	12	3,29	2,77
3	2,92	2,35	8	3,19	2,65	13	3,35	2,83
4	3,00	2,44	9	3,22	2,69	20	3,42	2,91
5	3,06	2,51	10	3,25	2,72			

Таблиця 31

Критичні значення ( $m$ ) для максимуму абсолютного значення  $k-1$  випадкових величин із загальним коефіцієнтом кореляції  $\rho=n/(b+n)$  для  $\alpha=0,05$

$\rho$	k-1											
	1	2	3	4	5	6	7	8	9	10	11	12
0,100	1,7	2,0	2,2	2,3	2,4	2,4	2,5	2,5	2,6	2,6	2,6	2,7
0,125	1,7	2,0	2,2	2,3	2,4	2,4	2,5	2,5	2,6	2,6	2,6	2,7
0,200	1,7	2,0	2,2	2,3	2,4	2,4	2,5	2,5	2,6	2,6	2,6	2,7
0,250	1,7	2,0	2,2	2,3	2,3	2,4	2,5	2,5	2,5	2,6	2,6	2,6
0,300	1,7	2,0	2,1	2,3	2,3	2,4	2,5	2,5	2,5	2,6	2,6	2,6
1/3	1,7	2,0	2,1	2,2	2,3	2,4	2,4	2,5	2,5	2,6	2,6	2,6

Продовження табл. 31

$\rho$	$k-1$											
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>	<i>12</i>
<b>0,375</b>	1,7	2,0	2,1	2,2	2,3	2,4	2,4	2,5	2,5	2,5	2,6	2,6
<b>0,400</b>	1,7	2,0	2,1	2,2	2,3	2,4	2,4	2,5	2,5	2,5	2,6	2,6
<b>0,500</b>	1,7	2,0	2,1	2,2	2,3	2,3	2,4	2,4	2,5	2,5	2,5	2,6
<b>0,600</b>	1,7	1,9	2,1	2,2	2,2	2,3	2,3	2,4	2,4	2,4	2,5	2,5
<b>0,625</b>	1,7	1,9	2,1	2,2	2,2	2,3	2,3	2,4	2,4	2,4	2,5	2,5
<b>2/3</b>	1,7	1,9	2,1	2,1	2,2	2,3	2,3	2,3	2,4	2,4	2,4	2,4
<b>0,700</b>	1,7	1,9	2,1	2,1	2,2	2,2	2,3	2,3	2,3	2,4	2,4	2,4
<b>0,750</b>	1,7	1,9	2,0	2,1	2,2	2,2	2,2	2,3	2,3	2,3	2,3	2,4
<b>0,800</b>	1,7	1,9	2,0	2,1	2,1	2,2	2,2	2,2	2,3	2,3	2,3	2,3
<b>0,875</b>	1,7	1,9	1,9	2,0	2,0	2,1	2,1	2,1	2,2	2,2	2,2	2,2
<b>0,900</b>	1,7	1,8	1,9	2,0	2,0	2,0	2,1	2,1	2,1	2,1	2,1	2,2

## ПОЗНАЧЕННЯ, прийняті в навчальному посібнику

### *Вибіркові показники*

- $n$  – об'єм вибірки (кількість спостережень, вимірів і таке інше);
- $f$  – кількість ступенів вільності;
- $y$  – вибіркова дата (значення спостереження, виміру та ін.);
- $\bar{y}$  – вибіркова середня арифметична;
- $H$  – вибіркова сума квадратів відхилень;
- $s^2$  – вибірковий середній квадрат відхилень (дисперсія, варіанса);
- $s$  – вибіркове середнє квадратичне відхилення;
- $s_{\bar{y}}$  – помилка репрезентативності (помилка вибіркового середнього);
- $R$  – розмах варіювання (різниця між максимальним та мінімальним значеннями у вибірці).

### **Показники, що відносяться до комплексів вибірок**

- $x$  – незалежна змінна (в дисперсійному аналізі може використовуватися для позначення рівня фактора);
- $N$  – об'єм комплексу (загальна кількість спостережень, вимірів та ін.);
- $k$  – загальна кількість вибірок у комплексі (в однофакторному дисперсійному аналізі дорівнює кількості градацій фактора), у багатфакторному регресійному аналізі – кількість факторів;
- $g$  – кількість градацій фактора;
- $SS$  – сума квадратів відхилень;
- $S^2$  – середній квадрат відхилень;
- $D$  – різниця між вибірковими середніми;
- $D_{\alpha}$  – в методах множинних порівнянь – стандартне (мінімальне) значення для визнання різниці між середніми значущою;
- $r$  – коефіцієнт кореляції;
- $\eta$  – кореляційне відношення (в кореляційному аналізі) або показник сили впливу (в дисперсійному аналізі);
- $\gamma$  – показник лінійності зв'язку;
- $A, B, C$  – позначення факторів;
- $p$  – в багатфакторному регресійному аналізі – кількість рівнів фактора;
- $b_0$  – вільний член рівняння регресії (може позначатися літерою  $a$ );
- $b_i$  – коефіцієнти рівняння регресії ( $i$  відмінне від 0, можуть використовуватися літери  $b, c$  і т.д.).

### *Генеральні параметри*

- $\alpha$  – рівень значущості;
- $\mu$  – середня арифметична генеральної сукупності;
- $\Delta$  – надійний інтервал;
- $\beta$  – коефіцієнти рівняння регресії для генеральної сукупності.

## ЛИТЕРАТУРА

1. **Адлер Ю.П.** Введение в планирование эксперимента. – М.: Металлургия, 1969. – 356 с.
2. **Адлер Ю.П.,** Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. – М.: Наука, 1976. – 382 с.
3. **Зайцев Г.Н.** Математическая статистика в экспериментальной ботанике. – М.: Наука, 1984. – 424 с.
4. **Кассандрова О.Н., Лебедев В.В.** Обработка результатов наблюдений. – М.: Наука, 1970. – 104 с.
5. **Лакин Г.Ф.** Биометрия. Учебное пособие для университетов и педагогических институтов. – М.: Высш. шк., 1973. – 343 с.
6. **Лакин Г.Ф.** Биометрия. – 4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
7. **Ликеш И., Ляга Й.** Основные таблицы математической статистики. – М.: Финансы и статистика, 1985. – 356 с.
8. **Лисенков А.Н.** Математические методы планирования многофакторных медико-биологических экспериментов. – М.: Медицина, 1979. – 343 с.
9. **Максимов В.Н.** Многофакторный эксперимент в биологии. – М.: Изд-во Моск. ун-та, 1980. – 280 с.
10. **Маркова Е.В., Лисенков А.Н.** Планирование эксперимента в условиях неоднородностей. – М.: Наука, 1973. – 219 с.
11. **Маркова Е.В., Лисенков А.Н.** Комбинаторные планы в задачах многофакторного эксперимента. – М.: Наука, 1979. – 345 с.
12. **Методические** указания к математической обработке результатов экспериментов по физиологии растений / С.Ф. Негруцкий, Л.П. Фильчаков. – Донецк: Изд-во Донец. ун-та, 1984. – 12 с.
13. **Методические** указания к применению дисперсионного анализа для статистической обработки экспериментальных данных по физиологии растений / Л.П. Фильчаков. – Донецк: Изд-во Донец. ун-та, 1992 – 70 с.
14. **Методические** указания к применению непараметрических методов при статистической обработке результатов экспериментов по физиологии растений / Л.П. Фильчаков, М.И. Бойко, З.И. Синельщикова. – Донецк: Донец. ун-т, 1986. – 24 с.
15. **Плохинский Н.А.** Биометрия. – М.: Изд-во Моск. ун-та, 1970. – 367 с.
16. **Плохинский Н.А.** Алгоритмы биометрии. – М.: Изд-во Моск. ун-та, 1980. – 150 с.
17. **Рокицкий П.Ф.** Биометрическая статистика. - 3-е изд., испр. – Минск: Вышэйш. шк., 1973. – 320 с.
18. **Урбах В.Ю.** Биометрические методы. – М.: Наука, 1964. – 410 с.
19. **Урбах В.Ю.** Статистический анализ в биологических и медицинских исследованиях. – М.: Медицина, 1975. – 295 с.
20. **Хан Г., Шаниро С.** Статистические методы в инженерных задачах. – М.: Мир, 1969. – 395 с.

## ЗМІСТ

<b>ВСТУП.....</b>	<b>4</b>
<b>5. КОРЕЛЯЦІЙНИЙ АНАЛІЗ .....</b>	<b>71</b>
Обчислення показників парної кореляції.....	73
Кореляційне відношення.....	75
Множинна кореляція.....	80
<b>6. РЕГРЕСІЙНИЙ АНАЛІЗ.....</b>	<b>84</b>
Основні положення та застосування в статистичній обробці експериментальних даних.....	84
Вирівнювання емпіричних рядів регресії.....	84
<b>Метод ковзкої середньої.....</b>	<b>85</b>
Обчислення коефіцієнтів рівняння лінійної регресії.....	88
Обчислення коефіцієнтів рівнянь регресії при нелінійних залежностях .....	91
<b>Гіперболічна залежність .....</b>	<b>91</b>
<b>Показникова (експоненціальна) функція .....</b>	<b>92</b>
<b>Степенева функція .....</b>	<b>92</b>
<b>Параболічні залежності.....</b>	<b>93</b>
<b>Логістична залежність.....</b>	<b>94</b>
Обчислення коефіцієнтів рівнянь множинної регресії.....	96
Адекватність рівняння регресії. Інтерполяція та екстраполяція.....	96
<b>7. НЕПАРАМЕТРИЧНІ МЕТОДИ СТАТИСТИЧНОЇ ОБРОБКИ.....</b>	<b>102</b>
Встановлення вірогідності різниці між двома вибірками (Т-критерій Уайта) .....	103
Х-критерій Ван дер Вардена .....	104
Серійний критерій.....	105
Критерій Колмогорова–Смирнова.....	107
Критерій знаків .....	110
Критерій Вілксона (Z) для супряжених пар ознак .....	111
Кореляція рангів.....	112
Рангова оцінка сили впливу .....	115
Множинні порівняння, засновані на критерії Краскела – Уолліса ..	117
<b>Порівняння всіх вибірок.....</b>	<b>117</b>
<b>Порівняння дослідних вибірок з контрольною.....</b>	<b>118</b>
<b>ПЛАНИ, ЩО .....</b>	<b>124</b>
<b>ЗАСТОСОВУЮТЬСЯ .....</b>	<b>124</b>

<b>9. ФАКТОРНІ ПЛАНИ <math>2^k</math></b> .....	<b>132</b>
ВИБІР ОСНОВНОГО РІВНЯ ТА ІНТЕРВАЛІВ ВАРІЮВАННЯ ФАКТОРІВ.....	132
ПОБУДОВА ПЛАНІВ ЕКСПЕРИМЕНТІВ $2^k$ .....	137
ВЛАСТИВОСТІ ПОВНОГО ФАКТОРНОГО ЕКСПЕРИМЕНТУ .....	140
СТАТИСТИЧНА ОБРОБКА РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТІВ ТИПУ $2^k$ .....	141
<b>Оцінка значущості коефіцієнтів регресії.....</b>	<b>142</b>
<b>Обчислення коефіцієнтів рівняння регресії. ....</b>	<b>144</b>
<b>Встановлення адекватності рівняння регресії .....</b>	<b>147</b>
<b>10. НЕПОВНИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ</b> .....	<b>151</b>
Складання планів дослідів неповного факторного експерименту .....	151
Вибір напівреплік. Генеруюче відношення та визначальні контрасти .....	153
Метод послідовної реалізації та метод перевалу .....	155
Репліки значної неповноти. Узагальнюючий визначальний контраст .....	156
<b>11. ПЛАНИ ЕКСПЕРИМЕНТІВ ДЛЯ ВСТАНОВЛЕННЯ НЕЛІНІЙНИХ ЗАЛЕЖНОСТЕЙ.....</b>	<b>159</b>
Експерименти типу $3^k$ .....	159
<b>Обчислення та перевірка значущості коефіцієнтів рівняння регресії</b>	<b>161</b>
Плани експериментів типу $5^k$ .....	163
<b>12. СИМПЛЕКС–ПЛАНУВАННЯ. ПОБУДОВА ПЛАНІВ. ПРИЙНЯТТЯ РІШЕНЬ ПРО НАПРЯМ РУХУ ДЛЯ ПОШУКУ ОПТИМУМУ</b> .....	<b>166</b>
<b>13. ЛАТИНСЬКІ КВАДРАТИ. ПОБУДОВА ПЛАНІВ. СТАТИСТИЧНА ОБРОБКА РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТУ.....</b>	<b>172</b>
Статистична обробка результатів експерименту .....	173
ДОДАТКИ.....	177
ПОЗНАЧЕННЯ,.....	205
ЛІТЕРАТУРА .....	206
ПРИСЕДСЬКИЙ ЮРІЙ ГЕОРГІЙОВИЧ .....	209



Навчальне видання

**Приседський Юрій Георгійович**

**Статистична обробка результатів біологічних  
експериментів**

Редактор:

Технічний редактор: