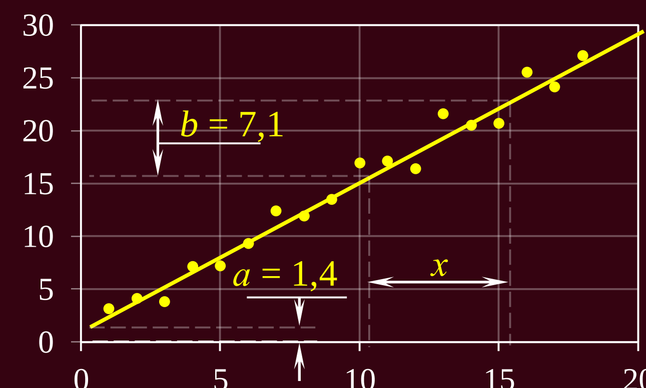
ДОСЛІДНА СПРАВА В АГРОНОМІЇ

2

ДОСЛІДНА СПРАВА В АГРОНОМІЇ

КНИГА ДРУГА

За редакцією професора,
доктора сільськогосподарських наук
А. О. Рожкова



Харків – 2016

Міністерство освіти і науки України
Харківський національний аграрний
університет ім. В. В. Докучаєва
Національний університет біоресурсів
і природокористування України
Інститут рослинництва ім. В. Я. Юр'єва НААН

ДОСЛІДНА СПРАВА В АГРОНОМІЇ

Навчальний посібник

КНИГА ДРУГА

Статистична обробка результатів агрономічних досліджень

За редакцією професора,
доктора сільськогосподарських наук
А. О. Рожкова

Харків «Майдан» 2016

УДК 631.5:001.891(075.8)
ББК П14СЯ7
Д 70

Рекомендовано до друку рішенням ученої ради
Харківського національного аграрного університету ім. В. В. Докучаєва
(протокол № 10 від 27.11.2015 р.)

Автори:
А. О. Рожков, В. К. Пузік, С. М. Каленська, Л. М. Пузік,
С. І. Попов, Н. М. Музафаров, В. Я. Бухало, Є. А. Криштоп

Рецензенти:
В. В. Кириченко, д-р с.-г. наук, академік НААН України, професор, директор Інституту рослинництва ім. В. Я. Юр'єва НААН;
А. В. Бикін, д-р с.-г. наук, чл.-кор. НААН України, професор, завідувач кафедри агрохімії та якості продукції рослинництва НУБіП;
В. Т. Саблук, д-р с.-г. наук, професор, завідувач лабораторії захисту рослин від шкідників та хвороб Інституту біоенергетичних культур і цукрових буряків

Д 70 **Дослідна** справа в агрономії: навч. посібник: у 2 кн. – Кн. 2. Статистична обробка результатів агрономічних досліджень / А. О. Рожков, В. К. Пузік, С. М. Каленська та ін. – Х.: Майдан, 2016. – 342 с.
ISBN 978-966-372-609-0.

Розглянуто завдання математичної статистики в агрономічних дослідженнях, застосування найбільш поширених статистичних методів обробки дослідних даних, а також методів кореляційного, факторного, коваріаційного та інших видів аналізу. На прикладах показано механізм і порядок проведення математичної обробки результатів досліджень різними методами.

Розраховано на студентів, аспірантів, докторантів агрономічних спеціальностей та на співробітників наукових і навчальних установ, які проводять експериментальну роботу.

УДК 631.5:001.891(075.8)
ББК П14СЯ7

ISBN 978-966-372-609-0

Зміст

Присвячується до 200-річчя Харківського НАУ ім. В. В. Докучаєва

...существуют вопросы, которые всегда
возбуждают живой интерес,
на которые не существует моды.
Один из таких вопросов «про хлеб насущный».

К. А. Тімірязєв

ПЕРЕДМОВА

Польовий дослід – найбільш важливий інструмент дослідження в агрономії. Будь-які припущення, які стосуються питань землеробства, які навіть на перший погляд здаються бездоганними та логічно обумовленими, без перевірки в польовому досліді не мають ніякої практичної цінності. Вони можуть бути використані тільки як гіпотеза для обґрунтування теми дослідження під час планування експерименту і не більше.

Польовий дослід тісно пов'язаний з питаннями «про хлеб насущный». Від того, наскільки обґрунтовані результати ми отримуємо в польових дослідах, залежить ефективність їх впровадження у виробництво. Без перебільшення можна сказати, що польовий дослід і продовольча безпека нашої держави досить взаємопов'язані.

Польовий дослід – найважливіший і найскладніший інструмент пізнання в агрономії. Його складність визначається значною кількістю варіабельних чинників, закономірності змін яких нам доволі часто невідомі. Ряд цих чинників не підлягає регулюванню з нашого боку. Жодне виробництво не має такого великого різноманіття параметрів, як сільськогосподарське. Це і мінливість клімату, погоди, популяційна мінливість рослин, варіабельність і строкатість ґрунтової родючості, елементи технології вирощування, різні для кожного окремого випадку, які ще й досі не узагальнені.

Ефективність і якість наукової роботи, глибина та результативність наукових досліджень визначаються їх методологічним рівнем. Найбільш вичерпно і точно роль методики в науково-дослідницькій роботі визначив професор М. О. Павлов: «Наука рухається поштовхами залежно від успіхів, які досягаються методикою. З кожним кроком методики вперед, ми підіймаємося на вищу сходинку, з якої нам відкривається ширший горизонт з невідомими раніше предметами».

У теперішній час найбільш відомим посібником з методики польового досліду є навчальний посібник Б. О. Доспєхова, виданий у 1985 р. У ньому автор стисло та в доступній формі викладає аспекти методики польового досліду та методів статистичного аналізу результатів досліджень. Ця робота перевидавалася п'ять разів (1965–1985 рр.) тиражами в декілька десятків тисяч кожного видання.

На наш погляд, найбільш повним посібником з усіх питань методики польового досліду є посібник М. Ф. Деревицького «Опытное дело в растениеводстве». Автор був не тільки теоретиком найвищого класу в питаннях методики дослідної справи, але й висококласним спеціалістом у галузях селекції, насінництва, землеробства та рослинництва. Він також брав участь у розробці техніки та інвентарю для сільськогосподарської науково-дослідницької роботи. Багато часу приділяв викладацькій роботі, працюючи у вищих навчальних закладах колишнього Радянського Союзу.

Достатньо повно питання методики польового досліду та методів статистичного аналізу результатів досліджень висвітлені в роботах П. П. Літуна, В. Г. Вольфа, В. О. Єщенка, П. Г. Найдіна, В. Г. Сироти, В. Ф. Савицького, М. А. Єгорова, В. А. Семенова, Г. А. Левітського, В. Н. Перегудова, А. А. Любіщєва, Т. Літла і Ф. Хілза та ін.

Особливістю нашого видання є те, що його автори значну увагу зосередили на поданні матеріалу, який би допоміг експериментатору не тільки правильно спланувати дослід, але й на належному рівні провести його за сучасними методами дослідження для одержання достовірних даних. Особлива увага приділяється проведенню статистичних розрахунків результатів досліджень, закладених за різними методиками. На прикладах показана техніка проведення статистичного аналізу як простих однофакторних, так і складних багатофакторних дослідів, проведених за поширеними методиками.

Частина перша

МАТЕМАТИЧНІ МЕТОДИ ДОСЛІДЖЕНЬ В АГРОНОМІЇ

Основу математичних методів досліджень складають такі ключові поняття, як варіація, види мінливості, крива розподілу, закономірності нормального розподілу, гістограма, сума квадратів відхилень (*СК*), число ступенів волі (*СВ*), дисперсія, помилка, сукупність, вибірка, коваріація, кореляція, регресія, критерії *t, F* і *НІР*, рівень значущості, довірчий інтервал, параметр, статистична характеристика, статистична та робоча гіпотези, квантилі, процентилі.

Під час вивчення математичних методів необхідно засвоїти логічну послідовність аналізу досліджень: об'єкт, досліджувана ознака, її показник в одиницях вимірювання, його статистична обробка, висновок і агрономічна інтерпретація. Кожний показник досліджуваної ознаки внаслідок його мінливості несе елемент випадковості і тому називається випадковою змінною. Ця обставина пояснює практику широкого застосування методів варіаційної статистики в агрономічних дослідженнях. Усі ознаки мають кількісний або якісний характер. Аналогічно розрізняються види мінливості. Для вимірювання деяких ознак, які відображають якість продукції, використовують метричну шкалу, і тоді його мінливість класифікують як кількісну.

Відмінності об'єктів за формою або забарвленням характеризуються атрибутивною мінливістю, яка являє собою поширений випадок якісної мінливості. Її різновидом є альтернативна мінливість, у якій об'єкти розрізняються за наявністю (відсутністю) ознаки. Для класифікації таких даних використовують альтернативні групи, наприклад, рослини здорові та ушкоджені, зернівки білі та червоні тощо. Їх відображають у частках або відсотках (номінальна шкала).

Дані кількісної мінливості в цілих одиницях складають дискретну мінливість. Для дробових значень з різною точністю обчислення використовують неперервну мінливість. Графіком варіаційного ряду служить гістограма та емпірична крива розподілу. Емпіричні криві фактичних спостережень можуть відповідати теоретичним, основною з яких є крива нормального розподілу. Її визначають дисперсія (s^2), стандартне відхилення (s) та генеральна середня ($\bar{x}$). На закономірностях цього розподілу базується класична (параметрична) статистика даних. В усіх випадках відхилення від цього розподілу (неоднорідні вибірки) використовують критерії і характеристики непараметричної статистики: медіану та квартилі.

За допомогою характеристик вибірки (генеральна середня, середній квадрат відхилень і середнє відхилення) проводять оцінювання параметрів усієї сукупності. Для характеристики точності дослідження беруть відносні величини варіації і помилки, а приналежність вибірки до всієї сукупності та істотність різниць між вибірками визначають за довірчими інтервалами і критеріями. *НІР* і t розраховують для кожної пари порівнянь, а F-критерій – для дослідних комплексів у цілому.

Ключові позиції статистичної обробки займають розрахунки сум квадратів відхилень (*СК*) і ступенів волі (*СС*). Агрономічна трактовка результатів дослідження враховує насамперед їхню прикладну сутність, а статистичні методи визначають ступінь їх достовірності. Серед них найбільше поширення отримав аналіз розбіжностей (дисперсійний аналіз), коваріації та стохастичних зв’язків (кореляції та регресії). Основне завдання статистичної обробки полягає у визначенні помилки, і залежно від її величини з певною мірою надійності – в оцінюванні результатів досліджень.

1. Характеристики мінливості ознак

1.1. Види мінливості, їх обчислення та вираження

На інтенсивність і спрямування процесів, які відбуваються у ґрунті та рослинах, впливає значна кількість чинників, що зумовлює значну варіабельність досліджуваних ознак. Усі об’єкти досліджуваної сукупності розрізняють за певними ознаками. Окреме значення досліджуваної ознаки називають випадковою змінною, варіантою або датою.

Кількісну неперервну та дискретну мінливість необхідно відрізняти від якісної дискретної мінливості. Характер ознаки та вид мінливості не завжди збігаються з назвою. Зокрема, такі кількісні показники, як цукристість буряків, вміст олії у насінні соняшнику, які виражаються у відсотках, характеризуються кількісною мінливістю. Кількісну мінливість обчислюють за метричною, а якісну – за номінальною та порядковою шкалами (табл. 1).

Прикладами неперервної кількісної мінливості можуть бути: урожайність продукції, висота рослин, довжина міжвузлів (у см), діаметр стебла (в мм), вегетативна маса рослин (у г, кг, т); площа листків або поглинальної поверхні коріння (у см^2); щільність ґрунту (у г/см^3); його водопроникність (у мм/ч); вміст поживних речовин у ґрунті (в г/кг) тощо.

Таблиця 1

Умовна класифікація видів мінливості ознак

Вид мінливості	Характер змінної	Приклади ознак	Шкала обчислення
Кількісна	Неперервна	Маса, об’єм довжина	Метрична пропорційна (г, см, л)
	Дискретна	Кількість, чисельність	Метрична пропорційна (шт.)
		Строки, температура, pH	Метрична інтервальна (дні, часи, градуси, од.)
Якісна атрибутивна (альтернативна)	Дискретна	Забарвлення, форма, розмір, ураженість,	Номінальна або часткова (частка, відсоток, частота)
Якісна рейтингова	Дискретна	Будь-яка ознака в балах	Порядкова, або бальна (бал, ранг)

Приклади дискретної метричної мінливості: кількість рослин, стебел, чисельність шкідників (у шт./м^2); продуктивних і непродуктивних колосків у колосі, зерен у колосі (шт.), загальна та продуктивна кількість стебел, коренів, плодів, міжвузлів, листків (у шт. на рослину).

Прикладами дискретних змінних якісної номінальної мінливості є: розмір бульб картоплі (крупні, середні, дрібні); форма кошика соняшнику (слабоувігнута, лійкоподібна, плоска, опукла, сідлоподібна, дзвоникоподібна); забарвлення колоса (білий, червоний, чорний).

До цієї групи також відносяться ознаки альтернативної (двояко-можливої) мінливості, коли об'єкти розрізняються за наявністю або відсутністю ознаки. Зокрема, колосся: остисті та безості, цілі та ушкоджені; клубні картоплі: товарні (насіннєві) і нетоварні (технічні або столові); насіннєвий матеріал: кондиційний та некондиційний; колоски: продуктивні та непродуктивні. В основі номінальної шкали лежить розподіл загальної чисельності об'єктів на дві або більшу кількість груп і визначення частоти їх прояву або частки ознаки.

Порядкова (рейтингова) якісна мінливість визначається числами рангу, рейтингу або балу на основі візуального обстеження, дегустації, обліків або підрахунків. Кожний бал може відображати певну кількість або співвідношення рослин, шкідників, площу покриття, смакові переваги продукту (дегустаційна оцінка), характер цвітіння (зав'язування плодів), морозостійкість, зимостійкість або містить інший пояснювальний коментар. Наприклад, візуальна шкала вилягання посівів, у балах: 1 – дуже слабке; 2 – слабке; 3 – середнє; 4 – сильне; 5 – дуже сильне. У цій шкалі 1 означає незначне вилягання посівів (менше 5 % рослин), 2 – до 10 %; 3 – до 15 %; 4 – до 25 % і 5 – понад 25 % рослин. У балах можна відображати і результати кількісної мінливості, наприклад вміст обмінного калію в орному шарі ґрунту: 1 – низька забезпеченість (до 50 мг на 1 кг ґрунту), 2 – середня (до 150), 3 – підвищена (понад 150 мг K_2O на 1 кг ґрунту). Під час оцінки рангами їх сума дорівнює чисельності вибірки, оскільки рангом окремого значення є його порядковий номер у побудованому ряді.

1.2. Генеральна сукупність і вибірка

Зміст *вибіркового методу*. Цей метод дає можливість проводити оцінку сукупності досліджуваних об'єктів на основі їх вибірки. У статистиці розрізняють загальну (генеральну) сукупність, яку слід вивчити, та вибіркову сукупність об'єктів, що підлягають вимірюванню. На практиці їх скорочено позначають як сукупність і вибірка. Генеральна сукупність (ГС) може бути кінцевою, тобто відомого об'єму (N). Як правило, її об'єм занадто великий або невідомий і вона сприймається як передбачувана (абстрактна). Наприклад, ГС можуть складати 200 насінин нової, перспективної, з точки зору се-

лекціонера, лінії або 1 т нового сорту, окрема ділянка, а також цілі поля господарства або району засіяні цим сортом. Для оцінки довжини стебел пшениці з генеральної вибірки (площі посіву) можна відібрати вибірку в кількості від 10 до 100 рослин. Сукупність показників (статистичну сукупність), яку одержують у результаті проведення досліджень, «ущільнюють» і одержують певні показники статистичної оцінки сукупності, які характеризують вибірку. На їх основі складають довірчі інтервали параметрів генеральної сукупності.

Помилка вибірки. Мета статистичної оцінки полягає у встановленні величини помилки вибірки, яку називають випадковою помилкою, статистичною похибкою або помилкою експерименту: $e = s\sqrt{n}$, абсолютна і $E = \sqrt{\frac{s^2}{n}} \cdot \frac{100}{\bar{y}}$ або $e, \% = e \cdot 100 / \bar{y}$, відносна помилка. У цих формулах s^2 – дисперсія; s – стандартне відхилення; n – число всіх значень, або об'єм вибірки; $\bar{y}$ – середнє значення вибірки, або вибіркова середня. Помилку позначають буквою s з індексом середнього значення ознаки $\bar{x}$ або $\bar{y}$ ($s_{\bar{x}}$, $s_{\bar{y}}$). Вона дозволяє визначити точність досліду та достовірність висновків до певного рівня ймовірності, але не до 100 %! Будь-які дослідження містять помилку, вона закономірна, оскільки певна частина сукупності залежно від впливу комплексу чинників ніколи не відображає всю генеральну сукупність. Випадкова помилка не є наслідком недосконалості методів та інструментів дослідження, слабкою підготовкою або неуважністю дослідника. Будь-який метод, як і людський чинник, має свою помилку. Все це джерела помилок систематичного або грубого характеру. Вони не підлягають обліку, оскільки на відміну від випадкової помилки не мають математичних формул і, як наслідок, лише підвищують величину останньої. Взагалі розрізняють три види помилок вибірки досліду: випадкові – ті, які визначаються статистичними методами, систематичні (зміщення) та грубі, що виникають унаслідок грубих порушень методичного характеру.

Систематичні помилки спотворюють досліджуваний показник у той чи інший бік унаслідок дії певної сталої причини. Такою причиною може бути закономірне варіювання недосліджуваних чинників (родючість ґрунту), збій у роботі інструментів, приладів і знарядь тощо. Елімінувати (контролювати) їх дію можна шляхом застосування правильної методики (повторність, повторення, блоки, ділянки, їх розміри, направлення), а також відповідної математичної моделі. Крім того, важливо використовувати відрегульовані прилади та знаряддя.

Грубі помилки (промахи) носять суб'єктивний характер і виникають внаслідок порушення основних вимог до польового досліду, недбалості виконавців, їх незнанням або невиконанням базових вимог наукового дослідження. До таких помилок відносять різного роду порушення та плутанину під час проведення досліду і польових робіт. Наприклад, виконавець через недбалість двічі вніс добрива на одну й ту саму ділянку, не дотримався правильних строків сівби, не врахував правило єдиної логічної різниці. Також грубі помилки виникають унаслідок різних строків сівби, проведення обприскувань, збирання врожаю, застосування в досліді різних знарядь або елементів агротехніки, що недопустимо на варіантах одного повторення. Грубі помилки можуть бути допущені в оформленні документації досліду (невірно записані показники, «втрата» цифр, або їх занесення не в ті клітини таблиці тощо). Ці помилки в жодному випадку не можуть бути нівельовані (компенсовані) і залишається лише забракувати зіпсовані дані, окремі повторення або дослід у цілому. Уникнути грубих помилок можна лише продуманою організацією та проведенням польового досліду.

Для статистичної обробки та одержання об'єктивних висновків можна використовувати лише ті результати досліджень, які не містять систематичних і, тим більше, грубих помилок. Присутність випадкових помилок і можлива їх кількісна оцінка приводять до того, що всі висновки за результатами проведеного експерименту мають елемент ймовірності.

Головним критерієм для практичного використання результатів дослідження є точність експерименту, яка визначається випадковою помилкою. Ця помилка зумовлена неоднорідністю експериментального матеріалу та супутніх умов, а також різного роду помилками, які допускалися під час досліду. Все це створює додаткову неоднорідність фону для порівняння варіантів і збільшує помилку. Разом із тим точність досліду зростає з ростом урожайності дослідних культур. Серед дослідів з однаковою абсолютною помилкою (e) меншу відносну помилку (E) та, відповідно, більшу точність буде мати дослід із більшим середнім показником урожайності: $E = e \cdot 100 / \bar{y}$. Саме тому однією з основних вимог, що ставиться до польового досліду, та умовою його ефективності є висока продуктивність дослідної культури.

Компонування вибірки. Під час відбирання проб потрібно добре знати сукупність і дотримуватися принципів рендомізації та репрезентативності.

Рендомізація (від англ. *random* – вибраний навмання) – одержання випадкової вибірки даних з усього масиву. Кожна рослина й точка на ділянці (полі) повинні мати рівний шанс бути відібраними. Випадковість під час відбору даних може бути обмежена секціями поля або ділянки.

Репрезентативність (від фран. *representatif* – показовий, характерний) – міра можливості відновити уявлення про цілий масив (генеральну сукупність) за її вибіркою. Репрезентативна (правильна) вибірка повинна відповідати природі сукупності та мати такий обсяг, щоб можна було об'єктивно оцінити всю сукупність. Вибірки до 10 одиниць прийнято вважати мінімальними, до 30 – малими, від 30 до 500 – великими і понад 500 – дуже великими. Зі збільшенням індивідуальних відмінностей відбирають більшу кількість проб.

Випадкова вибірка придатна для всіх досліджень і не має обмежень щодо математичної статистики. Якщо ділянка поділена на рівні та пронумеровані осередки, то місця відбору проб визначають випадково на основі таблиці випадкових чисел (додаток А). Для обліків використовують випадкове накладання рамки площею 0,25 або 0,50 м^2.

Залежно від обсягу вибірки, характеру мінливості досліджуваної ознаки в загальній сукупності та прийнятої точності, характерну вибірку можуть забезпечити систематичний, суб'єктивний, локально-рендомізований, стратифікований та інші методи відбору проб.

Систематичний метод має перевагу над випадковим у зв'язку з меншим обсягом вибірки при відсутності спрямованого зміщення певної ознаки, як, наприклад, вектора родючості ґрунту в польовому досліді. Систематичний відбір проб проводять через рівні відстані, наприклад, 7-ма рослина в кожному 5-му рядку. Місце відбору першої проби визначають рендомізовано. За необхідності охоплення всієї ділянки, її ділять на страти у двох направленнях. Зразки відбирають по діагоналі або у правому куті осередків отриманої сітки. За умови існування певної закономірності варіювання родючості ґрунту цей метод дає більшу помилку, ніж випадковий.

Суб'єктивний метод орієнтується на «типові» для дослідника рослини та місця ділянки. Така вибірка, як правило, не є об'єктивно репрезентативною та дає зміщену оцінку загальної сукупності. Зокрема, в одному випадку може бути відібрана надмірна кількість екстремальних (граничних) показників, в іншому, навпаки, дослідник може ними знехтувати, унаслідок чого вибірка не буде характерною для всієї сукупності. Зі збільшенням обсягу вибірки помилка випадкового методу знижується, а суб'єктивного – зростає. Суб'єктивний метод ефективний для мінімальних і маленьких вибірок, складених для попередньої оцінки варіації генеральної сукупності.

Локально-рендомізований метод передбачає однорідність сукупності. Відбір об'єктів вибірки проводиться випадково всередині виділених сегментів ділянки (поля). Цей метод забезпечує збереження більшої кількості дослідних рослин, покращує доступ до них, сприяє підвищенню продуктивності праці. Однак обсяг вибірки виходить більшим, ніж за випадкового методу.

Стратифікаційний метод застосовують у тих випадках, коли всю сукупність можна поділити на різного розміру підсукупності (страти). Кількість вибірок у кожній страті повинно бути пропорційним її розміру. Усередині кожної страти відбір рендомізований. Цей метод забезпечує найбільш репрезентативну вибірку.

Обсяг вибірки можна спланувати до виходу в поле, наприклад, розрахувати за формулою $N = (2/E)^2$, де коефіцієнт 2 відповідає 95%-му рівню довіри. Для ймовірності 90 і 50 % контрольні значення статистичного критерію t становлять відповідно: 1,64 і 0,84; $V = s$, % – коефіцієнт варіації, який беруть з попередніх досліджень; E, % – помилка вибірки, яка умовно допускається дослідником. Для даних, відображених в абсолютних величинах, ця формула матиме вигляд: $N = (2 \cdot s\ /\ s_{\bar{y}})^2$, де s – стандартне відхилення, $s_{\bar{y}}$ – абсолютна помилка.

Показник s можна запропонувати на основі розмаху варіювання: $W = y_{max} - y_{min}$ мінімальної вибірки та коефіцієнта k за формулою: $s = W\ /\ k$. Для обсягу вибірки, що становить 2, 3, 4 і 6, k відповідно становить 1,13; 1,69; 2,06; 2,33 і 2,53.

У великих вибірках (від 30 до 500 одиниць) показник k коливається в межах від 4 до 6. Проте для приблизного значення s достатню точність забезпечить формула $s = W/6$, що випливає з правила «трьох s», відповідно до якого вся сукупність уміщується в шість ($\bar{x} \pm 3s$) стандартних відхилень, одержаних на основі вибірки (закономірності розподілу t-критерію).

За якісної альтернативної мінливості обсяг вибірки розраховують за формулою $N = 4s^2\ /\ s^2_p$, або $4p \cdot q\ /\ s^2_p$, де s_p – помилка частки ознаки; p – частка ознаки у вибірці; q – ймовірність його відсутності; s – показник мінливості: $s^2 - pq$. Коефіцієнт 4 (2^2) – передумова 95 %-ї ймовірності.

Таким чином, будь-яке наукове дослідження від простого спостереження до складного експерименту залежить від об'єктивності вибірки. Вона повинна бути рендомізованою, репрезентативною та достатньою (повною), тобто випадковою за відбором, репрезентативною за складом і повною за обсягом.

2. Теоретичні розподіли

Емпіричні або фактичні розподіли значень ознак отримують при обстеженні сукупності. У більшій або меншій мірі вони можуть відповідати теоретичному розподілу. Теоретичні розподіли визначаються параметрами, постійними величинами (константами), а вибірки – характеристиками. Залежно від кількості параметрів розрізняють одно-, дво- та багатопарні розподіли. В агрономічних дослідженнях найбільш поширеними є двопараметричні розподіли: нормальне і біноміальне та однопараметричний розподіл Пуассона.

Якщо емпіричний розподіл відповідає теоретичному, значущість висновків значно зростає. При значному розходженні фактичної та теоретичної кривих використовують непараметричні методи оцінки.

Якщо емпіричні й теоретичні розподіли характеризують реальні значення ознаки, то контрольні розподіли стосуються стандартизованих або нормативних показників, наприклад, розподіли F, t і χ^2. Їх граничні значення слугують базовими критеріями оцінки результатів досліджень на основі вибірки.

Критерії F і t відображають параметричні методи, оскільки умовою їх застосування є нормальний розподіл значень сукупності, що не є обов'язковим для критерію χ^2. Зазвичай його застосовують як непараметричний метод перевірки відповідності фактичних результатів очікуваним.

Нормальний розподіл. У сільськогосподарській практиці приходиться мати справу з безліччю випадкових величин, які часто підпорядковуються різноманітним законам розподілів. Серед цих законів найбільш поширеним і вагомим є закон нормального розподілу.

Незважаючи на випадковий характер окремих значень досліджуваної ознаки в сукупності, ймовірність їх появи в побудованому (ранжованому) ряду зростає від граничних показників (найменших і найбільших) до середніх або від країв до центру. Цю закономірність вважають «нормою» розподілу показників будь-якої змінної, що і визначило назву – «нормальний розподіл».

Нормальний розподіл ймовірностей неперервної випадкової величини відображається формулою щільності:

$$P = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

де P – (частота) ймовірність прояву окремих показників змінної; x – випадкова змінна в інтервалі від $-\infty$ до $+\infty$; μ – генеральна середня або математичне очікування для x; σ^2 – дисперсія або математичне очікування квадрата відхилення x от μ: $(x-\mu)^2$; σ – середнє (стандартне) відхилення ($\pi \approx 3{,}14$; $e \approx 2{,}72$).

Графік цієї функції – це дзвоникоподібна асимптотична крива лінія, симетрична по обидва боки від центру з верхівкою над генеральною середньою $x = \mu$ та перегином у точках $x = \mu \pm \sigma$. Положення та форма кривої повністю визначаються двома параметрами: μ і σ^2. Для скороченого позначення кривої використовують букву $N = \mu$ і σ^2. Наприклад, N (15,5; 10,8) свідчить, що генеральна середня дорівнює 15,5, а дисперсія – 10,8. Зі збільшенням середньої крива зміщується вбік, а з ростом мінливості ознаки становиться пологою. І навпаки, за низької варіабельності показника в сукупності крутизна зростає, а крива приймає мечеподібну форму (рис. 1).

Розмах коливань від μ в один чи інший бік залежить від показника σ і зазвичай вміщується у межах трьох стандартних відхилень. Продовження кривої за межі $\mu \pm 3\sigma$ практично помічається лише за великої кількості спостережень, і цими «рідкими» значеннями вже можна знехтувати.

Ймовірність відхилення випадкової змінної нормального розподілу від очікуваного параметра (μ) не більш ніж на одно, два або три стандартних відхилення дорівнюють відповідно – $P(|y-\mu| \le \sigma) =$

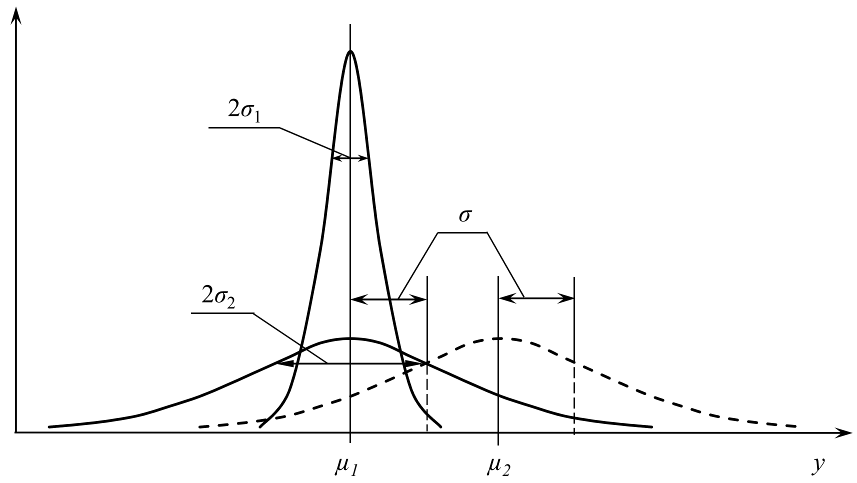


Рис. 1. Криві нормального розподілу: $\sigma_1 < \sigma_2$ і $\mu_1 < \mu_2$

$= 0{,}6827$; $P(|y-\mu| \le 2\sigma) = 0{,}9545$; $P(|y-\mu| \le 3\sigma) = 0{,}9973$. На цій основі формулюють закономірності нормального розподілу: в інтервалі від $\mu - \sigma$ до $\mu + \sigma$ знаходиться 68,27 % значень усієї сукупності, а ймовірність влучання окремих показників в інтервали $\mu \pm 2\sigma$ та $\mu \pm 3\sigma$ становить відповідно 95,45 і 99,73 %. Останню закономірність називають правилом «трьох сигм» (рис. 2).

Площу під кривою, обмежену від середнього на t стандартних відхилень, виражену у відсотках усієї площі, називають статистичною надійністю або рівнем ймовірності P, тобто ймовірності появи досліджуваного показника, який знаходиться в районі $\mu \pm t\sigma$. Ймовірність того, що значення досліджуваного показника знаходиться поза зазначеними межами, зветься рівнем значущості P_1. Він вказує ймовірність відхилення від установлених меж варіювання випадкового показника $P_1 = 1 - P$. Отже, чим більше рівень ймовірності, тим менше рівень значущості і навпаки.

У практиці агрономічних досліджень цілком достатньо користуватися ймовірністю 95 і 99 %, яким відповідає 5 %-ний і 1 %-ний рівні значущості. Ці ймовірності отримали назву довірчих ймовірностей, тобто значень, яким можна довіряти та користуватися. Приймаючи ймовірність 0,95, ризик зробити помилку становить 0,05, або лише 5 %. За ймовірності 0,99 цей ризик зменшується до 0,01, або 1 %.

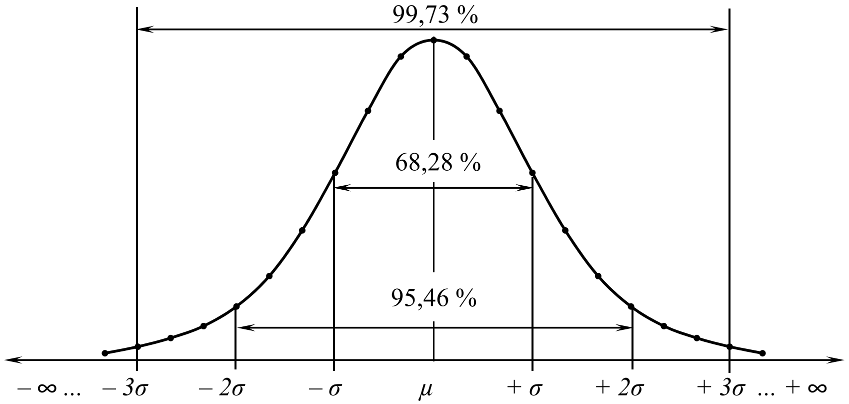


Рис. 2. Відсоток спостережень і площа, обмежена кривою нормального розподілу, для різних значень σ

Вибір довірчої ймовірності (рівня значущості) для тих чи інших досліджень визначаються практичними міркуваннями та важливістю висновків. Ймовірність 0,95 (95 %) і рівень значущості 0,05 (5 %) вважаються цілком достатніми для більшості досліджень.

Зазвичай результати спостережень польових і вегетаційних досліджень розподіляються згідно із симетричною кривою нормального розподілу, коли частоти варіантів відхиляються в той чи інший бік від середнього показника або рівні між собою, тобто симетричні. Однак на практиці іноді результати спостережень формують асиметричний (зміщений) розподіл, який істотно відрізняється від нормального.

Розрізняють правосторонню (позитивну) асиметрію, за якої збільшуються частоти правої сторони, та лівосторонню (негативну), коли зростають частоти показників лівої частини варіаційної лінії, або тих, які є меншими за медіану частотного розподілу показників (рис. 3).

Причиною асиметричного розподілу може бути неправильна (непропорційна) відібрана вибірка, коли до неї увійшла більшість чи меншість показників з більшим або меншим їх значенням. Також асиметричний розподіл частот може бути у випадку коли вплив певних чинників зміщує частоти досліджуваного показника у той чи інший бік від середнього значення.

Коли певні причини сприяють частішій появі середніх і крайніх показників вибірки, утворюється так званий *позитивний ексцесивний розподіл*, який має вигляд гострої піраміди з розширеною основою. Іноді спостерігається неправильний ексцесивний розподіл, коли в центрі розподілу частот утворюється не вершина, а западина, і крива розподілу частот нараховує дві верхівки (рис. 4).

Крива розподілу з двома чи більшою кількістю верхівок вказує на те, що у вибірку потрапили показники декількох сукупностей

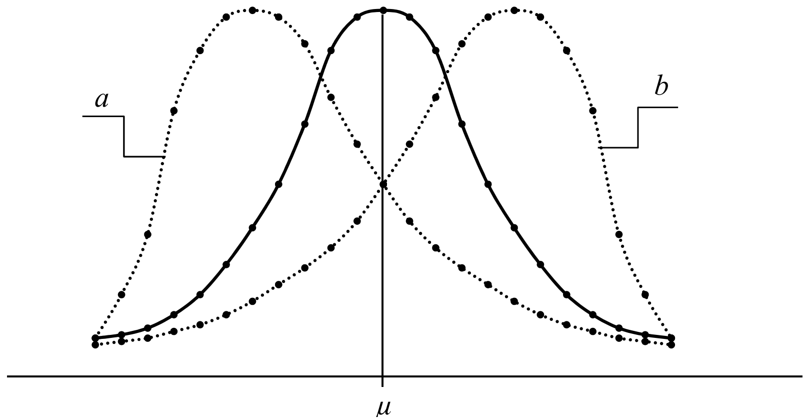


Рис. 3. Приклади асиметричного розподілу: *a* – позитивний, *b* – негативний

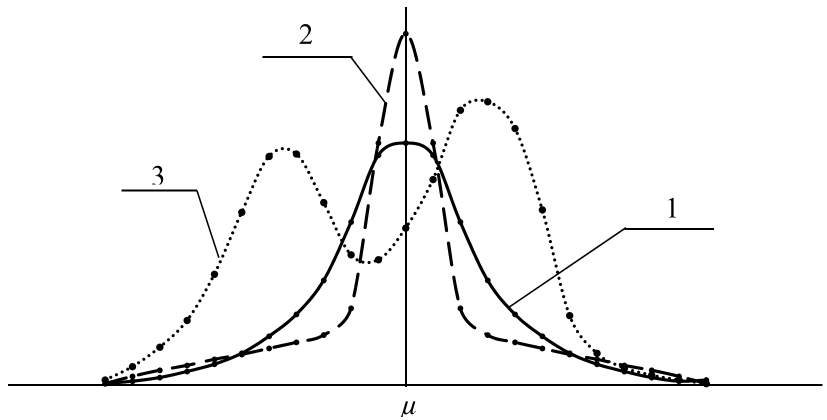


Рис. 4. Приклади ексцесивного розподілу: 1 – типова крива нормального розподілу; 2 – позитивний ексцес; 3 – негативний ексцес

з власними середніми показниками. Наприклад, у вибірку потрапили різні сорти рослин, які помітно відрізняються один від одного або вибірку рослин проводили на ділянках різних варіантів досліджень.

Нормальний розподіл найчастіше трапляється в експериментальній роботі. Головна його особливість полягає в тому, що він є граничним законом, до якого наближаються інші закони розподілу.

Біноміальний розподіл. Біноміальний закон розподілу описує ймовірність настання певної події *A* *m* раз у *n* незалежних випробуваннях за умови, що ймовірність *p* настання події *A* у кожному випробуванні залишається постійною.

В основі біноміального розподілу лежить альтернативне проявлення якісної ознаки: вона може проявлятися або ні. Наприклад, окремо взятий коренеплід може бути пошкодженим хворобами або бути здоровим (якісна ознака), тоді проба з декількох коренеплодів буде містити певну кількість здорових коренеплодів (кількісна ознака), а значна кількість проб однакового обсягу формують вже вибірку, для якої можна побудувати гістограму розподілу. Ймовірність відбору враженого коренеплоду становить *p*, а ймовірність альтернативної події – відбір здорового коренеплоду дорівнює $q = 1 - - p$. Якщо шанси відбору здорових і вражених коренеплодів однакові $p = q = 0{,}5$, то більшість проб буде мати близько половини альтернативних подій (порівну ушкоджених і здорових коренеплодів); розподіл прийме симетричну форму. Якщо ймовірності альтернативних подій різні, спостерігається той чи інший ступінь асиметрії їхнього розподілу. Біноміальний розподіл описується таким рівнянням:

$$P = \sum \left(\frac{n}{k}\right) \cdot p^k \cdot q^{n-k},$$

де *n* – обсяг вибірок; *k* – кількість проб, що мають певну ознаку; (n / k) – біноміальний коефіцієнт; співвідношення факторіалів: $n! / k! \cdot (n - k)!$, тут $n!$, $к!$ і $(n - k)!$ – добутки натурального ряду чисел від 1 до *n*, *k* і $(n - k)$ відповідно.

Приклад. При відборі проб по 10 рослин пшениці твердої ярої, в середньому лише одна рослина має два колоси. Побудуємо полігон розподілу ймовірностей відбору рослин з двома колосами в одній пробі з 10 рослин.

Для $m = 0$ (вибірка з 10 рослин не містить рослин з двома колосами), $P_{m,\,n}$ (ймовірність такої комбінації) буде дорівнювати:

$$P = \sum \left(\frac{n}{k}\right) \cdot p^k \cdot q^{n-k} = 1 \cdot 1 \cdot 0{,}34867844 = 0{,}34867844.$$

Для $m = 1$ (вибірка з 10 рослин містить одну рослину з двома колосами), $P_{m,\,n}$ (ймовірність такої комбінації) буде дорівнювати:

$$P = \sum \left(\frac{n}{k}\right) \cdot p^k \cdot q^{n-k} = 10 \cdot 0{,}1 \cdot 0{,}38742049 = 0{,}38742049.$$

За цим самим принципом визначаємо ймовірність решти можливих комбінацій відбору рослин за наявністю відмінної ознаки – два колоси на одній рослині. Після цього будуємо полігон розподілу ймовірностей відбору рослин з двома колосами в одній пробі (рис. 5).

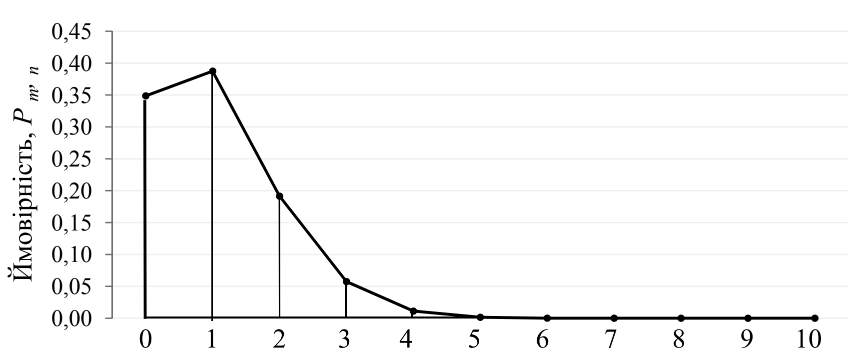


Рис. 5. Ймовірність відбору вражених рослин *m* із вибірки *n* (біноміальний розподіл)

З рисунка видно, що ймовірність відбору до однієї проби п’яти та більше рослин пшениці ярої з двома колосами мізерна за умови, що в середній пробі міститься одна така рослина.

Біноміальний розподіл має два параметри: *n* і *p*. Математичні очікування (генеральна середня та дисперсія) для показників ознаки дорівнюють відповідно: $\mu = n \cdot p$ и $\sigma^2 = n \cdot p \cdot q$.

Розподіл Пуассона. Цей розподіл являє собою окремий випадок біноміального розподілу, коли в біномі $(p + q)^n$ значення *p* дуже маленьке, а значення *q* наближається до нескінченності, тобто для дуже великої вибірки ймовірність появи рідкісної ознаки *P* мізерна. Тому його називають розподілом рідкісних подій для дуже великих вибірок. Він виражається функцією:

$$P_y = \frac{e^{-\mu} \cdot \mu^y}{y!},$$

де P_y – ймовірність значення y; y – число рідкісних подій, які мають місце в кожній великій групі; μ – середнє число рідкісних подій; $y!$ – факторіал y; (факторіал 0 дорівнює одиниці); e (основа натуральних логарифмів) ≈ 2,72.

Розподіл Пуассона визначається одним параметром – середньою. Дисперсія цього розподілу дорівнює середній, тобто $s^2 = a$. Звідси випливає, що всі теоретичні розподіли можна побудувати лише на основі однієї вибіркової середньої.

Приклад 2. Для проведення аналізу структури врожаю було відібрано 1000 рослин пшениці. Ймовірність того, що серед відібраних рослин пшениці опиниться рослина жита, становить 0,003. Знайдемо ймовірність того, що у відібраній партії рослин пшениці буде: *а* – дві; *б* – менше двох і *в* – більше двох рослин жита.

Проведемо розрахунки. Оскільки вибірка велика (n = 1000), а ймовірність події дуже маленька (p = 0,003) і події, що розглядаються, незалежні, то цей розподіл описується формулою Пуассона. У нашому прикладі μ (число рідкісних подій) становить: $\mu = n \cdot p = 3$.

Ймовірність того, що до вибірки потрапить дві рослини жита становитиме: $P_{1000\,(2)} = 3^2 \cdot e^{-3} / 2 = 9 / 2 \cdot e^{-3} = 4{,}5 \cdot 0{,}0498 \approx 0{,}224$.

Ймовірність того, що до вибірки потрапить менше двох рослин жита становитиме: $P_{1000\,(0)} + P_{1000\,(1)} = e^{-3} + 3 \cdot e^{-3} = 4 \cdot 0{,}0498 \approx 0{,}199$.

Події «вибірка містить більше двох рослин жита» і «вибірка містить не більше двох рослин жита» (ймовірність альтернативної події – q) протилежні. У цьому випадку P становитиме:

$$P = 1 - q = 1 - [P_{1000(0)} + P_{1000(1)} + P_{1000(2)}] =$$
$$= 1 - [0{,}224 + 0{,}199] = 0{,}577.$$

Приклад 3. Припустимо, що ймовірність ураження середньої вибірки рослин пшениці (n = 100 шт.) хворобою a дорівнює 0,10; b – 0,08; c – 0,06 і d – 0,04. За аналогією з попереднім прикладом розраховуються ймовірності відбору до проби (n = 100) різної кількості рослин пшениці, уражених хворобами (*a, b, c, d*). Після цього будується полігон розподілу ймовірностей (рис. 6).

Якщо n – достатньо високий і наближається до нескінченності, а значення p спрямовується до нуля так, що добуток $n \cdot p$ наближається до постійного числа, то закон наближається до біноміального

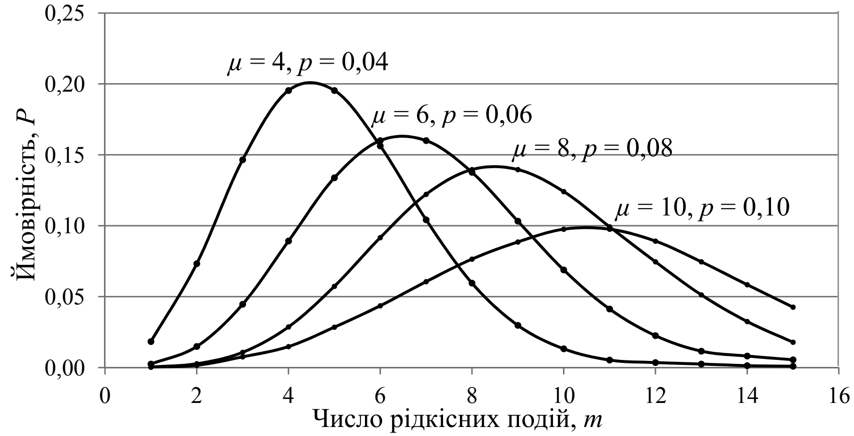


Рис. 6. Ймовірність появи певної рідкісної ознаки (*a, b, c* і *d*) за різних параметрів λ (розподіл Пуассона)

закону розподілу. Із рисунка видно, що чим більша ймовірність p, тим ближче крива розташовується біля осі m, тобто більш полога.

Біноміальний і закон розподілу Пуассона відображають розподіл ймовірностей дискретної випадкової величини. Розподіл рідкісних подій може служити контрольним розподілом під час проведення математичної обробки низьких показників засміченості насіннєвого зерна в партіях, кількості шкідників на 1 м^2 ділянки, рослин буряків у рядках, які цвітуть у перший рік – «цвітуха».

Розподіл Стьюдента (t-розподіл). За невеликої кількості спостережень (n < 20) показники зазвичай близькі і значні відхилення проявляються рідко. Це легко пояснюється законом нормального розподілу, відповідно до якого ймовірність маленьких відхилень значно більша, ніж значних відхилень. Зокрема, ймовірність відхилень, які перевищують ± 2σ (два стандартних відхилення) становить 0,05, або 5 %, а відхилень ± 3σ – 0,01, або лише 1 %. Зазвичай повторність у польовому досліді не перевищує 6–7-м повторень, тому дуже значних відхилень досліджуваної ознаки на паралельних ділянках не буде. Саме тому середнє відхилення маленької вибірки s у більшості випадків буде меншим, ніж усієї генеральної сукупності. Отже, у цих випадках покладатися на критерії нормального розподілу при формуванні висновків не можна.

Найбільше практичне значення в експериментальній роботі мав запропонований у 1908 р. англійським математиком В. Госсетом *t*-розподіл, названий на честь вченого розподілом Стьюдента (*student* з англ. студент – це псевдонім В. Госсета).

Розподіл Стьюдента (*t*-розподіл) для вибіркових середніх визначається рівнянням:

$$t = \frac{\overline{x} - \mu}{S_{\overline{x}}},\ f(t) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}}\, e^{\frac{-t^2}{2}},$$

де $\overline{x}$ – вибіркові середні; μ – генеральна середня (вибірки відбирались з однієї сукупності); $s_{\overline{x}}$; – помилки вибіркових середніх; σ – стандартне відхилення.

Теоретичні значення *t*-критерію для різних *n* наведено в дод. Б. На відміну від кривої нормального розподілу *t*-крива покриває більшу кількість значень по краях і менше поблизу центру (рис. 7). Зі збільшенням обсягу вибірки крива *t*-розподілу наближається до кривої нормального розподілу. Тому закономірності *t*-розподілу

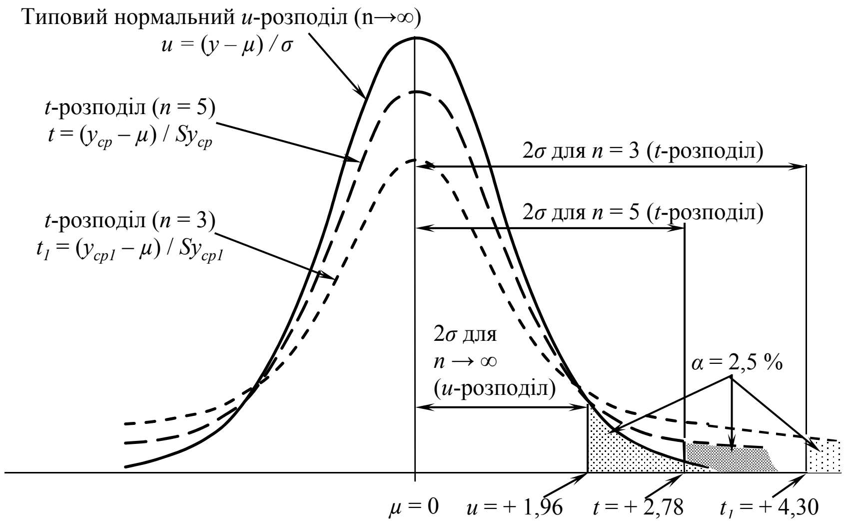


Рис. 7. Співвідношення між нормальним ($n = \infty$) і *t*-розподілом Стьюдента при $n = 3$ і $n = 5$. $U = \pm 1{,}96$, $t = 2{,}78$ і $t_1 = 4{,}30$ – межі довірчих 95 %-х інтервалів для значень *u* і *t*, або 0,95 квантилі (виділені ділянки під кривими), які включають 95 % усіх значень *u*, *t* і t_1

формулюються за аналогією з нормальним розподілом, тобто у межах $\overline{y} \pm s$; $\overline{y} \pm 2s$ і $y \pm 3s$ розміщено відповідно 68, 95 і 99 % значень усієї сукупності. Це правило поширюється і на розподіл вибіркових середніх, що дозволяє встановити три інтервали для генеральної середньої: $\overline{y} \pm s_{\overline{y}}$; $\overline{y} \pm 2s_{\overline{y}}$ і $\overline{y} \pm 3s_{\overline{y}}$ з імовірністю 68, 95 і 99 %.

Наукові дослідження можна проводити на різних рівнях статистичної надійності або з різним рівнем ймовірності (*P*): 50, 90, 95, 99 і 99,9 %. Останні три рівні ймовірності називають довірчими ймовірностями. За рівнів ймовірності 95, 99 і 99,9 % ризик допустити помилку становить відповідно 5, 1 і 0,1 %. Іншими словами, ймовірність того, що досліджувані показники «вийдуть» за межі довірчих інтервалів, становитиме відповідно 5, 1 і 0,1 %.

Імовірність помилки в статистиці називають рівнем значущості і позначають літерою α. Усі межові (табличні) показники критеріїв супроводжуються рівнем значущості, наприклад $t_\alpha = t_{05}$, t_{01}, t_{001}. Для заданого рівня ймовірності *P* рівень значущості $\alpha = 100 - P$.

Унаслідок симетричності кривої справедливе наступне рівняння: $t_\alpha = |-t_{1-\alpha}|$; $t_{\alpha/2} = |-t_{1-\alpha/2}$. Значення *t* дається за модулем. Оскільки крива покриває як позитивні, так і негативні значення, розрізняють *t* одно- і двосторонній. На одному рівні ймовірності показник першого нижчий, ніж другого: $t_{\alpha/2}$ (односторонній) = t_α (двосторонній).

Розподіл Стьюдента (*t*-розподіл) має важливе значення для характеристики малих вибірок. Він дозволяє визначити довірчий інтервал, який накриває середню сукупність μ, і перевірити ту чи іншу гіпотезу відносно генеральної сукупності. Для цього необов'язково знати параметри сукупності μ і σ, достатньо знати їхні оцінки $\overline{x}$ і *s* для певного обсягу вибірки.

Розподіл Пірсона (χ^2-розподіл). У 1900 р. математик Карл Пірсон запропонував універсальний і ефективний спосіб перевірки згоди між фактичними і модельними показниками. Запропонований ним «критерій *хі*-квадрат» – один з важливіших і найбільш часто використовуваний статистичний критерій. Більшість задач, пов'язаних з оцінкою невідомих параметрів моделі та перевірки згоди моделі й емпіричних показників, можна перевірити за його допомогою. На основі розподілу *хі*-квадрат (χ^2) побудований один з найбільш потужних критеріїв згоди – критерій χ^2 Пірсона. Фактичний χ^2 розраховується за формулою $\chi^2 = \sum(f - F)^2 / F$, де *f* – фактичні частоти; *F* – очікувані частоти (чисельності).

Критерій χ^2 служить тестом для оцінки чисельності. На відміну від непараметричного χ^2 область можливого застосування F- і t-критеріїв значно ширша, а точність оцінки вища. Кількість ступенів свободи для *хі*-квадрата визначається за формулою:

$$n = (r - 1) \cdot (c - 1),$$

де r – кількість рядків у таблиці частот; c – кількість стовпців у таблиці частот.

Приклад 4. У досліді із сортовипробування протягом трьох років порівнювали два перспективні сорти пшениці ярої. Існує думка, що сорт 1 більш стійкий до ураження хворобами. Спростуємо або підтвердимо це за допомогою χ^2-критерію. Дані відносно ураженості рослин досліджуваних сортів різними хворобами (фактичні частоти) представлені в табл. 2.

Таблиця 2

Фактичний розподіл частот ураженості рослин під час обстеження дослідних ділянок однакової площі (10 м^2)

Сорти	Хвороба *А*	Хвороба *В*	Хвороба *С*	Разом
1	4	8	13	25
2	12	15	11	38
Разом	16	23	24	63

Теоретично очікуємо, що частоти розподіляються рівномірно між досліджуваними сортами. Побудуємо таблицю теоретичних частот (табл. 3). Для цього перемножуємо відповідні суми показників за рядами і стовпцями і одержаний показник ділимо на загальну кількість показників n.

Таблиця 3

Теоретичний розподіл частот досліджуваної ознаки

Сорти	Хвороба *А*	Хвороба *В*	Хвороба *С*	Разом
1	(25 · 16) / 63 = 6,3	(25 · 23) / 63 = 9,1	(25 · 24) / 63 = 9,5	25
2	(38 · 16) / 63 = 9,7	(38 · 23) / 63 = 13,9	(38 · 24) / 63 = 14,5	38
Разом	16	23	24	63

Далі визначаємо $(Ф - Т)^2 / Т$ для кожної порівнюваної категорії і результати заносимо в підсумкову табл. 4.

Таблиця 4

Підсумкова таблиця розрахунків χ^2-крітерію

Сорти (категорія 1)	Показник ушкодження (категорія 2)		Фактичні	Теоретичні	$(Ф - Т)^2 / Т$
1	Хвороба	*А*	4	6,3	0,83
		В	8	9,1	0,13
		С	13	9,5	1,29
2	Хвороба	*А*	12	9,7	0,55
		В	15	13,9	0,09
		С	11	14,5	0,84
Сума					3,73

Критерій $\chi^2 = \sum(Ф - Т)^2 / Т = 3{,}73$. Кількість ступенів волі (n) у нашому прикладі становить: $n = (r - 1) \cdot (c - 1) = (3 - 1) \cdot (2 - 1) = 2$, де r – кількість рядів; c – кількість стовпців. Теоретичне значення χ^2 при $n = 2$ і рівні помилки 0,05 становить 5,99 (табл. 2 додатків). Оскільки фактичне значення χ^2 менше $\chi^2_{теор}$ (3,73 і 5,99 відповідно) для заданого рівня помилки – 0,05, то нульова гіпотеза не спростовується, а це свідчить про те, що досліджувані сорти за стійкістю до враженості хворобами знаходяться на однаковому рівні – $\chi^2_{теор}$ не «виходить» за граничні межі розподілу частот ймовірності – 0,95 (95 %) (рис. 8).

Розподіл *хі*-квадрат з n ступенями волі – це розподіл суми квадратів n випадкових величин, що мають нормальний розподіл з нульовим математичним очікуванням і одиничним стандартним відхиленням. Критерій *хі*-квадрат дає змогу порівнювати розподіл частот незалежно від того, розподілені вони нормально або ні. Форми розподілу χ^2 для різної кількості ступенів свободи наведені на рис. 8.

Цей розподіл було отримано на основі багаторазових вибірок з однієї сукупності та розрахунку значень χ^2 за формулою:

$$\chi^2 = \frac{\sum(y - \overline{y})^2}{\sigma^2} = \frac{(n-1)\cdot s^2}{\sigma^2},$$

де y – окремі значення; $\overline{y}$ – середні значення; s^2 – вибіркова дисперсія; σ^2 – генеральна дисперсія; n – обсяг вибірки.

Крива χ^2 розподілу має одну верхівку, викривлену вправо (висока щільність невисоких і витягнутий «хвіст» великих значень). Зі збільшенням кількості ступенів свободи відбувається її асимптотич-

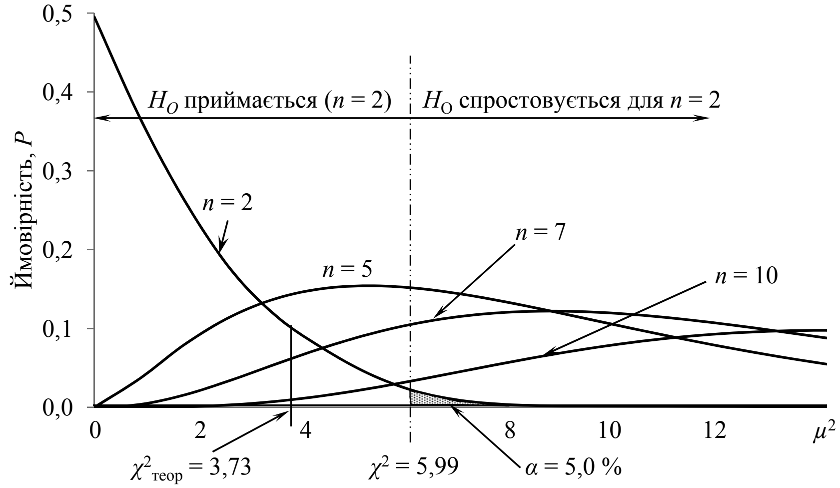


Рис. 8. Розподіл Пуассона для різних показників ступенів свободи n.

не наближення до нормальної кривої. Контрольні значення χ^2 представлені у дод. Б. Розподіл χ^2 часто використовують у генетичному аналізі, а також для оцінки підрахунків як критерію відповідності фактичних частот гіпотетично очікуваним.

Розподіл Фішера-Снедекора (F-розподіл). У ряді задач, які розв'язуються математичною статистикою, зокрема у дисперсійному та кореляційному аналізі, використовується «F-розподіл», названий за першою літерою прізвища англійського математика Р. Фішера.

Розподіл F називають розподілом Фішера-Снедекора. Саме Рональд Фішер перший дослідив розподіл відношень двох вибіркових дисперсій, але предметом його вивчення був розподіл не відношень дисперсій, а логарифмічної величини $Z = \log \sqrt{s_1^2/s_2^2}$. Дещо пізніше американський статистик Дж. Снедекор запропонував перейти безпосередньо до розподілу відношення $F = s_1^2 \,/\, s_2^2$, що виявилося значно зручніше для практичного використання в розрахунках. Цей розподіл він назвав на честь Фішера «F-розподілом». Пізніше цей вид розподілу почали називати іменем Фішера-Снедекора.

F-крива цього розподілу викривлена вправо з однією верхівкою (рис. 9). Окремими випадками F-розподілу є розподіл Пірсона та Стьюдента.

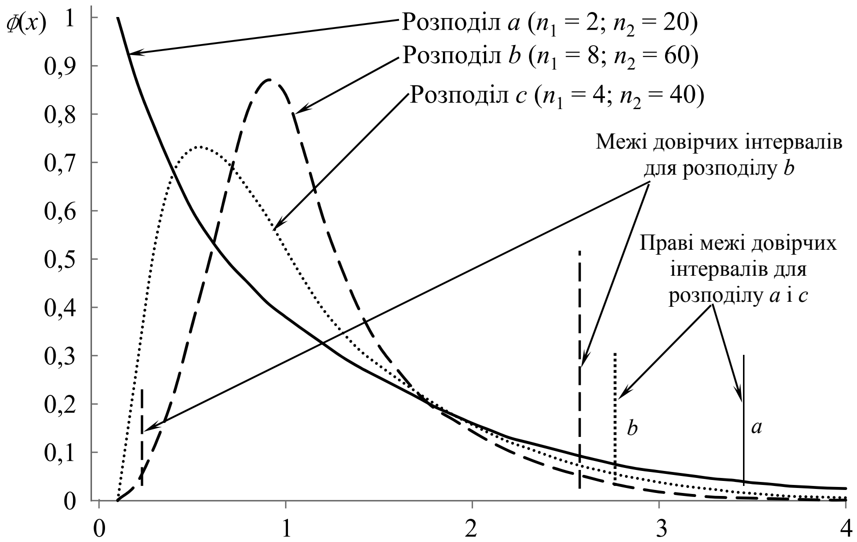


Рис. 9. Розподіл Фішера-Снедекора для різних показників ступенів свободи дисперсії чисельника n_1 і знаменника n_2

Контрольні значення F отримані на основі відношення дисперсій незалежних вибірок з однієї сукупності (дод. Б). F-критерій знаходять за двома числами ступенів свободи: для дисперсії чисельника ($k_1 = n_1 - 1$) і дисперсії знаменника ($k_2 = n_2 - 1$), де n_1 і n_2 – обсяги двох вибірок. Якщо k_1 прагне до нескінченності, то F-розподіл переходить у розподіл Пірсона. Якщо $k_1 = 1$, а k_2 = кількість ступенів свободи t-розподілу, то $F = t^2$. Кількість ступенів свободи для двох незалежних вибірок: $k = n_1 + n_1 - 2$. Під час порівняння двох вибірок у чисельник ставлять більшу дисперсію, звідки $F \geq 1$. Якщо $s_1{}^2 \approx s_2{}^2$ (дисперсії майже не відрізняються), то перевагу віддають t-критерію. Для порівняння залежних вибірок придатний лише t-критерій. Проте F-критерій має перевагу при проведенні оцінки дослідних даних, оскільки незалежно від характеру та числа варіантів від дозволяє оцінити істотність різниць дослідного комплексу в цілому. Фактичне значення F отримують зі співвідношення дисперсії досліджуваного чинника варіантів до дисперсії помилок. Тому у випадку відсутності ефекту варіантів і значної частки випадкової варіації воно може бути меншим від одиниці.

Альтернативний розподіл **–** розподіл дискретної випадкової величини, яка має два різноякісних показники (два класи). В одній пробі (спостереженні) міститься одна варіанта ($m = 1$), одно з двох можливих значень. Ймовірності кожного з них можуть бути рівними ($p = q$) або не рівними ($p < q; p > q$). Наприклад: рослини здорові та вражені, зернівки склоподібні та не склоподібні, коноплі з чоловічим суцвіттям (матіркою) і чоловічим (плоскінню) тощо. Визначення констант достатньо прості та не потребують побудови варіаційного ряду.

Важливішою характеристикою є частка (p) варіант певного виду (А), представлених загальним числом n_A у межах вибірки обсягом n:

$$p = n_A / n.$$

Якщо вихідні окремих випробувань відобразити числами 0 або 1 (що аналогічно відбору проб з обсягом $m = 1$), частка варіант збігається із середнім арифметичним, розрахованим для всіх значень:

$$M = \sum x / n.$$

Приклад 5. Результат відбору рослин пшениці ярої показав, що у відібраній групі (100 рослин) було 70 рослин одного сорту (А) і 30 – іншого (В). У нашому прикладі ми маємо справу з альтернативною ознакою (сорт А або сорт В). Таким чином, одержуємо вибірку $n = 100$ (x – значення сорту А, y – значення сорту В):

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy.

Частка рослин сорту А становить: $p = n_A / n = 70 / 100 = 0{,}7$, що збігається із середнім арифметичним для всього масиву:

$$M = \sum x / n = 70 / 100 = 0{,}7.$$

Для альтернативного розподілу можуть бути застосовані ті самі формули розрахунку вибіркових параметрів, що і для біноміального розподілу. Середня (частка рослин сорту А): $M = m \cdot p = 1 \cdot 0{,}7 = 0{,}7$.

Стандартне відхилення (при $m = 1$) становитиме:

$$s = \sqrt{m \cdot p \cdot q} = \sqrt{p \cdot q} = \sqrt{0{,}7 \cdot 0{,}3} = 0{,}46.$$

Помилка середньої (помилка частки рослин сорту А):

$$m = S / \sqrt{n} = 0{,}46 / \sqrt{100} = 0{,}046.$$

Довірчий інтервал для альтернативних ознак (їхніх часток, відсотків і частот) будується за допомогою φ-перетворення Фішера, що дає більш точні границі, особливо якщо частки сильно відрізняються. Спочатку замість показника частки (відсотка) однієї ознаки об'єктів беруть значення φ, знайдене за формулою: $\varphi = 2 \cdot \arcsin \sqrt{p}$. Далі визначають помилку: $m_\varphi = 1/\sqrt{n}$, довірчі границі $\varphi_{лів.} = \varphi - t \cdot m_\varphi$, $\varphi_{пр.} = \varphi + t \cdot m_\varphi$, після чого розраховані значення знову переводять у відсотки.

Знайдемо довірчі інтервали для рослин пшениці ярої сорту А $p = 0{,}7$ за рівня значущості $\alpha = 0{,}05$: $\varphi(70\ \%) = 1{,}982$, $m_\varphi = 1 / \sqrt{100} = 0{,}1$. $\varphi_{лів.} = 1{,}982 - 1{,}98 \cdot 0{,}1 = 1{,}784$, $\varphi_{лів.} = 1{,}982 + 1{,}98 \cdot 0{,}1 = 2{,}18$, $p_{лів.}(1{,}784) = 60{,}6\ \%$, $p_{пр.}(2{,}18) = 78{,}6\ \%$.

Таким чином, частка рослин пшениці ярої сорту А в генеральній сукупності (у посіві) становитиме мінімум 60,6 %, максимум – 78,6 %.

3. Емпіричні розподіли частот і способи їхнього зображення

Частотою називається кількість однакових значень досліджуваної ознаки в сукупності або чисельності класів (груп) вибірки. Оцінка розподілу частот шляхом розподілу вибірки у вигляді неперервного варіаційного ряду $y_1 ... y_n$ – малоефективна унаслідок незначної, одиничної і навіть нульової частоти зустрічі окремих значень. Значне прикладне значення має інтервальний та варіаційний ряд.

Усю сукупність розбивають на класи і розраховують групові середні. У результаті групування отримують графіки частот, наглядне представлення варіаційного ряду. Емпірична крива дозволяє підібрати відповідний теоретичний розподіл для оцінки результатів дослідження. Розрізняють гістограму (стовпчасто-злиту діаграму) і полігон (ламану діаграму) або емпіричну криву (згладжену лінію).

У гістограмі ширина кожного стовпця дорівнює груповому інтервалу, а висота – частоті класу. На осі абсцис відкладають границі класів або групові середні, а на осі ординат – частоти. Вершини стовпців (у центрі) з'єднують відрізками прямої лінії та отримують полігон, а при плавному переході вершин («згладжування» ламаної лінії) – криву емпіричного розподілу.

Групування великої вибірки при кількісній мінливості включає чотири етапи, в ході яких визначають: 1) кількість класів; 2) розмах варіювання та міжкласовий інтервал; 3) границі інтервалів; 4) частоти або чисельність класів і їхні середні.

Приклад 1. У фазу повної стиглості провели визначення довжини верхнього міжвузля 90 рослин тритикале ярого. Досліджуваний параметр коливався в діапазоні від 12 до 39 см (табл. 5).

Таблиця 5

Матриця вихідних даних довжини верхнього міжвузля рослин тритикале ярого, см (n = 90)

Рядок, i	Колонка (стовбець), j								
	1	2	3	4	5	6	7	8	9
1	27	19	29	23	17	28	33	25	31
2	22	26	27	30	25	21	27	27	24
3	26	23	39	26	22	36	19	29	17
4	30	28	24	29	20	28	26	31	26
5	27	34	27	16	24	16	31	20	23
6	12	25	20	27	15	24	37	29	22
7	25	28	26	18	27	28	30	24	25
8	27	32	29	25	13	29	22	32	28
9	26	28	21	28	30	26	27	35	27
10	23	36	25	33	29	23	39	26	30

У боковику таблиці розміщені номери рядків (i = 10), у головці – стовбці (колонки, j = 9), у графах – дані довжини міжвузля рослин.

Визначаємо кількість класів (k): $k = \sqrt{n}$ при $n \leq 100$ або $lg\ n$ при $n > 100$. Для $n = 90$ $k = \sqrt{90} \approx 9{,}5$. У цьому прикладі доцільно взяти 10 класів, тобто $k = 10$.

Визначаємо розмах варіювання W і груповий інтервал i. $W = y_{max} - y_{min}$ або $y_n - y_1$ – для вирівняного ряду, тобто 39 – 12 = 27. Візьмемо $W = 28$, оскільки прийнято збільшувати різницю на одиницю виміру (у даному випадку на 1), для того щоб уникнути надалі округлень. У нашому прикладі $i = W / k = 28 / 10 \approx 3$ см. За можливості інтервалами будуть цілі числа для зручності розрахунків. Вони повинні бути однаковими для всіх класів. Виняток можуть становити структурування сукупності, як, наприклад, вікова структура багаторічного насадження. Отже, кількість класів становить 10 з інтервалом у 3 см, що відповідає правилу: добуток числа класів k на інтервал i повинен перевищувати розмах варіації: $k \cdot i = 30$, $W = 28$.

Далі формуємо класи (ранги). Установлюємо нижні (НМ) і верхні (ВМ) межі. Для границь класів прийнято вибирати більш точні значення, ніж у вихідних показників: для цілих до 0,5, для десятих – до 0,05. Виключення становлять ряди дискретних змінних, представлених лише цілими числами. У нашому випадку також доцільно взяти цілі числа для виділення меж класів.

Оскільки в цьому прикладі граничні показники довжини міжвузлів становлять 12 і 39 та з урахуванням того, що $k \cdot i = 30$, нижньою границею меншої групи доцільно взяти 11 см. При цьому усі визначення довжини міжвузля входять у діапазон від 11 до 41 см і жоден із показників не перевищує встановлені межі. Отже, за нижньої границі 11 см і за $i = 3$ см одержуємо такі 10 класів: 1 – 11...13 см; 2 – 14...16 см; 3 – 17...19 см; 4 – 20...22 см; 5 – 23...25 см; 6 – 26...28 см; 7 – 29...31 см; 8 – 32...34 см; 9 – 35...37 см і 10 – 38...40 см.

Границі класів мають значення, яких немає серед вихідних даних, виражених цілими числами. Далі визначаємо середні значення кожної групи. Вони будуть дорівнювати: 12, 15, 18, 21, 24, 27, 30, 33, 36 і 39 см (табл. 6).

Далі розподіляють дати за класами (групами) і визначають частоти. Рекомендується починати з першої цифри ряду або колонки вихідної матриці і продовжувати у певному порядку: за рядами або стовбцями. У вирівняному ряду розподіл показників значно спрощується. Кожний показник закреслюють після його фіксування у відповідній групі. Для підрахунку дат використовують спосіб «конвертів» (квадрат з двома діагоналями) або штрихів (вертикальна риска). При кількості групи від 5 і більше кожну п'ятірку дат виділяють: коса риска перетинає 4 вертикальних (спосіб штрихів). Повний конверт (4 вершини + 4 сторони + 2 діагоналі) включає 10 дат (показників). Чотири крапки означають 4, а квадрат – 8 значень і т. ін. Правильність розстановки показників за групами перевірять за рівнянням $\sum f = n$.

Відносні суми накопичених частот використовують для побудови кумуляти, кривої накопичення частот (рис. 10).

Таблиця 6

Згрупований розподіл частот даних вимірювання довжини верхнього міжвузля 90 рослин тритикале ярого

Клас	Границі класів НМ...ВМ	Групова середня	Кількість класів (частота)		Сума накопичених частот	
			абсолют-на f	відносна f, %	абсолют-на F	відносна F, %
1	11...13	12	2	2,22	2	2,22
2	14...16	15	3	3,33	5	5,55
3	17...19	18	5	5,56	10	11,11
4	20...22	21	9	10,00	19	21,11
5	23...25	24	17	18,89	36	40,00
6	26...28	27	28	31,11	64	71,11
7	29...31	30	15	16,67	79	87,78
8	32...34	33	5	5,56	84	93,34
9	35...37	36	4	4,44	88	97,78
10	38...40	39	2	2,22	90	100,00
$\sum f = n = 90$ $\sum f$, % = 100						

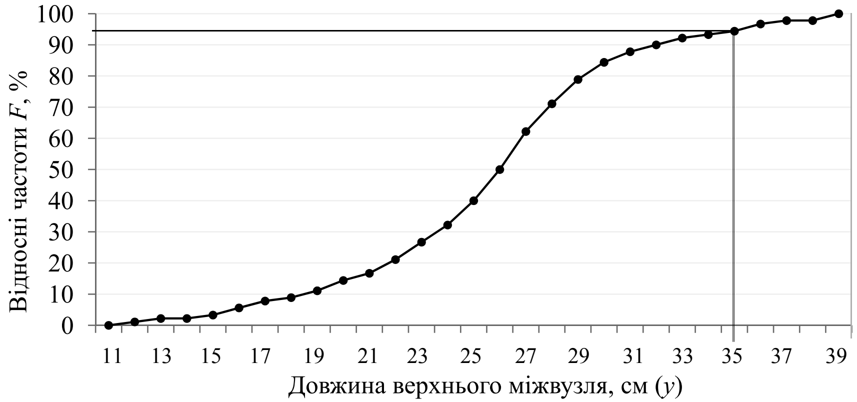


Рис. 10. Графік накопичення відносних частот F і відповідних процентилів з виділенням 0,95 квантилі (95 % рослин із довжиною міжвузлів ≤ 35 см)

Криві емпіричного розподілу будують на основі розрахункової таблиці. Масштаб графіка повинен відповідати правилу «золотого перетину»: відношення бази (ширини) графіка до висоти повинно становити 7 : 5 або 8 : 5. На осі абсцис відкладають середні за класами, а на осі ординат – відповідні їм частоти (рис. 11).

Висновок. Нерівномірність розподілу показників довжини верхнього міжвузля рослин тритикале ярого (менші значення з лівої сторони трапляються частіше) свідчить про нерівномірність вибірки. Незважаючи на це, емпіричний розподіл частот можна вважати таким, що відповідає нормальному розподілу. За основним варіантом групування переважаюча (біля 31 %) довжина верхнього міжвузля тритикале становила в середньому 27 см (6 група). Близько 8 % рослин мають довжину верхнього міжвузля більше 33 см. Майже 67 % показників варіювало в діапазоні 23–32 см (5, 6 і 7-й класи).

Тенденція досліджуваного показника групуватися навколо центра розподілу частот, статистичною характеристикою якого є середня арифметична $\bar{x}$, називається центральною тенденцією.

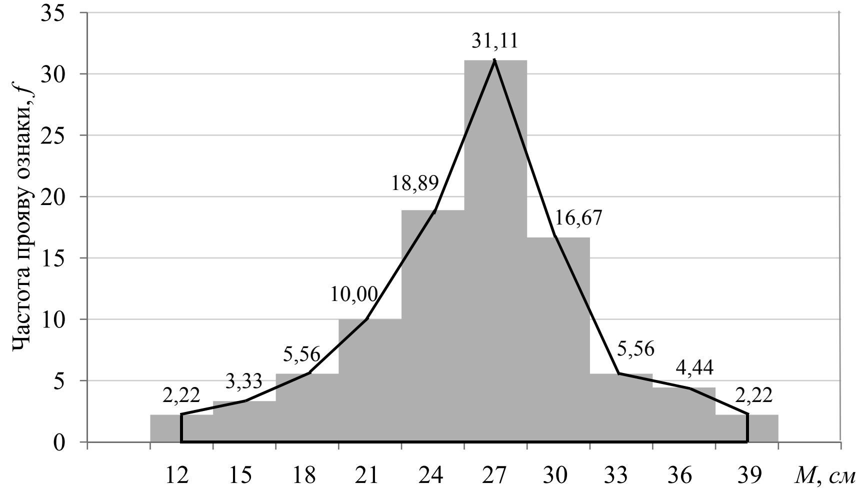


Рис. 11. Гістограма і полігон (багатокутник) розподілу показників довжини верхнього міжвузля рослин тритикале ярого (f – абсолютні частоти окремих значень; M – середньо групові показники довжини верхнього міжвузля рослин)

Поряд із середньою арифметичною, важливою статистичною характеристикою емпіричного розподілу є стандартне відхилення *s* – міра розкидання показників конкретної вибірки навколо середнього арифметичного показника. Квадрат стандартного відхилення s^2 називається середнім квадратом або дисперсією. Стандартне відхилення та дисперсія є найбільш поширеними характеристиками розсіювання мінливих ознак. Зі збільшенням дисперсії або стандартного відхилення зростає міра розсіювання значень ознак навколо центрального показника і навпаки, чим менше дисперсія та стандартне відхилення, тим менша мінливість значень ознак.

Середнє арифметичне та стандартне відхилення – це основні статистичні характеристики, які задають емпіричний розподіл частот. Цих характеристик цілком достатньо, щоб на основі знання закономірностей теоретичних розподілів побудувати емпіричний розподіл і встановити певну закономірність цього розподілу. Таким чином, головна цінність статистичних характеристик – можливість за допомогою простих показників виявити особливості емпіричних розподілів.

4. Характеристики вибірки

4.1. Кількісна мінливість

Статистична обробка даних вибірки проходить у кілька етапів:

– вирівнювання (ранжування) невеликих вибірок при n < 30 (безперервний варіаційний ряд) або групування великих вибірок (інтервальний варіаційний ряд) з побудовою гістограми;
– обчислення характеристик вибірки;
– проведення оцінки сукупності за допомогою характеристик вибірки.

У свою чергу, статистичні характеристики вибірки поділяються на три групи: 1) за характером центральної тенденції, визначають центр розміщення значень показника: арифметичну середню та медіану (статистичний показник який поділяє варіаційний ряд вибірки на дві рівні частини); 2) залежно від розсіювання або різного роду відхилень від центру статистичної сукупності; 3) за формою

кривої – коефіцієнтами викривлення (позитивної та негативної симетрії) і позитивного або негативного ексцесу (гостра або плоска верхівка).

4.1.1. Параметричні характеристики

Характеристики центральної тенденції. Середня арифметична – середнє значення малої або незгрупованої вибірки: $y = \{\sum_{i=1}^{n} y_1\} / n$, де y – окремі значення ознаки (у вітчизняній практиці з двох символів «x» і «y» перевагу віддають «x») (далі за текстом спрощено до y); n – обсяг вибірки або кількість усіх значень; $\sum_{i=1}^{n}$ – знак суми всіх значень від 1-го до n-го (надалі в тексті $\sum$).

Середньозважена арифметична середня – середнє значення згрупованої вибірки або середньозважена декількох вибірок:

$$\overline{y} = \frac{\sum_{i=1}^{k} f_i \cdot \overline{y}\mathrm{i}}{n},$$

де k – число класів або груп; f_i – чисельність окремих груп; $\overline{y}_i$ – середні групові значення; $\sum_{i=1}^{k}$ – знак сумування добутку $f_i \cdot \overline{y}_i$ від 1-ї до k-ї групи.

У випадку знаходження середньозважених кількох вибірок f буде відображати їхню чисельність, а $\overline{y}$ – середньоарифметичні. Припустимо, що середня врожайність соняшнику з площі 10, 6 і 4 га становить відповідно 1,5; 2,0 і 2,5 т/га. Середньозважена врожайність становитиме (1,5·10 + 2,0·6 + 2,5·4) / 20 = 1,85 т/га. Проста середня врожайність у цьому випадку викривить результат: (1,5 + 2,0 + 2,5) / 3 = 2,0 т/га.

Характеристики варіації. До таких характеристик відносять максимальне відхилення, дисперсію s^2, а також стандартне відхилення.

Максимальне відхилення, або *розмах варіювання,* $W = y_{max} - y_{min}$ або $y_n - y_1$ для розподіленого ряду. Цей показник дає перше уявлення про варіації показників у сукупності. На його основі розраховують кількість класів для групування великої вибірки, а також приблизний показник стандартного відхилення.

Дисперсія (*середній квадрат відхилення*) – сума квадратів відхилень ($C_{кв}$) окремих значень від середньої, поділена на число відхилень мінус одиниця (кількість ступенів свободи):

$$S^2 = \sum(y - \overline{y})^2 / n - 1 = C_{кв} / n - 1.$$

При розрахунку суми квадратів відхилень, насамперед великих вибірок, доцільно використати довільне число (А), зручне для розрахунків і близьке до середнього показника вибірки. При використанні А відпадає необхідність визначення середньої, а якщо А = 0 – і визначення відхилень.

Оскільки сума квадратів відхилень від будь-якого А завжди більша, ніж від $\bar{y}$, то в розрахункову формулу вводять зі знаком мінус поправку:

$$\text{при А} \neq 0\ C_{кв} = \sum(y - \bar{y})^2 = \sum(y - A)^2 - [\sum(y - A)]^2 / n,$$
$$\text{при А} = 0\ C_{кв} = \sum(y - \bar{y})^2 = \sum y^2 - (\sum y)^2 / n,$$

де вирази $[\sum(y - А)]^2 / n$ и $(\sum y)^2 / n$ – поправка, яка позначається буквою C або Δ.

Стандартне відхилення являє собою квадратний корінь з дисперсії: $s = \sqrt{s^2}$. Його використовують для визначення помилки вибіркової середньої та ступеня варіації.

Абсолютна помилка середньої визначається за формулою:

$s_{\bar{y}} = s / \sqrt{n} = \sqrt{s^2/n}$, і виражається в одиницях середньої.

Точність досліду (*відносну помилку*) розраховують за формулою $E = S_{\bar{y}} \times 100 / \bar{y}$. За $E < 3$ точність висока, 5 – добра, від 5 до 7 – достатня і понад 7 – недостатня (дослід вибраковується).

Коефіцієнт, або *ступінь варіації* (відносне стандартне відхилення): $V = s \cdot 100 / \bar{y}$ характеризує варіацію значень сукупності та точність досліду. Варіабельність вважають незначною, якщо $V < 10$ %, середньою, якщо $V = 10...20$ %, і сильною при $V > 20$ %.

4.1.2. Непараметричні характеристики

Медіана (*Me*) – центральний показник вирівняного ряду показників. Вона поділяє ряд на дві рівні частини, тобто одна половина показників більша, а друга менша показника медіани. Якщо обсяг вибірки (n) – непарне число, то медіаною буде одне центральне значення: $Me = y_{(n+1)/2}$. Якщо n – парне, медіаною буде півсума двох центральних значень вирівняного ряду: $Me = (y_{0,5n} + y_{0,5n+1}) / 2$. Отже, на відміну від середньої арифметичної, медіана знаходиться на основі одного або двох значень усієї вибірки. Вона більш ефективна для статистичного аналізу невеликих вибірок ($n < 30$). Навіть якщо лише

5 % показників виходять за границі $y \pm 3s$, медіана краще відображає центральну тенденцію сукупності, ніж середня арифметична.

Мода (*Mo*) – показник, який найчастіше трапляється в сукупності. На відміну від середньої арифметичної мода визначається одним значенням, яке можна виявити лише в умовах великої вибірки. Якщо розподіл показників сукупності багатоверхівковий, то це свідчить про те, що вона налічує декілька мод. Цей показник використовується для оцінки асиметричних розподілів.

Для симетричних кривих: $y = Me = Mo$; для правозміщених (з від'ємною асиметрією): $\bar{y} > Me > Mo$; для лівозміщених: $\bar{y} < Me < Mo$.

Квантилі (*Q*) (з лат. *quantum – скільки*) служать характеристиками центральної тенденції. Медіана ділить ранжируваний ряд пополам, а *квантилі*, *децилі* та *процентилі* відповідно – на 4, 10 і 100 частин. Зліва та справа від кожного граничного числа має бути однакова кількість значень, у зв'язку з чим квантилів завжди на одиницю менше кількості частин (квартилів). Наприклад, вибірка ($n = 23$):

$$\mathbf{1}_{(min)}, 2, 2, 2, 3, \mathbf{3_a}, 3, 4, 4, 4, 4, \mathbf{4_a}, 4, 5, 5, 5, 5, \mathbf{6_a}, 6, 6, 6, 7, \mathbf{8}_{(max)}$$

складається з чотирьох квартилів (1-й включає показники від $1_{(min)}$ до $3_а$; 2-й – від $3_а$ до $4_а$; 3-й – від $4_а$ до $6_а$ і останній – від $6_а$ до $8_{(max)}$) і трьох квантилів (перший (Q_1) = $3_а$ другий (Q_2) = $4_а$ і третій (Q_3) = $6_а$). У такому випадку другий квантиль (Q_2) збігається з медіаною вибірки $Q_2 = Me = 4_а$.

Як статистичні характеристики варіації використовують *квартильну відстань*: $КВ = Q_3 - Q_1$ і *квартильне відхилення*: $Q = КВ / 2$.

Центиль (сотка) – будь-яка з точок на ранжованому розподілі значень, кожна з яких містить 1/100 усього масиву. Цей показник застосовують у контрольних розподілах, так: t_{01} – 99 %-ний квантиль визначає 99 значень t, а F_{05} – 95 %-ний квантиль визначає 95 значень F.

4.1.3. Порядок проведення статистичної обробки

Увесь порядок статистичної обробки показників невеликої вибірки ($n < 30$) за кількісної мінливості складається з таких етапів:

1. Розрахунок середньої вибіркової $\bar{y} = \sum y / n$;
2. Визначення відхилень від середньої $(y - \bar{y})$;
3. Розрахунок суми квадратів відхилень від середньої:

$$C_{кв} = \sum(y - \bar{y})^2;$$

4. Визначення середнього квадрата відхилень (дисперсії): $s^2 = \frac{C_{кв}}{(n-1)}$ і стандартного (контрольного) відхилення $s = \sqrt{s^2}$;
5. Розрахунок помилки вибірки (відхилення $\bar{y} - \mu$): $s_y = \sqrt{s^2/n}$;
6. Оцінка параметрів генеральної сукупності на основі довірчих інтервалів L.

Для знаходження непараметричних характеристик – медіани та квартильного відхилення – необхідно провести вирівнювання вихідних показників. На основі характеристик: середньої вибірки $\bar{y}$, контрольного відхилення s, помилки середньої $s_{\bar{y}}$ і заданого рівня ймовірності одержують довірчі інтервали. Довірчим L називають інтервал, який покриває певну вибірку (параметр) із заданою надійністю. Наприклад, 99 %-ний довірчий інтервал для всієї сукупності або будь-якого середнього значення y_i:

$$L_{y(99)} \rightarrow \bar{y} - 2 \cdot s \leq y_i \leq \bar{y} + 2 \cdot s \text{ або } L_{y(99)} = \bar{y} \pm 2 \cdot s.$$

Для генеральної середньої μ 99 %-ний довірчий інтервал знаходиться в межах:

$$L_{\mu(99)} \rightarrow \bar{y} - 2 \cdot s_{\bar{y}} \leq \mu \leq \bar{y} + 2 \cdot s_{\bar{y}} \text{ або } L_{\mu(99)} = \bar{y} \pm 2 \cdot s_{\bar{y}}.$$

У першому інтервалі з імовірністю 99 % розміщуватимуться будь-які значення сукупності (99 % усіх значень), у другому – генеральна середня або середня всієї сукупності.

Для 68 і 95 %-ної ймовірності перед s і $s_{\bar{y}}$ слід ставити коефіцієнти 1 і 2 відповідно. Для невеликих вибірок ($n < 30$) коефіцієнти уточнюють в таблицях t-розподілу. Квантиль беруть на основі рівня значущості α, % = 100 – P (P – ймовірність довіри) і числа ступенів свободи. Зокрема, для 95 і 99 %-ної ймовірності коефіцієнти 2 і 3 замінюють значення показників t_{05} і t_{01}.

Для великих вибірок ($n > 30$) можна виділити два важливих етапи: групування даних і статистичної обробки на основі групових середніх і частот. За наявності програмного забезпечення достатньо лише ввести показники в комп'ютер і отримати результати. Агрономічні висновки дослідник робить сам на основі статистичного аналізу відповідно до тематики, моделі експерименту та рівня професіоналізму. Агрономічні висновки важливіші за статистичні. Дослідник може скорегувати висновки з урахуванням ступеня варіації і практичної вагомості розбіжностей.

Приклад 1. У двох вибірках по 7 рослин у кожній, їх висота становить: у вибірці A – 65, 70, 75, 80, 75, 85, 85 см; у вибірці B – 75, 65, 80, 75, 65, 75 і 70 см. У третій вибірці C з 21-ї рослини, яка була відібрана з тієї самої сукупності, отримали показники висоти рослин, аналогічні першій вибірці: 65, 70, 75, 80 і 85. Частота прояву показників по групах становила відповідно 4, 3, 8, 5 і 1. Знайдемо медіани досліджуваних вибірок, для чого їх систематизуємо (ранжуємо):

вибірка A – 65, 70, 75, **75**, 80, 85, 85;

вибірка B – 65, 65, 70, **75**, 75, 75, 80;

вибірка C – 65, 65, 65, 65, 70, 70, 70, 75, 75, 75, **75**, 75, 75, 75, 75, 80, 80, 80, 80, 80, 85.

Усі три вибірки представлені непарною кількістю показників, тому їх медіана – це центральний показник вибірки. Для всіх вибірок медіана становить 75 – виділене число в ранжованих рядах вибірок. Далі розраховуються середні арифметичні значення: прості для вибірок A і B та середньо групову для вибірки C:

$$\bar{y}_1 = \sum y / n = 76{,}4 \text{ см}; \bar{y}_2 = \sum y / n = 72{,}1 \text{ см};$$
$$\bar{y}_3 = \sum f \cdot y / n = (70 \cdot 3) + (75 \cdot 8) + (80 \cdot 5) + (65 \cdot 4) + (85 \cdot 1) / 21 =$$
$$= (210 + 600 + 400 + 260 + 85) / 21 = 74{,}0 \text{ см}.$$

В усіх вибірках показники медіани та середніх арифметичних розрізняються. Моду доцільно визначити для більшої вибірки – C. Це показник 75, частка якого в усій вибірці становить майже 40 %.

Серед показників оцінки вибірки найменшу довіру для характеристики її центральної тенденції заслуговує мода. Проте в деяких випадках мода має найвище значення (у дослідах з вивчення впливу інсектицидів, фунгіцидів тощо).

Медіани вибірок A і B однакові ($Me_1 = Me_2 = 75$), однак мінливість показників першої вибірки більша від другої. Зокрема, максимальна розбіжність між максимальним і мінімальним показником у першій і другій вибірці становить $W_1 = W_2 = y_{max} - y_{min} = 85 - 65 = 20$ см.

Ця оцінка варіації встановлена на основі одного максимального відхилення (65 і 85 см). Більш точною буде оцінка варіації на основі всіх відхилень від середньої. Для визначення середньоквадратичного відхилення необхідно розрахувати суму квадратів відхилень $C_{кв}$ і дисперсію s^2:

$C_{кв1} = \sum(70,0 - 76,4)^2 + (80,0 - 76,4)^2 + (75,0 - 76,4)^2 + (65,0 -$
$- 76,4)^2 + (85,0 - 76,4)^2 + (75,0 - 76,4)^2 + (85,0 - 76,4)^2 = 336,32;$
$s^2_1 = C_{кв1} / (n - 1) = 336,32 / (7 - 1) = 56,05; s_1 = \sqrt{s_1^2} = 7,49$ см;
$C_{кв2} = \sum(65,0 - 72,1)^2 + (65,0 - 72,1)^2 + (70,0 - 72,1)^2 + (75,0 -$
$- 72,1)^2 + (75,0 - 72,1)^2 + (75,0 - 72,1)^2 + (80,0 - 72,1)^2 = 192,87;$
$s^2_2 = C_{кв2} / (n - 1) = 192,87 / (7 - 1) = 32,15; s_2 = \sqrt{s_2^2} = 5,67$ см.

Отже, всі оцінки показують більшу варіабельність висоти рослин першої вибірки: для вибірки $A - W = 20$ см; $s = 7,49$ см; для вибірки $B - W = 15$ см; $s = 5,67$ см.

Непараметрична характеристика вибірки – квартильне відхилення, дорівнює напіврізниці між третім та першим квартилями: $Q = (Q_3 - Q_1) / 2$. Q_1 і Q_3 – граничні показники на стиках першого та другого, третього та четвертого квартилів. Якщо відношення $n / 4$ – дробове число, то $Q_1 = y_{n/4+1}$, а $Q_3 = y_{n-n/4}$. Якщо, $n / 4$ – ціле число, то $Q_1 = y_{n/4}$, а $Q_3 = y_{n-n/4}$. Індекси при у: $n/4$, $n/4 + 1$ і $n - n/4$ – позначають порядкові номери і показників вирівняного ряду.

Наприклад, якщо $n = 33$, лівим квартилем буде 9-те значення ряду: $33 / 4 = 8$; $8 + 1 = 9$, $Q_1 = y_9$; а правою – 25-е: $33 - 33 / 4 = 25$; $Q_3 = y_{25}$. Квартильна відстань – $КВ = Q_3 - Q_1$. За $n = 22$, $Q_1 = 22 / 4 = 5$; $5 + 1 = 6$; $22 - 5 = 17$, отже $КР = y_{17} - y_6$. Якщо $n = 80$, то $80 / 4 = 20$; $80 - 20 = 60$, отже, $КВ = y_{60} - y_{20}$.

Для третьої вибірки C ($n = 21$) $Q_1 = 21 / 4 + 1 = 6$; $Q_3 = n - n / 4 = 21 - 5 = 16$. Таким чином, $КВ = y_{16} - y_6 = 80 - 70 = 10$ см; $Q = 0,5\ КВ = 0,5 \cdot 10 = 5$ см.

Приклад 2. Після визначення забур'яненості посівів буряків на 21 ділянці (площа ділянки – 1 м^2) був збудований вирівняний ряд показників: 2, 4, 6, 7, 9, **10**, 12, 13, 16, 18, **21**, 22, 24, 25, 28, **30**, 32, 33, 36, 37 і 40 шт/м^2. У цьому прикладі лівим (нижнім) квартилем (Q_1) буде показник 10 ($Q_1 = y_i = y\ (n + 1) / 4 = y\ (21 + 1) / 4 = y_6 = 10$). Верхній квартиль ($Q_3$) – показник 30 ($Q_3 = y_{n-n/4} = y_{21-5} = y_{16} = 30$). Медіана в цьому прикладі збігається із середнім квартилем: $Q_2 = Me = y_{2i} = y_{11} = 21$. Квартильне відхилення становить: $КВ = Q_3 - Q_1 = 30 = 20$ шт/м^2.

Порівняємо квартильну оцінку з іншими оцінками варіації: максимальним, абсолютним і стандартним відхиленнями.

$\bar{y}$ (середня вибіркова) $= \sum y / n = 425 / 21 = 20,2$ шт/м^2;
Me (медіана) $= y_{11} = 21$ шт/м^2;
$C_{кв} = \sum y^2 - (y)^2 / n = (2^2 + 4^2 + ... + 40^2) - (425)^2 / 21 = 2758;$

$s^2 = C_{кв} / (n - 1) = 2758 / 20 = 137,9;$
s (стандартне відхилення) $= \sqrt{s^2} = \sqrt{137,9} = 11,7$ шт/м^2;
W (розбіжність між показниками) $= y_{max} - y_{min} = 40 - 2 = 38$ шт/м^2.

Визначимо стандартне відхилення простим способом на основі коефіцієнта $k = W / s$. За обсягу вибірки $n = 2$, $k = 1,13$ і далі відповідно ($n - k$): 5 – 2,33; 10 – 3,08; 15 – 3,11; 50 – 4,5; 100 – 5,02 і 500 – 6,07. За $n = 21$, $k = 3,27$, тобто $s = W / k = 38 / 3,27 = 11,73 \approx 11,7$, що відповідає стандартному відхиленню, розрахованому за стандартною формулою.

Висновок. Облік забур'яненості посівів проведено на низькому рівні, оскільки помилка вибірки становила:

$$E = s / \sqrt{n} \cdot 100 / \sqrt{\bar{y}} = 11,7 / \sqrt{21} \cdot 100 / 20,2 = 12,6\ \%.$$

В умовах сильної варіабельності показника $V = s / \bar{y} \cdot 100 = 58$ % потрібно збільшувати обсяг вибірки для зниження помилки.

Приклад 3. Проведемо статистичну обробку показників довжини верхнього міжвузля рослин тритикале ярого (табл. 5). Оскільки окремі значення вимірювалися з точністю до цілих одиниць, то кількісна варіабельність, неперервна за характером ознаки, стає дискретною:

12, 13, 15, 16, 16, 17, 17, 18, 19, 19, 20, 20, 20, 21, 21, 22, 22, 22, 22, 23, 23, 23, **23(Q_1)**, 23, 24, 24, 24, 24, 24, 25, 25, 25, 25, 25, 25, 25, 26, 26, 26, 26, 26, 26, 26, 26, **26(Q_2)**, 27, 27, 27, 27, 27, 27, 27, 27, 27, 27, 27, 28, 28, 28, 28, 28, 28, 28, 28, 29, 29, 29, **29(Q_3)**, 29, 29, 29, 30, 30, 30, 30, 30, 31, 31, 31, 32, 32, 33, 33, 34, **35($Q_{0,95}$)**, 36, 36, 37, 39, 39.

У представленому варіаційному ряді знайдемо три квартилі, а також 0,95 квантиль (процентиль 95 %). Лівий (нижній) квантиль становить: $Q_1 = y_{n/4} = y_{23} = 23$; середній: $Q_2 = y_{n/2} = y_{45} = 26$; правий (верхній): $Q_3 = y_{3n/4} = y_{68} = 29$. Квартиль 0,95 (95 %-ний процентиль) знаходимо за допомогою рис. 10 (с. 32): $Q_{0,95} = 35$ см.

Оскільки вибірка представлена парним числом ($n = 90$), то медіана – це середній показник двох центральних показників варіаційного ряду: $Me = (y_{45} + y_{46}) / 2 = 26 + 27 / 2 = 27$. Квартильна відстань становить: $КВ = Q = (Q_3 - Q_1) / 2 = (29 - 23) / 2 = 3$ см. Точність (E), визначена при використанні непараметричних характеристик, достатньо висока: $E = 100s / \sqrt{n} \cdot \bar{y} = 100 \cdot 3 / \sqrt{90} \cdot 26,1 = 300 / 248 = 1,21$ % за $V = 3 \cdot 100 / 26,1 = 11,5$ %.

Розрахуємо стандартні характеристики варіабельності вибірки: дисперсію s^2 та середнє відхилення s на основі групових середніх M і їхніх частот f (табл. 7).

Таблиця 7

Допоміжна таблиця для розрахунку квадратів після групування вибірки. Нижня границя першого інтервалу ($НГ_1$) = 11

M	M^2	f	fM	fM^2
12	144	2	24	288
15	225	3	45	675
18	324	5	90	1620
21	441	9	189	3969
24	576	17	408	9792
27	729	28	756	20412
30	900	15	450	13500
33	1089	5	165	5445
36	1296	4	144	5184
39	1521	2	78	3042
		$\sum f = n = 90$	$\sum fM = 2349$	$\sum fM^2 = 63927$

Розрахуємо дисперсію згрупованої вибірки:

$$s^2 = C_{кв} / n - 1;\ C_{кв} = \sum f \cdot M^2 - (f \cdot M)^2 / n;$$
$$C_{кв} = 63927 - (2349)^2 / 90 = 63927 - 61309 = 2618;$$
$$s^2 = 2618 / (90 - 1) = 29{,}4;\ s = \sqrt{s^2} = \sqrt{29{,}4} = 5{,}4 \text{ см}.$$

Для одночасної оцінки варіації і точності застосовуємо ступінь варіації $V = 100 \cdot s / \bar{y} = 100 \cdot 5{,}4 / 26{,}1 = 20{,}7$ %.

Другою оцінкою точності служить відносна помилка середньої:

$$s_{\bar{y}}, \% = E = 100 s_{\bar{y}} / \bar{y};\ s_{\bar{y}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{29{,}4}{90}} = 0{,}57 \text{ см};$$

E (відносна помилка середньої) = $100 \cdot 0{,}57 / 26{,}1 = 2{,}18$ %.

Отже, розрахунки, проведені на основі параметричних характеристик незалежно від сильної варіабельності показника ($V = 20{,}7$ %), показують, що представлене дослідження з оцінки довжини верхніх міжвузлів рослин тритикале ярого виконано з достатньо високою точністю (2,2 %). Порівнюючи результати оцінок параметричних

і непараметричних характеристик вибірки, можна побачити, що перша дає більш об'єктивну її характеристику, тобто є доцільнішою для використання.

4.2. Якісна номінальна варіабельність

Під час проведення агрономічних досліджень часто приходиться мати справу з якісною (атрибутивною) мінливістю ознак. Це може бути різне забарвлення суцвіть, плодів, насінин; форма листків або рослини; типи стебла, листків, суцвіть тощо. Для якісної мінливості характерна обмежена кількість градацій показника. Якісна мінливість зветься альтернативною, якщо досліджувані показники представляють собою одну з двох можливостей (альтернатив) – наявність або відсутність ознаки. Наприклад, рослини уражені та здорові, зернівки борошнисті та склоподібні, колоси остисті та безості, стебла прямостоячі або сланкі, материнські та батьківські форми рослин і т. ін.

Групування показників спостережень за якісного варіювання полягає в розподілі сукупності на складові групи з різними якісними показниками.

Основними статистичними показниками (параметрами) якісної мінливості є частка показника p, аналог вибіркової середньої, дисперсія s^2 і показник варіабельності s. За альтернативної мінливості частка показника становить відношення числа об'єктів, які володіють цією ознакою (n_1), до загального обсягу вибірки (n):$p = n_1/n$. Тоді частка об'єктів, які не володіють певною ознакою, дорівнює $q = n_2/n$ ($n_1 + n_2 = n$, $p + q = 1$). Якщо число градацій певної ознаки більше двох ($n > 2$), то відповідні частки градацій ознаки позначаються p_1, p_2, p_3...p_k. Кожна з них ураховує частку об'єктів з певною ознакою (n_1, n_2, n_3...n_k) відносно загального обсягу вибірки n.

Показник варіабельності якісної ознаки s характеризує мінливість величин ряду відносно одна одної. Якщо $k = 2$, показник варіабельності отримують з формули $s = \sqrt{pq}$. Наприклад, при $p = 0{,}20$ і $q = 0{,}80$ він становить: $s = \sqrt{p \cdot q} = \sqrt{0{,}20 \cdot 0{,}80} = 0{,}40$ (40 %).

Якщо $k > 2$, то $s = \sqrt[k]{p_1 . p_2 . p_3 ._{...}\ p_k}$ (k – число градацій ознаки). Розрахунки показника мінливості при $k > 2$ доцільно проводити за видозміненою формулою: $lg\, s = (lgp_1 + lgp_2 + ... + lgp_k) / k$.

Максимальна мінливість якісної ознаки спостерігається при рівних їх частках: $p_1 = p_2 = ... = p_k$. Якщо якісна мінливість альтернативна, то $p = q = 0,5$. Якщо k = 3, 4, 5, 6 і 7, то максимальна мінливість якісної ознаки (s_{max}) становить відповідно: 0,33 (33 %); 0,25 (25 %); 0,20 (20 %); 0,17 (17 %) і 0,14 (14 %). Користуючись величинами s_{max}, можна розрахувати коефіцієнт варіації якісних ознак.

Коефіцієнт варіації якісних ознак – це фактичний показник мінливості, виражений у відсотках до максимально можливої варіабельності: $V_p = 100 \cdot s / s_{max}$. Він характеризує відносну ступінь мінливості досліджуваних ознак і широко застосовується для порівняльної оцінки різних сукупностей.

Для оцінки частки ознаки всієї сукупності потрібно визначити помилку вибірки. *Помилка вибірки* s_p – ступінь відхилення частки ознаки вибіркової сукупності p від її частки в усій генеральній сукупності унаслідок неповної репрезентативності вибірки. Її розраховують за формулою: $s_p = s / \sqrt{n}$, де s – показник варіабельності якісної ознаки; n – обсяг вибірки. Для альтернативної якісної мінливості, коли значення $s = \sqrt{pq}$, помилку частки ознаки розраховують за формулою: $s_p = s / \sqrt{n} = \sqrt{pq} / \sqrt{n} = \sqrt{pq / n}$. Тут p і q можуть бути виражені в долях одиниці або відсотках.

За аналогією з кількісною мінливістю, за якісної мінливості, ймовірність зустріти p (або q) у межах інтервалу $p \pm s_p$ становить 68 %, у межах інтервалу $p \pm 2s_p$ – 95 %, у межах інтервалу $p \pm 3s_p$ – 99 %.

Приклад 4. У досліді з вивчення впливу інсектицидів на ураженість рослин ячменю ярого на контрольному варіанті було відібрано 300 рослин (n = 300), з яких 43 були вражені хворобами. Яка частка вражених рослин на всьому полі?

Оскільки $n = 300$, $n_1 = 43$, то частка вражених рослин становить: $p = 43 / 300 = 0,14$ (14 %). Частка здорових рослин дорівнює: $q = 1 - 0,14 = 0,86$ (86 %). Далі розраховується помилка вибірки: $s_p = \sqrt{pq / n} = \sqrt{0,14 \cdot 0,86 / 300} = 0,02$. Якщо взяти $\alpha = 0,05$, то частка вражених рослин на всьому полі (генеральна вибірка) буде знаходитись в інтервалі: l_{95} $0,14 \pm 2 \cdot 0,020 = 0,10...0,18$.

Висновок. Враженість рослин ячменю ярого на всьому полі буде варіювати в межах від 10 до 18 %. Статистична ймовірність цього твердження становить 95 %.

5. Статистичні гіпотези і тести

5.1. Нульова та альтернативна гіпотези

Статистичні методи перевірки гіпотез – надійна база прийняття рішень за певної невизначеності, яка зумовлена випадковою мінливістю явищ. Їх приймають, коли необхідно використати вибіркове спостереження для вирішення питання стосовно істотності різниці між вибірковими середніми, для підтвердження чи спростування факту приналежності певного варіанта до конкретної сукупності та відповідності між емпіричним і теоретичним розподілом частот.

Статистичною гіпотезою називають певне припущення стосовно закономірності та параметрів розподілу випадкової величини в сукупності. Їхня основна мета полягає в порівнянні статистичних характеристик, які проводять оцінку параметрів законів розподілу, тобто в перевірці певних статистичних гіпотез. Відмінності між середніми вибірок можуть бути як істинними, так і випадковими. Звідси випливає необхідність перевірки їх статистичної достовірності, або істотності. Порівняння двох вибірок передбачає підтвердження або спростування факту, що вони відібрані з однієї сукупності, а встановлені відмінності між ними – це випадкові чи закономірні відхилення. У більшості випадків слід перевірити гіпотезу щодо відсутності реальної відмінності між фактичними і теоретично передбачуваними спостереженнями. Цю гіпотезу називають «нульовою» гіпотезою (H_0).

Статистична обробка результатів експерименту має доказати (підтвердити) нульову гіпотезу H_0 (між варіантами немає істотної різниці і вони належать до однієї генеральної сукупності) або спростувати (відхилити) її (між варіантами існують закономірні відмінності і вони належать до різних сукупностей). Спростування нульової гіпотези свідчить про існування іншої альтернативної гіпотези (H_A). Об'єктивність нульової гіпотези перевіряється обчисленням статистичних критеріїв перевірки для заданого рівня значущості.

Нульову та альтернативну гіпотези символізують такі записи: H_0: $d = 0$ ($\mu_1 = \mu_2$) або $\bar{y}_1 \approx \bar{y}_2$ і H_A: $d \neq 0$. Альтернативною може бути двостороння робоча гіпотеза H_1: $\bar{y}_1 \neq \bar{y}_2$ і дві односторонніх – H_1: $\bar{y}_1 > \bar{y}_2$ або H_1: $\bar{y}_1 < \bar{y}_2$. Наприклад, якщо порівнюється врожайність рослин залежно від застосування різних варіантів системи живлення порівняно з контролем (без добрив), то не викликає сумніву, що робоча гіпотеза буде односторонньою: внесення добрив буде забезпечувати підвищення врожайності рослин. Прикладом двосторонньої робочої гіпотези може бути порівняння впливу на врожайність рослин двох нових сівалок, оскільки ефективність цих сівалок невідома. За нульову гіпотезу також можна взяти відомий факт або конкретно сформульовану пропозицію. Наприклад, за допомогою оптимізації системи застосування добрив потрібно збільшити врожайність зерна з 4 до 6 т/га. У цьому випадку за H_0 приймають $y_0 = 4$, а за конкуруючу (альтернативну) гіпотезу H_A (H_1): $y_1 = 6$ (або $y_1 > 4$).

Рівень значущості визначається поставленими задачами досліду: він характеризує ступінь ризику припуститися помилки, спростовуючи нульову гіпотезу. Чим менший рівень значущості, тим менша ймовірність спростувати нульову гіпотезу або допустити помилку І роду, але тим більше ймовірність припуститися помилки ІІ роду, коли нульова гіпотеза не спростовується. Рівень значущості не вимірює ступінь ризику, пов'язаного з прийняттям невірної гіпотези (помилка ІІ роду), він контролює лише помилку І роду.

Для підвищення ефективності науково-дослідної роботи, тобто для встановлення більшого числа істотних різниць, доцільно віддавати перевагу одностороннім гіпотезам. Граничні значення одностороннього критерію менші, ніж двостороннього за однакового рівня значущості. Ступінь ризику в цьому випадку зменшується вдвічі ($\alpha/2$). Наприклад, граничні значення t_{05} за $n = 10$ для дво- та одностороннього критеріїв відповідно становить 2,26 і 1,81 (дод. Б).

5.2. Алгоритм статистичного тесту

Алгоритм статистичного тесту порівняння двох вибірок на основі статистичних показників включає чотири етапи.

1. Інтерпретація результатів польового досліду із застосуванням статистичних методів. Формулювання нульової і робочих гіпотез. Наприклад, два способи сівби забезпечують формування різної урожайності основної продукції рослин: H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$;
2. Вибір метода перевірки H_0 і розрахунок фактичного критерію, наприклад $t = \bar{y}_1 - \bar{y}_2 / s_d$;

3. Порівняння розрахункового критерію з табличним, наприклад, $t < t_{01}$ і $t > t_{05}$;
4. *Висновок*. Дослідник може припуститися помилок, коли відхиляє правильну та приймає неправильну H_0, тобто приймає випадкову різницю як істотну і, навпаки, істотну – випадковою. Крім цих двох неправильних висновків, існує і два правильних, тобто є чотири можливі варіанти загальної оцінки.

А. Вибірки відібрані з однієї генеральної сукупності. Різниці між середніми випадкові:

1) прийнята правильна нульова гіпотеза, тобто різниці між середніми неістотні ($\bar{y}_1 \approx \bar{y}_2$);
2) прийнята неправильна альтернативна гіпотеза, тобто між вибірками є істотна різниця ($\bar{y}_1 \neq \bar{y}_2$). Такий висновок класифікується помилкою першого роду (α). Її ймовірність дорівнює прийнятому рівню значущості: $\alpha = 100 - P$ (P – рівень надійності).

Б. Вибірки відібрані з різних сукупностей, тобто між середніми має бути різниця:

1) нульова гіпотеза спростовується та приймається правильна робоча (альтернативна гіпотеза) H_A: $\bar{y}_1 \neq \bar{y}_2$, яка свідчить про існування достовірної різниці між середніми вибірок;
2) приймається неправильна нульова гіпотеза H_0: $\bar{y}_1 \approx \bar{y}_2$, тобто допущена помилка другого роду β.

Ймовірність прийняття альтернативної гіпотези H_A: $(1 - \beta)$ називається *робастістю*. Цей показник характеризує нечуттєвість до різного роду відхилень, яка пов'язана з тими чи іншими невідомими причинами. Це можуть бути свідомі спроби «підігнати» вибірку до того, поки вона не потрапить до статистики, помилки оформлення та інше. Робастість – показник чуттєвості, потужності експерименту.

Оскільки помилку β «відслідкувати» доволі важко, її практично не враховують у науковому дослідженні. Тому ймовірність помилки α називають ймовірністю помилки. Прийняття нульової гіпотези не виключає присутність істотної різниці в сукупності, а лише свідчить про відхилення альтернативної (робочої) гіпотези в цьому досліді.

У науковій практиці іноді трапляється «помилка ІІІ роду», коли для неправильно сформульованої проблеми робиться правильний статистичний висновок (спростовується чи приймається H_0). Ця помилка свідчить про непрофесіоналізм дослідника і припущення помилок під час формування моделі експерименту. Прикладом правильного статистичного висновку за невірно сформульованою проблематикою досліджень може бути висновок про відсутність істотної різниці між двадцятьма показниками (мала вибірка) глибини загортання насіння пшениці визначеними на ділянках із різною системою передпосівного обробітку ґрунту (1-й варіант – культивація на глибину посіву, 2-й варіант – дискування на глибину 15–17 см). Подібний висновок суперечить агрономічній логіці – помилка полягає у грубих прорахунках моделі експерименту. Якщо передпосівна культивація є типовим агрозаходом для оцінки якості сівби пшениці, то дискування для цього не використовують.

5.3. Локальна та інтервальна оцінки параметрів розподілу

Статистична оцінка вибіркової сукупності може бути представлена одним числом, точкою (локальна оцінка) або певним інтервалом (інтервальна оцінка), в якому з певною ймовірністю може знаходитись досліджуваний параметр. Зокрема, середня вибірки ($\bar{y}$) – є локальною оцінкою генеральної середньої μ, а вибіркова дисперсія s^2 – оцінкою генеральної дисперсії σ^2. *Локальну оцінку* генеральної середньої можна подати у вигляді $\bar{y} \pm s_{\bar{y}}$ ($s_{\bar{y}}$ – помилка середньої вибірки), а це свідчить, що $\bar{y}$ – оцінка генеральної середньої μ з помилкою $s_{\bar{y}}$.

Інтервальна оцінка характеризується двома числами – границями інтервалу, який покриває площу розподілу показників досліджуваного параметру.

Довірчий інтервал – певний відрізок, який охоплює частину масиву вибірки з певною ймовірністю. Центр цього відрізка – вибіркова оцінка точки, а границі (довірчі межі) визначаються середньою помилкою ($s_{\bar{y}}$) і заданим рівнем ймовірності.

У загальному вигляді довірчий інтервал для генеральної середньої виражається так:

$$\bar{y} - t \cdot s_{\bar{y}} \le \mu \le \bar{y} + t \cdot s_{\bar{y}}, \text{ або } \bar{y} \pm t \cdot s_{\bar{y}},$$

де $t \cdot s_{\bar{y}}$ – границі помилки середньої за певної кількості ступенів волі та заданого рівня значущості.

Показники t-критерію для заданих рівнів ймовірності та різних ступенів свободи наведено в дод. Б.

Приклад 1. При визначенні врожайності зерна кукурудзи розраховані такі показники вибірки: $\bar{y}$ (середня врожайність зерна вибірки) – 7,55 т/га; $s_{\bar{y}}$ (помилка середньої) – 0,12 т/га; n – 7. Визначимо 95 %-ний і 99 %-ний довірчі інтервали для генеральної середньої.

У дод. Б для 7 – 1 = 6 ступенів свободи $t_{05} = 2,45$ і $t_{01} = 3,71$, тобто відхилення досліджуваної ознаки в той чи інший бік з імовірністю 95 і 99 % становить відповідно $2,45 s_{\bar{y}}$ і $3,71 s_{\bar{y}}$.

Таким чином, довірчі інтервали для цієї вибірки з рівнями ймовірності 95 і 99 % становлять відповідно:

$$P_{95}: \bar{y} \pm t_{05} \cdot s_{\bar{y}} = 7,55 \pm 2,45 \cdot 0,12 = 7,55 \pm 0,29$$
(інтервал від 7,26 до 7,84 т/га),
$$P_{99}: \bar{y} \pm t_{01} \cdot s_{\bar{y}} = 7,55 \pm 3,71 \cdot 0,12 = 7,55 \pm 0,45$$
(інтервал від 7,10 до 8,00 т/га).

Ймовірність того, що певний показник сукупності вийде за межі інтервалу 7,26 – 7,84 т/га становить 5 %, а інтервалу 7,10 – 8,00 т/га – лише 1 %. Границі інтервалу – початок $\bar{y} - t_{05} \cdot s_{\bar{y}}$ і кінець – $\bar{y} + t_{05} \cdot s_{\bar{y}}$ звуться *довірчими границями* інтервалу.

Приклад 2. У двох вибірках (n = 8) визначені середні показники врожайності коренеплодів буряків цукрових ($\bar{y}$) та їх помилки ($s_{\bar{y}}$). Для першої вибірки: $\bar{y}_1 \pm s_{\bar{y}_1} = 18,6 \pm 0,4$; для другої вибірки: $\bar{y}_2 \pm s_{\bar{y}_2} = 19,8 \pm 0,6$. Потрібно перевірити нульову гіпотезу (H_0: $\bar{y}_1 - \bar{y}_2 = d = 0$), тобто визначити, чи є істотна різниця між середніми показниками двох вибірок для 95 %-ного рівня ймовірності.

Для семи ступенів свободи (n – 1 = 8 – 1) $t_{05} = 2,37$ (дод. Б). Для першої і другої вибірки довірчі інтервали становлять:

$$\bar{y}_1 \pm t_{05} \cdot s_{\bar{y}_1} = 18,6 \pm 2,37 \cdot 0,4 = 18,6 \pm 0,95,$$
або від 17,65 до 19,55 т/га;
$$\bar{y}_2 \pm t_{05} \cdot s_{\bar{y}_2} = 19,8 \pm 2,37 \cdot 0,6 = 19,8 \pm 1,42,$$
або від 18,38 до 21,22 т/га.

Довірчі інтервали для генеральних середніх перекривають один одного (рис. 12), що свідчить про те, що різницю між середніми вибірок $d = \bar{y}_1 - \bar{y}_2 = 1,2$ т/га не можна переносити на генеральні середні μ_1 і μ_2, оскільки різниця між ними може дорівнювати 0. Таким чи-

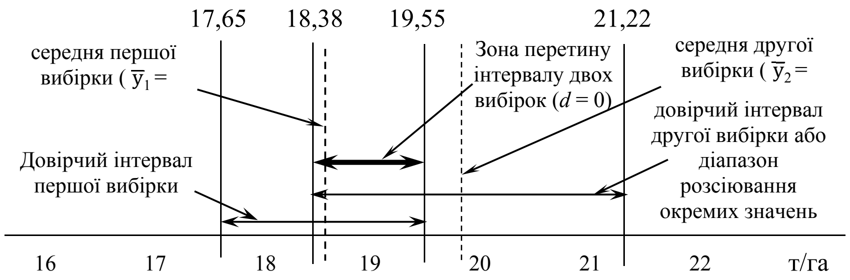


Рис. 12. Довірчі інтервали двох вибірок для рівня ймовірності – 95 %

ном, нульова гіпотеза H_0 не спростовується і між середніми вибірки істотної різниці з імовірністю 95 % немає.

Існує другий спосіб інтервальної оцінки генеральних параметрів сукупності. За формулою $s_d = \sqrt{s\bar{x}_1 + s\bar{x}_2}$ визначають помилку різниці середніх, потім за вищезазначеним способом знаходять довірчі інтервали для генеральної різниці середніх. Якщо довірчі інтервали двох вибірок перекриваються, то H_0: $d = 0$ не спростовується, якщо ж довірчі інтервали розмежовані та не перекриваються, то різниця між середніми істотна (H_0 спростовується).

Різниця між середніми d для прикладу 2 становить: $d = \bar{y}_1 - \bar{y}_2 =$ $= 19{,}8 - 18{,}6 = 1{,}2$ т/га, а помилка різниці середніх: $s_d = \sqrt{S^2_{\bar{y}_1} + S^2_{\bar{y}_2}} =$ $= \sqrt{0{,}4^2 + 0{,}6^2} = 0{,}72$.

Для 14 ступенів свободи ($n_1 + n_2 - 2 = 8 + 8 - 2 = 14$), t_{05} і t_{01} становить відповідно – 2,15 і 2,98 (дод. Б). Довірчі інтервали для генеральної різниці з рівнем значущості 05 і 01 % становитимуть:

$$\alpha_{5\%}: d \pm t_{05} \cdot s_d = 1{,}2 \pm 2{,}15 \cdot 0{,}72 = 1{,}2 \pm 1{,}55;$$
$$\alpha_{1\%}: d \pm t_{01} \cdot s_d = 1{,}2 \pm 2{,}98 \cdot 0{,}72 = 1{,}2 \pm 2{,}15.$$

Із розрахунків видно, що нульова гіпотеза не спростовується.

Добуток t-критерію та помилки різниці середніх sd зветься найменшою істотною різницею *НІР*, яка за кордоном позначається *LSD* (*least significant difference*). Якщо фактична різниця між середніми вибірок перевищує *НІР*, то нульова гіпотеза H_0 спростовується, якщо різниця виходить за межі *НІР*, то приймається робоча (альтернативна гіпотеза – H_A).

Найменша істотна різниця широко використовується для побудови довірчих інтервалів і перевірки статистичних гіпотез. Довірчий

інтервал для різниці генеральних середніх визначається за співвідношенням: $d - HIP \leq D \leq d + HIP$ або $d \pm HIP$.

За показником стандартного відхилення s проводиться оцінка інтервалу для певного значення та всієї сукупності:

$$\bar{y} - t \cdot s \leq \mu \leq \bar{y} + t \cdot s.$$

У середині цього інтервалу з певним рівнем імовірності знаходяться значення генеральної середньої μ та всі індивідуальні значення досліджуваного показника.

Наприклад, якщо для прикладу *1* стандартне відхилення s становить 0,25 т/га, то довірчі інтервали для рівнів імовірності 0,95 і 0,99 (95 і 99 %) становлять:

$$\alpha_{05}: \bar{y} \pm t_{05} \cdot s = 7{,}55 \pm 2{,}45 \cdot 0{,}25 = 7{,}55 \pm 0{,}61,$$
або від 6,94 до 8,16 т/га;
$$\alpha_{01}: \bar{y} \pm t_{01} \cdot s = 7{,}55 \pm 3{,}71 \cdot 0{,}25 = 7{,}55 \pm 0{,}93,$$
або від 6,62 до 8,48 т/га.

Отже, з імовірністю 95 % можна стверджувати, що всі окремі показники врожайності зерна кукурудзи знаходяться в діапазоні від 6,94 до 8,16 т/га, а з імовірністю 99 % – від 6,62 до 8,48 т/га.

Добуток $2t \cdot s$ називають *областю розсіювання* індивідуальних значень. У даному прикладі для 95 %-ного рівня ймовірності діапазон розсіювання окремих значень досліджуваної ознаки становить 1,22 т/га.

Для того щоб за оцінкою вибірки відзначити довірчий інтервал для генеральної середньої (μ), потрібно знати помилку цієї ознаки. Отже, для визначення будь-якої оцінки вибірки потрібно знати її середню помилку. Стандартні формули помилок для деяких характеристик вибірки представлені у табл. 8.

5.4. Оцінка істотності різниці між середніми з вибірки за t-критерієм

Особливості проведення статистичного тесту залежать від того, чи однакові обсяги вибірки, чи пов'язані досліджувані показники вибірок між собою загальними умовами. Наприклад, місцем або об'єктом дослідження варіанта. Прикладом незалежних вибірок

Таблиця 8

Формули середніх помилок оцінок вибірок

Вид оцінки вибірки	Середня помилка оцінки вибірки
Середня вибіркових	$s_{\bar{y}} = s / \sqrt{n} = \sqrt{s^2/n}$
Частка ознаки p	$s_p = \sqrt{pq / n}$
Стандартне відхилення	$s_s = s / \sqrt{2n}$
Коефіцієнт варіації	$s_v = V / \sqrt{2n}$
Різниця між середніми вибірок d	$s_d = \sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}}$
Різниця між частками вибірок d_p	$s_{dp} = \sqrt{s^2_{p_1} + s^2_{p_2}}$
Коеф. лінійної кореляції за малих r	$s_r = \sqrt{\frac{(1-r^2)}{(n-2)}}$

може бути порівняння ураженості рослин пшениці озимої на двох різних ділянках – контрольній та з внесенням добрив. Залежними слід вважати вибірки однієї сукупності, коли її об'єкти підлягають різним випробуванням (двом або більше) у часовому або просторовому інтервалі. Прикладом залежних вибірок може бути порівняння площі листків рослин однієї ділянки до проведення позакореневого підживлення полімерними добривами та через певний інтервал (2–3 тижні).

Якщо порівнюються середні двох незалежних вибірок, в яких одиниці спостережень першої вибірки не пов'язані ніякими загальними умовами з одиницями вимірювання другої вибірки, то за t-критерієм оцінюється істотність різниці середніх ($d = \bar{y}_1 - \bar{y}_2$). Для пов'язаних вибірок оцінюється істотність середньої різниці ($\bar{d} = \sum d / n$).

Порівняння середніх двох незалежних вибірок проводять за t-критерієм (Стьюдента):

$$t = d / S_d,$$

де $d = (\bar{y}_1 - \bar{y}_2)$ – різниця між середніми двох вибірок; $s_d = \sqrt{2s^2/n}$; s^2 – середньозважена для двох вибірок – $s^2 = \frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{n_1+n_2-2}$ або однакова ($s^2_1 = s^2_2$) (для статистичної оцінки дослідів беруть дисперсію помилок – s^2_z).

Залежно від того, розрізняються обсяги вибірок та їх дисперсії або ні, існують такі формули розрахунку s_d:

1) $s_d = \sqrt{\frac{s^2}{n}} + \sqrt{\frac{s^2}{n}} = \sqrt{\frac{2s^2}{n}}$, тут $n_1 = n_2 = n$, $s_1^2 \approx s_2^2 = s^2$;

2) $s_d = s \cdot \sqrt{s^2 \cdot \frac{n_1+n_2}{n_1 \cdot n_2}}$, тут $n_1 \neq n_2$, $s_1^2 \neq s_2^2$;

3) $s_d = \sqrt{\frac{n_1+n_2}{n_1 \cdot n_2} \cdot \frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{n_1+n_2-2}}$ або $s_d = \sqrt{s^2 \cdot \frac{n_1+n_2}{n_1 \cdot n_2}}$,

тут s^2 – середньозважена дисперсія ($n_1 \neq n_2$, $s_1^2 \neq s_2^2$);

4) $s_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, тут $n_1 \neq n_2$, $s_1^2 \neq s_2^2$.

Перші три формули використовують для порівняння дослідних середніх, а останню – для двох вибірок.

Табличне значення або критичний квантиль t беруть, виходячи з числа ступенів свободи СС = $2n - 2$ або СС = $n_1 + n_2 - 2$, якщо $n_1 \neq n_2$ і рівня довіри $1 - \alpha / 2$ (двостороння гіпотеза H_A: $\bar{y}_1 \neq \bar{y}_2$) або $1 - \alpha$ (одностороння H_A: $\bar{y}_1 > \bar{y}_2$ або $\bar{y}_1 < \bar{y}_2$).

Приклад 3 (незалежні вибірки). У досліді із сортовивчення порівнювали рівень урожайності насіння двох сортів ріпака ярого. Потрібно визначити, чи істотно розрізняється врожайність насіння досліджуваних сортів A і B.

Середня врожайність сорту A на 15-ти ділянках становить 3,05 т/га, а сорту B на 11-ти ділянках – 4,16 т/га. Дисперсія сорту A (s_A^2) дорівнює 1,22, сорту B (s_B^2) – 1,64. Оскільки в цьому прикладі $n_A \neq n_B$ і $S_A \neq S_B$, то розраховується середньозважена дисперсія:

$$s^2 = \frac{(n_A-1)\cdot s_A^2 + (n_B-1)\cdot s_B^2}{n_A + n_B - 2} = \frac{(15-1)\cdot 1{,}22 + (11-1)\cdot 1{,}64}{11+15-2} = 1{,}395;$$

$$s = \sqrt{s^2} = \sqrt{1{,}395} = 1{,}18;$$

$$s_d = s \cdot \sqrt{\frac{n_A + n_B}{n_A \cdot n_B}} = 1{,}18 \cdot \sqrt{\frac{15+11}{15 \cdot 11}} = 1{,}18 \cdot 0{,}397 = 0{,}47;$$

За формулою (4)

$$s_d = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} = \sqrt{\frac{1{,}22}{15} + \frac{1{,}64}{11}} = \sqrt{0{,}08 + 0{,}15} = 0{,}48;$$

$$t_{факт} = (\bar{y}_A - \bar{y}_B) / s_d = (4,16 - 3,05) / 0,48 = 2,31.$$

Оскільки $t_{факт} > t_{05}$ (2,31 > 2,06), то нульова гіпотеза H_0 спростовується, що свідчить про існування достовірної різниці за врожайністю насіння між досліджуваними сортами.

Аналогічну оцінку різниці між сортами дає і критерій *НІР*. Зокрема, за фактичної різниці між середніми 1,11 т/га ($d = \bar{y}_A - \bar{y}_B = 4,16 - 3,05 = 1,11$ т/га), $НІР_{05} = t_{05} \cdot s_d = 2,06 \cdot 0,47 = 0,97$, що значно менше d, тобто різниця істотна.

Разом із тим, за більш строгої оцінки ($p = 0,99$), різниця за врожайністю між сортами неістотна, адже $НІР_{01} = t_{01} \cdot s_d = 2,80 \cdot 0,47 = 1,32$, що значно вище за середню різницю d.

Інтервальну оцінку істотності різниці проводять шляхом зіставлення інтервалів для генеральних середніх:

для сорту A: $\bar{y}_A \pm t_{05} \cdot s\bar{y}_A$; t_{05} ($CC = 14$) = 2,15; t_{01} ($CC = 14$) = 2,98; $s\bar{y}_A = \sqrt{s_A^2 / n_A} = 0,30$. Таким чином, довірчий інтервал мінливості показників урожайності сорту A з імовірністю 95 % становить: 3,05 ± ± 2,14 · 0,3 = 3,05 ± 0,64 (l_{95}: від 2,41 до 3,69 т/га) і з ймовірністю 99 % – 3,05 ± 2,98 · 0,3 = 3,05 ± 0,89 (l_{99}: від 2,16 до 3,94 т/га);

для сорту B: $\bar{y}_B \pm t_{05} \cdot s\bar{y}_B$; $\bar{y}_B = 4,16$; t_{05} ($CC = n - 1 = 10$) = 2,23; $s\bar{y}_B = \sqrt{s_A^2 / n_A} = \sqrt{1,64 / 11} = 0,39$. Отже, довірчий інтервал мінливості показників урожайності сорту B з імовірністю 95 % становить: 4,16 ± 2,23 · 0,39 = 4,16 ± 0,87 (l_{95}: від 3,29 до 5,03 т/га), з імовірністю 99 % – 4,16 ± 3,17 · 0,39 = 4,16 ± 1,24 (l_{99}: від 2,92 до 5,40 т/га).

Для обох рівнів ймовірності довірчі інтервали двох вибірок перекривають один одного, хоча для α_{05} і незначно. Отже, на відміну від t-критерію та $НІР_{05}$, інтервальна оцінка не виявила істотної різниці між урожайністю насіння сортів ріпака через меншу її ефективність. Саме тому доцільно покластися на локальну оцінку.

Для проведення статистичної оцінки двох вибірок часто використовують F-критерій (Фішера-Снедекора), який є відношенням більшої дисперсії до меншої (у конкретному випадку $F = s_B^2 / s_A^2$). Розрахований (фактичний) критерій порівнюють з теоретичним для заданих рівнів ступенів свободи дисперсії чисельника – $n_B - 1 = 10$ і дисперсії знаменника – $n_A - 1 = 14$. Для заданих рівнів ступенів свободи обох дисперсій $F_{05} = 2,6$ (дод. В). Отже, $s_A^2 = 1,22$, $s_B^2 = 1,64$, а $F_{факт.} = s_B^2 / s_A^2 = 1,64 / 1,22 = 1,34$.

Оскільки $F_{факт.} < F_{05}$, то нульова гіпотеза не спростовується з імовірністю 95 %, що свідчить про відсутність істотної різниці між вибірками (вони мають однакову варіабельність досліджуваного показника). Крім того, F-критерій для оцінки пари середніх неефективний (це продемонстровано в прикладі).

Отже, у разі порівняння середніх двох непов'язаних вибірок саме оцінка їх середніх за t-критерієм та $НІР_{05}$ і $НІР_{01}$ найбільш об'єктивно характеризує їхню варіабельність і на її основі можна зробити висновок щодо істотності різниці між вибірками.

Помилку різниці середніх d для залежних вибірок визначають різницевим методом. Сутність цього методу полягає у тому, що оцінюється не різниця середніх ($d = \bar{y}_1 - \bar{y}_2$), а істотність середньої різниці ($\pm d$), хоч ця величина і виражена одним числом.

Для визначення помилки середньої різниці різницевим методом спочатку знаходять різницю між пов'язаними парами показників, потім середню різницю $\bar{d}$ і помилку середньої різниці за формулою:

$$s_d = \sqrt{\frac{\sum(d - \bar{d})^2}{n(n-1)}} \text{ або } s_d = \sqrt{\frac{\sum d^2 - (\sum d)^2 / n}{n(n-1)}}.$$

Після цього визначають фактичний t-критерій за формулою $t_ф = \bar{d} / s\bar{d}$, порівнюють його з теоретичним для заданого рівня ймовірності та роблять висновок щодо існування різниці між показниками двох залежних вибірок.

Кількість ступенів свободи для теоретичного t-критерію визначають за рівнянням $CC = n - 1$, де n – кількість пов'язаних пар вибірок.

Приклад 4. У семи господарствах області порівнювали два сорти сої за вмістом білка в насінні. У кожному господарстві відбір зразків насіння проводили на сусідніх ділянках. Отже, ми маємо справу з пов'язаними вибірками, тому обробку результатів слід проводити методом попарних порівнянь. Вихідні дані вмісту білка в насінні сої сортів A і B та їх попарні порівняння представлені в табл. 9.

$$s_d = \sqrt{\frac{\sum d^2 - (\sum d)^2 / n}{n \cdot (n-1)}} = \sqrt{\frac{79,68 - 22,8^2 / 7}{7 \cdot (7-1)}} = 0,36.$$

$$t_{факт} = \bar{d} / s\bar{d} = 3,26 / 0,36 = 9,06.$$

За кількості ступенів свободи $(n - 1) = 6$, $t_{05} = 2,45$, $t_{01} = 3,71$. Оскільки $t_{факт.} > t_{теор.}$ навіть за строгої оцінки ($P = 99$ %), досліджу-

Таблиця 9

Оцінка різниці між вмістом білка в насінні сої різних сортів на сусідніх ділянках

Господарство, n	Вміст білка, %		Різниця d	Квадрат різниці d^2
	сорт A	сорт B		
1	50,8	47,1	+ 3,7	13,69
2	48,6	44,4	+ 4,2	17,64
3	49,3	46,2	+ 3,1	9,61
4	45,8	41,4	+ 4,4	19,36
5	51,1	47,9	+ 3,2	10,24
6	47,7	46,0	+ 1,7	2,89
7	49,8	47,3	+ 2,5	6,25
Сума	343,1	320,3	22,8	79,68
Середнє	49,0	45,8	3,26	–

вані сорти істотно відрізняються один від одного за вмістом білка в насінні.

Розрахунок *НІР* також підтверджує цей висновок:

$$HIP_{05} = t_{05} \cdot s_{\overline{d}} = 2,45 \cdot 0,36 = 0,88\ \%;$$
$$HIP_{01} = t_{01} \cdot s_{\overline{d}} = 3,71 \cdot 0,36 = 1,34\ \%.$$

Оскільки середня різниця між вмістом білка значно перевищує HIP_{01} (3,26 > 1,34), то нульова гіпотеза H_0 спростовується і між середніми показниками є істотна різниця з імовірністю 99 %.

Якщо представлені дані білковості насіння сої оцінити між собою як незалежні спостереження, то розрахунки проводимо так:

$$s_{\overline{y}_A} = s / \sqrt{n} = \sqrt{s^2/n};$$
$$s_A^2 = \frac{\sum(Y - \overline{y})^2}{n-1} = [(50,8 - 49,0)^2 + (48,6 - 49,0)^2 + \ldots +$$
$$+ (49,8 - 49,0)^2] / 7 - 1 = 3,41;$$
$$s = \sqrt{s^2} = \sqrt{3,41} = 1,84;$$
$$s_{\overline{y}_A} = s / \sqrt{n} = 1,84 / \sqrt{7} = 0,69\ \%.$$

За цим принципом визначаємо $s_{\overline{y}_B}$:

$$s_B^2 = \frac{\sum(Y - \overline{y})^2}{n-1} = [(47,1 - 45,8)^2 + (44,4 - 45,8)^2 + \ldots +$$
$$+ (47,3 - 45,8)^2] / 7 - 1 = 4,98;$$

$$s = \sqrt{s^2} = \sqrt{4,98} = 2,23;$$
$$s_{\overline{y}_B} = s / \sqrt{n} = 2,23 / \sqrt{7} = 2,23 / 2,65 = 0,84\ \%.$$

Помилка різниці середніх d становить:

$$s_d = \sqrt{s^2_{\overline{y}_A} + s^2_{\overline{y}_B}} = \sqrt{0,69^2 + 0,84^2} = \sqrt{1,19} = 1,09;$$
$$t_{\text{факт}} = \frac{3,26}{1,09} = 2,99.$$

Фактичне значення t-критерію менше теоретичного на 1 %-му рівні значущості, тобто різницю між середніми двох вибірок не можна вважати істотною з імовірністю 99 %. Цей метод свідчить, що статистичний метод оцінки, який визначається умовами спостережень, не можна змінювати довільно.

Заходи визначення істотності між середніми двох пов’язаних рядів за допомогою t-критерію часто використовуються для порівняльної оцінки методів аналізу. Якщо два порівнювані методи забезпечують однакові результати, то за великої кількості спостережень середня різниця повинна становити $\overline{d} = 0$. За невеликої кількості спостережень $\overline{d} \neq 0$, отже, завжди є необхідність перевірки нульової гіпотези щодо відсутності постійного розходження.

Оцінка різниці вибіркових середніх рідкісних подій. Оцінку істотності середніх рідкісних подій, які підпорядковуються розподілу Пуассона, проводять за формулою:

$$t_\phi = (\overline{y}_1 - \overline{y}_2) / [\sqrt{\overline{y}_1 + \overline{y}_2}],$$

де $\overline{y}_1$ і $\overline{y}_2$ – число рідкісних подій у порівнюваних великих сукупностях.

Оцінка різниці між частками вибірки. Оцінку істотної різниці між частками за якісної мінливості проводять за тим же принципом, що і за кількісної мінливості – за t-критерієм:

$$t = d / s_d = (p_1 - p_2) / \sqrt{s^2_{p1} + s^2_{p2}},$$

де p_1 і p_2 – частки вибірок; s^2_{p1} і s^2_{p2} – помилки часток.

Цю формулу можна використовувати для визначення критерію істотності різниці між вибірковими частками, якщо порівнюються дві сукупності з однаковим обсягом вибірок ($n_1 = n_2$). Якщо $n_1 \neq n_2$, помилку різниці середніх визначають за формулою:

$$s_d = \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}.$$

Перевірка гіпотези щодо належності «сумнівного» показника до сукупності. На практиці іноді приходиться стикатися з випадками, коли частина показників вибірки значно відрізняється від більшості показників. Це може бути наслідком допущення грубих і випадкових помилок. Робастість у цьому випадку недопустима. Вирішити питання стосовно того, вибракувати сумнівні показники вибірок чи ні, можна лише на основі перевірки гіпотези про належність «сумнівних» найбільш екстремальних (крайніх) показників y_1 і y_n до даної сукупності, яке здійснюється на основі τ-критерію (*тау* критерію). Фактичне значення τ-критерію, яке є відношенням різниці між «сумнівним» і сусіднім показником до загального діапазону варіювання показників, порівнюють з теоретичним τ-критерієм на 5 %-му або 1 %-му рівні значущості. Якщо $\tau_{факт} > \tau_{теор}$, то «сумнівний» показник слід відкинути, якщо навпаки – $\tau_{факт} \le \tau_{теор}$, то рідкісний показник залишається у вибірці (нульова гіпотеза не спростовується). Теоретичні значення τ-критерію залежно від обсягу вибірки *n* для певного рівня значущості наведено в табл. 10.

Таблиця 10

Контрольні значення τ-критерію для перевірки «сумнівних» дат

n	τ-критерій для α		*n*	τ-критерій для α	
	05	01		05	01
4	0,96	0,99	14	0,40	0,50
5	0,81	0,92	16	0,37	0,47
6	0,69	0,80	18	0,35	0,45
7	0,61	0,74	20	0,33	0,43
8	0,55	0,68	22	0,32	0,41
9	0,51	0,64	24	0,31	0,40
10	0,48	0,60	26	0,30	0,39
12	0,43	0,54	30	0,28	0,37

Для розрахунку фактичного значення τ-критерію показники розміщують у порядку зростання, таким чином «сумнівні» дати опиняються по обидва краї вирівняного ряду. Сумнівними може

бути один або обидва крайні показники, тобто y_1 і y_n. За необхідності їх порівнюють із сусідніми показниками, тобто відповідно з y_2 і y_{n-1}.

Критерій τ розраховується за формулою:

для найменшого показника – $y_1 \tau = (y_2 - y_1) / (y_{n-1} - y_1)$;
для найбільшого – $y_n \tau = (y_n - y_{n-1}) / (y_n - y_2)$.

Різниця $y_{n-1} - y_1$ і $y_n - y_2$ показує діапазон мінливості показників варіаційного ряду без сумнівних показників – відповідно y_n і y_1.

Приклад 5. У вибірці з дев’яти показників урожайності зерна проса: 0,93; 1,84; 1,91; 1,93; 2,04; 2,14; 2,20 і 2,53 т/га сумнівними є крайні показники – y_1 (0,93) і y_9 (2,53). Потрібно перевірити гіпотезу щодо належності цих показників до однієї сукупності.

За представленими формулами розраховуємо значення фактичного τ-критерію та порівнюємо їх із теоретичними:

для $y_1 - \tau_ф = (y_2 - y_1) / (y_{n-1} - y_1) = (1,84 - 0,93) / (2,20 - 0,93) = 0,717$;
для $y_n - \tau_ф = (y_n - y_{n-1}) / (y_n - y_2) = (2,53 - 2,20) / (2,53 - 1,84) = 0,478$.

У цьому прикладі для y_1 $\tau_ф > \tau_{05}$ (0,717 > 0,512), таким чином, з імовірністю 95 % можна вважати, що y_1 (0,93) виходить за межі випадкових коливань і цей показник доцільно виключити з варіаційного ряду показників при розрахунку середнього значення. За більш строгого підходу (рівень ймовірності 99 %), цей показник також слід бракувати, оскільки $\tau_ф$ перевищує τ_{01} (0,717 > 0,635).

Фактичний τ-критерій максимального показника врожайності варіаційного ряду є нижчим за τ_{05} і τ_{01} (0,478 < 0,512 і < 0,635), таким чином, він не є результатом грубих помилок і підстав для його бракування немає.

Приклад 6. На семи паралельних ділянках досліду визначали збір білка сої: 520; 613; 627; 636; 641; 647 і 663 кг/га. Потрібно взнати, чи є підстави для браковки сумнівних (крайніх) показників представленого варіаційного ряду. Знаходимо фактичні значення τ-критерію та порівнюємо з теоретичним:

для $y_1 - \tau_ф = (y_2 - y_1) / (y_{n-1} - y_1) = (613 - 520) / (647 - 520) = 0,732$;
для $y_n - \tau_ф = (y_n - y_{n-1}) / (y_n - y_2) = (663 - 647) / (663 - 613) = 0,320$.

При $n = 7$ теоретичне значення τ-критерію становить 0,610 – для τ_{05} і 0,740 – для τ_{01}. У цьому прикладі y_n не можна виключати з вибірки при визначенні середнього показника $\bar{y}$, оскільки $\tau_ф < \tau_{теор.}$ для обох рівнів ймовірності.

Фактичний критерій τ для y_1 вищий, ніж τ_{05} (0,732 > 0,610), отже, для заданого рівня ймовірності (p_{95}) цей показник доцільно не враховувати при розрахунках середнього збору білка, разом із тим, за більш строгого підходу (рівень ймовірності – 99 %), його враховують при визначенні $\bar{y}$.

Визначити, чи належать крайні показники до сумнівних дат чи ні, можна також методом довірчих інтервалів. Показник підлягає бракуванню, якщо він виходить за межі $\bar{y} \pm 2s$ (для ймовірності 95 %) і за більш строгого підходу – за межі $\bar{y} \pm 3s$ (ймовірність 99 %).

При орієнтовних розрахунках стандартне відхилення s можна визначити за формулою: $s = k \cdot (y_n - y_1)$. Коефіцієнт k залежить від чисельності вибірки та становить: за n = 2–3 – 0,75; 4–5 – 0,50; від 6 до 10 – 0,33; від 11 до 25 – 0,25 і від 25 до 100 – 0,20.

Бракування сумнівних показників – це вимушений захід, до якого можна вдаватися лише у виключних ситуаціях. Дотримання методичних вимог на високому рівні, правильне виконання всіх складових експерименту на кожній стадії та професіоналізм науковця унеможливлюють допущення грубих помилок, а отже, і бракування сумнівних показників.

Оцінка відповідності між фактичним і теоретичним розподілом за критерієм χ^2. Цей критерій застосовують у випадках, якщо необхідно визначити зв'язок між емпіричним і теоретичним розподілами або двома емпіричними. Критерій χ^2 поширений в генетичному аналізі, коли потрібно перевірити чи є встановлене відхилення від теоретичного розщеплення (1:1; 1:3; 1:5; 7:4:2 і т. п.) закономірним і чи не виходить воно за межі випадкових варіацій.

Якщо позначити фактичні показники цієї групи об'єктів через f_1, $f_2, f_3, ..., f_n$, а теоретичні через $F_1, F_2, F_3, ..., F_n$, то відхилення фактичних показників від теоретичних становитимуть $f_1 - F_1, f_2 - F_2, f_3 - F_3, ...,$ $f_n - F_n$. Загальним показником відхилення фактичних показників від теоретичних буде сума відношень квадратів різниць між частотами фактичного і теоретичного розподілів до частот теоретичного розподілу для цієї групи:

$$\chi^2 = (f_1 - F_1)^2 / F_1 + (f_2 - F_2)^2 / F_2 + ... + (f_n - F_n)^2 / F_n \text{ або } \chi^2 = \\ = \sum[(f - F)^2 / F].$$

Якщо фактичні і теоретичні очікувані дані повністю збігаються, то $\chi^2 = 0$ і ймовірність запропонованої гіпотези стовідсоткова ($P =$ $= 1,0$). Якщо між фактичними і теоретичними даними є різниця, то χ^2 буде відхилятися від нуля тим більше, чим більші розбіжності між фактичними і теоретичними показниками. Граничні показники χ^2, за яких нульова гіпотеза не спростовується, знаходять за дод. Г. Якщо $\chi^2_{факт} < \chi^2_{05}$, то нульова гіпотеза щодо відповідності між фактичними і теоретичними показниками не спростовується, а якщо $\chi^2_{факт} > \chi^2_{05}$, то спростовується.

Оцінка розбіжностей між дисперсіями за F-критерієм Фішера. Цим критерієм оцінюється істотність розбіжностей між певними варіаціями показників двох дисперсій однієї сукупності. Критерієм F завжди є число не менше одиниці, оскільки в чисельник завжди ставиться більша дисперсія. Критичні (теоретичні) значення критерію F для заданого рівня ймовірності та з урахуванням числа ступенів свободи порівнюваних дисперсій знаходять за дод. Б. Якщо $F_{факт} > F_{теор.}$, то між порівнюваними дисперсіями є істотна різниця, а якщо $F_{факт} \leq F_{теор.}$, то різниця неістотна, при цьому нульова гіпотеза (H_0) щодо рівності порівнюваних дисперсій не спростовується. На цій основі (порівняння дисперсій) побудований поширений статистичний метод оцінки результатів експерименту – дисперсійний аналіз.

5.5. Порівняння вибірок за допомогою непараметричних критеріїв

Параметричні статистичні критерії (t, F та ін.) пов'язані із законом нормального розподілу та застосовуються для оцінки різниці між генеральними параметрами за вибірковими показниками порівнюваних сукупностей. Значною перевагою параметричних критеріїв служить їхня міцна статистична база, а недоліком – ускладнення розрахунків, непридатність до розподілів, які сильно відхиляються від нормального, а також дослідження якісних показників.

Поряд із параметричними критеріями для орієнтовної оцінки різниці між вибірками (особливо невеликими) застосовуються так звані непараметричні критерії, спрямовані насамперед на дослідження співвідношень рангів вихідних значень показників. *Ранг* (клас) – це число натурального ряду, яким позначається порядковий номер кожного представника впорядкованої сукупності вибірки. Ця заміна дозволяє порівняти вибірки як за кількісними, так і за якісними ознаками, значення яких не мають числового вираження, але які можна розташовувати в певному порядку. Конструкції непараметричних критеріїв відрізняються простотою.

Вся процедура складається з трьох етапів: 1) впорядкування та розміщення показників у певному порядку; 2) підрахунок сум рангів відповідно до правил цього критерію; 3) порівняння отриманого показника з теоретичними значеннями критерію. Оскільки замість самих значень показників використовують ранги, всі непараметричні методи вивчають одне питання – наскільки рівномірно показники різних вибірок «перемішані» між собою. Якщо варіанти різних вибірок більш-менш систематично чергуються у вирівняному ряді, значить, вони розподілені схожим чином і відмінностей між сукупностями немає. Якщо ж вибірки перетинаються неповно (змішуються лише краями розподілів або одна поглинає другу), то це свідчить, що вони відібрані з різних генеральних сукупностей (з різними дисперсіями або зі зміщеними центрами).

Серед непараметричних критеріїв широко розповсюдженні два: Уілкоксона-Манна-Уітні (доволі точний, проте складний у розрахунках) і Q-критерій Розенбаума (нескладний у розрахунках, однак не досить точний).

Критерій Уілкоксона-Манна-Уітні. Цей метод порівняння двох вибірок є найточнішим серед інших непараметричних критеріїв. Він полягає в тому, що всі варіанти порівнюваних сукупностей розміщують у певному порядку (ранжирують) в одному загальному ряді. Кожному показнику надається свій ранг, порядковий номер. При цьому однаковим значенням вибірки повинен відповідати один і той же середній ранг (клас). Далі ранги показників складають окремо за кожною вибіркою:

$$R_1 = \Sigma r_i,\ R_2 = \Sigma r_j,\ i = 1, 2, ..., n_1, j = 1, 2, ..., n_2;\ n = n_1 + n_2$$

і розраховують показник критерію:

$$t = (U - 0{,}5 \cdot n_1 \cdot n_2) / \sqrt{(n_1 \cdot n_2 \cdot (n + 1) / 12)}\,,$$

де $U = \max(U_1, U_2)$ – максимальне значення з двох показників:

$$U_1 = n_1 \cdot n_2 + 0{,}5 n_1 \cdot (n_1 + 1) - R_1,$$

$$U_2 = n_1 \cdot n_2 + 0{,}5 n_2 \cdot (n_2 + 1) - R_2.$$

Якщо вибірка достатньо велика ($n > 20$), розрахований показник порівнюється з табличним значенням t-критерію для $df = \infty$ и $\alpha = 0{,}1$ (тобто для верхньої 95 %-ї області нормального розподілу).

Цей метод є досить точним для вибірок обсягом більше 10. Якщо $n < 10$, то потрібно користуватися таблицями Уілкоксона-Манна-Уітні (див. дод. Д).

Приклад 7. Порівняємо вміст пігментів фотосинтезу (хлорофілу *a*) у листках рослин пшениці твердої ярої у фазу кущіння (перша вибірка) і фазу колосіння (друга вибірка) (E, мг/г):

у фазу кущіння ($n_1 = 5$): 8,0; 8,8; 9,2; 9,0; 11,7; $\bar{y} = 9{,}3$; $s = 2{,}8$;
колосіння ($n_2 = 6$): 12,3; 10,9; 9,5; 11,3; 9,2; 10,2; $\bar{y} = 10{,}6$; $s = 1{,}2$.

Високі коефіцієнти варіації (відповідно 30 і 11 %) свідчать про те, що розподіл показників, вірогідно, не відповідає нормальному розподілу. Отже, порівнювати середні вибірок слід за допомогою непараметричного критерію. Розмістимо показники сукупності в порядку зростання:

для першої вибірки: 8,0; 8,8; 9,0; 9,2; 11,7;
для другої вибірки: 9,1; 9,5; 10,2; 10,9; 11,3; 12,3.

Далі впорядковуємо всі значення разом таким чином, щоб значення кожної вибірки розміщувалися в двох окремих рядах ($E_{кущ.}$, $E_{кол.}$). Таке розміщення спрощує присвоєння рангів (ряди $R_{кущ.}$ і $R_{кол.}$) і визначення їхньої суми (R):

№	1	2	3	4	5	6	7	8	9	10	11	R
$E_{кущ.}$	8,0	8,8	9,0		9,2					11,7		
$E_{кол.}$				9,1		9,5	10,2	10,9	11,3		12,3	
$R_{кущ.}$	1	2	3		5					10		21
$R_{кол.}$				4		6	7	8	9		11	45

$$U_1 = 5 \cdot 6 + 0{,}5 \cdot 5 \cdot (5 + 1) - 21 = 24;$$
$$U_2 = 5 \cdot 6 + 0{,}5 \cdot 6 \cdot (6 + 1) - 45 = 6;$$
$$U = \max(U_1, U_2) = U_1 = 24;\ \sum n = 5 + 6 = 11;$$
$$t = (U - 0{,}5 n_1 \cdot n_2) / \sqrt{(n_1 \cdot n_2 \cdot (n + 1)/12)} =$$
$$= (24 - 0{,}5 \cdot 5 \cdot 6) / \sqrt{5 \cdot 6 \cdot 11/12} = 9 / \sqrt{27{,}6} = 1{,}71.$$

Отримане значення (1,71) менше від теоретичного t_{05}, що свідчить з імовірністю 95 % про те, що між вибірками немає істотної різниці. Зовсім інший висновок отримуємо, якщо для порівняння скористатися табличним критерієм Уілкоксона-Манна-Уітні. У цьому прикладі доцільно скористатися цим критерієм, оскільки вибірки малі (n_1 = 5 і n_2 = 6). За цих значень *n* $t_{(0,05,\ n1,\ n2)}$ = 18. Розраховане значення (U_1 = 24) більше від табличного (18), що свідчить про наявність істотної різниці між вибірками.

Критерій Q Розенбаума. Цей критерій, як і попередні, проводить оцінку достовірності двох емпіричних розподілів, але, на відміну від них, фактично не потребує обчислень. Порівняємо дві вибірки показників, які характеризують масу 20-ти рослин ячменю ярого без проведення підживлень (вибірка *A*) і після позакореневого підживлення рослин полімерними добривами (вибірка *B*). Варіаційний ряд показників вибірки *A*: 143; 127; 164; 203; 138; 156; 159; 176; 190; 133; 142; 159; 174; 198; 141; 152; 181. Варіаційний ряд вибірки *B*: 173; 241; 260; 159; 186; 160; 193; 181; 208; 219; 155; 188; 172; 193; 204; 166; 174. Максимальне та мінімальне значення вибірки *A* становить відповідно 203 і 127, вибірки *B* – 260 і 155. Далі визначаємо порядковий номер порівнюваних сукупностей. Першою має бути вибірка *B* з найбільшим показником – 260.

Для визначення фактичного *Q*-критерію знаходимо число показників другої вибірки, які перевищують максимальне значення першої вибірки (203): Q_1 = 5 (це показники 241, 260, 208, 219 і 204). Після цього визначаємо число показників першої вибірки, які менші від мінімального показника другої вибірки (155): Q_2 = 7 (це показники 143, 127, 138, 133, 142, 141 і 152). Далі розраховуємо критерій *Q* Розенбаума як суму показників: $Q = Q_1 + Q_2 = 5 + 7 = 12$. За таблицею, наведеною в дод. Е, знаходимо критичне значення $Q_{(0,05,\ n1,\ n2)} = 7$. Оскільки фактичний показник (12) більший від теоретичного (7), робимо висновок про існування достовірної різниці між вибірками, а отже, про вплив підживлень рослин полімерними добривами на підвищення їхньої маси.

Слід мати на увазі, що цей метод не точний, а орієнтовний, що дає тільки приблизний результат. Він є ефективним лише за умови існування значної різниці між вибірками.

6. Дисперсійний аналіз

6.1. Сутність і алгоритм дисперсійного аналізу

Існує цілий ряд статистичних методів, які дозволяють визначати силу, напрям і закономірності впливу чинників на результат у генеральній або вибірковій сукупності. Найбільш поширеним з них є дисперсійний аналіз (метод), який у зарубіжній літературі зветься *ANOVA* – «*Analisis of variance*». Цей метод необхідний для визначення впливу окремих чинників на результати експерименту, а також для подальшого планування аналогічних експериментів.

Уперше дисперсійний аналіз був запропонований англійським математиком Рональдом Фішером для обробки результатів агрономічних досліджень з виявлення умов, за яких вирощуваний сорт рослин формує максимальну врожайність.

Сутність методу дисперсійного аналізу полягає у визначенні впливу окремих дисперсій (загальної, варіантів, повторень, помилок), і в подальшому визначенні сили (частки) впливу досліджуваних чинників (оцінки ролі кожного з чинників або їхнього сумісного впливу) на результативні показники (показник).

В основі дисперсійного аналізу лежить припущення про те, що одна частина змінних може розглядатися як причини (чинники, незалежні змінні), а друга – як наслідки (залежні змінні). Незалежні (факторні) змінні іноді називають регулюючими чинниками саме тому, що в експерименті дослідник має змогу управляти ними та аналізувати отриманий результат.

Факторними (варіативними) називають ті ознаки, які впливають на досліджуване явище, *результативними* – ознаки, які змінюються під впливом факторних показників.

Дисперсійний метод полягає в аналізі відхилень усіх показників досліджуваної сукупності від середнього арифметичного. За міру відхилення беруть дисперсію – середній квадрат відхилень. Відхилення, які спричиняються впливом варіативної ознаки, порівнюють з показниками відхилень, викликаних випадковими помилками. Якщо відхилення, викликані варіативними показниками, більші, ніж випадкові відхилення, то вважають, що варіативний чинник істотно впливає на результативний показник.

За допомогою дисперсійного аналізу досліджують вплив однієї або кількох незалежних змінних на одну залежну змінну (одномірний аналіз) або на декілька залежних змінних (багатомірний аналіз).

У звичайному випадку незалежні змінні набувають лише дискретних значень та належать до номінальної або порядкової шкали. У цьому разі також говорять про факторний аналіз. Якщо незалежні змінні належать до інтервальної шкали або до шкали відношень, їх називають коваріаціями, а відповідний аналіз – коваріаційним.

Важливим методичним значенням для застосування дисперсійного аналізу є правильне формування вибірки. Залежно від поставленої мети вибіркові групи можуть формуватися довільно, незалежно одна від одної (контрольна та експериментальна групи для вивчення певного показника). Такі вибірки звуться незалежними.

Досить часто результати впливу чинників досліджують в одній і тій самій сукупності (наприклад, в одних і тих самих рослин) до та після певного впливу (обробка посівів пестицидами, підживлення в різні фази розвитку тощо), такі вибірки звуться залежними.

Дисперсійний аналіз належить до групи параметричних методів, і тому його слід застосовувати тільки тоді, коли доведено, що цей розподіл є нормальним.

У статистиці існують певні критерії контролю так званої «нормальності» показників і підстави для їхнього застосування немає до тих пір, доки обсяги вибірок не будуть достатньо великими. Незалежність передбачає відсутність зв'язку між показниками помилок за варіантами. Оскільки сусідні ділянки в польовому досліді схильні до значної залежності між собою, ніж з іншими ділянками, то важливо запобігти розміщенню однакових варіантів на сусідніх ділянках.

Для проведення дисперсійного аналізу можна використовувати і якісні (колір, форма), і кількісні ознаки (кількість рослин, урожайність, маса рослин тощо).

Умови проведення дисперсійного аналізу такі:
– завданням дослідження є визначення сили впливу одного чи кількох чинників на результат або визначення сили сумісного впливу комплексу чинників;
– досліджувані чинники повинні бути незалежні (не пов'язані) між собою;
– підбір груп для проведення досліджень слід проводити рендомізовано (випадковий відбір).

Вихідним матеріалом для дисперсійного аналізу служать дані досліджень трьох і більше вибірок, які можуть бути як однаковими, так і різними за обсягом, як пов'язаними, так і непов'язаними. За кількістю регульованих чинників дисперсійний аналіз буває однофакторним (досліджується вплив одного чинника на результати експерименту), двофакторним (досліджується вплив двох чинників) і багатофакторним (дозволяє оцінити не тільки вплив кожного з досліджуваних чинників окремо, але і їхню взаємодію).

Класичний дисперсійний аналіз проводять у декілька етапів:
– побудова дисперсійного комплексу;
– розрахунок середніх квадратів відхилень;
– обчислення дисперсій;
– порівняння дисперсії варіантів (факторіальної) та дисперсії помилок (залишкової);
– оцінка результатів за допомогою теоретичних значень розподілу Фішера-Снедекора (дод. В).

Види дисперсійного аналізу. Дисперсійний аналіз схематично можна поділити на декілька категорій. Цей поділ здійснюється залежно від того, по-перше, скільки чинників бере участь у цьому досліді, по-друге, скільки змінних міститься в цьому масиві і, по-третє, як співвідносяться між собою значення вибірок.

Однофакторний дослід, у якому вивчається вплив одного чинника, може мати два різновиди:
– аналіз непов'язаних (різних) вибірок. Наприклад, під час визначення врожайності різних сортів рослин нульову гіпотезу можна сформулювати таким чином: середня врожайність одного сорту не відрізнятиметься від середньої врожайності іншого, тобто вона не залежить від сортоособливостей;
– аналіз пов'язаних вибірок, тобто двох вимірювань, проведених з одним масивом рослин у різних умовах. Наприклад, визначення впливу певного агротехнічного чинника на динаміку росту однієї групи рослин або наростання площі листя рослин одного масиву (сукупності) у динаміці розвитку тощо.

У випадку, якщо одночасно досліджується вплив двох або більше чинників, маємо справу з багатофакторним дисперсійним аналізом (комплексом), який також можна класифікувати за типом вибірки. Зазвичай в агрономії застосовують одно- та двофакторні, максимум трифакторні дисперсійні комплекси. Багатофакторні комплекси можна досліджувати, послідовно аналізуючи одно- або двофакторні комплекси, які виділяють з усієї досліджуваної сукупності.

Залежно від природи досліджуваного чинника розрізняють дві принципові моделі дисперсійного аналізу.

Перша модель: градаціями чинника або варіантами, є фіксовані норми висіву, дози добрив, способи сівби тощо. Нульова гіпотеза (H_0) записується так: $\mu_1 = \mu_2 = \mu_n$, тобто варіанти досліду істотно не розрізняються і складають одну генеральну сукупність.

Друга модель: градаціями чинника служать випадкові вибірки. Дисперсія варіантів дорівнює нулю ($s_v = 0$). Наприклад, мінливість врожайності одного сорту в різних регіонах вирощування.

Для багатофакторних дослідів допускаються змішані моделі, наприклад сорти та регіони вирощування. Подібна ситуація складається в однофакторному польовому досліді за наявності повторень, контрольованих джерел варіації.

Послідовність проведення розрахунків простого дисперсійного аналізу *(однофакторного польового досліду)* така:

1. Підготовка таблиці для вихідних даних, сум і середніх.
2. Розрахунок загального числа дат (спостережень): $N = l \cdot n$, де l – кількість варіантів у досліді, n – кількість повторень.
3. Визначення сум за кожним варіантом: $V_i = \sum y_i$, середніх за варіантами ($\bar{y}_i = \sum y_l / n$) і загальної середньої по досліду ($\bar{y} = \sum y_{ln} / N$).
4. Розрахунок корегуючого чинника (поправки): $C(\Delta) = (\sum y)^2 / N$.
5. Визначення загальної суми квадратів відхилень (загальної варіації): $C_y = \sum y^2 - C(\Delta)$.
6. Піднесення сум за варіантами в квадрат (V^2) і знаходження їхньої суми: $\sum V^2$.
7. Знаходження варіації варіантів, для чого суми їхніх квадратів ділять на кількість повторень експерименту (n) і віднімають корегуючий чинник (C): $C_v = \sum V^2 / n - C$.
8. Піднесення сум за повтореннями у квадрат (P^2) і знаходження їхньої суми: $\sum P^2$.
9. Знаходження варіації повторень, для чого суми квадратів повторень ділять на кількість повторень експерименту (n) і віднімають корегуючий чинник (C): $C_p = \sum P^2 / n - C$.
10. Визначення остаточної (випадкової) варіації, яка становить різницю між загальною сумою квадратів відхилень (C_y) та сумою квадратів відхилень варіантів і повторень: $C_z = C_y - C_v - C_p$.
11. Визначення частки варіантів, повторень і помилок у загальній варіабельності (у відсотках). Для варіантів – $C_v \cdot 100 / C_y$, для повторень – $C_p \cdot 100 / C_y$, для помилок – $C_z \cdot 100 / C_y$.
12. Поділ загального числа ступенів свободи на складові: $N - 1$ (загальне) $= (v - 1) + (p - 1) + (v - 1) \cdot (p - 1)$, де $(v - 1)$ – кількість ступенів свободи варіантів; $(p - 1)$ – кількість ступенів свободи повторень, $(v - 1) \cdot (p - 1)$ – кількість ступенів свободи помилок (залишку). Кількість ступенів свободи помилок легше визначати як різницю між загальним числом ступенів свободи та числом ступенів свободи варіантів і повторень: $Cв_z$ (число ступенів свободи помилок) $= (N - 1) - [(v - 1) + (p - 1)]$.
13. Визначення середніх квадратів (дисперсії) варіантів, повторень та помилок. Для цього розраховані для них суми квадратів ділять на відповідні показники числа ступенів свободи: s_v^2 (середній квадрат дисперсії варіантів) $= C_v / (v - 1)$; s_p^2 (середній квадрат дисперсії повторень) $= C_p / (p - 1)$; s_z^2 (середній квадрат дисперсії помилок) $= C_p / [(v - 1) \cdot (p - 1)]$.
14. Розрахунок фактичного F-критерію: s_v^2 / s_z^2 і визначення його табличного значення відповідно до числа ступенів свободи чисельника і знаменника для потрібного рівня ймовірності (95 або 99 %).
15. Зведення всіх розрахунків у підсумкову таблицю дисперсійного аналізу.
16. Зіставлення фактичного F-критерію з табличним. Якщо фактичний F-критерій більший від табличного, то нульова гіпотеза (H_0) спростовується, а це свідчить про те, що між варіантами є істотна різниця. Якщо $F_ф < F_m$, то різниці між варіантами немає, отже, вони належать до однієї сукупності. Теоретичні значення критерію F для прийнятого в дослідженнях рівня значущості знаходять за дод. В. В основному обирають 5 %-й, а за більш строгого підходу 1 %-й або навіть 0,1 %-й рівень значущості.

Для оцінки точності досліду додатково розраховують ступінь варіації: $V = \sqrt{S_z^2} \cdot 100 / \bar{y}$ і точність досліду (відносну помилку середніх): $E = \sqrt{S_z^2 / n} \cdot 100 / \bar{y}$. Якщо $V < 7$ %, а $E < 3$ %, точність досліду вважається високою.

Якщо обробляють однофакторні статистичні комплекси у вегетаційному досліді (без повторень), то загальна мінливість показників, яка вимірюється загальною сумою варіантів C_y, розділяється на два компоненти: варіювання між варіантами (вибірками) C_v і всередині вибірок – C_z. Отже, у загальній формі варіабельність показника можна подати так: $C_y = C_v + C_z$.

Тут варіація між вибірками (варіантами) представляє ту частину загальної дисперсії, яка зумовлена дією досліджуваних чинників, а дисперсія всередині вибірок характеризує випадкове варіювання досліджуваної ознаки, тобто помилку експерименту.

Загальне число ступенів свободи $(N – 1)$ також поділяють на дві частини: ступені свободи для варіантів $(l – 1)$ та помилок $(N – l)$, а саме: $N – 1 = (l – 1) + (N – l)$.

За наявності загальних принципів можливі різні моделі або конкретні схеми дисперсійного аналізу, які відображають умови та методику проведення експерименту. Загальну схему дисперсійного аналізу однофакторних комплексів представлено в табл. 11.

Важливо відзначити, що всі суми квадратів більші від нуля. Від'ємне значення суми свідчить про те, що в ході розрахунків припустилися помилки, яку потрібно знайти і виправити.

Для кожного виду експерименту є певна математична модель або схема проведення дисперсійного аналізу. Так, в однофакторному польовому досліді, поставленому методом неорганізованих повторень, варіативність показників пов'язана з двома складниками: варіантами та випадковими помилками. У досліді, поставленому методом організованих повторень, виділяють уже три варіативні компоненти: варіанти, повторення та помилки. У разі закладання експериментів методом латинського квадрата або прямокутника джерел мінливості показників буде чотири: варіанти, стовпці, ряди та помилки.

Зі збільшенням факторності (досліджуваних чинників) зростає чисельність джерел варіювання досліджуваної ознаки. Зокрема, у двофакторному польовому досліді, поставленому методом організованих повторень, загальна сума квадратів варіантів (C_v) складається з трьох компонентів: C_A (варіювання чинника A), C_B (варіювання чинника B) і C_{AB} (варіювання взаємодії AB), а в трифакторному – вже із семи компонентів (джерел): C_A (варіювання чинника A), C_B (варіювання чинника B), C_C (варіювання чинника C), C_{AB} (варіювання взаємодії AB), C_{AC} (варіювання взаємодії AC), C_{BC} (варіювання взаємодії BC), C_{ABC} (варіювання взаємодії ABC). Загальна сума квадратів для дво- і трифакторних польових дослідів представлена такими рівняннями:

Таблиця 11

Алгоритм дисперсійного аналізу однофакторних комплексів (дослідів)

Вид досліду	Суми квадратів (чисельник) і ступені свободи (знаменник)				
	загальна	(повторень) рядів	стовпців	варіантів	помилок
Вегетаційні та польові досліди, поставлені методом неорганізованих повторень	$\frac{C_v}{(N-1)}$	–	–	$\frac{C_v}{(l-1)}$	$\frac{C_v}{(N-l)}$
Польові та вегетаційні досліди, поставлені методом: а) організованих повторень;	$\frac{C_y}{(N-1)}$	$\frac{C_p}{(n-1)}$	–	$\frac{C_v}{(l-1)}$	$\frac{Cz}{(n-1)\cdot(l-1)}$
б) латинського квадрата;	$\frac{C_y}{(N-1)}$	$\frac{C_p}{(n-1)}$	$\frac{C_c}{(n-1)}$	$\frac{C_v}{(l-1)}$	$\frac{Cz}{(n-1)\cdot(l-1)}$
в) латинського прямокутника	$\frac{C_y}{(N-1)}$	$\frac{C_p}{(n-1)}$	$\frac{C_c}{(n-1)}$	$\frac{C_v}{(l-1)}$	$\frac{Cz}{(n-1)\cdot(l-1)}$

$$C_Y = (C_A + C_B + C_{AB}) + C_p + C_z;$$
$$C_Y = (C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}) + C_p + C_z.$$

Відповідно до наведених компонентів варіювання результативної ознаки, розщеплюють і загальне число ступенів свободи кожної компоненти.

Як правило, багатофакторні досліди ставлять методом розщеплених ділянок. Цей метод також застосовують для закладання двофакторних дослідів. Вибір цього методу зумовлений як характером експерименту (пріоритет визначення впливу певних комбінацій чинників), так і технологічною можливістю постановки досліду. Наприклад, якщо планується провести трифакторний експеримент з вивчення впливу попередників, способів сівби та позакореневих підживлень, технологічно поставити цей дослід, не виділивши попередників ділянками головного порядку, фактично не можливо. Логічно поставити цей дослід методом розщеплених ділянок, виділивши ділянками головного порядку варіанти попередника.

Головним недоліком методу розщеплених ділянок є неможливість порівняння різних комбінацій досліджуваних чинників з однаковою точністю, що пов'язано з розміщенням варіантів на розщеплених ділянках. Загальна помилка досліду у цьому випадку включає помилку для варіантів, розміщених на ділянках першого порядку (C_{ZI}), помилку для варіантів, розміщених на ділянках другого порядку (C_{ZII}) і под. Ці помилки надалі використовуються для оцінки ефективності головних ефектів та їхньої взаємодії. Загальна сума квадратів для дво-, три- і чотирифакторних польових дослідів, поставлених методом розщеплених ділянок, становить відповідно:

$$C_Y = (C_A + C_B + C_{AB}) + C_p + C_{ZI} + C_{ZII};$$
$$C_Y = (C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}) + C_p + C_{ZI} + C_{ZII} + C_{ZIII};$$
$$C_Y = (C_A + C_B + C_C + C_D + C_{AB} + C_{AC} + C_{AD} + C_{BC} + C_{BD} + C_{CD} + C_{ABC} + C_{ABD} + C_{ACD} + C_{BCD} + C_{ABCD}) + C_p + C_{ZI} + C_{ZII} + C_{ZIII} + C_{ZIIII}.$$

Отже, у дослідах, поставлених методом розщеплених ділянок, оцінка істотності різниці спиратиметься не на одну загальну суму квадратів помилок, як у дослідах, поставлених методом рендомізованих повторень, а на суми квадратів, розрахованих для ділянок першого, другого та наступних порядків. Представлені схеми свідчать про значне ускладнення проведення розрахунків при збільшенні кількості досліджуваних чинників. Так, якщо у двофакторному досліді, поставленому методом розщеплених ділянок, кількість джерел впливу на варіабельність результатів експерименту становить 6, то у трифакторному досліді – вже 11, а в чотирифакторному – взагалі 20.

Для спрощення розрахунків сум квадратів вихідні показники можна перетворювати шляхом віднімання від них умовного початку A – цілого числа, близького до середнього показника в досліді $\bar{y}$. Зміна початку відліку не впливає на співвідношення сум квадратів і в той же час дозволяє працювати з маленькими цифрами.

Дисперсійний аналіз дає змогу визначити частку кожного чинника в загальній мінливості (дисперсії) показника, яку приймають за 100 % або одиницю. Так, частки варіантів, повторень та помилок визначаються відповідно за співвідношеннями: $\eta^2_V = C_V / C_Y$; $\eta^2_P = C_P / C_Y$ і $\eta^2_Z = C_Z / C_Y$. Загальна сума часток впливу варіантів, повторень та помилок дорівнює одиниці ($\eta^2_V + \eta^2_P + \eta^2_Z = 1$). У двофакторних і багатофакторних дослідах частки впливу окремих досліджуваних чинників також визначають за цим самим принципом. Так, у двофакторному польовому досліді частки чинників A, B і їхньої взаємодії AB розраховують відповідно за співвідношеннями: $\eta^2_A = C_A / C_Y$; $\eta^2_B = C_B / C_Y$; $\eta^2_{AB} = C_{AB} / C_Y$. Сума цих часток показує загальну частку впливу варіантів: $\eta^2_V = \eta^2_A + \eta^2_B + \eta^2_{AB}$.

Широкого розповсюдження дисперсійний метод набув завдяки важливим перевагам порівняно з методом попарних порівнянь за t-критерієм, а саме:

– цим методом можна обробляти прості та складні, однорічні та багаторічні, однофакторні та багатофакторні досліди;
– замість визначення помилок для кожного варіанта, в дисперсійному аналізі використовується загальна помилка середніх, яка спирається на більше число спостережень і тому є об'єктивнішою для оцінки;
– дисперсійний аналіз дозволяє зручно, у вигляді істотних різниць подати результати статистичної обробки.

У наступній (другій) частині на конкретних прикладах буде показано техніку проведення статистичних розрахунків результатів дослідів, поставлених методом розщеплених ділянок.

6.2. Оцінка істотності різниць між середніми

Зіставлення теоретичного і фактичного критерію Фішера (F) дає підставу зробити висновок щодо існування достовірної різниці між порівнюваними варіантами досліду, проте оцінка цього співвідношення має лише узагальнювальний характер, тому не можна отримати відповідь на питання про те, різниця між якими саме середніми показниками є достовірною. Отже, визначення фактичного критерію Фішера у дослідах із достовірною різницею між варіантами не є кінцевим етапом статистичного аналізу. У разі спростування нульової гіпотези (H_0) необхідно визначити, між якими саме варіантами є достовірна різниця.

Найчастіше оцінку значущості різниці між середніми показниками проводять за *найменшою істотною різницею* (*НІР*) (табл. 12).

Існує три способи порівняння середніх за варіантами на базі *НІР:*
- порівняння всіх можливих комбінацій попарних порівнянь середніх показників за варіантами між собою;
- порівняння досліджуваних варіантів із простим (один) або складним (середнє двох або більше варіантів) контролем;
- порівняння середніх за варіантам із середнім по досліду.

Таблиця 12

Можливі пари порівняння для середніх варіантів у досліді на базі *НІР*

База порівняння	Кількість порівнянь (пар)	Формула узагальненої помилки різниці s_d	Примітка
Будь-яке середнє $\bar{y}_i$ (кожна з кожною)	$l(l-1)/2$	$s_d = \sqrt{\frac{s_Z^2}{n}}$	s_Z^2 – дисперсія помилки
Простий стандарт (один) ($n_{st} = n$); $\bar{y}_{st} = \sum y_{st}/n$	$l-1$	$s_d = \sqrt{\frac{2s_Z^2}{n}}$	n (n_{st}) – повторність дослідних варіантів
Складний (середній) стандарт $n_{st} = kn$ ($k \geq 2$); $\bar{y}_{st} = \sum y_{st}/kn$	$l-k$	$s_d = \sqrt{s_Z^2 \cdot [\frac{1}{n} + \frac{1}{nk}]}$ за k = 2, а $n_1 \neq n_2$ $s_d = \sqrt{s_Z^2 \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}$	k – кількість контрольних варіантів
Середнє по досліду $\bar{y}_0 = \sum y/n$	l	$s_{\overline{d}} = \sqrt{s_Z^2 \cdot \frac{l-1}{n \cdot l}}$	$s_{\overline{d}}$ – помилка середньої різниці

Критерій $НІР = t_{sd}$ показує мінімальну помилку для різниці двох вибіркових середніх. Якщо фактична різниця $d \geq НІР$, то вона істотна, а якщо $d < НІР$ – неістотна.

Зазвичай урожайність досліджуваних варіантів порівнюють із контрольним варіантом, проте іноді потрібно порівняти середні показники за варіантами із середнім по досліду. У цьому разі помилку різниці середніх визначають за формулою: $s_{\overline{d}} = \sqrt{s_Z^2 \cdot (l-1)/(n \cdot l)}$. Оскільки $s_{\overline{y}} = \sqrt{s_n^2/n}$, формула набуває вигляду: $s_{\overline{d}} = s_{\overline{y}} \cdot \sqrt{(l-1)/l}$.

Якщо доводиться порівнювати варіанти, які спираються на різну кількість повторень, помилку різниці середніх розраховують за формулою:

$$s_{\overline{d}} = \sqrt{s_Z^2 \cdot (n_1 + n_2) / n_1 \cdot n_2}\ ,$$

де s_Z^2 – дисперсія помилок, яка береться з підсумкової таблиці дисперсійного аналізу, а n_1 і n_2 – кількість повторень у порівнюваних групах (варіантах).

Після визначення помилки різниці середніх (s_d) розраховують найменшу істотну різницю (*НІР*) в абсолютних і відносних показниках для потрібного рівня значущості:

– для α_{05}: $НІР_{05} = t_{05} \cdot s_d$; $НІР_{05}\ \% = t_{05} \cdot s_d \cdot 100 / \bar{y}$;
– для α_{01}: $НІР_{05} = t_{01} \cdot s_d$; $НІР_{01}\ \% = t_{01} \cdot s_d \cdot 100 / \bar{y}$;
– для α_{001}: $НІР_{05} = t_{001} \cdot s_d$; $НІР_{001}\ \% = t_{001} \cdot s_d \cdot 100 / \bar{y}$.

Значення *t*-критерію для заданого рівня ймовірності та числа ступенів свободи дисперсії помилок знаходять у дод. Б. Індекси для *НІР* та критерію *t* – це показники рівня значущості (5; 1 і 0,1 %-й). Нагадаємо, що 5 %-му рівню значущості відповідає 95 %-й рівень ймовірності, 1 %-му – 99 %-ий і 0,1 %-му – 99,9 %-й.

Залежно від потреби у встановленні точності статистичних висновків, кількості варіантів досліду та вимог щодо порівняльної оцінки середніх показників за варіантами (з контролем, між собою, із середнім показником у досліді), існують різні методичні підходи для визначення *НІР*.

Якщо кількість варіантів у досліді незначна (2–4), то для їхнього порівняння можна покластися на одне значення *НІР*. При цьому максимальна точність порівняння досягається лише для однієї пари середніх. Зі збільшенням кількості порівнянь при використанні одного значення *НІР* ризик помилитися під час проведення оцінки і визнати неістотні різниці істотними помітно зростає.

У дослідах зі значною кількістю варіантів (більше шести) і з кількістю повторень не менше чотирьох, обґрунтовано застосування критерію Перегудова. Утричі збільшена помилка середньої – $s_{\overline{y}}$ і є критерієм істотності. Якщо фактична різниця перевищує $3s_{\overline{y}}$, то вона істотна на 5 %-му рівні, а якщо $d < 3s_{\overline{y}}$ – різниця неістотна.

У дослідах з незначною кількістю варіантів (до шести) і кількістю повторень до чотирьох цю оцінку застосовувати недоцільно, оскільки вона дає збільшене число істотних різниць. Це легко зрозуміти, якщо розглянути механізм виникнення критерію $3s_{\overline{y}}$. У процесі дисперсійного аналізу визначається однакова для всіх помилка середніх ($s_{\overline{y}} = s_{\overline{y}_1} = s_{\overline{y}_2} = s_{\overline{y}_n}$) і, відповідно, єдина для всіх помилка різниці середніх:

$$s_d = \sqrt{s_{\overline{y}_1}^2 + s_{\overline{y}_2}^2} = \sqrt{2 \cdot s_{\overline{y}}^2} = 1{,}414 s_{\overline{y}}.$$

Якщо кількість ступенів свободи дисперсії помилок дорівнює 16, то t_{05} становить 2,12, звідси найменша істотна різниця на 5 %-му рівні значущості дорівнює: $HIP_{05} = t_{05} \cdot s_d = 2{,}99 s_{\overline{y}}$, або $\approx 3 s_{\overline{y}}$. Зі збільшенням числа ступенів свободи з 16 до 30 t_{05} зменшується лише з 2,12 до 2,04, або менше ніж на 4,0 %. Тож навіть якщо число ступенів свободи дисперсії помилок становить 30, $HIP_{05} = t_{05} \cdot s_d = 2{,}04 \times \times 1{,}414 s_{\overline{y}} = 2{,}9 s_{\overline{y}} \approx 3 s_{\overline{y}}$.

Таким чином, для дослідів з кількістю ступенів свободи дисперсії помилок ≥ 16 утричі збільшена $s_{\overline{y}}$ – це фактично і є HIP_{05}. Якщо кількість ступенів свободи дисперсії залишку < 16, коефіцієнт перед $s_{\overline{y}}$ зростає, і найдужче в тому разі, якщо кількість ступенів свободи залишку зменшується нижче семи, тобто у дослідах з незначною кількістю варіантів. Звідси випливає, що застосування критерію Перегудова ($s_{\overline{y}}$) обґрунтоване для дослідів з числом ступенів свободи дисперсії помилок понад 16. Якщо кількість ступенів свободи менше 16, то використання критерію $s_{\overline{y}}$ для оцінки неправомірне, оскільки така оцінка дає збільшену кількість істотних різниць.

Для підвищення точності порівняння середніх використовують рангові критерії, які підвищують поріг найменшої істотної різниці за поступового віддалення середніх показників за варіантами один від одного у вирівняному ряду (дод. Ж, И).

Наприклад, статистична оцінка врожайності дев'яти гібридів кукурудзи в досліді із сортовивчення виявила такі показники узагальненої помилки середніх і різниці середніх: $s_{\overline{y}} = 1{,}96$ і $s_d = 2{,}43$. Кількість ступенів свободи для варіації залишку (C_Z) становить 15. Чим далі одне від одного розміщені середні показники за варіантами впорядкованого ряду рангових груп, тим більшим є показник HIP_{05} для їх порівняння (табл. 13).

Перший метод підвищення точності порівняння дослідних середніх був розроблений Джоном Тьюкі (ранговий критерій для чисельних порівнянь q). Фактичний критерій q одержують з відношення різниці між будь-якою парою середніх впорядкованого ряду до

Таблиця 13

Диференціація показника HIP_{05} на основі критеріїв: Тьюкі-Ньюмана (q), Дунетта (D) і Дункана (C) порівняно з однією HIP на основі t-критерію для впорядкованого ряду середніх ($n_z – 1 = 15$; $s_{\overline{y}} = 1{,}72$; $s_d = 2{,}43$)

Ранг середнього у впорядкованому ряді	HIP_{05}			
	$t_{05} \cdot s_d$	$q \cdot s_{\overline{y}}$	$D \cdot s_d$	$C \cdot t_{05} \cdot s_d$
2	5,18	5,18	5,93	5,18
3	5,18	6,31	6,34	5,44
4	5,18	7,02	6,63	5,59
5	5,18	7,52	6,85	5,70
6	5,18	7,91	7,02	5,80
7	5,18	8,22	7,17	5,80
8	5,18	8,50	7,29	5,85

Примітка. Для сусідніх середніх $r = 1$ (критерій D) і $r = 2$ (критерій q і C).

загальної помилки середніх (до помилки досліду): $q = |y_1 - y_2| / s_x$. Табличний показник q-критерію визначається рівнем значущості (05 або 01), кількістю ступенів свободи помилки та позицією (рангом r) середнього стосовно до бази порівняння. Заключна частина оцінки аналогічна t-критерію, тобто за $q_ф > q_{теор}$ різниця між середніми істотна. На базі q-критерію розраховують критичний поріг Тьюкі-Ньюмана HSD (*honest significant difference*), який є своєрідною аналогією HIP. Якщо різниця між варіантами більша від показника HSD, то варіанти істотно відрізняються один від одного, якщо ж $HSD \geq d$, варіанти не відрізняються між собою.

Оскільки $HSD = q \cdot s_{\overline{y}}$, а у дослідах з двома варіантами $q = t_{05}\sqrt{2}$, то $HSD = HIP_{05} = \sqrt{2} \cdot t_{05} \cdot s_{\overline{y}}$. Якщо в досліді більше двох варіантів, то $D > HIP_{05}$. Для дослідів з 6–15-ма варіантами при 4–6-кратній повторності значення показника HSD у 1,5–1,8 раза більше за HIP_{05}. Це свідчить про те, що показник HSD є більш надійною основою для оцінки середніх показників за варіантами, оскільки він базується не тільки на кількості ступенів свободи дисперсії залишку (помилок), але і враховує кількість варіантів у досліді.

Ранговий критерій Дункана (C) служить для оцінки середніх показників за варіантами, тобто залежно від значущості розподіляє їх по варіантах на різні класи (групи) і є свого роду поправкою, яка нівелює фактор поступового викривлення оцінки порівняння середніх показників за варіантами зі збільшенням різниці між ними. Якщо різниця між середніми за варіантами удвічі перевищує HIP_{05}, то оцінку цієї різниці та розподіл показників на рангові групи доцільно проводити з використанням рангового критерію Дункана. Зі збільшенням числа ступенів свободи дисперсії помилок і показника рангу критерій Дункана поступово зростає (див. дод. И).

Для чисельних порівнянь і окремих середніх різної повторності, і їхніх груп з різною чисельністю на основі контрастів був уведений ранговий критерій Шефе та відповідний показник найменшої істотної різниці: $HIP_\alpha = s_x \cdot (v - 1)F_\alpha$. Зараз більш поширений критерій Ньюмана, який, на відміну від критерію Шефе, має вищий поріг достовірності. Спрощеною версією методу Шефе є метод контрастів. Водночас цей метод забезпечує достатньо точну оцінку різниці між середніми показниками за варіантами.

Приклад 1. У порівняльному досліді, поставленому у чотирьох повтореннях, урожайність шести нових сортів пшениці твердої ярої була вищою порівняно зі стандартом. Урожайність стандарту (y_{st}) становила 3,2 т/га, першого нового сорту (y_1) – 5,3, другого (y_2) – 4,3, третього (y_3) – 3,7, четвертого (y_4) – 5,3, п’ятого (y_5) – 4,7 і шостого (y_6) – 3,4 т/га. За F-критерієм установлено достовірність різниць між варіантами. Дисперсія помилок (s_Z^2) становить 0,08. Кількість ступенів свободи для варіантів дорівнює $l - 1 = 6$. Визначимо істотність різниць контрастним методом. Першим контрастом (ортогональним порівнянням) C_1 перевіряють істотність перевищення врожайності зерна шістьма новими сортами порівняно із сортом-стандартом. Стандарт одержує ортогональний коефіцієнт 6, а кожний новий сорт – коефіцієнт – 1. $C_1 = (6 \cdot 3{,}2) - (5{,}3 + 4{,}3 + 3{,}7 + 5{,}3 + 4{,}7 + 3{,}4) = -7{,}5$. Суму квадратів відхилень визначають за формулою: $СКС_1 = nC_1^2/k_1^2$, де k_1^2 – сума квадратів ортогональних коефіцієнтів першого контрасту. У наведеному прикладі $k_1^2 = 6^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 = 42$.

$СКС_1 = n \cdot C_1^2 / k_1^2 = 4 \cdot (-7{,}5)^2 / 42 = 5{,}35$. Оскільки кількість ступенів свободи для кожного порівняння дорівнює 1, то $s_1^2 = СКС_1 = 5{,}35$, і тоді $F_ф = 5{,}35 / 0{,}08 = 66{,}8$. Якщо кількість ступенів свободи дисперсії чисельника дорівнює 1, а знаменника (дисперсії помилок) – 18, теоретичний $F_{05} = 4{,}41$ (див. дод. В). Оскільки $F_{факт.} > F_{05}$, то врожайність зерна нових сортів пшениці ярої істотно перевищує стандартний сорт.

Другим контрастом порівнюють урожайність зерна першого нового сорту з рештою сортів. У цьому випадку цей сорт позначається коефіцієнтом + 5, а решта нових сортів по – 1. Оскільки $C_2 = (5 \cdot 5{,}3) - (4{,}3 + 3{,}7 + 5{,}3 + 4{,}7 + 3{,}4) = 5{,}1$, а $k_1^2 = 5^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 = 30$, то $СКС_2 = 4 \cdot 5{,}1^2 / 30 = 3{,}47$. Фактичний критерій F становить: $3{,}47 / 0{,}08 = 43{,}4$. Для першого нового сорту $F_{факт.} > F_{05}$, що дає підставу рекомендувати цей сорт до подальшого виробничого випробування як один з найперспективніших у досліджуваній серії. Подібним чином проводять порівняння решти нових сортів.

Іноді істотність різниць між середніми перевіряють за F-критерієм, який визначають для кожної порівнюваної пари середніх за формулою:

$$F_{факт.} = (d^2 / s^2) \cdot (n_1 \cdot n_2) / (n_1 + n_2),$$

де d – різниця між середніми; s^2 – дисперсія помилок (залишку); n_1, n_2 – кількість повторень порівнюваних груп.

Якщо $n_1 = n_2$ формула набуває вигляду: $F_{факт.} = d^2 \cdot n / (2s^2)$.

Фактичне значення критерію F порівнюють з теоретичним для числа ступенів свободи дисперсії чисельника, що дорівнює 1, і потрібного числа ступенів свободи дисперсії помилок. Якщо $F_{факт.} > F_{теор.}$, то різниця між середніми істотна, а якщо $F_{факт.} < F_{теор.}$ – неістотна, а це свідчить, що нульова гіпотеза (H_0) не спростовується.

Під час проведення дисперсійного аналізу істотні різниці між середніми зазвичай визначають за HIP_{05}. Для орієнтовних висновків можна застосовувати критерій Перегудова ($3s_{\bar{y}}$), а для більш точної оцінки, особливо у дослідах зі значною кількістю варіантів і мінімальною кількістю повторень (3–4), доцільно застосовувати критерій Дункана (C), Тьюкі-Ньюмана (q) або Дуннетта (D). Якщо в досліді розглядають лише два варіанти, то для визначення достовірності різниці між ними розраховувати HIP не потрібно, достатньо лише порівняти $F_{факт.}$ і $F_{теор}$.

Усі варіанти агротехнічних дослідів і дослідів із сортовивчення за результатами оцінки поділяють на три групи:

I група – відхилення середніх показників за варіантами від контролю більші за HIP зі знаком «+»;

II група – відхилення не виходять за межі ± HIP;

III група – відхилення середніх показників за варіантами від контролю більші за *НІР* зі знаком «–».

Висновок. Якщо єдиний критерій *F* для оцінки достовірності різниці між варіантами однофакторного досліду вважається нормою, то використання лише одного показника *НІР* може знизити оцінку окремих середніх, особливо у дослідах з великою кількістю варіантів. На базі випадкової варіації, яка визначається для досліду в цілому, знаходять загальну помилку різниці середніх. Тому варіанти одного досліду під час проведення статистичної оцінки розглядаються за винятком однієї пари статистично, тобто пов'язаними вибірками. У дійсності варіанти досліду не залежать один від одного. Єдиний субстрат або однорідний ґрунт не можна вважати сполучною ланкою, оскільки поряд з іншими однаковими умовами вони служать фоном для порівняння варіантів. Залежність статистичного походження корегують шляхом використання критеріїв, які враховують віддаленість середніх одне від одного в упорядкованому ряді. Зі збільшенням різниці між середніми зростає ранг і найменша істотна різниця. Це дозволяє знайти *НІР* для кожної пари середніх упорядкованого ряду і здійснити в такий спосіб більш точну оцінку варіантів досліду.

6.3. Перетворення

В агрономічних дослідженнях незалежність порівнянь досягається випадковим розосередженням варіантів у досліді та методично правильним відбором середніх проб до вибірки. Якщо є підстави передбачати неоднорідність дисперсій за вибірками, про що свідчать значні розбіжності між показниками за варіантами досліду, наприклад під час визначення враженості посівів шкідниками або хворобами, то вихідні показники доцільно трансформувати для зручності подальших розрахунків. Перетворення показників дає змогу зменшити діапазон їхнього варіювання, усунути неоднорідність дисперсій за вибірками та порівняти показники більш точно.

Існує чимало способів перетворень показників, серед яких найпоширенішими є: 1) логарифмічні, які передбачають переведення кожного показника в $lg\ y$ (або в $lg\ (y + 1)$, якщо деякі дати мають нульове значення); 2) переведення показників y в $\sqrt{y}$ або в $\sqrt{y+1}$,

якщо частина дат має нульове або близьке до нуля значення; 3) переведення показників y в «кут-арксинус $\sqrt{\text{відсоток}}$» за таблицею, наведеною в дод. И, якщо результати визначень виражені у відсотках, наприклад схожість насіння, виживаність рослин, ушкодження рослин шкідниками та хворобами. Переведення можна не проводити, якщо всі значення містяться в межах від 15 до 85. Але якщо є значення, близькі до 0 або 100, коли варіація сильно знижується, необхідні перетворення, які дозволять провести порівняння результатів більш точно.

Переведені дати статистично обробляють методом дисперсійного аналізу і після оцінки істотності різниці знову переводять назад до первинних показників. Середні показники, розраховані після перетворення, будуть дещо відрізнятися від середніх, отриманих за вихідними датами, але ця різниця, як правило, незначна, і більш правильним середнім буде значення, отримане зворотнім переходом. Техніку дисперсійного аналізу з різними варіантами перетворення вихідних показників наведено в наступній (другій) частині.

7. Кореляція, регресія та коваріація

7.1. Лінійна кореляція і регресія

Існує безліч наукових питань, пов'язаних із з'ясуванням тісноти зв'язків між певними чинниками, які визначають характер розвитку рослин, формування їхньої продуктивності, показників якості тощо. Дисперсійний аналіз лише дає відповідь на питання про те, є чи немає істотної різниці за певними показниками між досліджуваними варіантами. Так, аналізуючи врожайність двох сортів методом дисперсійного аналізу, визначаємо істотність різниці між ними. Водночас викликає інтерес установлення закономірностей формування врожайності залежно від ряду супутніх агротехнічних, абіотичних та інших чинників. Саме для вирішення цих питань передбачений кореляційний аналіз. За його допомогою можна з'ясувати, наскільки сильний зв'язок між температурою повітря та врожайністю або між кількістю опадів і врожайністю. Існує величезний спектр важливих питань, відповіді на які може дати кореляційний аналіз. Завдяки йому можна визначити характер і тісноту зв'язків між виляганням посівів і нормою висіву насіння, системою застосування добрив та якістю насіння або його врожайністю, установити залежність між польовою схожістю і строками сівби, між виживаністю рослин та густотою посіву і под.

Кореляція (від лат. *correlatio* – співвідношення) термін, який широко застосовується в агрономії для визначення взаємозв'язку, взаємної відповідності, співвідношення понять, предметів.

Форми прояву взаємозв'язків досить різноманітні. Як загальні види виділяють функціональний (повний) і кореляційний (неповний) зв'язок. Функціональний зв'язок проявляється в тому разі, якщо кожному значенню *x* відповідає одне певне значення ознаки *y*. Кореляційний зв'язок – це такий зв'язок, за якого кожному певному значенню однієї ознаки *x* відповідає кілька значень іншої взаємопов'язаної з нею ознаки, наприклад зв'язок між вегетативною масою та врожайністю рослин, між нормою висіву та діаметром стебла та ін. Кореляційний зв'язок проявляється в середньому для масових спостережень, коли заданим значенням залежної змінної відповідає певний ряд імовірнісних значень незалежної змінної. Зв'язок зветься кореляційним, якщо кожному значенню факторної ознаки відповідає цілком певне невипадкове значення результативної ознаки.

Під час вивчення кореляційних зв'язків виникають три основні запитання: чи є зв'язок між досліджуваними ознаками? наскільки він сильний? який тип цього зв'язку? Для вирішення цих питань застосовують спеціальні статистичні методи – *кореляцію* і *регресію*.

Кореляція – статистичний метод, який дозволяє визначити, чи існує залежність між досліджуваними показниками та наскільки вона сильна.

Регресія – статистичний метод, який застосовують для опису характеру зв'язку між досліджуваними ознаками (позитивний або негативний, лінійний або нелінійний).

Наочним зображенням кореляційного зв'язку може бути кореляційне поле. Воно являє собою графік, на якому на осі абсцис відкладають значення *x*, по осі ординат – значення *y*, а точками показують комбінації *x* і *y*. За розміщенням точок можна судити про наявність зв'язку (рис. 13–14).

Отже, за формою кореляційний зв'язок може бути лінійним і криволінійним, за напрямом – прямим і зворотнім. Якщо досліджується

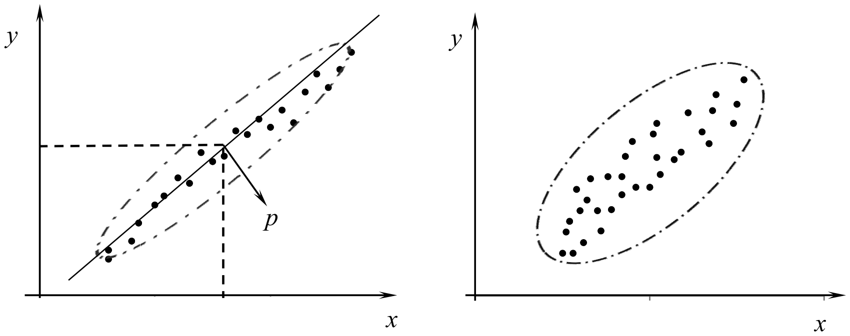


Рис. 13. Прямий зв'язок: вирівняний (зліва) і розсіяний (справа)

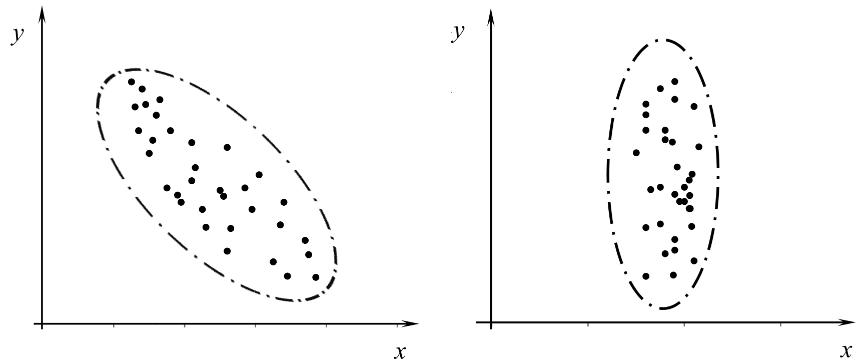


Рис. 14. Приклади кореляційних зв'язків: 1) існує негативна кореляція (зворотній зв'язок) зліва; 2) кореляційний зв'язок відсутній (справа)

тіснота зв'язків між двома ознаками, то кореляція і регресія є простою, якщо між трьома і більше – складною (множинною).

Кореляційний та регресійний аналізи мають важливе значення в агрономічних дослідженнях. Під регресією розуміють зміну функціонального показника (*y*) за певної зміни одного (проста регресія) або кількох факторіальних (множинна регресія) показників. Зв'язок між функціональними та факторіальними показниками виражається рівнянням регресії. Рівняння простої регресії схематично позначається так: $y = f(x)$, множинної – $y = f(x, z, v, w, ...)$. Якщо зв'язок між досліджуваними ознаками тісний, то за рівнянням регресії можна прогнозувати зміни функціональної ознаки за певних значень факторіальної. Тіснота зв'язків між досліджуваними ознаками визначається коефіцієнтом кореляції і кореляційним відношенням.

Прямолінійна залежність – це залежність, що має лінійний характер і виражається стандартним рівнянням прямої лінії: $y = a + bx$. Це рівняння називається рівнянням регресії y на x, а відповідна йому пряма лінія – вибірковою лінією регресії y на x. Пряма лінія, зображена на рис. 13, проходить через точку p, яка відповідає середнім показникам x і y та має нахил, який визначається в одиницях y на одиницю x. *Лінійна регресія* – це залежність, при якій за будь-якого значення факторіальної ознаки x її підвищення на одиницю вимірювання спричиняє підвищення функціональної ознаки на постійну величину. Якщо зі збільшенням аргументу (x) на сталу величину функціональна ознака (y) змінюється безсистемно, регресія називається криволінійною. Лінійна регресія показує, як у середньому змінюється функціональний показник (y) зі зміною показника аргументу (x). Кореляція та регресія зветься прямою (позитивною), якщо зі збільшенням показника однієї ознаки відбувається приріст другої, тобто зі збільшенням x зростає y. Якщо приріст показника аргументу спричиняє зменшення показника функції, то кореляція і регресія зветься оберненою, або від'ємною.

Для аналізу лінійної кореляції між двома ознаками проводиться певна кількість (n) парних вимірювань, кожне з яких дає пару чисел (x_1, y_1), (x_2, y_2),..., (x_n, y_n). Ці показники є базою для визначення фактичних показників коефіцієнта кореляції і регресії, виведення рівняння лінійної регресії, побудови теоретичної лінії регресії та оцінки значущості отриманих результатів.

Силу (тісноту) зв'язків між певними ознаками оцінюють за допомогою коефіцієнта кореляції r (ρ – для генеральної сукупності):

$$r = \frac{\overline{\sum} (x - \overline{x})\cdot(y - \overline{y})}{\sqrt{\sum(x - \overline{x})^2 \cdot \sum(y - \overline{y})^2}} \text{ – стандартна формула;}$$

$$r = \frac{\sum x \cdot y - \sum x \cdot \sum y / n}{\sqrt{[\sum x^2 - (\sum x^2)/n] \cdot [\sum y^2 - (\sum y)^2/n}}.$$

Поряд з вищенаведеним можна застосовувати спрощений запис чисельника та знаменника формули:

$$r = (\overline{xy} - \overline{x} \cdot \overline{y}) / (s_x \cdot s_y) \text{ або } r = s_{xy} / (s_x \cdot s_y),$$

де $\overline{xy}$ – середньоарифметичний добуток значень двох ознак ($\overline{xy} = \sum xy / n$); s_x і s_y – стандартні відхилення для x і y; s_{xy} – зчленоване стандартне відхилення.

На основі відношення $s_{xy} < s_x s_y$ установлено, що асимптотичними межами коефіцієнта кореляції є 0 і +1, якщо зв'язок прямий, і – 1 і 0, якщо зворотній, тобто він може нескінченно близько наближатися до 0 або до ± 1. Якщо кожному показнику аргументу x відповідає певний показник y, то кореляційний зв'язок переходить у функціональний. Для функціональних зв'язків $r = \pm 1$. Чим дужче r наближається до + 1 або до – 1, тим тісніший прямий кореляційний зв'язок. З наближенням r до нуля зв'язок послаблюється. Якщо $r = 0$, лінійий зв'язок між досліджуваними ознаками відсутній, водночас криволінійний зв'язок між ними може існувати. Вважається, що за $r < 0{,}2$ кореляційний зв'язок між ознаками дуже слабий, за $r = 0{,}2–0{,}5$ – слабкий; 0,5–0,7 – середній; 0,7–0,9 – сильний і за $r > 0{,}9$ – дуже сильний.

Теорія кореляції показує, що ступінь зв'язку у варіації двох ознак точніше вимірюється квадратом коефіцієнта кореляції (r^2). Наприклад, якщо $r = 0{,}4$, то не 40, а лише 16 % варіабельності залежної змінної y (результативної ознаки) пов'язано з мінливістю незалежної змінної x (факторіальної ознаки) ($0{,}4^2 = 0{,}16$, або 16 %), решта мінливості (1,0 – 0,16 = 0,84 %) зумовлена дією інших чинників.

Квадрат коефіцієнта кореляції (r^2) називають *коефіцієнтом детермінації* і позначають d_{yx}. Коефіцієнт детермінації характеризує частку варіації (дисперсії) результативної ознаки y, яка пояснюється регресією, у загальній варіації дисперсії y. Коефіцієнт детермінації, як і коефіцієнт кореляції, набуває значень від 0 до + 1. Чим ближче його значення по модулю до одиниці, тим тісніший зв'язок результативної ознаки y з досліджуваними факторами x. Наприклад, якщо коефіцієнт детермінації $r^2 = 0{,}8$, то рівнянням регресії пояснюється 80 % дисперсії результативної ознаки, а частка інших чинників становить лише 20 %. Якщо $r^2 = 0$, то це означає, що зв'язок між змінними регресійної моделі відсутній і замість неї для оцінки показників змінної ознаки з упевненістю можна використовувати прості середні її значень. І навпаки, якщо $r^2 = 1{,}0$, то це відповідає ідеальній моделі, коли всі точки спостережень містяться точно на лінії регресії, тобто сума квадратів їхніх відхилень дорівнює 0.

Показник коефіцієнта детермінації служить важливим критерієм оцінки якості лінійних і нелінійних моделей. Чим більша частка поясненої варіації, тим менша роль інших чинників, а отже, модель регресії добре апроксимує вихідні дані і такою регресійною моделлю можна скористатися для прогнозу значень результативної ознаки.

Метод розрахунку коефіцієнта кореляції залежить від виду шкали, до якої належать змінні. Так, для інтервальних змінних коефіцієнт кореляції розраховують за методом Пірсона або Спірмана, для рангових змінних r – методом Спірмана або Кендала:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})\cdot(y_i - \overline{y})}{(n-1)\cdot s_x \cdot s_y} \text{ – формула Пірсона,}$$

де x_i і y_i – значення двох змінних, $\overline{x}$ і $\overline{y}$ – їхні середні значення, s_x і s_y – їхні стандартні відхилення, n – кількість пар досліджуваних ознак.

Для змінних, які належать до порядкової або до інтервальної шкали за умови, що вони не підпорядковуються нормальному розподілу, розраховують рангову кореляцію за Спірманом. Для цього окремим значенням змінних присвоюють рангові місця, які надалі обробляють за допомогою відповідних формул.

Коефіцієнти рангової кореляції досить близькі до відповідних значень коефіцієнтів кореляції Пірсона (вихідні змінні підпорядковуються закону нормального розподілу).

Ще одним варіантом рангових коефіцієнтів кореляції є коефіцієнти Кендала (τ_b Кендала). За цим методом одну змінну подають у вигляді монотонної послідовності в порядку зростання величин; другій змінній надаються відповідні рангові місця. Кількість інверсій (порушень монотонності порівняно з першим рядом) використовують у формулі для кореляційних коефіцієнтів. Коефіцієнт Кендала застосовують переважно тоді, якщо у вихідних даних трапляються «викиди» й обидві змінні належать до рангової шкали.

Для оцінки достовірності коефіцієнта кореляції розраховують його помилку та критерій істотності. Стандартну помилку цього коефіцієнта визначають за формулою:

$$s_r = \sqrt{(1-r)^2 / n-2}\,,$$

де s_r – помилка коефіцієнта кореляції; r – коефіцієнт кореляції; n – чисельність вибірки, тобто кількість пар значень, за якими розрахований r.

Судячи з формули, підвищення коефіцієнта кореляції сприяє зменшенню його помилки, тобто підвищує його точність. Збільшення кількості порівнюваних пар (n) також підвищує точність визначення коефіцієнта кореляції і забезпечує зменшення s_r.

Критерій достовірності коефіцієнта кореляції розраховують за формулою:

$$t_r = r / s_r.$$

За $t_{rфакт.} \geq t_{rтеор.}$ кореляційний зв'язок достовірний, якщо навпаки – $t_{rфакт.} < t_{rтеор.}$, то зв'язок неістотний. Теоретичні значення критерію t (Стьюдента) для заданого рівня значущості (05 або 01 %) знаходять за таблицею в дод. Б для кількості ступенів свободи, що дорівнює $n - 1$.

При малих вибірках і значеннях r, близьких до 1, розподіл показників коефіцієнтів кореляції відрізняється від нормального. У зв'язку із цим критерій t не є надійною базою для оцінки значущості коефіцієнта кореляції, побудови довірчих інтервалів кореляції генеральної сукупності та порівняння коефіцієнтів. Вихід із цієї ситуації знайшов Р. Фішер. Він запропонував перетворювати коефіцієнт кореляції у показник z (зет), який розподіляється відповідно до закону нормального розподілу. Перехід від r до z здійснюють за допомогою таблиці наведеній в дод. Л. Стандартну помилку показника z знаходять за формулою:

$$s_z = 1/\sqrt{n-3}\text{ , (}n\text{ – кількість пар }x, y\text{).}$$

Критерій значущості для перетвореного показника z, різниці між z_1 і z_2 та довірчі межі цього показника визначають за рівняннями:

$$t_z = z/s_z;\; t_{z1-z2} = z_1 - z_2 / \sqrt{s_{z_1}^2 + s_{z_2}^2}\;;\; z \pm ts_z.$$

Після визначення довірчих меж за допомогою зворотніх перетворень (див. дод. Л) знаходять відповідні для z_{min} і z_{max} показники r_{min} і r_{max}.

Якщо $r_{факт.} \geq r_{теор.}$, то між аргументом і функцією є достовірний зв'язок (H_0 спростовується), а якщо $r_{факт.} < r_{теор.}$, то цей зв'язок неістотний (H_0 підтверджується). Для доведення значущості слабких зв'язків між x і y потрібно мати не менше 40 їхніх пар, середніх – від 12-ти до 40-ка, і сильних – від 6-ти до 12-ти пар.

Кореляційний коефіцієнт показує тісноту зв'язків між ознаками, але не дозволяє судити про те, як змінюється результативність функціонального показника за зміни аргументу (x) на одну і ту саму величину або на одиницю вимірювання, що має важливе теоретичне та практичне значення. Відповіді на поставлені запитання здатен дати *регресійний аналіз*, основним завданням якого є визначення формули кореляційної залежності, тобто рівняння прямої лінії.

Рівняння лінійної регресії залежності y від x має вигляд:

$$y = \overline{y} - b_{yx}\,(x - \overline{x}),$$

де $\overline{x}$ і $\overline{y}$ – середні арифметичні показників x і y; b_{yx} – коефіцієнт регресії y за x.

Коефіцієнт регресії – це одна з характеристик зв’язків між функціональною ознакою y і факторіальною ознакою (аргументом) x. Він показує, на скільки одиниць зросте або зменшиться значення y, якщо x зміниться на одиницю свого вимірювання. Геометрично b є кутовим коефіцієнтом нахилу прямої лінії. Якщо коефіцієнт регресії розраховується для y залежно від зміни x (b_{yx}), то він вимірюється в одиницях вимірювання функції (y), а якщо b розраховується для x залежно від варіації y (b_{xy}), то в одиницях вимірювання аргументу (x). Під час дослідження односторонньої залежності, наприклад кореляції між урожайністю рослин і температурою повітря впродовж вегетаційного періоду, розраховують тільки один коефіцієнт регресії для залежності зміни y (урожайності) від x (температури), тобто значення b_{yx}, оскільки визначення регресії зміни температури залежно від варіабельності врожайності (b_{xy}) позбавлене логічного сенсу. Коефіцієнт регресії розраховують за формулами:

$$b_{yx} = \frac{\sum(x-\overline{x})\cdot(y-\overline{y})}{\sum(x-\overline{x})^2} \text{ – для залежності зміни } y \text{ від } x;$$

$$b_{yx} = \frac{\sum(x-\overline{x})\cdot(y-\overline{y})}{\sum(y-\overline{y})^2} \text{ – для залежності зміни } x \text{ від } y.$$

Таким чином, коефіцієнт лінійної регресії – це число, яке показує, у якому напрямі та на скільки зміниться в середньому значення y при зміні значення x на одиницю вимірювання.

Добуток коефіцієнтів регресії дорівнює коефіцієнту кореляції:

$$r^2 = b_{yx} \cdot b_{xy}.$$

За допомогою цього рівняння можна перевіряти правильність розрахунків коефіцієнтів регресії.

Помилку (s_b) та критерій достовірності (t_b) коефіцієнта регресії знаходять за формулами:

$$s_{byx} = s_r \cdot \sqrt{\frac{\sum(y-\overline{y})^2}{\sum(x-\overline{x})^2}} \text{ – для регресії } y \text{ на } x;$$

$$s_{bxy} = s_r \cdot \sqrt{\frac{\sum(x-\overline{x})^2}{\sum(y-\overline{y})^2}} \text{ – для регресії } x \text{ на } y;$$

$$t_b = b/s_b.$$

Критерій достовірності коефіцієнта кореляції (t_r) також використовують для оцінки значущості коефіцієнта регресії (t_b), оскільки вони мають однакове значення – $t_r = t_b$. Істотність коефіцієнта регресії оцінюють за таблицею, наведеною в дод. Б, для числа ступенів свободи $n - 2$.

Точкова діаграма може показати сильну варіацію індивідуальних показників спостережень, що не дозволяє з високою точністю визначити будь-яке значення функціонального показника y для заданого показника x. Для цього необхідно нівелювати (вирівнювати) вплив випадкових відхилень і знайти теоретичну лінію функціональної залежності y від x, тобто розрахувати рівняння регресії середнього приросту або зменшення (для зворотної залежності) показника функції (y) зі збільшенням аргументу (x).

Усереднену течію функції знаходять за принципом визначення середньої арифметичної, що розміщується найближче стосовно до індивідуальних показників, через що сума квадратів їхніх відхилень від середнього показника буде найменшою. Для вирівнювання емпіричних рядів застосовують два способи: графічний і аналітичний.

Графічний метод полягає в тому, що на точковій діаграмі розподілу показників візуально проводять пряму лінію, яка теоретично має проходити якомога ближче до всіх точок (пар y, x). Це приблизний (грубий) метод, який можна застосовувати лише тоді, коли потрібно лише наближено виявити загальну тенденцію лінійної залежності функції (y) від аргументу (x). Тому для точного вирівнювання розкиду емпіричних показників, тобто знаходження оптимального положення прямої лінії, краще застосовувати аналітичний метод.

Сутність аналітичного методу полягає у визначенні середніх показників $\overline{x}$, $\overline{y}$, коефіцієнта регресії b_{yx} і після підставлення цих значень у рівняння лінійної регресії $y = \bar{y} + b_{yx} \cdot (x - \bar{x})$ визначенні формули рівняння прямої лінії, яке має стандартний вигляд для прямої лінійної залежності – $y = a + b \cdot x$. За рівнянням знаходять теоретично усереднені значення $\bar{y}$ для мінімального і максимального показника x. Визначені екстремальні показники (точка x_{min}, y_{min} і точка x_{max}, y_{max}) наносять на діаграму точкового розподілу показників, проводять через них пряму лінію – це і є пряма лінія регресійної залежності y від x.

Приклад аналітичного методу графічно зображено на рис. 15. Представлена пряма лінія максимально близько проходить біля всього масиву пар показників x, y і є ілюстрацією параметрів a і b для рівняння виду $y = 1{,}4 + 7{,}1y$.

У цьому прикладі коефіцієнт $a = 1{,}4$ – ордината лінії, початок відліку за $y = 0$. Як правило цей показник не має логічного сенсу. Якщо лінія рівняння регресії за будь-яких позитивних значень аргументу x перетинає вісь y нижче від 0, то показник a має від'ємне значення. У наведеному графіку максимальному значенню $x = 4$ відповідає парне значення $y = 1{,}4 + 7{,}1 \cdot 4 = 29{,}8$. Коефіцієнт регресії залежності y від x показує, на скільки в середньому зміниться значення функції y зі збільшенням аргументу на одиницю вимірювання. Оскільки $b = 7{,}1$, це свідчить, що зі збільшенням аргументу x на одиницю, наприклад з 2 до 3, приріст функції в середньому становитиме 7,1 одиниці.

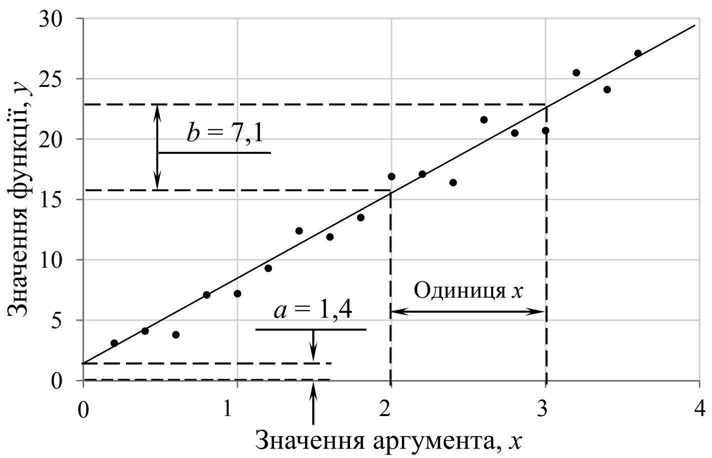


Рис. 15. Точковий розподіл пар показників x, y і розраховане для них рівняння лінійної регресії – $y = 1{,}4 + 7{,}1y$

7.2. Множинна та партикулярна лінійні кореляції і регресії

Під час розв'язання практичних задач дослідники стикаються з тим, що кореляційні зв'язки не обмежуються лише зв'язками між двома ознаками: результативною y і факторіальною x. Насправді результативна ознака залежить від значної кількості чинників. Наприклад, урожайність рослин тісно пов'язана з кількістю опадів за вегетацію, температурою повітря, дозою внесених добрив та ін.

За впливу багатьох чинників показники парної кореляції виявляються умовними та неточними. Кількісно оцінити вплив різних чинників на певний показник, визначити форму і тісноту зв'язків між результативною та факторіальними ознаками x_1, x_2,..., x_n можна методами множинної (багатофакторної) кореляції.

Багатофакторний кореляційно-регресійний аналіз зводиться до вирішення таких завдань:

– обґрунтувати взаємозв'язки чинників, які впливають на досліджуваний показник;
– визначити ступінь впливу кожного чинника на результативну ознаку шляхом побудови моделі-рівняння множинної регресії, яка дасть змогу установити, в якому напрямі та наскільки зміниться результативна ознака за зміни кожного чинника, що входить у модель;
– кількісно визначити тісноту зв'язків між функціональною ознакою і чинниками.

Включення в рівняння множинної регресії того чи іншого набору чинників пов'язано насамперед з уявленням дослідника про природу взаємозв'язку модельованого показника з іншими абіотичними або агротехнічними явищами. Чинники, які включаються до множинної регресії, повинні відповідати таким вимогам:

– бути кількісно вимірюваними;
– не повинні бути корельовані між собою і тим більше мати точний функціональний зв'язок.

Наявність між двома чинниками тісного лінійного зв'язку (парний коефіцієнт кореляції перевищує за абсолютним показником 0,7) зветься *колінеарністю*, а між декількома чинниками – *мультиколінеарністю*.

Мультиколінеарність – це нестрога лінійна залежність між факторними ознаками, яка приводить до таких небажаних наслідків:

– оцінки параметрів стають ненадійними, вони виявляють більші стандартні помилки, низьку значущість, у той же час модель є значущою, тобто значення множинного коефіцієнта кореляції завищене;
– незначні зміни вихідних даних приводять до істотної зміни оцінок параметрів моделі;
– оцінки параметрів моделі мають неправильні знаки або невиправдано завищені значення, що робить модель непридатною для аналізу та прогнозування;
– стає неможливим визначення ізольованого впливу чинників на результативну ознаку.

Найбільшою мірою за мультиколінеарність відповідатиме та ознака, яка найтісніше пов'язана з іншими чинниками моделі (вона має високі за модулем значення коефіцієнтів парної лінійної кореляції). Тому якщо фактори явно колінеарні, вони дублюють один одного і один з них доцільно виключити з регресії. Перевага при цьому віддається не тому чиннику, який більш тісно пов'язаний з результатом, а тому, який за достатньо тісного зв'язку з функціональним показником має найменш тісний зв'язок з іншими чинниками.

Найпростішою формою множинного зв'язку є лінійна залежність між трьома ознаками, коли одна з них, наприклад урожайність, розглядається як функція (y), а дві інші – як аргументи (x і z). Як міру тісноти зв'язків трьох ознак використовують часткові коефіцієнти кореляції, які позначаються як $r_{xy \cdot z}$, $r_{xz \cdot y}$, $r_{zy \cdot x}$, і множинні, які позначаються символами $r_{x \cdot yz}$, $r_{y \cdot xz}$, $r_{z \cdot xy}$.

Якщо змінні корелюють одна з одною, то на показнику коефіцієнта парної кореляції частково позначається вплив інших змінних. Наприклад, якщо між ознаками x і z існує тісний кореляційний зв'язок і крім того, y залежить від x, то y також буде корелювати із z. Цілком можливо, що кореляція між y і z не пряма, а апагогічна, яка виникає внаслідок впливу ознаки x. Тому потрібно дослідити партикулярну кореляцію між y і z, крім впливу x на y. Елементи, що виключаються, можуть закріплюватися як на середніх, так і на інших рівнях, що вибрані відповідно до ділянок вимірювання змінних які цікавлять дослідника, між якими визначається зв'язок у «чистій» формі. Тут слід ураховувати професійно-теоретичне міркування про досліджуване явище.

Математична статистика дозволяє визначати тісноту зв'язків між двома змінними показниками за постійного показника третьої ознаки, використовуючи парні коефіцієнти кореляції – r_{xy}, r_{xz} і r_{yz}. Партикулярні коефіцієнти кореляції розраховують за формулами:

$$r_{yx \cdot z} = r_{yx} - r_{yz} \cdot r_{xz} \,/\, \sqrt{(1 - r_{yx}^2) \cdot (1 - r_{xz}^2)}\ ;$$

$$r_{yz \cdot x} = r_{yz} - r_{yx} \cdot r_{zx} \,/\, \sqrt{(1 - r_{yx}^2) \cdot (1 - r_{zx}^2)}\ ;$$

$$r_{xz \cdot y} = r_{xz} - r_{yx} \cdot r_{yz} \,/\, \sqrt{(1 - r_{yx}^2) \cdot (1 - r_{yz}^2)}\ .$$

Індекси перед точками показують, між якими саме ознаками проводиться визначення кореляційного зв'язку. Індекс після точки – це третя ознака, вплив якої елімінується (усувається). Помилка і критерій значущості партикулярної кореляції визначаються за тими самими формулами, що і для парної кореляції:

$$sr_{xy \cdot z} = \sqrt{\frac{1 - r_{xy \cdot z}^2}{(n - 1)}}\ ; \quad t = r \,/\, s_r.$$

Теоретичні значення t-критерію знаходять для потрібного рівня значущості при кількості ступенів свободи $n - 3$.

За допомогою коефіцієнта множинної кореляції не можна зробити висновок про характер взаємозв'язків, тобто позитивну або негативну кореляцію між змінними. Тільки якщо всі коефіцієнти парної кореляції мають однаковий знак, тоді його можна віднести до коефіцієнта множинної кореляції і стверджувати про характер множинного зв'язку. Чим більше значення коефіцієнта кореляції наближається до одиниці, тим сильніший взаємозв'язок.

Подібно до парних коефіцієнтів кореляції, партикулярні коефіцієнти можуть набувати значень від –1 до +1. Часткові коефіцієнти детермінації знаходять шляхом піднесення в квадрат партикулярних коефіцієнтів кореляції:

$$d_{xy \cdot z} = r^2_{xy \cdot z};\ d_{xz \cdot y} = r^2_{xz \cdot y};\ d_{yz \cdot x} = r^2_{yz \cdot x}.$$

Якщо між двома ознаками кореляційний зв'язок відсутній, то коефіцієнт множинної детермінації дорівнює сумі парних коефіцієнтів детермінації. Так, якщо між ознаками x і z зв'язок відсутній, то множинний коефіцієнт детермінації можна розрахувати за формулою: $d^2_{y \cdot xz} = d^2_{yx} + d^2_{yz}$. Тобто він дорівнює сумі інтенсивності взаємозв'язків між y і x та між y і z. Отже, за відсутності кореляційного зв'язку між двома змінними аналіз зв'язків спрощується.

Для з’ясування техніки розрахунків і розуміння часткового коефіцієнта кореляції розглянемо на прикладі розрахунок даних з визначення парної кореляції між діаметром (y), масою (x) і довжиною (z) другого міжвузля рослин пшениці м’якої ярої. Парні коефіцієнти кореляції для цих ознак становлять:

$$r_{yx} = 0{,}764;\ r_{yz} = 0{,}575;\ r_{xz} = 0{,}499.$$

За наведеними рівняннями на базі парних коефіцієнтів розрахуємо партикулярні коефіцієнти кореляції:

$$r_{yx\bullet z} = 0{,}764 - 0{,}575 \cdot 0{,}499 \,/\, \sqrt{(1-0{,}764^2)\cdot(1-0{,}499^2)} = 0{,}753;$$

$$r_{yz\bullet x} = 0{,}575 - 0{,}764 \cdot 0{,}499 \,/\, \sqrt{(1-0{,}764^2)\cdot(1-0{,}499^2)} = 0{,}348;$$

$$r_{xz\bullet y} = 0{,}499 - 0{,}764 \cdot 0{,}575 \,/\, \sqrt{(1-0{,}764^2)\cdot(1-0{,}575^2)} = 0{,}114.$$

Партикулярний коефіцієнт кореляції між діаметром і масою міжвузля за однакової їхньої довжини ($r_{yx\bullet z} = 0{,}753$) показує, що лише незначна частка зв’язків цих структурних елементів у загальній кореляції ($r_{yx} - r_{yx\bullet z} = 0{,}764 - 0{,}753 = 0{,}011$) зумовлена впливом третього елемента (довжиною міжвузля). Те саме стосується тісноти зв’язку між діаметром міжвузля та його довжиною за однакової їхньої маси ($r_{yz} - r_{yz\bullet x} = 0{,}575 - 0{,}348 = 0{,}227$). Водночас партикулярний коефіцієнт кореляції між масою та довжиною міжвузля за однакового їхнього діаметра значно відрізняється від загального коефіцієнта кореляції ($r_{xz} - r_{xz\bullet y} = 0{,}499 - 0{,}114 = 0{,}385$), а це свідчить про те, що якщо підібрати міжвузля з однаковим діаметром, то зв’язок між довжиною та масою міжвузлів буде дуже слабкий ($r_{xz\bullet y} = 0{,}114$), оскільки значна частина цього зв’язку обумовлена варіабельністю діаметра міжвузля.

У деяких випадках партикулярний коефіцієнт кореляції може бути протилежним за знаком парному коефіцієнту кореляції. Наприклад, під час встановлення тісноти зв’язків між структурними елементами врожаю ячменю ярого: масою верхнього міжвузля (y), його довжиною (x) та діаметром (z) – отримали такі коефіцієнти парної кореляції (кількість порівнюваних пар – 50):

$$r_{yx} = 0{,}583;\ r_{yz} = 0{,}875;\ r_{xz} = 0{,}390.$$

У разі елімінування третьої ознаки, партикулярні коефіцієнти кореляції тісноти зв’язків між цими елементами становлять:

$$r_{yx\bullet z} = 0{,}583 - 0{,}875 \cdot 0{,}390 \,/\, \sqrt{(1-0{,}583^2)\cdot(1-0{,}390^2)} = 0{,}324;$$

$$r_{yz\bullet x} = 0{,}875 - 0{,}583 \cdot 0{,}390 \,/\, \sqrt{(1-0{,}583^2)\cdot(1-0{,}390^2)} = 0{,}866;$$

$$r_{xz\bullet y} = 0{,}390 - 0{,}583 \cdot 0{,}875 \,/\, \sqrt{(1-0{,}583^2)\cdot(1-0{,}875^2)} = -\,0{,}305.$$

Розраховані коефіцієнти кореляції між масою та довжиною верхнього міжвузля при елімінуванні діаметра міжвузля ($r_{yx\bullet z} = 0{,}324$), та між масою і діаметром міжвузля при елімінуванні його довжини ($r_{yz\bullet x} = 0{,}866$) є цілком закономірними. Встановлено високий кореляційний зв’язок між масою верхнього міжвузля та його довжиною при елімінуванні діаметра ($r_{yx} - r_{yx\bullet z} = 0{,}583 - 0{,}324 = 0{,}259$) і слабкий зв’язок між масою та діаметром верхнього міжвузля ($r_{yz} - r_{yz\bullet x} = 0{,}875 - 0{,}866 = 0{,}009$). Кореляція між довжиною та діаметром міжвузля при елімінуванні його маси вийшла негативною ($r_{xz\bullet y} = -\,0{,}305$), тоді як парний коефіцієнт кореляції вказує на позитивний зв’язок між цими ознаками. Отримані результати можуть здаватися помилковими, адже зі збільшенням довжини міжвузля його діаметр має зростати, однак розрахований частковий коефіцієнт кореляції є цілком логічним якщо врахувати те, що умовою для його розрахунку було елімінування показника маси міжвузля. І справді, якщо маса міжвузля не змінюється, то збільшення його діаметра може відбуватися лише зі зменшенням довжини. Зрозуміло, що одночасне збільшення довжини та діаметра міжвузля може відбуватися лише з підвищенням його маси.

Метод партикулярної кореляції дає змогу визначити партикулярні коефіцієнти кореляції за чотирма порівнюваними ознаками. Цей коефіцієнт, його ще називають коефіцієнт другого порядку, показує тісноту зв’язку між першою та другою ознаками за постійних (сталих) значень третьої і четвертої ознаки. Розрахунок коефіцієнта другого порядку проводять на основі партикулярних коефіцієнтів кореляції першого порядку за формулою:

$$r_{xy\cdot zv} = \frac{r_{xy\cdot v} - r_{xz\cdot v}\cdot r_{yz\cdot v}}{\sqrt{(1-r^2_{xz\cdot v})\cdot(1-r^2_{yz\cdot v})}},$$

де, $r_{xy\cdot v}$, $r_{xz\cdot v}$, $r_{yz\cdot v}$ – партикулярні коефіцієнти кореляції першого порядку, які розраховують на основі парних коефіцієнтів кореляції: r_{xy}, r_{xz}, r_{xv}, r_{yz}, r_{yv} і r_{zv}.

Множинний коефіцієнт кореляції визначає інтенсивність або тісноту зв'язків між трьома і більшою кількістю ознак (змінних). Він призначений для характеристики тісноти зв'язків однієї з ознак із сукупністю інших.

Коефіцієнт множинної кореляції використовується також як показник точності оцінки регресії. За ним можна визначити, наскільки факторіальні (змінні) ознаки зумовлюють кількісну варіацію залежної (функціональної) змінної. Якщо коефіцієнт множинної кореляції набуває значень, які близькі до одиниці, то варіація залежної (функціональної) змінної майже повністю визначається змінами пояснювальних (факторіальних) змінних, тобто включені в аналіз факторіальні змінні сильно впливають на залежну змінну.

Коефіцієнт множинної кореляції має показник, який не менший від абсолютної величини будь-якого коефіцієнта парної та партикулярної кореляції з таким же первинним індексом. Це не залежить від того, існує між факторіальними змінними причинний зв'язок чи ні. Також слід додати, що, на відміну від парного коефіцієнта кореляції, множинний коефіцієнт завжди має позитивне значення і розміщується в діапазоні $0 \le r_{y \cdot xz \ldots} \le 1$. Коефіцієнт множинної кореляції для будь-якого числа факторіальних (незалежних) змінних у загальному вигляді можна подати так:

$$r_{y \cdot 12 \ldots n} = \sqrt{r_{y1} \cdot b_1' + r_{y2} \cdot b_2' + \ldots + r_{ym} \cdot b_m'},$$

де $r_{y \cdot 12 \ldots n}$ – показник тісноти зв'язків між функціональною (залежною) ознакою (літера перед точкою) і сукупністю решти ознак (цифри після точки); r_{y1}, r_{y2}, r_{ym} – парні коефіцієнти кореляції; b_1', b_2', b_m' – коефіцієнти регресії між функціональною (y) та факторіальними ознаками (1, 2, m).

При встановлених показниках коефіцієнтів парної кореляції (r_{yx}, r_{yz}, r_{xz}) множинні коефіцієнти кореляції між трьома ознаками зручніше розраховувати за формулами:

$$r_{y \cdot xz} = \sqrt{\frac{r_{yx}^2 + r_{yz}^2 - 2 \cdot r_{yx} \cdot r_{yz} \cdot r_{xz}}{1 - r_{xz}^2}};$$

$$r_{x \cdot yz} = \sqrt{\frac{r_{yx}^2 + r_{xz}^2 - 2 \cdot r_{yx} \cdot r_{yz} \cdot r_{xz}}{1 - r_{yz}^2}};$$

$$r_{z \cdot xy} = \sqrt{\frac{r_{yz}^2 + r_{xz}^2 - 2 \cdot r_{yx} \cdot r_{yz} \cdot r_{xz}}{1 - r_{yx}^2}}.$$

Квадрат коефіцієнта множинної кореляції називають *коефіцієнтом множинної детермінації*. Він показує частку незалежних (факторіальних змінних) у варіації показників залежної (функціональної) змінної. Значимість множинної кореляції оцінюють за F-критерієм:

$$F = \frac{r^2}{1 - r^2} \cdot \left(\frac{n-k}{k-1}\right),$$

де n – обсяг вибірки, k – кількість ознак (у випадку, якщо вивчається тіснота зв'язків між трьома змінними, $k = 3$).

Теоретичне значення критерію F знаходять за таблицею, наведеною в дод. В, для кількості ступенів свободи чисельника $k - 1$, кількості ступенів свободи знаменника $n - k$ та для потрібного рівня значущості (95 чи 99 %). Якщо $F_{факт.} > F_{теор.}$, то нульова гіпотеза стосовно рівності множинного коефіцієнта кореляції у сукупності ($H_0 : R = 0$) спростовується, і якщо $F_{факт.} \le F_{теор.}$ – не спростовується.

Визначимо коефіцієнт множинної кореляції між масою зерна з колоса (y), кількістю зерен у колосі (x) і довжиною верхнього міжвузля (z). При кількості порівнюваних пар ($n = 20$) парні коефіцієнти кореляції для досліджуваних ознак становлять: між масою зерна (y) та озерненістю колоса (x) $r_{yx} = 0{,}917$; між масою зерна з колоса (y) та довжиною верхнього міжвузля (z) $r_{yz} = 0{,}620$; між озерненістю колоса (x) та довжиною верхнього міжвузля (z) $r_{yx} = 0{,}388$. На підставі парних коефіцієнтів кореляції розраховуємо коефіцієнт множинної кореляції між залежною змінною y і незалежними змінними x і z:

$$r_{y \cdot xz} = \sqrt{\frac{r_{yx}^2 + r_{yz}^2 - 2 \cdot r_{yx} \cdot r_{yz} \cdot r_{xz}}{1 - r_{xz}^2}};$$

$$r_{y \cdot xz} = \sqrt{\frac{(0{,}917)^2 + (0{,}620)^2 - 2 \cdot 0{,}917 \cdot 0{,}620 \cdot 0{,}388}{1 - 0{,}388^2}} = 0{,}961;$$

$$F_ф = \frac{0{,}961^2}{1 - 0{,}961^2} \cdot \left(\frac{20-3}{3-1}\right) = 103{,}3.$$

Теоретичне значення критерію F для α_{05} і α_{01} при числі ступенів свободи 20 – 3 = 17 становить відповідно 3,59 і 6,11 (див. дод. В).

Отже, взаємозв'язок між масою зерна з колоса, його озерненістю та довжиною верхнього міжвузля $r_{y \cdot xz} = 0{,}961$ достовірний за обох рівнів значущості, оскільки $F_{факт.} > F_{01}$. Коефіцієнт детермінації множинної регресії $r^2_{y \cdot xz} = (0{,}961^2) = 0{,}924$ свідчить, що в цьому прикладі варіабельність маси зерна з колоса пшениці на 92,4% пов'язана з дією незалежних факторіальних ознак: озерненістю колоса та довжиною верхнього міжвузля. При цьому лише 7,6 % змін (100 – 92,4) маси зерна з колоса відбувається за впливу інших чинників.

Будь-яка досліджувана ознака детермінується значним числом інших ознак, що діють сукупно. Тому досить важливо навчитися визначати варіативність певної залежної змінної (y) від кількох факторіальних (незалежних) змінних: x, z, ..., n у конкретних умовах.

Математичне рівняння для лінійної апроксимації залежності між трьома і більше змінними називається *множинним лінійним рівнянням регресії* і має такий загальний вигляд:

$$y = \alpha + b_{yx \cdot z...n} \cdot x + b_{yz \cdot x...n} \cdot z + ... + b_{yn \cdot x,\, z...n-1} \cdot n,$$

де y – залежна змінна; x, z, ..., n – незалежні факторіальні змінні, довільний початок відліку; $b_{yx \cdot z...n}$, $b_{yz \cdot x...n}$, $b_{yn \cdot x,\, z...n-1}$ – коефіцієнти партикулярної регресії.

Коефіцієнт b_1 показує, наскільки збільшиться y у разі підвищення ознаки x на одиницю вимірювання за незмінних значень інших факторіальних ознак; коефіцієнт b_2 показує, наскільки підвищиться показник y у разі збільшення показника z на одиницю вимірювання за незмінних значень інших факторіальних ознак і т. ін.

Пояснювальні змінні x, z,..., n справляють сумісний одночасний вплив на залежну змінну y. Завдання полягає в оцінці параметрів регресії за результатами вибіркових спостережень над змінними, включеними в аналіз. Для цього застосовують метод найменших квадратів, який дозволяє знайти таке положення площини регресії у просторі, за якого сума квадратів відхилень емпіричних точок від неї є мінімальною, а саме (для трьох порівнюваних ознак):

$$s(\alpha, b_x, b_z) = \sum_{i=1}^{n} (y - \alpha - b_1 x - b_2 z)^2 \rightarrow min.$$

Для виведення рівняння множинної регресії визначають коефіцієнти регресії b_1 і b_2 за формулами:

$$b_1 = \frac{s_{xy} \cdot s_z^2 - s_{zy} \cdot s_{xz}}{s_x^2 \cdot s_z^2 - s_{xz}^2}\, ;\; b_2 = \frac{s_{zy} \cdot s_x^2 - s_{xz} \cdot s_{xy}}{s_x^2 \cdot s_z^2 - s_{xz}^2}$$

або в зручнішому для розрахунків вигляді:

$$b_1 = \frac{\sum(z-\bar{z})^2 \cdot \sum(x-\bar{x}) \cdot (y-\bar{y}) - \sum(x-\bar{x}) \cdot (z-\bar{z}) \cdot \sum(y-\bar{y}) \cdot (z-\bar{z})}{\sum(x-\bar{x})^2 \cdot \sum(z-\bar{z})^2 - [\sum(x-\bar{x}) \cdot (z-\bar{z})]^2}\, ;$$

$$b_2 = \frac{\sum(x-\bar{x})^2 \cdot \sum(y-y) \cdot (z-\bar{z}) - \sum(x-\bar{x}) \cdot (z-\bar{z}) \cdot \sum(y-\bar{y}) \cdot (x-\bar{x})}{\sum(x-\bar{x})^2 \cdot \sum(z-\bar{z})^2 - [\sum(x-\bar{x}) \cdot (z-\bar{z})]^2}\, .$$

Для кращого розуміння зв'язку між досліджуваними ознаками встановлене рівнянням регресії відношення між ними краще подавати у вигляді відповідних графіків поверхні регресії. Для прикладу, у вигляді графіка покажемо розподіл теоретичних показників відповідно до рівняння регресії: $y = 7{,}3 + 0{,}8x + 0{,}25z$.

Підставляючи в рівняння множинної регресії певні значення глибини загортання насіння, наприклад 5 см, 6, 7, 8 і 9 см, та норми висіву насіння, наприклад 30 тис./га, 40, 50, 60 і 70 тис./га, отримуємо координати точок для побудови графіків множинної лінійної залежності (табл. 14).

За наведеними в табл. 14 даними будуємо лінійний графік регресії врожайності (y) від норми висіву (z) для п'яти фіксованих значень глибини загортання насіння (x) (рис. 16).

Одержані результати можна зобразити у вигляді тримірної поверхні відгуку лінійної множинної регресії врожайності (y) на гли-

Таблиця 14

Урожайність насіння соняшнику, залежно від глибини загортання (x) та норми висіву насіння (z), (т/га)

Глибина загортання насіння, см (x)	Норма висіву, тис./га (z)				
	30	40	50	60	70
5	1,88	2,13	2,38	2,63	2,88
6	1,96	2,21	2,46	2,71	2,96
7	2,04	2,29	2,54	2,79	3,04
8	2,12	2,37	2,62	2,87	3,12
9	2,20	2,45	2,70	2,95	3,20

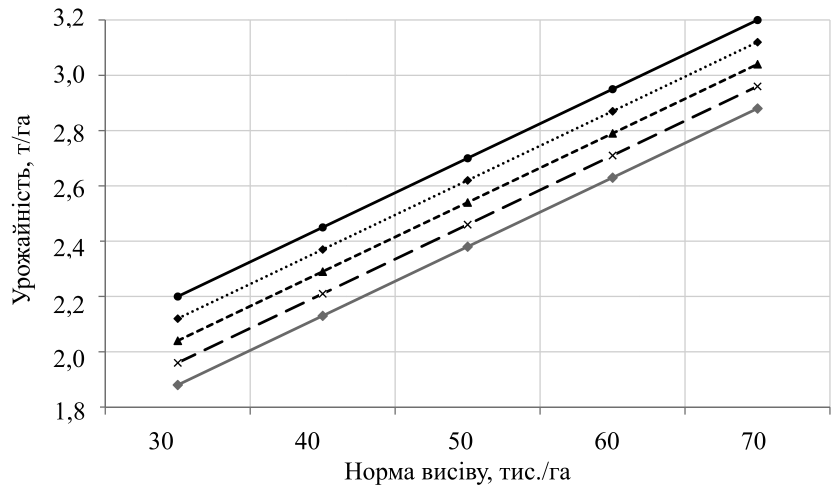


Рис. 16. Лінії регресії рівняння $y = 7{,}3 + 0{,}8x + 0{,}25z$ для залежності врожайності насіння соняшнику від глибини загортання (x) та норми висіву насіння (z). *Умовні позначення*: глибина загортання насіння: 5 см; 6 см; 7 см; 8 см; 9 см.

бину загортання (x) та норму висіву насіння соняшнику (рис. 17). На цій діаграмі врожайність насіння соняшнику відповідає певним комбінаціям факторіальних ознак x і z.

Поверхня регресії дає чітке уявлення стосовно комбінованого впливу досліджуваних факторіальних чинників (тут – глибини загортання та норм висіву насіння) на варіативність функціональної ознаки – врожайності насіння.

Важливо відзначити, що математичне рівняння як для парної, так і для множинної регресії має сенс лише в діапазоні змін фактичних значень досліджуваних ознак і лише тоді, коли кореляційний зв'язок не дорівнює нулю.

7.3. Криволінійна кореляція і регресія

Якщо між досліджуваними ознаками існують нелінійні відношення, то коефіцієнт кореляції в його прийнятій для лінійного зв'язку формі не може відображати інтенсивності зв'язків. Так,

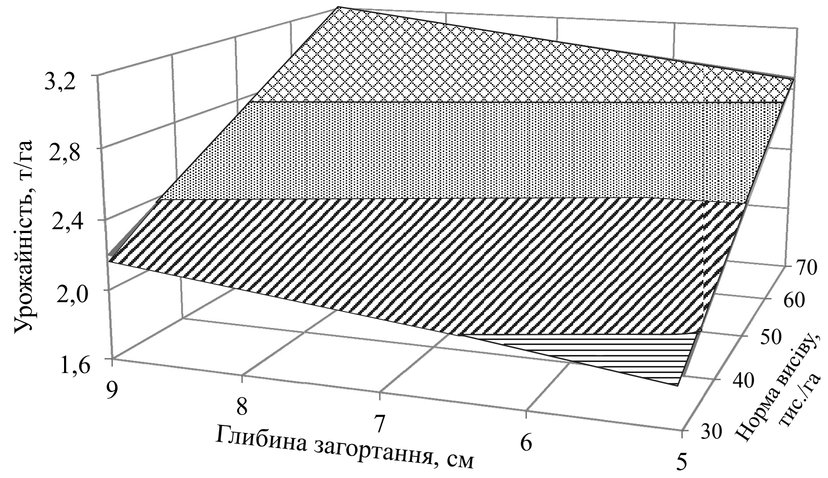


Рис. 17. Залежність урожайності насіння соняшнику від сумісного впливу норми висіву та глибини загортання насіння (площина регресії)
Умовні позначення: Врожайність насіння соняшнику в межах:
1,6-2,0 т/га; 2,0-2,4 т/га; 2,4-2,8 т/га; 2,8-3,2 т/га.

якщо розрахувати коефіцієнт кореляції для двох статистичних рядів:

2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30
1, 3, 7, 10, 12, 15, 17, 20, 16, 13, 11, 9, 8, 5, 3,

то він дорівнюватиме нулю ($r = 0$), хоча очевидно, що між обома рядами існує тісний зв'язок. Але з цього зовсім не випливає, що лінійний коефіцієнт кореляції у деяких випадках не приводить до потрібних результатів при нелінійному зв'язку рядів спостережень. Поряд з цим все ж таки є необхідність визначення достовірного показника тісноти зв'язків при нелінійних відношеннях. Таким показником зв'язку може служити індекс кореляції.

Універсальним показником тісноти зв'язку, незалежно від її форми, є кореляційне відношення η. Перед його розрахунком вибірку (парні спостереження) вирівнюють у порядку зростання за факторною ознакою x і формують 4–8 груп. Кожна група має містити не менше двох пар. За $n < 20$ для розрахунків використовують формулу:

$$\eta_{yx} = \sqrt{\frac{\sum(y-\bar{y})^2 - \sum(y-\bar{y}_x)^2}{(y-\bar{y})^2}},$$

де $\sum(y-\bar{y})^2$ – сума квадратів відхилень (*СК*) окремих значень результативної ознаки y від їхнього середнього значення (загальне варіювання); $\sum(y - \bar{y}_x)^2$ – *СК* внутрішньогрупових значень відхилень y від їхнього середнього $\bar{y}_x$ (внутрішньо-групове варіювання).

Для вибірок обсягом понад 30 ($n > 30$) проводиться групування показників. Статистичний матеріал при цьому є більш наочним. Після групування і розподілу показників визначають *СК* групового варіювання $\sum f(\bar{y}_x - \bar{y})^2$, *СК* загального варіювання $\sum f(y-\bar{y})^2$ і розраховують кореляційне відношення за формулою:

$$\eta_{yx} = \sqrt{\sum f(\bar{y}_x - \bar{y})^2 / \sum f(y-\bar{y})^2},$$

де f – чисельність груп; $\sum(\bar{y}_x - \bar{y})^2$ – *СК* середньогрупових значень від загального середнього (варіювання між групами). Воно характеризує ту частину варіативності результативної ознаки y, яка зумовлена варіативністю ознаки x; $\sum(y-\bar{y})^2$ – *СК* окремих значень від загального середнього (загальне варіювання ознаки y).

Граничними межами для кореляційного відношення є $0 \le \eta < 1$. За $\eta_{yx} = 0$ між досліджуваними ознаками зв'язку немає. У той же час $\eta_{yx} \neq 1$, оскільки якщо тіснота зв'язку дорівнює одиниці, то це свідчить про функціональну залежність. Чим ближче η_{yx} до одиниці, тим сильніша функціональна залежність y від x, і навпаки, чим ближче η_{yx} до нуля, тим слабше виражена ця залежність.

На величину кореляційного відношення впливає характер групування статистичного матеріалу. Чим більше груп виділено, тим менше значень залежної змінної (y) потрапляє в кожну групу, тим більше розсіюються групові середні стосовно загального середнього, тобто вплив випадкових і неврахованих чинників проявляється більшою мірою. Отже, міжгрупова дисперсія групових середніх $\bar{y}_i$ зі зростанням числа груп, як правило, збільшується, а загальна дисперсія залишається без змін. Зазвичай спостерігається така тенденція: зі зростанням числа груп за факторіальною змінною кореляційне відношення збільшується. При фіксованій кількості груп за пояснювальною змінною кореляційне відношення залежить також від групування значень залежної змінної. Як правило, це відношення тим більше, чим більш диференційоване групування за залежною змінною. Усе це потрібно мати на увазі під час використання кореляційного відношення як показника тісноти зв'язку.

Кореляційне відношення в квадраті (η^2_{yx}), яке отримують з відношення квадрата внутрішньогрупового варіювання до загального варіювання, найточніше показує частку мінливості результативної ознаки, зумовлену варіативністю факторної ознаки для будь-якої форми зв'язку. Це відношення називається індексом кореляції (коефіцієнтом детермінації) і виражається у відсотках:

$$d_{yx}, \% = \eta^2_{yx} \cdot 100.$$

Помилку (d_{yx}) та критерій істотності кореляційного відношення (η^2_{yx}) розраховують за формулами:

$$s_\eta = \sqrt{\frac{1-\eta^2}{n-2}};\ t_\eta = \eta / s_\eta.$$

Теоретичне значення критерію t_{05} і t_{01} знаходять за таблицею, наведеною в дод. Б, для числа ступенів свободи $n - 2$.

Під час опрацювання досліджуваного матеріалу методом дисперсійного аналізу квадрат кореляційного відношення розраховують як відношення суми квадратів відхилень для варіантів C_v до загальної суми квадратів C_y: $\eta^2_{yx} = C_v / C_y$, а оскільки $\eta_{yx} = \sqrt{\eta^2_{yx}}$, то $\eta_{yx} = \sqrt{C_v / C_y}$.

Для визначення ступеня наближення нелінійної залежності до прямолінійної використовують критерій Фішера (F), який шукають за формулою:

$$F_{факт.} = (\eta^2 - r^2) \cdot (n-k) / (1-\eta^2) \cdot (k_x - 2),$$

де η^2 – квадрат кореляційного відношення y за x; r^2 – квадрат коефіцієнта лінійної кореляції; n – обсяг вибірки; k_x – кількість груп ряду x.

Якщо $F_{факт.} < F_{теор.}$, то зв'язок переходить у лінійний і розрахунки проводять як для прямолінійної кореляції і регресії, а якщо $F_{факт.} \ge F_{теор.}$, то це свідчить про наявність нелінійного зв'язку між ознаками. Теоретичні значення F беруть з наведеної в дод. В таблиці для $k_x - 2$ ступенів свободи чисельника і $n - 2$ ступенів свободи знаменника.

Для прикладу визначимо характер кореляційної залежності між ознаками (прямолінійна або нелінійна) за умови, що $r = 0{,}762$; $\eta_{yx} = 0{,}854$; $n = 60$ і $k = 6$.

$$F = [(0{,}854^2 - 0{,}762^2) \cdot (60 - 6)] / [(1 - 0{,}854^2) \cdot (6 - 2)] = 7{,}37.$$

Для 4-х ступенів свободи чисельника ($k - 2 = 4$) і 58-ми ступенів свободи знаменника ($n - 2 = 58$) критерії F_{05} і F_{01} становлять відповідно 2,52 і 3,65. Оскільки $F_{факт.} > F_{01}$, то нульова гіпотеза (H_0) спростовується, а це свідчить про те, що користуватися лінійною кореляцією і регресією не можна, адже криволінійність підтверджується на 99 %.

Для вибору та обґрунтування типу нелінійної регресії немає універсального методу. Односторонню стохастичну залежність між досліджуваними ознаками можна описати рівнянням поліноміальної регресії: $y = b_0 + b_1x + b_2x^2 + b_3x^3 + ...$ або за допомогою гіперболічної регресії: $y = b_0^2 + b_1^2 \cdot 1/x$. Для описання криволінійної залежності застосовуються також ступенева, показова, логарифмічна і тригонометрична функції. Приклади різних типів криволінійної залежності представлені у вигляді кривих ліній регресії і відповідних рівнянь на рис. 18.

Про характер залежності між досліджуваними ознаками часто судять за зовнішнім виглядом емпіричного графіка регресії. Однак

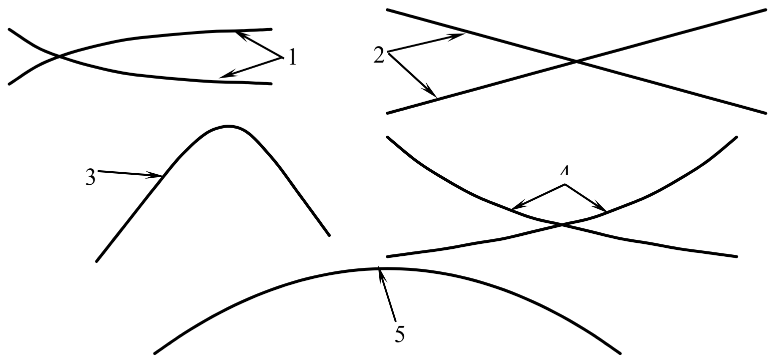


Рис. 18. Криві, які ілюструють різні типи залежності між ознаками: 1 – крива лінія типу $lg\, y = a - в_1 \cdot x$; 2 – прямолінійна залежність типу $y = a + в \cdot x$; 3 – крива типу $y = a - в_1 \cdot x + в_2 \cdot x^2 + в_3 \cdot x^3$; 4 – крива типу $lg\, y = a - в_1 \cdot x$; 5 – крива типу $y = a - в_1 \cdot x + в_2 \cdot x^2$.

за незначної кількості спостережень це призводить до незадовільних результатів, оскільки різкі зиґзаґи лінії регресії ускладнюють установлення закономірності. У кожному конкретному випадку потрібно перевіряти можливість застосування лінійної регресії на обмеженій ділянці розподілу змінних. Також необхідно звертати увагу на те, щоб оцінки регресії проводилися з високою ймовірністю надійності.

Розрізняють два класи нелінійних регресій. До першого класу належать так звані *квазілінійні регресії*, нелінійні стосовно залучених до аналізу факторіальних змінних x_k, але лінійні за невідомими параметрами регресії, які підлягають оцінюванню. Їхня перевага полягає в тому, що для них можливе безпосереднє застосування методу найменших квадратів, відповідно, залишаються в силі всі вихідні передумови лінійного регресійного аналізу та властивості оцінок параметрів регресії (незміщеність, гомоскедастичність і т. ін.). Застосовуються ті самі критерії значущості, аналогічно будуються довірчі інтервали та довірчі зони.

Другий клас нелінійних регресій характеризується нелінійністю за оцінюваними параметрами. Нелінійні регресії другого класу називають *істотно нелінійними регресіями*. Цей клас регресії досить розповсюджений в агрономічних дослідженнях. Однак тут істотним недоліком є неможливість застосування звичайного методу найменших квадратів. Для розв’язання системи нелінійних рівнянь вдаються до апроксимації параметрів шуканої залежності. На практиці також використовується лінійне перетворення функції регресії, яке дозволяє застосовувати до перетворених параметрів статистичні критерії лінійної регресії.

Зв’язок між експериментальними ознаками польових досліджень часто відповідає рівнянням, близьким до квадратичної параболи:

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

Близьким до цього рівняння є розподіл залежності врожайності рослин від норми висіву, глибини загортання насіння, строків сівби. Так, зі збільшенням норми висіву або глибини загортання насіння врожайність спочатку зростає до певного рівня (оптимізації досліджуваного чинника), а потім поступово зменшується. Рівняння для квадратичної параболи розраховується за співвідношенням:

$$y = \overline{y} + \frac{\sum(x-\overline{x})\cdot y}{\sum(x-\overline{x})^2}\cdot(x-\overline{x}) + \left[\frac{\sum(x-\overline{x})^2\cdot y - n\cdot C\cdot \overline{y}}{\sum(x-\overline{x})^4 - n\cdot C^2}\right]\cdot$$
$$\cdot\left[(x-\overline{x})^2 - \frac{\sum(x-\overline{x})^2}{n}\right].$$

На основі даних табл. 15 розрахуємо поліноміальне рівняння регресійної залежності між урожайністю зерна (y) та нормою висіву насіння ячменю ярого (x) за допомогою наведеної формули:

$$y = 31{,}3 + \frac{2{,}4}{7{,}0}\cdot(x-4{,}5) + \left[\frac{208{,}5 - 7\cdot 1\cdot 31{,}3}{9{,}01 - 7\cdot 1^2}\right]\cdot[(x-4{,}5)^2 - 1] =$$
$$= 31{,}3 + 0{,}34\cdot(x-4{,}5) - 5{,}27\cdot[(x-4{,}5)^2 - 1] =$$
$$= -71{,}68 + 47{,}76x - 5{,}28x^2.$$

Емпіричний і теоретичний (поліноміальне рівняння) розподіл показників залежності між урожайністю зерна та нормою висіву ячменю ярого зображено на рис. 19.

У разі, якщо важко підібрати теоретичне рівняння регресійної залежності для емпіричного розподілу показників взаємодії або якщо немає вагомих підстав уточнювати результати досліджень, вирівнювання емпіричного ряду можна провести за допомогою методу так званого змінного середнього. Відповідно до цього методу, для кожного значення факторіальної змінної x визначається середнє арифметичне із сусідніх значень функціональної змінної y. Залежно від характеру

Таблиця 15

Вплив норм висіву x (млн шт./га) на врожайність ячменю ярого y (ц/га)

x	y	$(x-\overline{x})$	$(x-\overline{x})^2$	$(x-\overline{x})^4$	$(x-\overline{x})\cdot y$	$(x-\overline{x})^2\cdot y$
3,0	28,5	– 1,5	2,25	3,38	– 42,75	64,13
3,5	31,3	– 1,0	1,0	1,0	– 31,3	31,3
4,0	31,4	– 0,5	0,25	0,125	– 15,7	7,85
4,5	33,1	0	0	0	0	0
5,0	33,8	0,5	0,25	0,125	16,9	8,45
5,5	32,2	1,0	1,0	1,0	32,2	32,2
6,0	28,7	1,5	2,25	3,38	43,05	64,58
$\sum x = 31{,}5$ $\overline{x} = 4{,}5$	$\sum y = 219$ $\overline{y} = 31{,}3$	$\sum(x-\overline{x}) = 0$	$\sum(x-\overline{x})^2 = 7{,}0$	$\sum(x-\overline{x})^4 = 9{,}01$	$\sum(x-\overline{x})\cdot y = 2{,}4$	$\sum(x-\overline{x})^2\cdot y = 208{,}5$

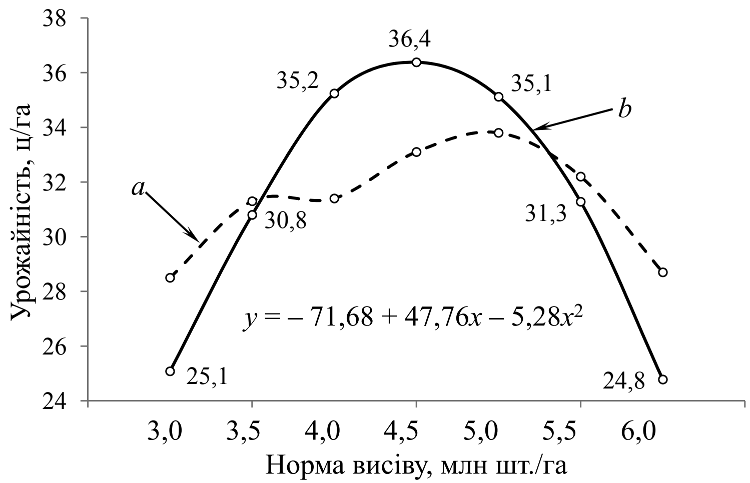


Рис. 19. Емпірична (a) і теоретична (b) лінії регресії врожайності зерна (y) від норми висіву ячменю ярого (x)

кривизни та діапазону варіабельності показників, середня арифметична може розраховуватися за трьома (для незначної варіативності) або п'ятьма (для значної варіативності) результативними показниками:

$$\overline{y}_i = (y_{i-1} + y_i + y_{i+1})\ /\ 3;$$
$$\overline{y}_i = (y_{i-2} + y_{i-1} + y_i + y_{i+1} + y_{i+2})\ /\ 5.$$

Центральне значення, як більш пріоритетне можна зробити більш впливовим під час знаходження середнього змінного, для цього y_i в обох рівняннях подвоюють:

$$\overline{y}_i = (y_{i-1} + 2y_i + y_{i+1})\ /\ 4;$$
$$\overline{y}_i = (y_{i-2} + y_{i-1} + 2y_i + y_{i+1} + y_{i+2})\ /\ 6.$$

Для прикладу проведемо визначення змінних середніх за трьома точками для даних, які характеризують зв'язок між глибиною загортання насіння і довжиною верхнього міжвузля пшениці твердої ярої (у сантиметрах) (табл. 16).

Для знаходження середніх змінних граничних показників їх подвоюють і додають (або віднімають) сусідній показник, після чого отриману суму ділять на три. Так, лівий крайній $y_{x-3,0} = (2\cdot 21 + 26)\ /\ 3 = 22{,}7$ см, а крайній правий – $y_{x-7,0} = (2\cdot 23 + 32)\ /\ 3 = 26{,}0$ см.

Таблиця 16

Показники довжини верхнього міжвузля пшениці твердої ярої, вирівняні методом середньої змінної за рівнянням:

$\bar{y}_i = (y_{i-1} + 2y_i + y_{i+1}) / 4$, см

x	3,0	3,5	4,0	4,5	5,0	5,5	6,0	6,5	7,0
y	21	26	27	33	36	37	34	32	23
$\bar{y}_i$	22,7	25,0	28,3	32,3	35,5	36,0	34,3	30,3	26,0

Правильність вирівнювання перевіряють, використовуючи коефіцієнт кореляції між емпіричними та вирівняними показниками y:

$$r = \sqrt{\frac{\sum(y-\bar{y})^2 - \sum(y_i - y)^2}{\sum(y-\bar{y})^2}} = \sqrt{1 - \frac{\sum(\bar{y}_i - y)^2}{\sum(y-\bar{y})^2}}.$$

Вирівнювання вважають достатнім, якщо $r > 0{,}95$, За $r < 0{,}95$ збіг між фактичними та вирівняними даними є недостатнім і потрібно провести повторне згладжування.

Для наведеного прикладу (табл. 16) r становить:

$$\sum(y-\bar{y})^2 = \sum y^2 - (\sum y)^2 / n = 8309 - 8040 = 269;$$

$$\sum(\bar{y}_i - y)^2 = (22{,}7 - 21{,}0)^2 + (25{,}0 - 26{,}0)^2 + \ldots + (26{,}0 - 23{,}0) = 19{,}3;$$

$$r = \sqrt{1 - \frac{\sum(\bar{y}_i - y)^2}{\sum(y-\bar{y})^2}} = \sqrt{1 - \frac{19{,}3}{269{,}0}} = \sqrt{0{,}93} = 0{,}96.$$

Розрахований показник $r = 0{,}96$, свідчить, що одержана теоретична лінія згладжування добре відображає фактичні показники (рис. 20).

Для множинної криволінійної кореляційної залежності вихідні параметри заносять до таблиці і для фіксованих значень факторіальних (незалежних) змінних x і z визначають найбільш ймовірні функціональні значення залежної змінної y. Одержані результати відображають графічно у вигляді поверхні регресії y за показниками x і z, яка наочно показує форму залежності функціональної ознаки від двох незалежних ознак – x і z.

Приклад складної криволінійної залежності між трьома ознаками: врожайністю насіння соняшника (y), діаметром кошику (x), та густотою рослин (z) наведено на рис. 21. Аналізуючи його дані, легко переконатися, що факторіальні ознаки мають значний сумісний

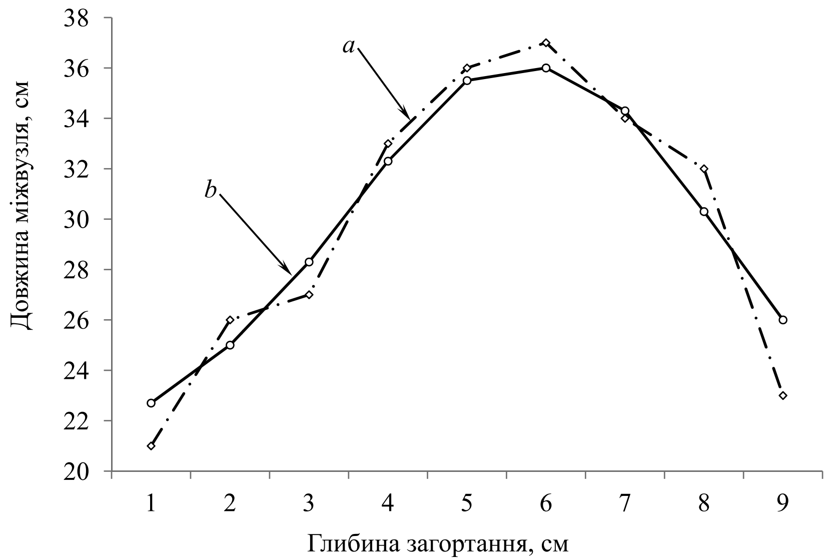


Рис. 20. Емпірична (a) та згладжена (b) лінії регресії функціональної ознаки y від незалежної x

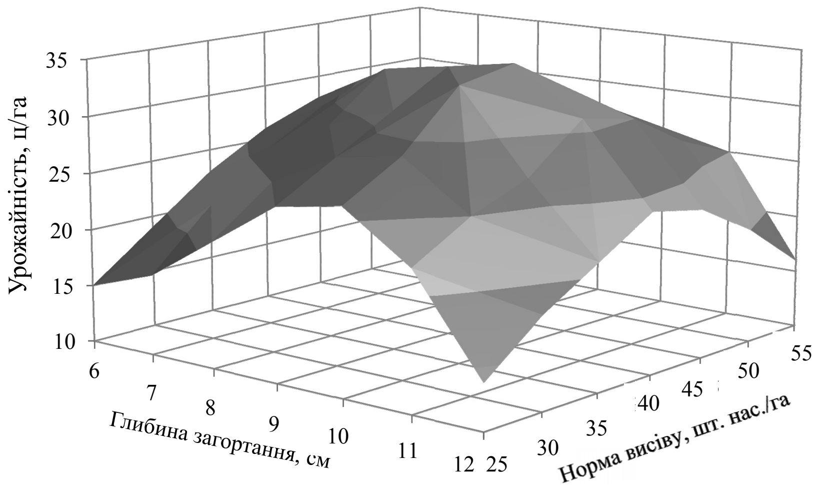


Рис. 21. Залежність врожайності насіння соняшнику від одночасного впливу діаметра кошику та глибини загортання насіння

вплив на мінливість урожайності соняшнику. Максимальна ефективність норм висіву проявлялася в діапазоні глибини загортання насіння від 8 до 10 см, у свою чергу, ефективність глибини загортання визначалася нормою висіву насіння. Максимальна врожайність насіння за всіх досліджуваних показників глибини загортання була за норми висіву 40 тис. нас./м^2. Легко переконатися, що оптимізація глибини загортання насіння сприяла нівелюванню негативного ефекту різкого зниження насіннєвої продуктивності за граничних норм висіву насіння.

Завдяки кореляційному аналізу формується значно більше уявлення про результати експерименту порівняно з дисперсійним аналізом, водночас він лише більш розширено аналітично констатує результати досліду: визначає напрям і тісноту зв'язків між ознаками. Водночас цей статистичний метод ні в якому разі не може бути альтернативою професійним навичкам дослідника, законам логістики, спеціальним знанням, які дозволяють розуміти сутність і механізм динамічних процесів розвитку рослин. Статистичні методи обробки результатів експерименту, зокрема кореляційний та регресійний аналізи, при правильному підході є важливим і незамінним аргументом підтвердження припущень дослідника, що з певною ймовірністю дозволяє будувати прогнози стосовно будь-яких ознак (варіантів), які не потрапили в поле зору експериментатора.

7.4. Оцінка кореляції при якісній мінливості

Методи кореляційно-регресійного аналізу було розроблено для характеристики тісноти зв'язків ознак кількісної мінливості. Однак іноді досить важливо визначити ще й тісноту зв'язків для цінних господарських якісних показників або залежності між якісними і кількісними показниками. З'ясувати ці питання можна за допомогою нестандартних формул.

Розглянемо на конкретних прикладах техніку визначення коефіцієнта кореляції між комбінаціями якісних і кількісних ознак.

Приклад 1. Визначимо тісноту зв'язку між обробкою посівів пшениці озимої фунгіцидами та ураженістю рослин бурою іржею. Кількість здорових (+) і ушкоджених (–) рослин на контрольному (–) та обробленому фунгіцидами (+) варіантах указано в табл. 17.

Таблиця 17

Кількість здорових і уражених бурою іржею рослин пшениці озимої залежно від проведення обробки посівів фунгіцидами, шт./м^2

Рослини	Варіант, оброблений фунгіцидом (+)	Контроль (–)	Сума
Здорові (+)	317(n_1)	61(n_2)	378 (N_1)
Уражені (–)	33(n_3)	294(n_4)	327 (N_2)
Сума	350(N_3)	355(N_4)	705 (N)

Оскільки у цьому прикладі ми маємо справу з альтернативною якісною мінливістю, то для розрахунку коефіцієнта кореляції скористаємося зручною формулою Юла (*Yula*):

$$r = (n_1 \cdot n_4 - n_2 \cdot n_3) \ / \ \sqrt{N_1 \cdot N_2 \cdot N_3 \cdot N_4},$$

де n_1 і n_4 – сукупності з однаковими знаками (+ + і – – відповідно); n_2 і n_3 – сукупності з різними знаками (+ – або – +); N_1 і N_2 – суми показників за рядами; N_3 і N_4 – суми показників за стовпцями.

$$r = (317 \cdot 294 - 61 \cdot 33) \ / \ \sqrt{378 \cdot 327 \cdot 350 \cdot 355} = 0{,}74;$$

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - 0{,}74^2}{705 - 2}} = 0{,}03; \ t_r = r \ / \ s_r = 0{,}74 \ / \ 0{,}03 = 24{,}7.$$

Теоретичне значення t_{05} і t_{01} при $n - 2 = 703$ становить відповідно 1,96 і 2,58 (див. дод. Б). Таким чином, обробка посівів пшениці озимої фунгіцидами забезпечує істотне зменшення ураженості рослин бурою іржею, оскільки фактичний критерій t значно перевищує теоретичний для обох рівнів значущості.

Коефіцієнт кореляції між якісними та кількісними ознаками однієї вибірки розраховують за формулою:

$$r = \frac{\bar{x}_1 - \bar{x}}{s} \cdot \sqrt{\frac{n_1}{n - n_1}},$$

де $\bar{x}$ – середнє значення кількісної ознаки; $\bar{x}_1$ – середнє значення якісної ознаки з наявністю кількісної; n – кількість усіх спостережень; n_1 – кількість об'єктів з певною якісною ознакою; s – стандартне відхилення для кількісної ознаки.

Приклад 2. Проведемо визначення тісноти зв'язків між масою зерна з одного колоса тритикале ярого (г/колоса) та його ураженістю клопом-черепашкою на основі вибірки з 200 шт. (табл. 18).

Для визначення коефіцієнта кореляції шукають загальне середнє $\bar{x}$, середні показники маси зерна з колоса для неушкодженого ($\bar{x}_2$) і ураженого ($\bar{x}_1$) колоса тритикале та стандартне відхилення s.

$$\bar{x} = \frac{\sum fx}{n} = \frac{175{,}3}{200} = 0{,}877;\ \bar{x}_1 = \frac{\sum f_1 x}{n_1} = \frac{63{,}8}{76} = 0{,}839;$$

$$\bar{x}_2 = \frac{\sum f_2 x}{n_2} = \frac{111{,}5}{124} = 0{,}899 \text{ г/колоса};$$

$$s = \sqrt{\frac{\sum fx^2 - (\sum fx)^2 / n}{n-1}} = \sqrt{\frac{159{,}4 - (175{,}3)^2 / 200}{200-1}} =$$

$$= \sqrt{\frac{5{,}75}{199}} = 0{,}17 \text{ г};$$

$$r = \frac{\bar{x}_1 - \bar{x}}{s} \cdot \sqrt{\frac{n_1}{n - n_1}} = \frac{0{,}839 - 0{,}877}{0{,}17} \cdot \sqrt{\frac{76}{200-76}} = -0{,}175;$$

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(-0{,}175)^2}{200-2}} = \sqrt{\frac{0{,}969}{198}} = 0{,}07;\ t_r = r / s_r = 2{,}5.$$

Таблиця 18

Маса зерна з колоса рослин тритикале ярого (г) і його ураженість клопом-черепашкою

Маса зерна з колоса, x (г)	Кількість колосків, шт.			f_1x	f_2x	fx	x^2	fx^2
	ураже-них, f_1	неушко-джених, f_2	разом, f ($f_1 + f_2$)					
1,2	3	9	12	3,3	10,8	14,4	1,44	17,3
1,1	6	15	21	6,6	16,5	23,1	1,21	25,4
1,0	9	23	32	9,0	23,0	32,0	1,00	32,0
0,9	15	29	44	13,5	26,1	39,6	0,81	35,6
0,8	21	22	43	16,8	17,6	34,4	0,64	27,5
0,7	14	19	33	9,8	13,3	23,1	0,49	16,2
0,6	8	7	15	4,8	4,2	9,0	0,36	5,4
Сума	**76**	**124**	**200**	**63,8**	**111,5**	**175,3**	**–**	**159,4**

При $n - 2 = 200 - 2 = 198$ ступенях свободи, теоретичні критерії t_{05} і t_{01} становлять відповідно 1,96 і 2,58 (див. дод. Б). Таким чином, між масою зерна з одного колоса та його ураженістю клопом-черепашкою існує слабий зворотній зв'язок, достовірний з імовірністю 95 % ($t_r > t_{05}$). Істотність цього зв'язку для 99 %-го рівня значущості не підтверджена.

7.5. Рангові коефіцієнти кореляції

Поряд з розглянутими лінійними та нелінійними коефіцієнтами кореляції існує цілий ряд показників тісноти зв'язку, який застосовують у випадках якщо досліджуваним ознакам не вдається однозначно надати тих чи інших абсолютних значень. До них належить *коефіцієнт рангової кореляції Спірмена*. Його застосування, на відміну від розглянутих коефіцієнтів кореляції, не пов'язано з передумовою нормальності розподілу вихідних даних.

Під час застосування методів рангової кореляції за основу беруть не точні кількісні оцінки значень ознак змінних, а ранги. Для цього елементи сукупності розташовують у певному порядку відповідно до якоїсь ознаки, притаманною їм неоднаковою мірою. Цей ряд називають упорядкованим. Сам процес упорядкування називають ранжуванням, а кожному члену цього ряду присвоюють ранг або рангове число (порядковий номер). Зокрема, елементу з найменшим показником присвоюється ранг 1, наступному за ним елементу – ранг 2 і т. ін. Напряму вирівнювання ряд значень не має, тобто складові ряду можна розташовувати як у порядку зростання, так і в порядку спадання. Якщо елемент має не одну, а дві ознаки x і y, то для дослідження їхнього впливу один на одного кожному елементу присвоюється два порядкові номери згідно з визначеним правилом ранжування. Далі переходять від кореляції ознак змінних x і y до вивчення зв'язку між ранговими числами шляхом визначення відповідності між двома послідовностями порядкових оцінок. Тобто визначається тіснота рангової кореляції. Оскільки досліджується зв'язок між двома змінними, коефіцієнт рангової кореляції є парним. Коефіцієнт рангової кореляції Спірмена розраховують за формулою:

$$r_s = 1 - 6\sum d^2 / n(n^2 - 1),$$

де d – різниця між рангами пов'язаних рядів x і y, тобто $d = x - y$; n – кількість парних визначень (спостережень).

Із формули видно, що для визначення коефіцієнта рангової кореляції необхідно знайти лише квадрати відхилень рангів. На практиці трапляються випадки, коли два або більше елементів сукупності мають однакові значення однієї ознаки і дослідник не може знайти істотної різниці між ними. Елементи, що мають цю властивість – відсутність переваги, – називають пов'язаними, а сформовану з них групу – зв'язкою. Метод, який застосовується для присвоювання порядкового номера пов'язаним елементам, зветься методом середніх рангів. Він полягає в усередненні рангів, які б мали ці елементи, якщо б вони розрізнялися.

Коефіцієнт рангової кореляції може мати значення в інтервалі від –1 до +1 ($-1 \le r_s \le +1$). Якщо ранги змінних однакові, то $r_s = 1$. У цьому разі має місце повна узгодженість між елементами двох послідовностей. Кожен елемент займає одне й те саме місце в обох рядах, що свідчить про повний позитивний кореляційний зв'язок рангів. Якщо $r_s = -1$, то елементи двох послідовностей розміщені у зворотному порядку і між ними немає узгодженості. Це означає повний негативний кореляційний зв'язок між рангами. І нарешті, якщо $r_s = 0$, то це свідчить про відсутність кореляції між рангами.

Приклад 3. У двох різних господарствах отримали дані врожайності 10-ти гібридів соняшнику. Потрібно визначити зв'язок сортів за врожайністю для цих господарств (табл. 19).

$$r_s = 1 - 6 \cdot \sum d^2 / n \cdot (n^2 - 1) = 1 - 6 \cdot 30 / 10 \cdot 99 = 1 - 0{,}18 = 0{,}820.$$

Величина $r_s = 0{,}820$ свідчить про тісний позитивний зв'язок між урожайністю досліджуваних гібридів соняшнику. Якщо $n = 10$, t-критерій фактичний становить:

$$t_r = \frac{r_s}{\sqrt{1 - r_s}} \cdot \sqrt{n-2} = \frac{0{,}820}{\sqrt{0{,}18}} \cdot \sqrt{8} = 1{,}952 \cdot 2{,}83 = 5{,}52.$$

Теоретичне значення t-критерію для односторонньої гіпотези становить: $t(n - 2; 1 - \alpha) = t(8; 1 - 0{,}05) = 1{,}89$. Оскільки $t_r > t_{05}$, нульова гіпотеза щодо відсутності різниці у рейтингу сортів за двома вибірками відхиляється. Отже, райони вирощування гібридів соняшнику мають різний вплив на їхню врожайність, тобто врожайність гібридів соняшнику має регіональні відмінності.

Таблиця 19

Робоча таблиця підготовки даних для оцінки зв'язку врожайності 10 сортів соняшнику за двома районами вирощування

Сорт	Ранг урожайності		$d = r(x) - r(y)$	d^2
	$r(x_1)$	$r(y_1)$		
1	4	2	2	4
2	3	1	2	4
3	6	7	– 1	1
4	8	9	– 1	1
5	2	5	– 3	9
6	9	10	– 1	1
7	10	8	2	4
8	1	3	– 2	4
9	7	6	– 1	1
10	5	4	1	1

Коефіцієнт рангової кореляції загалом служить досить точною характеристикою тісноти зв'язку досліджуваних змінних. Його перевага полягає в тому, що він не пов'язаний з передумовою нормальності розподілу вихідних даних. Коефіцієнт рангової кореляції тим більше наближається до коефіцієнта парної кореляції, чим менше кореляційний зв'язок між досліджуваними змінними відрізняється від лінійного і чим міцніший цей зв'язок. Для нормально розподіленої генеральної сукупності і великого обсягу вибірки ($n \ge 30$) між обома коефіцієнтами існує таке асимптотичне відношення:

$$r_{yx} = 2 sin\left(\frac{\pi \cdot r_s}{6}\right).$$

Метод рангової кореляції не потребує лінійної кореляції між змінними. Проте необхідно, щоб функція регресії, яка відображає цей зв'язок, була монотонною.

Особливо корисною рангова кореляція є під час визначення характеру зв'язків між явищами, які не підлягають кількісному оцінюванню. У таких випадках експериментатор на базі власного досвіду приписує елементам вибірки ранги за кожною з досліджуваних якісних ознак. Тобто рангова кореляція є завжди корисною для дослідження зв'язків там, де властивості ознак не підлягають точному кількісному вимірюванню, але дозволяють проводити порівняльну оцінку, завдяки якій установлюються послідовності рангів.

Коефіцієнт рангової кореляції τ-Кендала є альтернативою методу рангової кореляції r-Спірмана. Він призначений для визначення взаємозв'язку між двома ранговими змінними. Цей коефіцієнт базується на основі відношень типу «більше-менше», справедливість яких установлена під час побудови певних шкал. Інтерпретація результатів розрахунків коефіцієнта рангової кореляції τ-Кендала визначається як різниця ймовірностей збігу та інверсії в рангах.

Для одних і тих самих значень змінної значення коефіцієнта кореляції r-Спірмена буде завжди дещо більше, ніж значення коефіцієнта рангової кореляції τ-Кендала, тоді як рівень значущості залишиться однаковим або у коефіцієнта кореляції Кендала буде не набагато більшим.

Цей коефіцієнт розраховується за рангами обох вибірок. При цьому елементи вибірки розташовуються таким чином, щоб послідовність рангів однієї зі змінних являла собою натуральний ряд чисел 1, 2, 3, ..., n. Для кожного i-го члена послідовності рангів другої змінної встановлюємо числа p_i і q_i, які, відповідно, відображають прямий і зворотній порядок розташування наступних рангів. Далі знаходимо суми цих чисел $P = \sum p_i$ і $Q = \sum q_i$ та різницю між ними $S = P - Q$. Коефіцієнт рангової кореляції τ являє собою відношення цієї різниці до найбільшого можливого значення P і Q, тобто до найбільшої можливої суми p_i і q_i. Такого показника можна досягнути лише тоді, коли порядок рангів в обох послідовностях повністю збігається. Він становить: $S_{max} = n(n - 1) / 2$.

Коефіцієнт рангової кореляції Кендала (τ) розраховується за однією з еквівалентних формул:

$$\tau = S / S_{max} = 2S / n \cdot (n - 1) \text{ або}$$
$$\tau = 1 - [4Q / n \cdot (n - 1)] = [4P / n \cdot (n - 1)] - 1.$$

З нижньої формули видно, що для визначення τ достатньо мати або показник P, або Q. Зазвичай у формулу підставляють той показник, який має найменше значення.

Конкретних рекомендацій стосовно того, який із цих двох рангових коефіцієнтів слід застосовувати, на практиці немає. Ці коефіцієнти збудовані по-різному. У розрахунках r_s і τ однієї послідовності чисел зазвичай r_s-критерій більший, ніж τ-критерій. Але порівняння

цих коефіцієнтів не дає ніякої додаткової інформації щодо інтенсивності зв'язку.

Іншим простим показником тісноти зв'язку між двома статистичними рядами є *індекс Фехнера*. Для його визначення спочатку по кожному ряду знаходять середні ($\bar{x}$ і $\bar{y}$) і з'ясовують знаки відхилення $x_i - \bar{x}$ і $y_i - \bar{y}$. Кожна пара спостережень (x_i і y_i) буде характеризуватися однаковими або різними знаками (++, – –, + –, – +). Якщо кількість пар, що збіглися, позначити через v, а пари з різними знаками – через w, то індекс Фехнера можна виразити через формулу:

$$i = (v - w) / (v + w).$$

Половину відхилень, що дорівнюють нулю, відносять до v, другу – до w. Легко переконатися, що i змінюється в діапазоні від –1 до +1. Якщо $i > 0$, то кореляція позитивна, якщо $i < 0$ – негативна, а при $i = 0$ зв'язок відсутній.

Беззаперечна перевага індексу Фехнера – це простота розрахунків. Однак його значний недолік полягає в тому, що він враховує тільки кількість збігів і незбігів знаків відхилень. Тому його можна використовувати лише для приблизної оцінки зв'язку.

7.6. Коваріація

Програми наукових досліджень в агрономії передбачають одночасні спостереження за варіацією декількох ознак рослин у поєднанні з умовами довкілля й агротехніки. Наприклад, варіація довжини стебел і числа продуктивних колосків, збору білка та вмісту білка в насінні сої, висоти та маси рослин, кількості опадів, норм висіву, врожайності і т. ін.

Сумісне застосування кореляційного, регресійного та дисперсійного методів для уточнення експерименту одержало назву коваріаційного аналізу. Саме слово коваріація складається з початкових букв слів «кореляція» та «варіація».

Коваріаційний аналіз – це розділ аналізу даних, метою якого є визначення моделі зв'язку між залежним показником і певним набором кількісних та якісних показників. Тобто це свого роду синтез регресійного та дисперсійного аналізу.

Суть коваріаційного аналізу полягає у такому. Якщо між функціональною ознакою y і незалежною ознакою x є значний лінійний зв'язок, то цим методом можна статистично вирівняти умови проведення досліду стосовно до ознаки x і тим самим помітно знизити помилку досліду й отримати більше інформації про досліджуване явище. Коваріація дорівнює сумі добутків відхилень окремих значень двох ознак від їхніх середніх: $\sum(x-\bar{x})\cdot(y-\bar{y})$. Середню коваріацію ($cov$) розраховують за формулою:

$$cov = \sum(x-\bar{x})\cdot(y-\bar{y})\,/\,(n-1) \text{ або}$$
$$cov = (\sum xy - \sum x\cdot\sum y\,/\,n)\,/\,(n-1).$$

Коваріація вказує лише напрям зв'язку і кваліфікується позитивною, негативною або нульовою, коли зі збільшенням показників однієї ознаки значення іншої у середньому відповідно зростають, зменшуються або залишаються незмінними. Власне залежність ознак установлюють методом кореляції або ймовірнісного зв'язку.

Показник коваріації (cov) може набувати як позитивних, так і негативних значень. У широкому сенсі коваріація – це сукупність трьох статистичних показників: середніх арифметичних $\bar{x}$ і $\bar{y}$, сум квадратів відхилень $\sum(x-\bar{x})^2$ і $\sum(y-\bar{y})^2$ та суми добутків відхилень $\sum(x-\bar{x})\cdot(y-\bar{y})$. Розподіл цих показників за факторами варіювання і становить сутність коваріаційного аналізу.

Коваріаційний аналіз складається з трьох етапів:

- дисперсійний аналіз ряду x, y і добутків xy;
- розкладання дисперсії помилок (C_z) ряду y (залишок I) на суму квадратів відхилень, зумовлену регресією y за x, яку позначають C_b, і суму квадратів відхилень від регресії $C_{dy\cdot x}$ (залишок II);
- приведення експериментальних середніх ряду y до повної врівноваженості умов експерименту за рядом супутніх змінної x.

Отже, коваріаційний аналіз – це поширення методів дисперсійного аналізу при декількох змінних, а також кореляційного та регресійного аналізів на схеми польових, вегетативних і лабораторних експериментів.

Якщо між досліджуваною функціональною ознакою y і незалежною факторіальною змінною x існує лінійний зв'язок, то доцільно запланувати вимірювання показника x. Це дасть змогу отримати додаткову інформацію про досліджувану ознаку і використовувати

регресію для більш точної оцінки експерименту. Суму квадратів відхилень, зумовлену регресією y за x, визначають за формулою:

$$C_b = \frac{[\sum(x-\bar{x})(y-\bar{y})]^2}{\sum(x-\bar{x})^2}.$$

Суму квадратів відхилень від регресії знаходять за різницею як залишок: $C_{dy\cdot x}$ (помилка II) = C_z (помилка I) – C_b.

Вирівнювання результативної ознаки y проводять за формулою:

$$y_1 = y + b_{yx}(\bar{x} - x),$$

де y_1 – відкориговане значення функціональної ознаки; y – емпіричне значення показника; b_{yx} – коефіцієнт регресії y за x; $\bar{x} - x$ – різниця між середнім значенням факторіальної змінної по досліду $\bar{x}$ і фактичним її значенням x.

Зазвичай вирівнюють тільки підсумкові середні дані, тому в рівнянні регресії y і x будуть відповідати середнім показникам за варіантами досліду.

В агрономічних дослідженнях коваріаційний аналіз доцільно застосовувати для уточнення досліду у двох основних випадках:

1) якщо на результативну ознаку значною мірою впливають абіотичні, едафічні або технологічні чинники, які можна виміряти на початку досліду;
2) якщо у ході проведення досліду на досліджувану ознаку впливають незалежні від варіантів досліду, неконтрольовані причини: пошкодження хворобами та шкідниками, вилягання рослин через сильний вітер тощо.

Застосування коваріаційного аналізу передбачає незалежний від варіантів розподіл випадкового показника x. Якщо незалежна змінна x пов'язана з досліджуваною залежною y, то виключати її вплив не можна, оскільки це призводить до викривлення ефекту y. Наприклад, під час сортовивчення певні сорти можуть бути менш стійкі до хвороб, тож виключення цього впливу буде неправильним стосовно до більш стійких сортів. У дослідах із просапними культурами: кукурудзою, соняшником, буряками цукровими тощо, коли формування густоти посівів є результатом впливу досліджуваних чинників, не можна робити поправок на зрідженість.

Контрольні запитання та завдання

1. Дайте умовну класифікацію видів мінливості ознак.
2. Охарактеризуйте закон нормального розподілу.
3. Що являє собою розподіл Пуасона?
4. Яке значення має розподіл Стьюдента і Фішера на практиці?
5. З яких етапів складається техніка групування великої вибірки?
6. Перерахуйте параметричні та непараметричні характеристики вибірки.
7. Наведіть приклади локальної та інтервальної оцінки параметрів розподілу.
8. Назвіть найпоширеніші непараметричні критерії оцінки вибірки.
9. Яка основна мета дисперсійного аналізу?
10. Які виділяють види дисперсійного аналізу?
11. У чому полягає принципова різниця між функціональним і кореляційним зв'язком?
12. Яке значення мають кореляційний та регресійний аналізи в агрономічних дослідженнях?
13. Що показує коефіцієнт кореляції, детермінації і регресії?
14. Які завдання вирішує багатофакторний кореляційно-регресійний аналіз?
15. Яка різниця між квазілінійною та істотно нелінійною регресією?
16. Які формули застосовують для оцінки кореляції при якісній мінливості?
17. Які передумови для застосування рангових коефіцієнтів кореляції?
18. Охарактеризуйте основні методи рангової кореляції.
20. Яка основна мета коваріаційного аналізу? Назвіть його основні етапи.

Частина друга

ТЕХНІКА СТАТИСТИЧНОЇ ОБРОБКИ РЕЗУЛЬТАТІВ ДОСЛІДЖЕНЬ

Основним завданням статистичної обробки є узагальнення результатів досліджень, їхня математична оцінка, визначення певних закономірностей досліджуваних чинників, укладання об'єктивних висновків по досліду. Під час проведення стат истичного аналізу слід чітко дотримуватися особливостей його проведення, мети, яка ставиться дослідником, розумно використовувати допоміжні засоби: числові таблиці, номограми, комп'ютерні програми. Слід пам'ятати, що будь-яка сучасна комп'ютерна програма не може надати об'єктивну інформацію, якщо одержані дані сумнівні. Головний обов'язок дослідника полягає в пошуку достовірних експериментальних даних і методично правильному підході до проведення статистичної обробки одержаних даних. Існує великий вибір методів статистичної обробки, єдиною метою яких є правильна їх інтерпретація.

Перш ніж почати аналіз цифрового матеріалу та його статистичну обробку, необхідно провести їхню оцінку з агрономічного погляду. Агрономічна оцінка полягає у первинному оцінюванні результатів експерименту, з урахуванням комплексу чинників, що мають безпосередній вплив на отримані результати. Важливе значення на цьому етапі має інтуїтивне визначення та усування допущених під час оформлення первинного матеріалу помилок. Експерименти, проведені з порушенням методики, викривляють агрономічну сутність досліджуваних чинників і не мають практичної цінності.

Після проведення первинної агрономічної оцінки, ретельного аналізу методики і техніки проведення досліду, перевірки показників з первинної документації (щоденника, польового журнала) та усунення помилок проводять початкову цифрову обробку одержаних результатів експерименту.

1. Первинна обробка результатів досліду

Первинна обробка результатів досліду включає такі етапи:

1) приведення одержаних результатів до загальноприйнятих одиниць (урожайність, т/га; збір білка, кг/га; кількість рослин, шт./м^2; площа листя, тис. м^2/га і т. ін.);
2) якщо визначають урожайність, то її перераховують на стандартну вологість (для зернових – 14 %, для соняшнику – 8 % і под.);
3) складання таблиці результатів експерименту – визначення сум показників за варіантами, повтореннями. Розрахунок загальної суми показників, середніх показників за варіантами і дослідом у цілому.

Під час складання таблиці важливе точне визначення показників. Тут доцільно користуватися принципом десятої частки відсотка або трьох цифр. Так, за цим принципом, урожайність рослин, що не перевищує 10 т/га (більшість посівів) у таблицю заносять з точністю до 0,01, а якщо понад 10 т/га – то з точністю до 0,1 т/га. Тризначні показники (кількість рослин зернових з 1 м^2, натуру зерна та ін.) заносять до таблиці з точністю до 1 або 0,1 %.

Якщо у ході досліджень довелося вибракувати певні ділянки, але за умови, що кількість ділянок кожного варіанта не менше трьох, проводять відновлення вибракуваних, тобто тих ділянок досліду, що випали.

Завжди потрібно мати чітке уявлення про абсолютну помилку певних методів дослідження. Ураховуючи помилку вихідних даних, яка визначається варіативністю ознак і вимірюваної апаратури, необхідно точно розраховувати дані конкретного досліду. Логічно, що результати вимірювань не можуть бути точнішими, ніж досліджувані дані. Так, якщо показники врожайності занесені до таблиці з точністю до 0,1, то і статистичні розрахунки мають проводитися з цією точністю. Зайва точність нічого крім нераціональної витрати часу, не дасть.

Важливо, щоб під час проведення проміжних розрахунків кількість значущих цифр була на порядок вища, ніж їхня кількість у кінцевому результаті. У цьому разі є гарантія, що а ході самих розрахунків не будуть внесені істотні помилки. Після розрахунків отримані показники

округлюють до точності фактичних результатів експерименту. Під час округлення чисел необхідно дотримуватися таких правил:

1) якщо цифра, яка відкидається під час округлення, менша від 5, то остання цифра не змінюється (наприклад 9,238 скорочується до 9,2), а якщо більша від 5, то остання цифра збільшується на одиницю (наприклад 12,688 скорочується до 12,7);
2) якщо перед округленням за значущою цифрою стоїть 5, то її збільшують на одиницю, якщо вона непарна (наприклад 14,553 скорочується до 14,6), і залишають без змін, якщо вона парна або дорівнює нулю (наприклад, 16,451 скорочується до 16,4).

Деякі статистичні методи обробки експериментального матеріалу, наведені в цьому розділі, викликають певні труднощі в розрахунках. Найбільші труднощі виникають під час розрахунку сум квадратів відхилень $\sum(x - \bar{x})^2$ насамперед багатофакторних дослідів зі значною кількістю варіантів. Для практичних розрахунків доцільно застосовувати більш зручну формулу, яка не приводить до накопичення помилок округлення: $\sum(x - \bar{x})^2 = \sum x^2 - (\sum x)^2 / N$.

2. Визначення статистичних характеристик вибірки при кількісній мінливості ознаки

До кількісних належать ознаки, які можуть бути виміряні кількісно: урожайність зерна, кількість рослин і стебел на одиниці площі, висота та маса рослин, озерненість колоса, маса 1000 зерен, білковість зерна тощо. Розрізняють два види кількісної мінливості: дискретну (переривчасту) і неперервну. У першому випадку різниці між одиницями спостереження виражаються цілими числами, між якими немає переходів, як, наприклад, кількість рослин і стебел на квадратному метрі, кількість листків і міжвузлів на рослині, озерненість колоса тощо, у другому випадку значення ознаки виражені одиницями вимірювання довжини, маси, об’єму і т. ін.

Вибірки до 20-ти одиниць називають дуже маленькими, від 20-ти до 30-ти – маленькими і більше 30-ти одиниць – великими. Вибіркова сукупність являє собою ряд варіювальних показників досліджуваної ознаки, які записують у тій послідовності, у якій їх отримували. Статистичні характеристики розраховують за формулами, наведеними в табл. 20.

У цій таблиці x – це окремі значення ознаки в малих вибірках і групові середні у великих вибірках; $\bar{x}_1$ – переведені вихідні показники; A – довільний початок; f – частота, чисельність групи; n – обсяг вибірки; t – теоретичне значення критерію Стьюдента.

Для розрахунку середньої арифметичної і суми квадратів у табл. 20 зазначено декілька рівнозначних формул, які забезпечують фактично однаковий результат.

Щоб прискорити проведення розрахунків, вихідні показники доцільно перетворювати таким чином, щоб відкинути зайві цифри та

Таблиця 20

Формули для визначення статистичних характеристик вибірки при кількісній мінливості

Показник	Незгруповані дані	Згруповані дані
Середня арифметична	$\bar{x} = \frac{\sum x}{n} = A + \frac{\sum x_1}{n}$	$\bar{x} = \frac{\sum fx}{n} = A + \frac{\sum fx_1}{n}$
Дисперсія (розсіювання показників)	$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum x^2 - (\sum x)^2 / n}{n-1} = \frac{\sum x_1^2 - (\sum x_1)^2 / n}{n-1}$	$s^2 = \frac{\sum f(x-\bar{x})^2}{n-1} = \frac{\sum fx^2 - (\sum fx)^2 / n}{n-1} = \frac{\sum fx_1^2 - (\sum fx_1)^2 / n}{n-1}$
Стандартне відхилення	$s = \sqrt{s^2}$	$s = \sqrt{s^2}$
Коефіцієнт варіації	$V = 100s / \bar{x}$	$V = 100s / \bar{x}$
Помилка середньої	$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$	$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$
Відносна помилка середньої	$s_{\bar{x}}\,\% = \frac{s_{\bar{x}}}{\bar{x}} \cdot 100$	$s_{\bar{x}}\,\% = \frac{s_{\bar{x}}}{\bar{x}} \cdot 100$
Довірчий інтервал для середнього значення	$\bar{x} \pm t \cdot s_{\bar{x}}$	$\bar{x} \pm t \cdot s_{\bar{x}}$
Ступінь свободи	$n - 1$	$n - 1$

коми. Після проведення аналізу отримані дані «накладають» на вихідні емпіричні дані. Існують різні моделі перетворення фактичних показників, яким приділимо увагу в наступних розділах.

Розглянемо на конкретних прикладах техніку визначення статистичних характеристик вибірок різного розміру.

2.1. Малі вибірки (незгруповані дані)

Приклад 1. Під час визначення маси зерна з колоса стебла другого порядку рослин пшениці ярої отримано такі результати: 0,71; 0,75; 0,67; 0,69; 0,72; 0,77 і 0,73 г/колоса. Потрібно визначити середнє значення $\bar{x}$, помилку середньої $s_{\bar{x}}$, 95 і 99 %-й довірчі інтервали для середнього значення сукупності.

Проведення розрахунків. Фактичні результати доцільно перевести за відношенням $x_1 = xk - A = x \cdot 100 - 70$, тобто спочатку кожне значення помножити на 100, а потім відняти умовний середній показник ($A = 70$). У висновку отримуємо ряд простих цифр, зручних для розрахунків статистичних показників. Якщо розрахунки здійснюють на комп'ютері, переведення вихідних показників можна не проводити.

У табл. 21 наведено три варіанти розрахунків сум квадратів відхилень для цієї вибірки. Перший спосіб передбачає віднімання від фактичних показників загальної середньої $\bar{x}$, за другого способу статистичні розрахунки проводять з фактичними показниками, третій спосіб (найбільш раціональний) передбачає статистичну обробку з перетвореними показниками.

Подальші розрахунки доцільно проводити в такій послідовності:

$$\bar{x} = \sum x / n = 5{,}04 / 7 = 0{,}72;$$

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{0{,}007}{7-1} = 0{,}0012;$$

$$s = \sqrt{s^2} = \sqrt{0{,}0012} = 0{,}035;$$

$$V = \frac{s}{\bar{x}} \cdot 100 = \frac{0{,}035}{0{,}72} \cdot 100 = 4{,}86\ \%;$$

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{0{,}0012}{7}} = 0{,}013; \quad s_{\bar{x}}\,\% = \frac{s_{\bar{x}}}{\bar{x}} \cdot 100 = \frac{0{,}013}{0{,}72} \cdot 100 = 1{,}81\ \%.$$

Таблиця 21

Можливі способи розрахунків середньої арифметичної і суми квадратів відхилень

x	І спосіб		ІІ спосіб	ІІІ спосіб $x_1 = x - A$ ($A = 0{,}7$)		ІІІ спосіб $x_1 = x \cdot k - A$ ($A = 70$; $k = 100$)	
	$x - \bar{x}$	$(x - \bar{x})^2$	x^2	x_1	x_1^2	x_1	x_1^2
0,71	– 0,010	0,0001	0,5041	0,01	0,0001	1	1
0,75	0,030	0,0009	0,5625	0,05	0,0025	5	25
0,67	– 0,050	0,0025	0,4489	– 0,03	0,0009	– 3	9
0,69	– 0,030	0,0009	0,4761	– 0,01	0,0001	– 1	1
0,72	–	–	0,5184	0,02	0,0004	2	4
0,77	0,050	0,0025	0,5929	0,07	0,0049	7	49
0,73	0,010	0,0001	0,5329	0,03	0,0009	3	9
$\sum x = 5{,}04$	$\sum(x - \bar{x}) = 0$	$\sum(x - \bar{x})^2 = 0{,}0070$	$\sum x^2 = 3{,}6358$	$\sum x_1 = 0{,}14$	$\sum x_1^2 = 0{,}0098$	$\sum x_1 = 14$	$\sum x_1^2 = 98$
Серед-ня $\bar{x}$	$\sum x / n = 5{,}04 / 7 = 0{,}72$			$A + \frac{\sum x_1}{n} = 0{,}70 + 0{,}14/7 = 0{,}72$		$(A + \frac{\sum x_1}{n}) / k = (70 + \frac{14}{7}) / 100 = 72 / 100 = 0{,}72$	
Сума квадратів $\sum(x - \bar{x})^2$	0,0070		$\sum x^2 - (\sum x)^2 / n = 3{,}6358 - (5{,}04)^2 / 7 = 0{,}0070$	$\sum x_1^2 - (\sum x_1)^2 / n = 0{,}0098 - (0{,}14)^2 / 7 = 0{,}0070$		$[\sum x_1^2 - (\sum x_1)^2 / n] / k = [98 - (14)^2 / 7] / 100^2 = 0{,}0070$	

У кінці статистичного аналізу визначаємо довірчі інтервали для середнього значення вибірки при 95 і 99 %-му рівні ймовірності:

$\bar{x} \pm t_{05} \cdot s_{\bar{x}} = 0{,}72 \pm 2{,}45 \cdot 0{,}013 = 0{,}72 \pm 0{,}032$ (від 0,688 до 0,752 г);

$\bar{x} \pm t_{01} \cdot s_{\bar{x}} = 0{,}72 \pm 3{,}71 \cdot 0{,}013 = 0{,}72 \pm 0{,}048$ (від 0,672 до 0,768 г).

Теоретичні значення t_{05} і t_{01} знаходять за дод. Б для числа ступенів свободи $n - 1 = 7 - 1 = 6$.

Таким чином, середній показник цієї вибірки (маса зерна з колоса стебла другого порядку пшениці твердої ярої) з імовірністю 95 % коливається в межах від 0,688 до 0,752 г і з імовірністю 99 % – від 0,672 до 0,768 г. Імовірність допущення помилки в першому випадку

становить 5 %, а в другому – 1 %. Коефіцієнт варіації (V) становить 4,86 %. У цьому прикладі він оцінює помилку паралельних аналізів. Абсолютна помилка середньої ($s_{\bar{x}}$) становить 0,013 г, відносна ($s_{\bar{x}}$ %) – 1,81 %.

Приклад 2. У вегетативному досліді отримали такі показники кількості зерен пшениці твердої ярої за паралельними посудинами: 375, 411, 396, 417, 358, 386, 390, 405 шт./посудину. Потрібно визначити середню арифметичну ($\bar{x}$), помилку середньої ($s_{\bar{x}}$), відносну помилку різниці середніх ($s_{\bar{x}}$ %) і 95 %-й довірчий інтервал для середнього значення сукупності.

Проведення розрахунків. У цьому прикладі визначення середньої арифметичної і суми квадратів відхилень доцільно проводити за переведеними показники відповідно до рівняння: $x_1 = x - A = x - 400$. Розрахуємо середню арифметичну $\bar{x}$ і суми квадратів відхилень двома способами: за вихідними датами і переведеними показниками (табл. 22).

Таблиця 22

Визначення середнього показника ($\bar{x}$) і суми квадратів відхилень

x	Визначення за вихідними показниками x	Визначення за перетвореними показниками x_1	
	x^2	$x_1 = x - A$, ($A = 400$)	x_1^2
375	140625	– 25	625
411	168921	11	121
396	156816	– 4	16
417	173889	17	289
368	135424	– 32	1024
386	148996	– 14	196
390	152100	– 10	100
405	164025	5	25
$\sum x = 3148$	$\sum x^2 = 1240796$	$\sum x_1 = -52$	$\sum x_1^2 = 2396$
Середня $\bar{x}$	$\sum x / n = 393{,}5$	$A + \sum x_1 / n = 400 - 52 / 8 = 393{,}5$	
Сума квадратів $\sum(x - \bar{x})^2$	$\sum x^2 - (\sum x)^2 / n = 1240796 - (3148)^2 / 8 = 2058$	$\sum x_1^2 - (\sum x_1)^2 / n = 2396 - (52)^2 / 8 = 2058$	

Розрахунки слід проводити за тим самим принципом, що і в першому прикладі:

$$\bar{x} = \sum x / n = 3148 / 8 = 393{,}5;$$

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{2058}{8-1} = 294;$$

$$s = \sqrt{s^2} = \sqrt{294} = 17{,}1 \text{ шт./посудину};$$

$$V = \frac{s}{\bar{x}} \cdot 100 = \frac{17{,}1}{393{,}5} \cdot 100 = 4{,}35\ \%;$$

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{294}{8}} = 6{,}1 \text{ шт./посудину};$$

$$s_{\bar{x}}\,\% = \frac{s_{\bar{x}}}{\bar{x}} \cdot 100 = \frac{6{,}1}{393{,}5} \cdot 100 = 1{,}55\ \%.$$

У кінці статистичного аналізу визначаємо довірчі інтервали для середнього значення вибірки за 95 %-го рівня ймовірності:

$$\bar{x} \pm t_{05} \cdot s_{\bar{x}} = 393{,}5 \pm 2{,}37 \cdot 6{,}1 = 393{,}5 \pm 14{,}5 \text{ (від 379 до 408 шт.)}.$$

Теоретичні значення t_{05} знаходять за дод. Б для числа ступенів свободи $n - 1 = 8 - 1 = 7$.

Таким чином, середній показник досліджуваного ряду (кількість зерен пшениці твердої ярої в одній посудині) з імовірністю 95 % коливається в межах від 379 до 408 шт. зерен /посудину.

2.2. Великі вибірки (згруповані дані)

Якщо кількість показників досліду велика, їх слід подавати у вигляді систематизованого варіаційного ряду для подальшого поділу на певні групи. Кількість груп залежить від обсягу вибірки. Орієнтовно кількість груп має дорівнювати кореню квадратному із загального числа показників варіаційного ряду. Наприклад, при 30–40 спостереженнях доцільно виділити 6 груп, при 40–50 – 6–7 груп, при 50–60 – 7–8 груп, при 60–100 – 8–9 груп. Якщо кількість спостережень 100 і більше, виділяють 10–15 груп. Для дуже великих вибірок ($n > 400$) достатньо мати 14–15 груп, адже зі збільшенням їхньої кількості відбувається зайва деталізація вибірки, що значно ускладнює аналіз. Потрібно також розуміти, що невелика кількість груп (до шести) не дозволить визначити характерну особливість розподілу досліджуваної ознаки в сукупності.

Після з’ясування кількості груп (класів) визначають крок інтервалу, межі кожної групи та середні значення ознаки для кожної групи. Інтервали класів визначають за формулою:

$$i = R / k,$$

де R – діапазон варіабельності показників ($x_{макс} - x_{мін}$); k – кількість класів.

Межі класів потрібно визначити таким чином, щоб жодне значення не повторювалося у двох класах. Кінець кожного класу має бути меншим, ніж початок наступного на показник, що дорівнює прийнятій точності вимірювання. Наприклад, якщо перша група закінчується показником 15, то друга група повинна починатися показником 16; якщо перша група закінчується показником 15,5, то наступна має починатися з 15,6 і т. д.

За початок першої групи краще обирати таке число, яке забезпечить розміщення найменшого показника вибірки приблизно в центрі інтервалу першої групи. Наприклад, якщо довжина інтервалу для однієї групи становить 10 певних одиниць і досліджуваний показник варіює в межах від 14,7 до 96,1, то початок відліку першого інтервалу доцільно починати з 10-ти, при цьому початок інших груп починатиметься з 20-ти, 30-ти, 40-ка,..., 80-ти, 90-та.

За неперервної мінливості центральні показники кожної групи визначають шляхом додавання до початку кожної групи половини інтервалу. Так, для наведеного прикладу центральне значення першої групи становитиме 10 + 10 / 2 = 15, другої – 20 + 10 / 2 = 25 і т. д.

Частоту появи показників вибірки у кожній групі встановлюють шляхом розподілу емпіричних даних за класами. Для спрощення розподілу та уникнення можливих помилок під час пошуку показників які належать до одного класу, доцільно користуватися простими і водночас надійними способами, серед яких найпоширеніші – способи «штрихів» і «конвертиків».

Суть *методу «штрихів»* полягає в тому, що у вихідних даних закреслюють першу дату і заносять її до відповідної групи робочої таблиці, відмічаючи її вертикальною рискою. Далі закреслюють наступний показник вибірки і також заносять його до таблиці в певну групу. У робочій таблиці перші чотири дати кожного класу позначають вертикальними рисками, а п’яту – у вигляді діагоналі. Якщо кількість показників однієї групи перевищує 5, то всі наступні розподіляють за тим самим принципом, тобто 6-й, 7-й, 8-й і 9-й показники позначають вертикальними рисками, а 10-й – у вигляді діагоналі і т. д.

Метод «конвертиків» передбачає позначення емпіричних показників вибірки в робочий таблиці у вигляді конвертиків. Перші чотири показники кожної групи позначають точками у вигляді квадрата; наступні чотири показники позначають сторонами квадрата, які з'єднують попередньо нанесені точки, 9-й і 10-й показники позначають у вигляді діагоналей квадрата. Тож кожні 10 показників, наведені в одній групі, формують квадрат з виділеними діагоналями – ⊠. Сума частот усіх груп має дорівнювати обсягу вибірки ($\sum f = n$). Правильність розподілу показників перевіряють повторним складанням робочої таблиці.

Після розподілу показників за класами неперервний варіаційний ряд трансформується в дискретний. При цьому вихідні дані кожної групи прирівнюються до середньогрупових показників, які використовують для подальших статистичних розрахунків. Оскільки при цьому втрачається частина інформації, метод розрахунку статистичних характеристик за згрупованими даними не є абсолютно точним. Водночас, для великих вибірок похибки цього методу незначні і їх до уваги можна не брати.

Щоб краще наочно відобразити тенденції розподілу досліджуваної ознаки в загальній сукупності, варіаційні ряди зазвичай подають у вигляді ступінчастої гістограми або полігону розподілу, на якому середні групові з'єднують ламаною лінією. Графічне зображення варіаційного ряду називається кривою розподілу.

Розглянемо механізм групування даних і розрахунку статистичних показників неперервної мінливості на конкретному прикладі.

Приклад 1. У фазу повної стиглості зерна було виміряно довжину головного стебла жита озимого в сантиметрах (табл. 23). Потрібно згрупувати показники за класами і провести статистичний аналіз: визначити $\bar{x}$, s, V, $s_{\bar{x}}$, довірчі інтервали $\bar{x} \pm t_{05} \cdot s_{\bar{x}}$ і побудувати гістограму та полігон розподілу 120-ти рослин жита озимого за довжиною головного стебла.

Порядок розрахунків. Граничні показники довжини головного стебла жита озимого становлять 54,7 і 137,8 см. У цьому діапазоні довжина стебла може набувати будь-яких значень, отже, спостеріга-

Таблиця 23

Довжина головного стебла 120 рослин жита озимого, см

91,3	128,0	104,2	122,2	62,5	109,5	57,8	107,0	80,4	87,6
54,8	94,3	101,5	75,8	92,5	74,5	126,5	124,8	94,1	70,6
117,2	80,9	101,2	101,4	86,5	84,5	93,7	82,2	61,0	107,2
95,1	113,5	90,9	114,2	76,8	**137,8**	73,5	117,1	84,6	66,4
69,1	102,7	101,2	88,6	108,8	115,2	81,1	97,4	71,1	94,2
96,6	61,2	75,5	60,4	78,0	98,1	93,1	57,2	69,1	123,7
133,3	84,4	68,1	93,8	63,0	94,5	113,0	83,5	95,8	119,2
71,6	110,4	104,4	128,0	100,4	79,1	95,2	89,4	128,8	80,3
101,5	92,2	122,6	103,3	**54,7**	89,1	59,4	72,9	91,2	112,4
63,2	63,5	83,5	77,5	96,0	83,5	77,8	87,7	96,4	105,2
113,2	81,4	93,3	114,0	117,3	122,1	91,7	97,4	108,1	114,5
92,5	72,4	115,2	100,6	87,5	97,0	123,4	109,2	118,5	91,5

ємо неперервну кількісну мінливість, і для проведення розрахунків доцільно згрупувати дані за певними інтервалами.

У цьому прикладі показники доцільно розподілити на дев'ять класів. Показник інтервалу визначають шляхом поділу розмаху варіювання ($R = x_{max} - x_{min}$) на кількість класів (k). Для зручності за величину інтервалу доцільно брати ціле число, навіть якщо кількість класів при цьому буде дещо більшою або меншою від визначених вище орієнтовних чисел. Оскільки в цьому прикладі розмах зміни показників становить 81,7 см ($R = x_{max} - x_{min}$ = 137,8 – 54,7 = 83,1 см), а кількість класів – 9, доцільно обрати груповий інтервал цілим числом – 10 см, оскільки якщо інтервал становитиме 9 см, то крайні показники вибірки можуть не потрапити до жодної з груп. Як зазначено вище, початок першої групи слід вибрати так, щоб найменший показник вибірки потрапив приблизно в середину першого інтервалу. Таким чином, у цьому прикладі початок першого групового інтервалу становитиме 54,7 – 10 / 2 = 49,7 ≈ 50 см. Початок другої групи буде починатися з 60 (50 + 1i), третьої з 70 (50 + 2i) і т. д. Відповідно до цього центром першої групи буде число (50 + 60) / 2 = 55, другої – (60 + 70) / 2 = 65 і т. д. Кінець кожної групи визначається відніманням від початку наступної групи показника точності вимірювання (для цього прикладу – 0,1 см). Тож перша група буде закінчуватися числом 60 – 0,1 = 59,9; друга – 70 – 0,1 = 69,9; третя –

Таблиця 24

Розподіл вихідних показників за групами

Група		Метод «штрихів»	Метод «конвертиків»	Частота	Групові середні
1	50,0–59,9			5	55
2	60,0–69,9			11	65
3	70,0–79,9			14	75
4	80,0–89,9			19	85
5	90,0–99,9			26	95
6	100,0–109,9			18	105
7	110,0–119,9			15	115
8	120,0–129,9			10	125
9	130,0–139,9			2	135
Сума				120	

80 – 0,1 = 79,9 і т. д. Межі кожної групи показників показані з лівої сторони табл. 24.

Для визначення частот показників у кожній групі емпіричні дані вносимо до свого класу, використовуючи один з розглянутих вище способів («штрихів» чи «конвертиків»). Правильність розподілу показників перевіряємо їхнім повторним рознесенням до аналогічної таблиці.

Після розподілу показників за групами можна легко визначити характер їхньої мінливості. У цьому прикладі найбільша кількість показників входить в інтервал 90,0–99,9 см (п'ята група). Група з найбільшою кількістю показників називається модальною (у цьому прикладі – п'ята група). Лімітами (краями) наведеного розподілу є показники першої і дев'ятої груп.

Далі визначаємо середньоарифметичний показник вибірки $\bar{x}$ і суму квадратів відхилень (табл. 25). Після цього розраховуємо статистичні характеристики вибірки: дисперсію s^2, стандартне відхилення s, коефіцієнт варіації V, абсолютну і відносну помилку вибіркової середньої ($s_{\bar{x}}$ *і* $s_{\bar{x}}$ % відповідно) та довірчий інтервал для генеральної середньої:

$$\bar{x}\text{ (середньозважена)} = \sum fx / n = 11260 / 120 = 93{,}8;$$

Таблиця 25

Групування вихідних показників, визначення середньої арифметичної та суми квадратів при неперервній мінливості

Група	Частота, f	Центральний показник, x	Розрахунок суми квадратів: fx	x^2	fx^2
50,0–59,9	5	55	275	3025	15125
60,0–69,9	11	65	715	4225	46475
70,0–79,9	14	75	1050	5625	78750
80,0–89,9	19	85	1615	7225	137275
90,0–99,9	26	95	2470	9025	234650
100,0–109,9	18	105	1890	11025	198450
110,0–119,9	15	115	1725	13225	198375
120,0–129,9	10	125	1250	15625	156250
130,0–139,9	2	135	270	18225	36450
Сума	120	–	11260	–	1101800
Середня $\bar{x}$			$\sum fx / n = 11260 / 120 = 93{,}8$		
Сума квадратів $\sum f(x-\bar{x})^2$			$\sum fx^2 - (\sum fx)^2 / n = 1101800 - (11260)^2 / 120 = 45237$		

$$s^2\text{ (дисперсія)} = \sum f(x-\bar{x})^2 / (n-1) = 45237 / (120-1) = 380{,}1;$$
$$s\text{ (стандартне відхилення)} = \sqrt{s^2} = \sqrt{380{,}1} = 19{,}5\text{ см};$$
$$V\text{ (коефіцієнт варіації)} = 100 \cdot s / \bar{x} = 100 \cdot 19{,}5 / 93{,}8 = 20{,}8\ \%;$$
$$s_{\bar{x}}\text{ (абсолютна помилка середньої)} = s / \sqrt{n} = 19{,}5 / \sqrt{120} = 1{,}78\text{ см};$$
$$s_{\bar{x}}\ \%\text{ (відносна помилка середньої)} = 100 \times s_{\bar{x}} / \bar{x} = 100 \cdot 1{,}78 / 93{,}8 = 1{,}9\ \%;$$
$$\bar{x} \pm t_{05} \cdot s_{\bar{x}} = 93{,}8 \pm 1{,}96 \cdot 1{,}78 = 93{,}8 \pm 3{,}5\text{ (від 90,3 до 97,3 см)};$$
$$\bar{x} \pm t_{01} \cdot s_{\bar{x}} = 93{,}8 \pm 1{,}96 \cdot 2{,}58 = 93{,}8 \pm 5{,}1\text{ (від 88,7 до 98,9 см)}.$$

Теоретичне значення критерію $t_{05} = 1{,}96$ знаходимо за дод. Б при кількості ступенів свободи $n - 1 = 120 - 1 = 119$.

Таким чином, з ймовірністю 95 % середній показник вибірки буде знаходитися в інтервалі від 90,3 до 97,3 см, і з ймовірністю 99 % в інтервалі від 88,7 до 98,9 см. Абсолютна помилка вибіркової середньої становить 1,78 см, відносна – 1,9 %, коефіцієнт варіації довжини головного стебла жита озимого – 20,8 %.

Для кращого розуміння статистичних розрахунків доцільно збудувати гістограму та полігон розподілу рослин жита озимого за довжиною головного стебла (рис. 22).

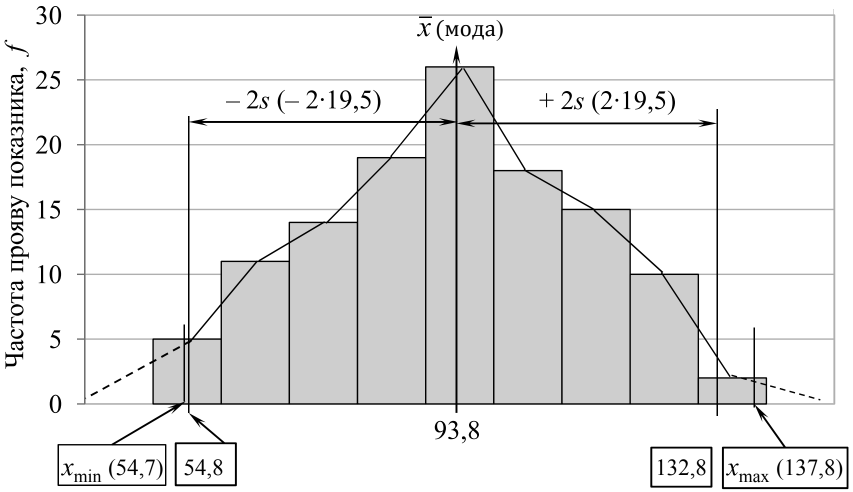


Рис. 22. Гістограма та полігон розподілу часток показників довжини головного стебла 120-ти рослин жита озимого

На горизонтальній осі абсцис відмічають межі класів, а на вертикальній осі ординат – частоту *f*. У підсумку отримуємо ступеневий графік розподілу частот у вигляді стовпчиків, висота яких пропорційна частотам (*f*), а ширина – інтервалу *i*. Поєднавши лініями центральні показники груп, отримуємо криву розподілу (полігон).

3. Визначення статистичних характеристик вибірки під час вивчення якісних ознак

Ознаки, які не підлягають кількісному вимірюванню, належать до якісних (атрибутивних). Це може бути наявність або відсутність певних ознак, різні види хвороб, колір зерна або квіток, форма плодів, реакція на певні агротехнічні елементи тощо. Найчастіше в дослідженні якісних ознак трапляється альтернативний розподіл, тобто коли є дві можливі градації ознаки (наявність або відсутність ознаки). Прикладом альтернативної мінливості може бути колір зерна – білий або чорний, ураженість рослин – здорові чи уражені, тип стебла – прямостояче або сланке та ін.

Типові формули для визначення статистичних характеристик при якісний мінливості наведено в табл. 26. У цій таблиці p_1, p_2,..., p_k і q – частки ознаки в сукупності, n_1, n_2, ..., n_k – кількість класів, k – кількість градацій ознаки, N – обсяг вибірки, t – теоретичне значення критерію Стьюдента.

Сутність статистичної характеристики при якісній мінливості полягає у розподілі вихідних показників за групами, визначенні середнього показника частки, мінливості ознаки та довірчого інтервалу для певного рівня ймовірності, у межах якого розміщений показник частки в генеральній сукупності. Розраховуючи коефіцієнт варіації, слід мати на увазі, що максимально можлива мінливість ознак за альтернативного розподілу (дві градації ознак) становить 50,0 %, при трьох градаціях ознак – 33,3 %, чотирьох – 25,0 %, п'яти – 20,0 %, шести – 16,7 %, семи – 14,3 %.

На конкретних прикладах покажемо послідовність визначення статистичних характеристик при якісній мінливості.

Приклад 1. Серед 300-т проаналізованих рослин кукурудзи 40 були уражені іржею. Визначити 95 і 99 %-й довірчі інтервали для генеральної частки уражених рослин у сукупності.

Таблиця 26

Типові формули для розрахунку статистичних характеристик вибірки при якісній мінливості

Показник	Формула
Частка ознаки при альтернативному розподілі (k =2)	$p = n_1 / N$; $q = 1 - p$; $p_1 = n_1 / N$; $p_2 = n_2 / N$; $p_k = n_k / N$
Стандартне відхилення при альтернативному розподілі (k =2) при $k > 2$	$s = \sqrt{pq}$ $s = \sqrt[k]{p_1 \cdot p_2 \cdot \ldots p_k}$ $lg\, s = (lg\, p_1 + lg\, p_2 + \ldots + lg\, p_k) / k$
Коефіцієнт варіації	$V_p = 100 \cdot s / s_{max}$
Помилка частки	$s_p = s / \sqrt{n} = \sqrt{pq/n}$
Довірчий інтервал для частки ознаки в сукупності	$p \pm t \cdot s_p$
Ступені свободи	$n - 1$

Проведення розрахунків. У цьому прикладі спостерігаємо альтернативний розподіл досліджуваної ознаки на дві групи: здорові та уражені рослини. Перша група – вражені рослини ($n_1 = 40$), друга група – здорові рослини ($n_2 = N - n_1 = 300 - 40 = 260$).

Далі визначаємо частки уражених p та здорових q рослин кукурудзи, стандартне відхилення частки s, коефіцієнт варіації V_p, та помилку вибіркової частки s_p. Частка уражених p і здорових q рослин становитиме:

$$p = n_1 / N = 40 / 300 = 0{,}133 \ (13{,}3\ \%);\ q = 1 - p = 0{,}867 \ (86{,}7\ \%);$$
$$s \text{ (стандартне відхилення частки)} = \sqrt{pq} = \sqrt{0{,}133 \cdot 0{,}867} = 0{,}34 \ (34\ \%);$$
$$s_{max} \text{ (максимально можлива мінливість)} = 0{,}50, \text{ оскільки } k = 2;$$
$$V_p \text{ (коефіцієнт варіації)} = 100 \cdot s / s_{max} = 100 \cdot 0{,}34 / 0{,}50 = 68{,}0\ \%;$$
$$s_p \text{ (помилка вибіркової частки)} = \sqrt{pq/n} = \sqrt{0{,}133 \cdot 0{,}867/300} =$$
$$= 0{,}02 \ (2\ \%).$$

Для наведеного альтернативного розподілу з обсягом вибірки $n = 300$ довірчій інтервал генеральної частки уражених іржею рослин у сукупності з ймовірністю 95 і 99 % становитиме:

$$p \pm t_{05} \cdot s_p = 0{,}133 \pm 1{,}96 \cdot 0{,}02 = 0{,}133 \pm 0{,}039 \text{ (від 9,4 до 17,2 \%)};$$
$$p \pm t_{01} \cdot s_p = 0{,}133 \pm 2{,}58 \cdot 0{,}02 = 0{,}133 \pm 0{,}052 \text{ (від 8,1 до 18,5 \%)}.$$

Теоретичне значення t_{05} і t_{01} знаходимо за таблицею дод. Б для $n - 1 = 300 - 1 = 299$ ступенів свободи.

Отже, генеральна частка рослин кукурудзи, уражених іржею, у досліджуваній сукупності з ймовірністю 95 % буде варіювати в діапазоні від 9,4 до 17,2 %, а з ймовірністю 99 % – від 8,1 до 18,5 %.

Приклад 2 (для $k > 2$). У вибірці з 1000-чі зернівок було 758 зернівок пшениці твердої ярої, 180 – пшениці м'якої і 62 – ячменю ярого. Визначити процентну частку кожної групи зерен у сукупності та їхні довірчі інтервали в генеральній сукупності з ймовірністю 95 і 99 %.

Проведення розрахунків. Спочатку визначаємо частку кожної групи зерен у сукупності:

$$p_1 \text{ (частка зерен пшениці твердої ярої)} = n_1 / N =$$
$$= 758 / 1000 = 0{,}758 \ (75{,}8\ \%);$$
$$p_2 \text{ (частка зерен пшениці м'якої ярої)} = n_2 / N =$$
$$= 180 / 1000 = 0{,}180 \ (18{,}0\ \%);$$

$$p_3 \text{ (частка зерен ячменю ярого)} = n_3 / N =$$
$$= 62 / 1000 = 0{,}062 \ (6{,}2\ \%).$$

Далі знаходимо стандартне відхилення частки (s), коефіцієнт варіації (V_p), помилку частки (s_p) і довірчі інтервали з 5 і 1 %-м рівнем значущості:

$$lg\, s = (lg\, p_1 + lg\, p_2 + lg\, p_3) / k = (lg\, 75{,}8 + lg\, 18{,}0 + lg\, 6{,}2) / 3 =$$
$$= (1{,}880 + 1{,}255 + 0{,}792) / 3 = 1{,}309;$$
$$s \text{ (стандартне відхилення частки)} = antilg\, 1{,}309 = 0{,}1943 \approx 0{,}194 \ (19{,}4\ \%);$$
$$\text{при } k = 3 \text{ максимально можлива мінливість } (s_{max}) \text{ дорівнює } 0{,}333,$$
$$\text{звідси } V_p \text{ (коефіцієнт варіації)} = 100s / s_{max} =$$
$$= 0{,}194 \cdot 100 / 0{,}333 = 58{,}3\ \%;$$
$$s_p \text{ (помилка частки)} = s / \sqrt{n} = 0{,}194 / \sqrt{1000} = 0{,}0061 \approx 0{,}006 \ (0{,}6\ \%).$$

Для представленого якісного розподілу з обсягом вибірки 1000 зернівок довірчий інтервал генеральної частки зерен пшениці твердої ярої у сукупності з імовірністю 95 і 99 % становитиме:

$$p \pm t_{05} \cdot s_p = 0{,}758 \pm 1{,}96 \cdot 0{,}006 = 0{,}758 \pm 0{,}011 \text{ (від 74,7 до 76,9 \%)};$$
$$p \pm t_{01} \cdot s_p = 0{,}758 \pm 2{,}58 \cdot 0{,}006 = 0{,}758 \pm 0{,}015 \text{ (від 74,3 до 77,3 \%)}.$$

Теоретичне значення t_{05} і t_{01} знаходимо за таблицею, наведеною в дод. Б, для $n - 1 = 1000 - 1 = 999$ ступенів свободи.

За аналогією знаходимо довірчі інтервали генеральної частки зерен пшениці м'якої ярої (n_2) і ячменю ярого (n_3) у генеральній сукупності і ймовірністю 95 і 99 %.

Для зерен пшениці м'якої ярої:

$$p \pm t_{05} \cdot s_p = 0{,}180 \pm 1{,}96 \cdot 0{,}006 = 0{,}180 \pm 0{,}011 \text{ (від 16,9 до 19,1 \%)};$$
$$p \pm t_{01} \cdot s_p = 0{,}180 \pm 2{,}58 \cdot 0{,}006 = 0{,}180 \pm 0{,}015 \text{ (від 16,5 до 19,5 \%)};$$

Для зерен ячменю ярого:

$$p \pm t_{05} \cdot s_p = 0{,}062 \pm 1{,}96 \cdot 0{,}006 = 0{,}062 \pm 0{,}011 \text{ (від 5,1 до 7,3 \%)};$$
$$p \pm t_{01} \cdot s_p = 0{,}062 \pm 2{,}58 \cdot 0{,}006 = 0{,}062 \pm 0{,}015 \text{ (від 4,7 до 7,7 \%)}.$$

Отже, на основі результатів вибірки можна вважати, що генеральна частка зерен пшениці твердої ярої у досліджуваній сукупності з імовірністю 95 % варіюватиме в діапазоні від 74,7 до 76,9 % та з імовірністю 99 % – у діапазоні від 74,3 до 77,3 %.

Імовірність того, що частка зерен пшениці м'якої ярої у генеральній сукупності буде міститися в інтервалі від 16,9 до 19,1 %, становить 95 % і в інтервалі від 16,5 до 19,5 % – 99 %. Частка зерен ячменю ярого з імовірністю 95 % буде міститися в інтервалі від 5,1 до 7,3 %, а з імовірністю 99 % – в інтервалі від 4,7 до 7,7 %.

4. Оцінка відповідності між емпіричними і теоретичними розподілами за критерієм χ^2

У практиці агрономічних досліджень досить часто виникає необхідність з'ясувати наскільки отриманий емпіричний матеріал підтверджує теоретичний розподіл, наскільки фактичні дані збігаються з теоретичними. Постає завдання статистичної оцінки різниці між розподілом фактичних і теоретичних показників, установленням того, в яких випадках і якою мірою можна вважати цю різницю достовірною та навпаки, коли її слід вважати неістотною. В останньому випадку зберігається гіпотеза, на основі якої розраховано теоретично очікувані показники. Таким варіаційно-статистичним засобом перевірки гіпотези є метод хі-квадрат (χ^2). На практиці цей показник називають «критерій відповідності» Пірсона. За його допомогою можна з тією чи іншою ймовірністю судити про міру відповідності експериментальних показників теоретичним.

З формальних позицій порівнюється два варіаційних ряди, дві сукупності: одна – емпіричний розподіл, друга представляє вибірку з тими самими параметрами, що й емпірична вибірка, але її частотний розподіл точно відповідає вибраному теоретичному закону (нормальному, біноміальному тощо), якому передбачувано підпорядковується характер розподілу досліджуваного показника.

Загальний вигляд формули критерію згоди такий:

$$\chi^2 = \sum(f - F)^2 / F,$$

де f – фактична частота спостережень, F – теоретично очікувана частота прояву для цього класу.

Нульова гіпотеза передбачає, що достовірних різниць між порівнюваними розподілами немає. Для оцінки істотності цих різниць слід звертатися до спеціальної таблиці критичних значень χ^2 (дод. Г) і, порівнявши розрахований показник з теоретичним, вирішувати, істотно чи неістотно відрізняється емпіричний розподіл від теоретичного. У такий спосіб гіпотеза про відсутність різниці буде або спростована, або не спростована. Якщо розрахований показник χ^2 дорівнює табличному або перевищує його, такий емпіричний розподіл достовірно відрізняється від теоретичного. Отже, гіпотеза про відсутність цих різниць буде спростована. Якщо $\chi^2_{факт.} < \chi^2_{теор.}$, нульова гіпотеза ($H_0$) залишається в силі. Зазвичай прийнято вважати допустимим рівень значущості $\alpha = 0{,}05$, оскільки в цьому випадку лишається тільки 5 % шансів, що нульова гіпотеза правильна і, відповідно, є достатньо підстав (95 %), щоб від неї відмовитись.

Певну проблему становить правильність визначення числа ступенів свободи df, для яких із таблиці беруть значення χ^2-критерію. Для визначення числа ступенів свободи від загальної кількості класів (k) слід відняти кількість обмежень, тобто число параметрів, які використовують для розрахунку теоретичних частот. Залежно від типу розподілу досліджуваної ознаки, формула для розрахунку числа ступенів свободи буде змінюватися. Для альтернативного розподілу ($k = 2$) у розрахунках бере участь лише один параметр (обсяг вибірки), відповідно кількість ступенів свободи становитиме $df = k - 1$. Для поліноміального розподілу формула аналогічна: $df = k - 1$. Для перевірки відповідності варіаційного ряду розподілу Пуассона використовують уже два параметри – обсяг вибірки і середнє значення, яке кількісно збігається з дисперсією; кількість ступенів свободи $df = k - 2$. Під час перевірки відповідності емпіричного розподілу варіантів нормальному або біноміальному розподілу число ступенів свободи беруть як кількість фактичних класів мінус три умови побудови рядів – обсяг вибірки (k), середнє ($\bar{x}$) і дисперсію (s), ($df = k - 3$).

У типових умовах застосування критерію відповідності число ступенів свободи визначають за формулою: $(c - 1) \cdot (k - 1)$, де c – кількість рядків, k – кількість колонок у таблиці.

Одразу слід застережити, що критерій χ^2 можна застосовувати лише для вибірок обсягом не менше 25-ти варіантів, а частота окремих класів має бути не менше чотирьох (при менш строгому підході за мінімум приймають три спостереження).

Критерій χ^2 широко використовується в генетичному аналізі відповідності розщеплення гібридів теоретично очікуваному розподілу (приклад 1), для оцінки тісноти зв'язків (незалежні чи пов'язані) у розподілі об’єктів сукупності (приклади 2–5), визначення міри відповідності фактичного розподілу досліджуваної ознаки нормальному (приклади 6, 7) та оцінки відповідності двох емпіричних розподілів між собою (оцінка однорідності розподілів).

Приклад 1. Припустимо, що в результаті схрещування двох сортів жита одержали 533 зернини, характерних для одного сорту, і 167 зернин – для другого. Потрібно визначити, чи відповідають результати досліду теоретично очікуваному співвідношенню зерен одного сорту до другого як 3:1. Співвідношення 3:1 береться в ролі нульової гіпотези (H_0), яку потрібно перевірити.

Проведення розрахунків. Загальна сума зерен жита обох типів становить 700 шт. (533 + 167). При співвідношенні 3:1 теоретично очікувані частоти F становлять:

$$F_1 = 700 \cdot 3 / 4 = 525;\ F_2 = 700 \cdot 1 / 4 = 175;$$
$$\sum(F_1 + F_2) = 525 + 175 = 700.$$

Підставляючи емпіричні та теоретично очікувані частоти у формулу χ^2, отримують:

$$\chi^2 = \sum(f - F)^2 / F = (533 - 525)^2 / 525 + (167 - 175)^2 / 175 = 0{,}49.$$

Теоретичне значення χ^2 = 3,84 знаходять за дод. Г для $(c - 1) \times (k - 1) = (2 - 1) \cdot (2 - 1) = 2$ ступенів свободи.

Розрахунки критерію χ^2 також можна проводити в послідовності, наведеній в табл. 27.

$$\chi^2 = \sum(f_1 - F_1)^2 / F_1 + \sum(f_2 - F_2)^2 / F_2 = 0{,}12 + 0{,}37 = 0{,}49.$$

Таблиця 27

Розрахунок критерію χ^2

Показник	Зерна характерні для		Сума
	сорту *A*	сорту *B*	
Очікуваний розподіл (H_0)	3	1	4
Фактичні частоти (f)	533	167	700
Очікувані частоти (F)	525	175	700
Різниця ($f - F$)	8	– 8	–
Квадрат різниці $(f - F)^2$	64	64	–
Відношення $(f - F)^2 / F$	0,12	0,37	χ^2 = 0,49

Висновок. Оскільки $\chi^2_{факт.} < \chi^2_{теор.}$, різниця між фактичними і теоретичними частотами неістотна, тобто нульова гіпотеза (H_0) не спростовується.

Приклад 2. У табл. 28 наведено одержані в досліді з ячменем ярим показники кількості листків на стеблах рослин. Визначити, чи відповідає емпіричний розподіл частот f теоретично очікуваному у співвідношенні 10 : 5 : 3 : 1.

Таблиця 28

Визначення теоретичних частот (*F*) і критерію відповідності (χ^2) спів відношення кількості листків рослин ячменю ярого

Показник	Кількість листків на стеблі				Сума
	7	6	5	4	
Очікуваний розподіл (H_0)	10	5	3	1	19
Фактичні частоти (f)	497	263	141	37	938
Очікувані частоти (F)	493,7	246,8	148,1	49,4	938
Різниця ($f - F$)	3,3	16,2	– 7,1	– 12,4	–
Квадрат різниці $(f - F)^2$	10,89	262,44	50,41	153,76	–
Співвідношення $(f - F)^2 / F$	0,02	1,06	0,34	3,11	χ^2 = 4,53

Проведення розрахунків. Техніку проведення розрахунку критерію χ^2 показано в табл. 28. Очікувані частоти визначають шляхом множення теоретично очікуваної частки стебел у сукупності на загальну кількість спостережень. Зокрема, частка стебел із сімома листками у загальній сукупності має становити: $F_1 = 938 \cdot 10 / 19 = 493{,}7$ шт., із шістьма листками – $F_2 = 938 \cdot 5 / 19 = 246{,}8$ шт., з п’ятьма листками – $F_3 = 938 \cdot 3 / 19 = 148{,}1$ шт., і з чотирма листками – $F_4 = 938 \cdot 1 / 19 = 49{,}4$ шт.

$$\chi^2 = \sum(f_1 - F_1)^2 / F_1 + (f_2 - F_2)^2 / F_2 + (f_3 - F_3)^2 / F_3 + (f_4 - F_4)^2 / F_4 = 4{,}53.$$

Теоретичне значення χ^2_{05} = 7,81 знаходять за дод. Г для $(c - 1) \times (k - 1) = (2 - 1) \cdot (4 - 1) = 3$ ступенів свободи.

Висновок. Оскільки $\chi^2_{факт.} < \chi^2_{05}$, достовірної різниці між фактичними і теоретично очікуваними частотами за співвідношенням 10:5:3:1 немає (нульова гіпотеза не спростовується).

Приклад 3. Фактичну кількість листків різного типу рослин тритикале ярого стебел різних порядків наведено в табл. 29. Очікуване співвідношення частот становить – 1 : 3 : 1 : 1 : 3 : 1 : 1 : 3 : 1. Визначимо за χ^2-критерієм відповідність теоретичного розподілу емпіричному.

Послідовність проведення розрахунків наведено в табл. 29. Очікувані частоти F визначають як добуток теоретично очікуваної частки листків конкретного типу та їхньої загальної кількості. Зокрема, частка листків, розміщених під кутом 45 ° до стебла, становить 1/15, тож очікувана кількість цих листків дорівнюватиме: для $F_1 = 318 \cdot 1/15 = = 21{,}2$ шт.; для $F_2 = 318 \cdot 3/15 = 23{,}6$ шт. і т. д.

$$\chi^2 = \sum(f_1 - F_1)^2 / F_1 + (f_2 - F_2)^2 / F_2 + ... + (f_9 - F_9)^2 / F_9 =$$
$$= 2{,}87 + 16{,}51 + ... + 1{,}81 = 30{,}44.$$

Теоретичне значення $\chi^2_{05} = 15{,}51$ знаходять за дод. Г для $(c - 1) \times \times (k - 1) = (2 - 1) \cdot (9 - 1) = 8$ ступенів свободи.

Висновок. Оскільки $\chi^2_{факт.} > \chi^2_{05}$, між очікуваними і фактичними частотами (за співвідношенням 1 : 3 : 1 : 1 : 3 : 1 : 1 : 3 : 1) існує достовірна різниця (нульова гіпотеза не спростовується).

Приклад 4. Припустимо, що в географічних дослідженнях було обстежено 150 полів (табл. 30). Потрібно визначити, чи є істотна різ-

Таблиця 29

Розрахунок теоретичних частот (F) і критерію відповідності (χ^2) кількості листків різних типів рослин тритикале ярого

Стебла рослин	Кут між листком і стеблом	Очікуваний розподіл, H_0	Емпіричні частоти, f	Очікувані частоти, F	Різниця $(f - F)$	Квадрат різниці $(f - F)^2$	Співвідношення $(f - F)^2 / F$
Головні	Кут 45 °	1	29	21,2	7,8	60,8	2,87
	Еректоїдні	3	96	63,6	32,4	1049,8	16,51
	Горизонтальні	1	26	21,2	4,8	23,0	1,08
Першого порядку	Кут 45 °	1	18	21,2	–3,2	10,2	0,48
	Еректоїдні	3	53	63,6	–10,6	112,4	1,77
	Горизонтальні	1	16	21,2	–5,2	27,0	1,27
Другого порядку	Кут 45 °	1	17	21,2	–4,2	17,6	0,83
	Еректоїдні	3	48	63,6	–15,6	243,4	3,82
	Горизонтальні	1	15	21,2	–6,2	38,4	1,81
Сума		15	318	318	–	–	$\chi^2 = 30{,}44$

ниця за рівнем забур'яненості полів кукурудзи на зерно залежно від впливу попередника.

Проведення розрахунків. Необхідно перевірити нульову гіпотезу, відповідно до якої вибір попередника не впливає на забур'яненість посівів і коливання співвідношень середньо- та слабозабур'янених полів у кожній колонці табл. 30 є випадковим.

На основі нульової гіпотези для кожної клітини табл. 30 розраховують очікувані значення F. Для визначення очікуваних частот загальну кількість полів у кожній групі множать на очікувану частку слабозабур'янених (53,3 %) або середньозабур'янених (46,7 %) полів.

Таблиця 30

Фактична та очікувана забур'яненість посівів кукурудзи на зерно після різних попередників (*A* – чистий пар; *B* – буряки цукрові)

Попередник	Забур'яненість: слабка f	F	$(f - F)$	середня f	F	$(f - F)$	Сума f	F	Відсоток слабо забур'янених полів
A (ч. пар)	54	40	14	21	35	– 14	75	75	72,0
B (буряк)	26	40	– 14	49	35	14	75	75	34,7
Сума	80	80	0	70	70	0	150	150	53,3

Очікувана кількість слабозабур'янених полів після чистого пару становитиме $F_1 = 75 \cdot 53{,}3 / 100 = 40$ шт., середньо забур'янених – $F_2 = 75 \cdot 46{,}7 / 100 = 35$. Після буряків цукрових очікувана кількість слабозабур'янених полів кукурудзи на зерно становитиме $F_3 = 75 \times \times 53{,}3 / 100 = 40$, середньо забур'янених – $F_4 = 75 \cdot 46{,}7 / 100 = = 35$ шт.

Далі визначають різниці між фактичними і теоретично очікуваними частотами $(f - F)$ і критерій χ^2.

$$\chi^2 = \sum(f_1 - F_1)^2 / F_1 + (f_2 - F_2)^2 / F_2 + (f_3 - F_3)^2 / F_3 + (f_4 - F_4)^2 / F_4 =$$
$$= 14^2 /40 + (-14)^2 /35 + (-14)^2 /40 + 14^2 /35 =$$
$$= 4{,}9 + 5{,}6 + 4{,}9 + 5{,}6 = 21{,}0.$$

Теоретичне значення $\chi^2_{05} = 3{,}84$ знаходять за дод. Г для $(c - 1) \times \times (k - 1) = (2 - 1) \cdot (2 - 1) = 1$ ступеня свободи.

Висновок. Оскільки $\chi^2_{факт.} > \chi^2_{05}$, попередники істотно впливали на рівень забур'яненості посівів кукурудзи на зерно (нульова гіпотеза спростовується).

Застосування χ^2-критерію з таблицями цього типу (табл. 29) вимагає, щоб кожне зі спостережень мало не менше п'яти повторень. У разі, якщо теоретичні чисельності невеликі, для розрахунку χ^2 усі різниці $(f - F)$ обчислюють з урахуванням поправки Ійтса, наближуючи їх до цілого числа.

Приклад 5. На 310-ти ділянках (у багатофакторному досліді) визначено врожайність зерна пшениці озимої і попередньо розраховано (у фазу кущіння) забур'яненість посівів (табл. 31). Потрібно з'ясувати, чи залежить урожайність зерна від рівня забур'яненості посівів.

Проведення розрахунків. На основі нульової гіпотези (H_0) розраховуються очікувані частоти для кожного з 10-ти можливих варіантів часток прояву ознаки. Кількість полів, на яких отримали низьку врожайність за низької забур'яненості, має становити $F_1 = 114 \cdot 20{,}6 : 100 = 23{,}5$ шт., за помірної (2 бали) – $F_2 = 71 \cdot 20{,}6 / 100 = 14{,}6$ шт., середньої (3 бали) – $F_3 = 45 \cdot 20{,}6 /100 = 9{,}3$ шт., сильної (4 бали) – $F_4 = 36 \cdot 20{,}6 / 100 = 7{,}5$ шт. і дуже сильної (5 балів) – $F_5 = 44 \cdot 20{,}6 : 100 = 9{,}1$ шт. Кількість полів із середньою врожайністю зерна за низької забур'яненості (1 бал) за теоретичним очікуванням має ста-

Таблиця 31

Теоретично очікувана кількість полів (*F*) пшениці озимої із заданими параметрами врожайності та забур'яненості на базі фактичних даних

Забур'яненість	Урожайність				Сума	Відсоток
	низька		середня			
	f	*F*	*f*	*F*		
Низька	3	23,5	111	90,5	114	36,8
Помірна	6	14,6	65	56,4	71	22,8
Середня	9	9,3	36	35,7	45	14,5
Сильна	15	7,5	21	28,5	36	11,6
Дуже сильна	31	91	13	34,9	44	14,2
Сума	64	64	246	246	310	–
Відсоток	20,6	–	79,4	–	100	100

новити $F_5 = 79{,}4 \cdot 114 / 100 = 90{,}5$ шт., за помірної (2 бали) – $F_7 = 79{,}4 \times \times 71 / 100 = 56{,}4$ шт., середньої (3 бали) – $F_8 = 79{,}4 \cdot 45 / 100 = 35{,}7$ шт., сильної – $F_9 = 79{,}4 \times 36 / 100 = 28{,}5$ шт. і дуже сильної – $F_{10} = 79{,}4 \times \times 44 / 100 = 34{,}9$ шт.

Розрахунки різниці між фактичними та очікуваними частотами $(f - F)$ подають у вигляді окремої таблиці (табл. 32).

Таблиця 32

Різниця між фактичною і теоретичною (очікуваною) кількістю полів з різним рівнем урожайності зерна та забур'яненістю

Рівень урожайності	Забур'яненість					Сума
	низька	помірна	середня	сильна	дуже сильна	
Низький	– 20,5	– 8,6	– 0,3	7,5	21,9	0
Середній	20,5	8,6	0,3	– 7,5	– 21,9	0
Сума	0	0	0	0	0	0

Після цього розраховують фактичне значення χ^2:

$$\chi^2 = (-20{,}5)^2 /23{,}5 + (-8{,}6)^2 /14{,}6 + (-0{,}3)^2 /9{,}3 + 7{,}5^2 / 7{,}5 + + 21{,}9^2 / 9{,}1 + + 20{,}5^2 /90{,}5 + 8{,}6^2 /56{,}4 + 0{,}3^2 /35{,}7 + (-7{,}5)^2 /28{,}5 + + (-21{,}9)^2 /34{,}9 = 105.$$

Теоретичне значення $\chi^2_{05} = 9{,}5$ знаходять за дод. Г для $(c - 1) \times \times (k - 1) = (2 - 1) \cdot (5 - 1) = 4$ ступенів свободи.

Висновок. Доведено істотний вплив рівня забур'яненості посівів на врожайність зерна пшениці озимої, оскільки $\chi^2_{факт.} > \chi^2_{05}$.

Приклад 6. За довжиною верхнього міжвузля 130-ти рослин пшениці твердої ярої маємо такий розподіл:

X (довжина міжвузля) 25 26 27 28 29 30 31 32 33 34;
f (фактична частота) 3 6 11 19 28 33 14 9 5 2.

Потрібно визначити, чи підпорядковується розподіл рослин пшениці ярої за довжиною верхнього міжвузля нормальному розподілу.

Проведення розрахунків. Розрахунки краще проводити за формою табл. 33. За даними фактичного розподілу частот f необхідно визначити теоретично очікувані частоти F, що відповідають нормальному розподілу, а потім визначити ступінь відповідності фактичних і очікуваних частот за χ^2-критерієм. На першому етапі потрібно знайти середньоарифметичний показник згрупованих даних $\bar{x}$ і стандартне відхилення s:

Таблиця 33

Обчислення теоретичних частот (*F*) і критерію відповідності фактичного розподілу рослин пшениці ярої нормальному (χ^2)

X	f	$t = \frac{X - \bar{x}}{s}$	$\Phi(t)$	F	$f - F$	$(f - F)^2$	$(f - F)^2 / F$
25	3 } 9	2,34	0,0258	1,78 } 7,14	–	–	–
26	6	1,81	0,0775	5,36	1,86	3,46	0,48
27	11	1,28	0,1758	12,10	– 1,10	1,21	0,10
28	19	0,74	0,3034	20,96	– 1,96	3,84	0,18
29	28	0,21	0,3902	26,96	1,04	1,08	0,04
30	33	0,32	0,3790	26,19	6,81	46,38	1,77
31	14	0,85	0,2780	19,20	– 5,2	27,04	1,41
32	9	1,38	0,1540	10,64	– 1,64	2,69	0,25
33	5 } 7	1,91	0,0644	4,45 } 5,82	1,18	1,39	0,24
34	2	2,45	0,0198	1,37	–	–	–
Сума	–	–	–	≈ 30	–	–	$\chi^2 = 4{,}47$

$$\bar{x} = \sum f \cdot X / n = (f_i \cdot X_1 + f_2 \cdot X_2 + \ldots + f_l \cdot X_l) / n =$$
$$= ((25 \cdot 3) + (26 \cdot 6) + \ldots + (34 \cdot 2)) / 130 = 29{,}4;$$
$$s = \sqrt{\sum f \cdot (X - \bar{x})^2 / (n - 1)} =$$
$$= \sqrt{\frac{3 \cdot (25{,}0 - 29{,}4)^2 + 6 \cdot (26{,}0 - 29{,}4)^2 + \ldots + 2 \cdot (34{,}0 - 29{,}4)^2}{(130 - 1)}} =$$
$$= 1{,}88.$$

Далі розраховують нормоване відхилення в частках стандартного відхилення: $t = (X - \bar{x}) / s$.

За показником t (дод. М) знаходять $\Phi(t)$ – імовірність повторюваності цього значення x у нормально розподіленій сукупності.

Наступним кроком є розрахунок теоретичного ряду частот F, що відповідає цьому обсягу вибірки n, $\bar{x}$ і s при певному інтервалі i за формулою:

$$F = \Phi(t) \cdot n \cdot i / s.$$

У цьому прикладі значення $n \cdot i / s$ становить: $130 \cdot 1 / 1{,}88 = 69{,}1$.

У табл. 33 граничні показники об’єднують за принципом, відповідно до якого значення F має бути більшим від п’яти, розраховують різниці $(f - F)$, знаходять квадрати одержаних показників. Далі визначають теоретичні частоти $(f - F)^2 / F$, одержані дані додають, визначаючи χ^2. Теоретичне значення $\chi^2_{05} = 11{,}07$ знаходять у таблиці, наведеній у дод. Г, для $(k - 3) = (8 - 3) = 5$ ступенів свободи варіації.

Висновок. Розподіл рослин пшениці твердої ярої за довжиною верхнього міжвузля не відрізняється від нормального, оскільки $\chi^2_{факт.} < \chi^2_{05}$ (нульова гіпотеза не спростовується). Цей висновок наочно демонструє рис. 23.

На практиці доволі часто розподіл часток прояву досліджуваного показника відрізняється від нормального, зокрема крива лінія розподілу може мати декілька верхівок. Розглянемо приклад фактичного розподілу частот, що істотно відрізняється від нормального (нульова гіпотеза спростовується).

Приклад 7. Установлено розподіл 200 рослин кукурудзи за кількістю листків на одній рослині:

x (кількість листків) 18 19 20 21 22 23 24 25 26 27 28 29;

f (фактична частота) 6 11 15 31 20 14 11 28 34 13 10 7.

Слід визначити, чи підпорядковується розподіл частот рослин кукурудзи нормальному розподілу за кількістю листків на одній рослині.

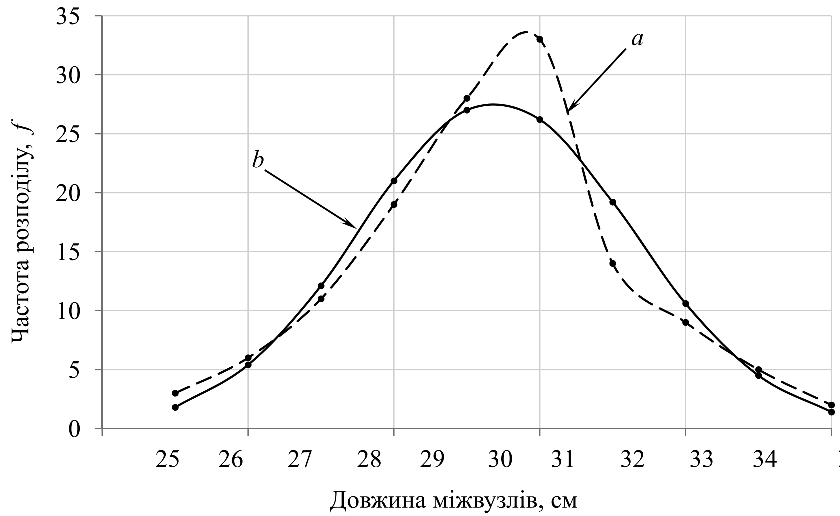


Рис. 23. Фактичний (*a*) і теоретичний (*b*) розподіл рослин пшениці твердої ярої за довжиною верхнього міжвузля

Проведення розрахунків. Розрахунки проводять за тим самим принципом, що і в попередньому прикладі. Спочатку знаходять середньоарифметичний показник згрупованих даних ($\bar{x}$) та стандартне відхилення s:

$$\bar{x} = \sum f \cdot x / n = [(6 \cdot 8) + (11 \cdot 19) + ... + (7 \cdot 29)] / 200 = 23{,}6;$$

$$s = \sqrt{\sum f(x-\bar{x})^2 / (n-1)} =$$
$$= \sqrt{[6 \cdot (18-23{,}6)^2 + 11 \cdot (19-23{,}6)^2 + \cdots + (29-23{,}6)^2] / 199} =$$
$$= 2{,}76.$$

Далі шукають нормоване відхилення в частках стандартного відхилення $t = (x - \bar{x}) / s$. Результати аналізу заносять до табл. 34.

За показником t (дод. М) знаходять імовірність повторюваності цього показника x у нормально розподіленій сукупності. Далі розраховують теоретичний ряд частот F, який відповідає цьому обсягу вибірки за обраного інтервалу – i:

$$F = \Phi(t) \cdot n \cdot i / s.$$

У цьому прикладі коефіцієнт $n \cdot i / s$ становить $200 \cdot i / 2{,}76 =$ $= 2{,}75$. Далі визначають різниці $(f - F)$ і квадрати $(f - F)^2$. До останньої колонки табл. 69 заносять показники $(f - F)^2 / F$ і знаходять їхню суму χ^2.

Таблиця 34

Визначення теоретичних частот (*F*) і критерію відповідності фактичного розподілу рослин кукурудзи типовому (нормальному) χ^2

x	f	$t = (x - \bar{x}) / s$	$\Phi(t)$	F	$f - F$	$(f - F)^2$	$(f - F)^2 / F$
18	6	2,03	0,0508	3,68	2,32	5,38	1,46
19	11	1,67	0,0989	7,17	3,83	14,67	2,05
20	15	1,30	0,1714	12,43	2,57	6,60	0,53
21	31	0,94	0,2505	18,16	12,84	164,87	9,08
22	20	0,58	0,3372	24,45	– 4,45	19,80	0,81
23	14	0,22	0,3894	28,23	–14,23	202,49	7,17
24	11	0,14	0,3951	28,64	–17,64	311,17	10,86
25	28	0,51	0,3503	25,40	2,60	6,76	0,27
26	34	0,87	0,2732	19,81	14,19	201,36	10,16
27	13	1,23	0,1872	13,57	–0,57	0,32	0,02
28	10	1,59	0,1127	8,17	1,83	3,35	0,41
29	7	1,96	0,0584	4,23	2,77	7,67	1,81
Сума	200	–	–	≈ 200	–	–	$\chi^2 = 44{,}63$

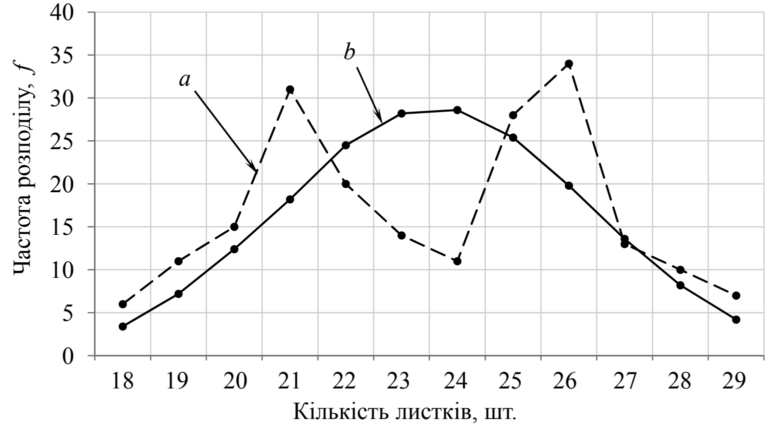


Рис. 24. Фактичний (*a*) і теоретичний (*b*) розподіл рослин кукурудзи за кількістю листків на одній рослині

У цьому прикладі $\chi^2_{05} = 16{,}92$, що відповідає $(k - 3) = 9$ ступеням свободи варіації (дод. Г).

Висновок. Оскільки $\chi^2_{факт.} > \chi^2_{05}$, нульова гіпотеза ($H_0$) спростовується і розподіл досліджуваної ознаки (кількості листків на одній рослині кукурудзи) істотно відрізняється від нормального (типового). Цей висновок добре демонструє рис. 24.

5. Порівняння двох середніх за *t*-критерієм

Порівняння двох вибіркових середніх або часток проводять шляхом перевірки нульової гіпотези, відповідно до якої між вибірковими середніми (або частками) відсутні істотні відмінності.

Перевірку нульової гіпотези здійснюють за допомогою статистичного *t*-критерію. Між двома середніми значеннями ($\bar{x}_1 - \bar{x}_2$) або частками ($p_1 - p_2$) є істотна різниця за умови, якщо $F_{факт.} > F_{теор}$. У разі, якщо $F_{факт.} \le F_{теор.}$, різниця між вибірковими середніми або частками неістотна (нульова гіпотеза H_0 не спростовується).

5.1. Оцінка істотності різниці середніх і середньої різниці за t-критерієм

У непов'язаних вибірках (досліджувані показники двох порівнюваних вибірок незалежні) за t-критерієм оцінюють істотність різниці середніх $d = \bar{x}_1 - \bar{x}_2$. У пов'язаних вибірках (одиниці обліку першої вибірки пов'язані з одиницями спостережень другої) за t-критерієм оцінюють істотність середньої різниці $d = (\sum d) / n$. Розрахунки статистичних характеристик під час оцінювання істотності різниці проводять за формулами, наведеними в табл. 35.

Якщо досліджувана ознака розподіляється за правилом рідкісних подій Пуассона (вибіркові середні і дисперсії однакові, тобто $\bar{x} = s^2$), критерій істотності середніх визначають за формулою:

$$t = x_1 - x_2 / \sqrt{x_1 + x_2},$$

Таблиця 35

Формули для визначення істотної різниці між середніми двох вибірок при кількісній мінливості

Показник	Непов'язані вибірки	Пов'язані вибірки
Середня арифметична	$\bar{x} = \sum x / n$	$\bar{x} = \sum x / n$
Різниця між середніми і середня різниця	$d = \bar{x}_1 - \bar{x}_2$	$\bar{d} = \sum d / n$
Помилка середньої арифметичної	$s_{\bar{x}} = s / \sqrt{n} = \sqrt{\frac{\sum(x-\bar{x})^2}{n \cdot (n-1)}} = \sqrt{\frac{\sum x^2 - (\sum x)^2/n}{n \cdot (n-1)}}$	Не розраховується
Помилка різниці середніх та середньої різниці	$s_d = \sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}}$	$s_d = \sqrt{\frac{\sum(d-\bar{d})^2}{n \cdot (n-1)}}$
Довірчий інтервал для генеральної різниці	$d \pm t_{теор} \cdot s_d$	$d \pm t_{теор} \cdot s_d$
Критерій істотності	$t_{факт.} = d / s_d$	$t_{факт.} = \bar{d} / s_{\bar{d}}$
Ступені свободи	$n_1 + n_2 - 2$	$n - 1$

Примітка: x – значення ознаки; d – різниці між пов'язаними параметрами; n – обсяг вибірки; $t_{теор.}$ – теоретичне значення t-критерію.

де $\bar{x}_1$ і $\bar{x}_2$ – кількість рідкісних подій у порівнюваних великих вибірках, для яких теоретичні значення t_{05} і t_{01} становлять відповідно 1,96 і 2,58.

Далі на прикладах 1 і 2 показано техніку проведення розрахунків для незалежних вибірок. У прикладах 3–5 розкрито техніку проведення розрахунків із пов'язаними вибірками, а в прикладі 6 – техніку розрахунків різниці середніх для рідкісних подій.

Приклад 1. На семи облікових площах кожної з двох дослідних ділянок, засіяних різними сортами вівса, визначено середню висоту рослин (табл. 36). Потрібно обчислити 5 %-ві довірчі інтервали для середньої висоти рослин і перевірити істотність різниці між досліджуваними сортами за показниками висоти рослин.

Таблиця 36

Статистичні розрахунки показників висоти рослин двох сортів вівса

Сорт 1 (перша вибірка)			Сорт 2 (друга вибірка)		
x	$x_1 = x - \bar{x}_1$	x_1^2	x	$x_1 = x - \bar{x}_2$	x_1^2
93	– 4	16	74	– 5	25
107	10	100	82	3	9
96	– 1	1	88	9	81
90	– 7	49	71	– 8	64
98	1	1	74	– 5	25
102	5	25	80	1	1
91	– 6	36	85	6	36
$\sum x = 677$ $\bar{x}_1 = 96{,}71$	$\sum x_1 = -2$	$\sum x_1^2 = 228$	$\sum x = 554$ $\bar{x}_2 = 96{,}71$	$\sum x_1 = 1$	$\sum x_1^2 = 241$
Сума квадратів $\sum(x - \bar{x})^2 = \sum x_1^2 - (\sum x_1)^2 / n$					
$228 - (2)^2 / 7 = 227{,}4$			$241 - (1)^2 / 7 = 240{,}9$		
Абсолютна помилка середньої $s_{\bar{x}} = \sqrt{\sum(x-\bar{x})^2 \cdot (n-1) / n \cdot (n-1)}$					
$s_{\bar{x}_1} = \sqrt{227{,}4 / (7-1) \cdot 7} = 2{,}33$			$s_{\bar{x}_2} = \sqrt{240{,}9 / (7-1) \cdot 7} = 2{,}39$		
Відносна помилка середньої $s_{\bar{x}}\% = s_{\bar{x}} \cdot 100 / \bar{x}$					
$s_{\bar{x}_1}\% = 2{,}33 \cdot 100 / 96{,}71 = 2{,}4$ %			$s_{\bar{x}_2}\% = 2{,}39 \cdot 100 / 79{,}14 = 3{,}0$ %		
Довірчий інтервал для генеральної середньої $\bar{x} \pm t_{05} \cdot s_{\bar{x}}$					
$96{,}71 \pm 2{,}45 \times 2{,}33 = 96{,}71 \pm 5{,}71$			$79{,}14 \pm 2{,}45 \times 2{,}39 = 79{,}14 \pm 5{,}86$		
Діапазон довірчого інтервалу					
від 91 до 102			від 73 до 85		

Проведення розрахунків. Оскільки пробні ділянки для визначення висоти рослин двох різних сортів вівса не пов'язані між собою, отримані показники обробляють за принципом непов'язаних вибірок.

Після визначення сумарних показників висоти рослин кожного сорту ($\sum x$ і $\sum x_1$), середньої висоти рослин кожного сорту ($\bar{x}_1, \bar{x}_2$), суми різниць ($\sum x_1$) кожного сорту та суми квадратів відхилень між варіантною з їхнім середнім показником ($\sum x_1^2$) для кожного сорту розраховують суму квадратів відхилень $(\sum x - \bar{x})^2$; абсолютну помилку середньої ($s_{\bar{x}}$ %), довірчий інтервал для генеральної середньої і $t_{\text{факт.}}$

Теоретичне значення $t_{05} = 2{,}45$ знаходять за дод. Б для $(n - 1) = 7$ ступенів свободи.

Критерій істотності $t_{\text{факт.}} = (\bar{x}_1 - \bar{x}_2) / \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2} = (96{,}71 - 79{,}14) : \sqrt{2{,}33^2 + 2{,}39^2} = 17{,}57 / \sqrt{5{,}43 + 5{,}71} = 5{,}26$. У цьому прикладі $t_{05} = 2{,}18$ (для $n_1 + n_2 - 2 = 12$ ступенів свободи) (див. дод. Б).

Висновок. Оскільки $t_{\text{факт.}} > t_{\text{теор.05}}$, нульова гіпотеза ($H_0$) спростовується і різниця між двома досліджуваними сортами вівса за висотою рослин істотна з імовірністю 95 %.

Приклад 2. У досліді з вивчення впливу двох попередників на забур'яненість посівів соняшнику одержано такий розподіл кількості бур'янів, шт./м^2 (табл. 37). Потрібно визначити, чи є істотна різниця між ефектом попередників на забур'яненість посівів соняшнику з рівнем імовірності 95 %.

Таблиця 37

Частота випадків забур'яненості посівів соняшнику з відзначеними показниками їхньої кількості на 1 м^2, %

x		5	10	15	20	25	30	35	40
f, %	Варіант *А*	3	9	11	20	33	15	7	2
	Варіант *В*	15	21	30	13	14	5	2	–

Розрахунки для згрупованих незалежних вибірок доцільно проводити за зразком табл. 38. Теоретичне значення $t_{05} = 1{,}98$ знаходять за дод. Б для $(n - 1) = 100 - 1 = 99$ ступенів свободи.

Критерій істотності $t_{факт.} = d / s_d = x_1 - x_2 / \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2} = 22{,}7 - 15{,}7 : \sqrt{0{,}78^2 + 0{,}77^2} = 7{,}0 / \sqrt{0{,}61 + 0{,}59} = 6{,}4$.

Значення $t_{05} = 1{,}97$ беруть з таблиці, наведеної в дод. Б, для $n_1 + n_2 - 2 = 100 + 100 - 2 = 198$ ступенів свободи.

Таблиця 38

Послідовність проведення статистичного аналізу згрупованих показників

x	Попередник А (перша вибірка)					Попередник В (друга вибірка)				
	f	$x_1 = (x - 25)/10$	fx_1	x_1^2	fx_1^2	f	$x_1 = (x - 15)/10$	fx_1	x_1^2	fx_1^2
5	3	–2,0	– 6,0	4,0	12,0	15	–1,0	– 15,0	1,0	15,0
10	9	–1,5	–13,5	2,3	20,7	21	–0,5	– 10,5	0,3	6,3
15	11	–1,0	–11,0	1,0	11,0	30	0	0	0	0
20	20	–0,5	–10,0	0,3	6,0	13	0,5	6,5	0,3	3,9
25	33	0	0	0	0	14	1,0	14,0	1,0	14,0
30	15	0,5	7,5	0,3	4,5	5	1,5	7,5	2,3	11,5
35	7	1,0	7,0	1,0	7,0	2	2,0	4,0	4,0	8,0
40	2	1,5	3,0	2,3	4,6	–	–	–	–	–
Сума	100	–	– 23,0	–	65,8	100	–	6,5	–	58,7
Середнє $\bar{x}_1 = 25 + (-23 / 100) \cdot 10 = 22{,}7$						$\bar{x}_2 = 15 + 6{,}5 / 100 \cdot 10 = 15{,}7$				
Сума квадратів $[\sum fx_1^2 - (\sum fx_1 / n)] \cdot 10^2$										
$[65{,}8 - (-23)^2 / 100] \cdot 10^2 = 6051$						$[58{,}7 - (6{,}5)^2 / 100] \cdot 10^2 = 5830$				
Абсолютна помилка середньої $s_{\bar{x}} = \sqrt{\sum(x - \bar{x})^2 \cdot (n-1) / n \cdot (n-1)}$										
$s_{\bar{x}_1} = \sqrt{6051 / 100 \cdot (100 - 1)} = 0{,}78$						$s_{\bar{x}_2} = \sqrt{5830/100 \cdot (100 - 1)} = 0{,}77$				
Відносна помилка середньої $s_{\bar{x}}\% = s_{\bar{x}} \cdot 100 / \bar{x}$										
$s_{\bar{x}_1}\% = 0{,}78 \cdot 100 / 22{,}7 = 3{,}4$ %						$s_{\bar{x}_2}\% = 0{,}77 \cdot 100 / 15{,}7 = 4{,}9$ %				
Довірчий інтервал для генеральної середньої $\bar{x} \pm t_{05} \cdot s_{\bar{x}}$										
$22{,}7 \pm 1{,}98 \cdot 0{,}78 = 32{,}3 \pm 1{,}54$						$15{,}7 \pm 1{,}98 \cdot 0{,}77 = 15{,}7 \pm 1{,}52$				
Діапазон довірчого інтервалу										
від 30,8 до 33,8						від 14,2 до 17,2				

Висновок. У цьому досліді $t_{факт.} > t_{05}$, отже, нульова гіпотеза (H_0) спростовується і між варіантами є істотна різниця з імовірністю 99 %.

Приклад 3. У досліді з вивчення динаміки росту другого листка рослин пшениці м'якої ярої одержано такі показники площі листків (табл. 39). Обліки проводили у фазу початку колосіння (варіант А) та у фазу молочної стиглості (варіант В). Вибірка – 10 рослин.

Проведення розрахунків. За формулами, наведеними у табл. 70 для пов'язаних вибірок, розраховуємо середню різницю ($\bar{d}$), помилку середньої різниці ($s_{\bar{d}}$) та довірчий інтервал для генеральної різниці.

Таблиця 39

Площа другого листка рослин пшениці м'якої ярої у фазу колосіння та молочної стиглості

Номер спостереження (рослини)	Площа другого листка		Різниця, $d = x_1 - x_2$	Квадрат різниці, d^2
	перший строк обліку, x_1	другий строк обліку, x_2		
1	16	23	–7	49
2	19	28	–9	81
3	25	29	–4	16
4	18	24	–6	36
5	21	26	–5	25
6	15	26	–11	121
7	19	32	–13	169
8	17	25	–8	64
9	16	23	–7	49
10	20	28	–8	64
Сума	186	264	–78	674
Середнє	18,6	26,4	–7,8	–

$$s_d = \sqrt{\frac{\sum d^2 - (\sum d)^2/n}{n \cdot (n-1)}} = \sqrt{\frac{674 - (78)^2/10}{10 \cdot (10-1)}} = \sqrt{\frac{(674{,}0 - 608{,}4)}{90}} =$$

$$= 90 = 0{,}85 \text{ см}^2.$$

$$t_{факт.} = d / s_d = 7{,}8 / 0{,}85 = 9{,}17.$$

Теоретичне значення t_{05} для $(n - 1) = (10 - 1) = 9$ ступенів свободи становить 2,23 (див. дод. Б). Довірчий інтервал для генеральної різниці $\bar{d} \pm t_{05} \times s_{\bar{d}} = 7{,}8 \pm 2{,}23 \times 0{,}85 = 7{,}8 \pm 1{,}9$. Діапазон довірчого інтервалу – від 5,9 до 9,7.

Висновок. У цьому прикладі $t_{факт.} > t_{05}$. Таким чином, нульова гіпотеза спростовується, що дає підставу стверджувати, що за більш пізніх обліків (у фазу молочної стиглості) площа другого листка буде істотно більшою, ніж у перший строк обліку (фазу початку колосіння).

Приклад 4. У 20-ти зразках зерна пшениці твердої ярої визначали масу 1000 насінин різними способами (перший – за методикою Н. А. Майсуряна (1970 р.), другий – на підставі ГОСТ 10842-89). Отримані показники наведено в табл. 40.

Таблиця 40

Маса 1000 насінин пшениці твердої ярої за різними методиками, г

Номер зразка зерна	Маса 1000 насінин		Різниця d	Номер зразка зерна	Маса 1000 насінин		Різниця d
	I	II			I	II	
1	38,2	39,6	–1,4	11	37,4	38,8	–1,4
2	39,0	37,4	1,6	12	38,3	40,0	–1,7
3	40,5	41,2	–0,7	13	40,1	39,6	0,5
4	39,8	40,7	–0,9	14	41,6	39,2	2,4
5	38,2	40,2	–2,0	15	37,0	38,3	–1,3
6	37,4	39,1	–1,7	16	39,9	40,6	–0,7
7	34,2	27,2	–3,0	17	41,2	39,4	1,8
8	39,1	28,6	0,5	18	39,8	37,7	2,1
9	38,8	39,1	–0,3	19	37,0	38,6	–1,6
10	40,4	38,8	1,6	20	40,5	41,4	–0,9

Розраховуємо середню різницю ($\bar{d}$), помилку середньої різниці ($s_{\bar{d}}$) та фактичне значення критерію F.

$$\sum d = d_1 + d_2 + ... + d_l = (-1{,}4) + 1{,}6 + ... + (-0{,}9) = 10{,}5 - 17{,}6 = 7{,}1.$$

$$\bar{d} = \sum d / n = 7{,}1 / 20 = 0{,}36;$$

$$\sum d^2 = (1{,}4)^2 + (1{,}6)^2 + ... + (-0{,}9)^2 = 48{,}67;$$

$$(\sum d^2) = (-7{,}1)^2 = 50{,}41;$$

$$s_d = \sqrt{\frac{\sum d^2 - (\sum d)^2/n}{n \cdot (n-1)}} = \sqrt{\frac{48{,}67 - 50{,}41 / 20}{20 \cdot (20-1)}} = \sqrt{0{,}12} = 0{,}35;$$

$$t_{факт.} = \bar{d} / s_d = 0{,}36 / 0{,}35 = 1{,}03 \text{ г.}$$

Теоретичне значення $t_{05} = 2{,}09$, таким чином $t_{факт}. < t_{05}$.

Висновок. Оскільки $t_{факт.} < t_{теор.}$ з імовірністю 95 %, між показниками маси 1000 насінин істотна різниця відсутня. Таким чином, обидві методики у цілому забезпечували аналогічну оцінку маси 1000 насінин пшениці м'якої ярої.

Приклад 5. У масовому географічному досліді в 50-ти господарствах досліджували вплив двох способів сівби на врожайність ячменю ярого. На кожному полі досліджувані варіанти розміщували один біля одного. Частоту розподілу різниць урожайності ячменю ярого вказано в табл. 41. Потрібно визначити, чи є істотна різниця між досліджуваними способами сівби за показником врожайності зерна.

Проведення розрахунків. Статистичний аналіз згрупованих різниць пов'язаних вибірок проводять у тій же послідовності, що і для незгрупованих вибірок, але з урахуванням частки прояву. Знаходять добуток показника різниці та частки їхнього прояву (fd), квадрат показників різниць (d^2), добуток квадратів різниці та частки їхнього прояву (fd^2). Далі визначають суми та середні показники стовбців (граф). Результати заносять до табл. 41.

На наступному етапі розраховують помилку середньої різниці ($s\bar{d}$) та критерій $t_{факт.}$:

$$s_d = \sqrt{\frac{\sum d^2 - (\sum d)^2/n}{n \cdot (n-1)}} = \sqrt{\frac{1800 - 324^2/\,50}{50 \cdot (50-1)}} = \sqrt{0{,}122} = 0{,}35 \text{ ц/га};$$

$$t_{факт.} = d\ /\ s_d = 6{,}5\ /\ 0{,}35 = 18{,}6.$$

При кількості ступенів свободи $(n - 1) = (50 - 1) = 49$ теоретичне значення $t_{05} = 2{,}01$ (див. дод. Б). Довірчий інтервал для генеральної різниці $d \pm t_{05} \cdot s\bar{d} = 6{,}5 \pm 0{,}35 \cdot 2{,}01 = 6{,}5 \pm 0{,}70$. Діапазон довірчого інтервалу – від 5,8 до 7,2 ц/га.

Таблиця 41

Статистичний аналіз показників урожайності зерна ячменю ярого за згрупованими різницями

Різниця $d = x_1 - x_2$, ц/га	Частота прояву, f	fd	d^2	fd^2
–10	1	– 50	100	100
– 9	2	– 18	81	162
– 8	4	– 32	64	256
–7	10	– 70	49	490
– 6	13	– 78	36	468
– 5	8	– 40	25	200
– 4	6	– 24	16	96
– 3	2	– 6	9	18
– 2	2	– 4	4	8
– 1	2	– 2	1	2
Сума	50	– 324	–	1800
Середнє	–	$\bar{d} = 6{,}5$	–	–

Висновок. У цьому прикладі $t_{факт.} > t_{05}$. Таким чином, нульова гіпотеза (H_0) спростовується і між порівнюваними пов'язаними вибірками є істотна різниця.

Приклад 6. У 200-х зразках пшениці озимої (100 зразків одного сорту і 100 – другого) визначили кількість пошкоджених жуком-кузькою зернівок (кожний зразок – по 100 зерен). У 100 зразках першого сорту знайшли 96 ушкоджених зернівок, а у 100 зразках іншого сорту – 65 шт. Визначити, чи є істотна різниця між порівнюваними сортами пшениці озимої за показником пошкоджених жуком-кузькою зерен.

Проведення розрахунків. Оскільки в цьому прикладі ми маємо справу з рідкісними подіями, що підпорядковуються розподілу Пуассона, оцінку значущості різниці вибіркових середніх розраховують за формулою: $t_{факт.} = (x_1 - x_2)\ /\ \sqrt{\bar{x}_1 + \bar{x}_2}$. Тут кількість рідкісних подій у порівнюваних сукупностях x_1 і x_2 відповідно становить 96 і 95 штук, отже, фактичний критерій t становитиме: $t_{факт.} = (96 - 65) : : \sqrt{96 + 65} = 31 / 12{,}7 = 2{,}44$.

Теоретичне значення $t_{05} = 1{,}97$ і $t_{01} = 2{,}60$ знаходять за дод. Б для $n_1 + n_2 - 2 = 100 + 100 - 2 = 198$ ступенів свободи.

Висновок. Оскільки $t_{факт.} > t_{05}$, між показниками ушкоджених зерен є істотна різниця з ймовірністю 95 %. За більш строгої оцінки (рівень імовірності – 99 %) різниця за кількістю ушкоджених зерен між двома сортами пшениці озимої відсутня.

5.2. Оцінка істотності різниці між вибірковими частками (альтернативна мінливість)

Порівняння вибіркових часток за альтернативної мінливості доцільно проводити з використанням методу χ^2. Якщо додатково потрібно визначити ще й довірчий інтервал різниці часток, застосовують параметричний критерій. При цьому істотну різницю за t-критерієм визначають за формулою:

$$t_{факт.} = d\ /\ s_d = (p_1 - p_2)\ /\ \sqrt{sp_1^2 + sp_2^2} = (p_1 - p_2)\ /\ \sqrt{\frac{p_1 \cdot p_1}{n_1} + \frac{p_2 \cdot q_2}{n_1}},$$

де p і q – вибіркові частки; s_{p1} і s_{p2} – помилки часток; n_1 і n_2 – обсяг порівнюваних вибірок.

Довірчий інтервал для різниці часток знаходять за співвідношенням $d \pm t \cdot s_d = (p_1 - p_2) \pm t \cdot s_d$.

Розглянемо на конкретному прикладі техніку статистичної оцінки між вибірковими частками за альтернативної мінливості.

Приклад 7. Серед 1000 проаналізованих рослин пшениці м'якої ярої і такої ж кількості рослин тритикале ярого ураженість летючою сажкою становила відповідно 17 і 3 % (170 і 30 шт.). Потрібно розрахувати критерій $t_{факт.}$, довірчий інтервал для різниці часток та визначити, чи є достовірна різниця між пшеницею твердою ярою та тритикале ярим за стійкістю до ушкодження летючою сажкою.

Проведення розрахунків. Визначаємо частку здорових (q) і ушкоджених (p) рослин кожної культури. Для пшениці твердої ярої:

$$p_1 = n_1 / N_1 = 170 / 1000 = 0{,}17\ (17\ \%);$$
$$q_1 = 1 - p_1 = 1 - 0{,}17 = 0{,}83\ (83\ \%);$$

для тритикале ярого:

$$p_2 = n_2 / N_2 = 30 / 1000 = 0{,}03\ (3\ \%);$$
$$q_2 = 1 - p_2 = 1 - 0{,}03 = 0{,}97\ (97\ \%).$$

Далі розраховуємо помилку вибіркової частки кожної вибірки:

$$sp1 = \sqrt{p_1 \cdot q_1 / N_1} = \sqrt{0{,}17 \cdot 0{,}83 / 1000} = 0{,}012, \text{ або } 1{,}2\ \%;$$
$$sp_2 = \sqrt{p_2 \cdot q_2 / N_2} = \sqrt{0{,}03 \cdot 0{,}97 / 1000} = 0{,}017, \text{ або } 1{,}7\ \%.$$

На наступному етапі визначають помилку різниці часток:

$$s_d = \sqrt{s \cdot p_1^2 + s \cdot p_2^2} = \sqrt{0{,}012^2 + 0{,}017^2} = 0{,}021, \text{ або } 2{,}1\ \%.$$

Довірчий інтервал різниці часток з імовірністю 95 % становить: $(p_1 - p_2) \pm t_{05} \cdot s_d = (0{,}17 - 0{,}03) \pm 1{,}96 \cdot 0{,}021 = 0{,}14 \pm 0{,}04$. Діапазон варіабельності довірчого інтервалу становить від 0,10 до 0,18. Теоретичне значення $t_{05} = 1{,}96$ знаходять за дод. Б для $n_1 + n_2 - 2 = 1000 + 1000 - 2 = 1998$ ступенів свободи.

$$t_{факт.} = (p_1 - p_2) / s_d = 0{,}17 - 0{,}03 / 0{,}021 = 6{,}67.$$

Висновок. Оскільки $t_{факт.} > t_{05}$, нульова гіпотеза (H_0) спростовується і між рослинами пшениці твердої ярої та тритикале ярого є істотна різниця за стійкістю до ушкодження летючою сажкою.

6. Дисперсійний аналіз даних вегетаційного досліду

Вегетаційні досліди зазвичай складаються з кількох незалежних варіантів. Як правило, у вегетаційних дослідах варіанти територіально не розмежовують в окремі повторення. У цих випадках статистичні розрахунки проводять методом дисперсійного аналізу як для непов'язаних вибірок. Якщо досліджувані варіанти об'єднують у територіально обмежовані повторення, то обробку статистичних даних вегетаційного досліду проводять так само, як у польовому досліді, закладеному методом рендомізованих повторень.

Дисперсійний аналіз результатів вегетаційного досліду передбачає перевірку нульової гіпотези (H_0), тобто визначення достовірності різниці між досліджуваними варіантами. Нульова гіпотеза записується як $H_0 / d = 0$.

Залежно від кількості досліджуваних чинників, виділяють однофакторні та багатофакторні вегетаційні досліди. На конкретних прикладах розглянемо техніку проведення дисперсійного аналізу вегетаційного досліду з різною кількістю досліджуваних чинників.

6.1. Однофакторний дослід

В однофакторному вегетаційному досліді, поставленому без територіального обмеження повторень, загальне варіювання досліджуваного показника поділяється на мінливість варіантів і помилок: $C_Y = C_V + C_Z$. Статистичну обробку результатів проводять у кілька етапів:

1. У першу чергу складають розрахункову таблицю, до якої вносять загальну суму показників, суму показників за варіантами, середнє значення досліджуваного показника за кожним варіантом та в цілому по досліду (табл. 42).
2. За формулами, наведеними в табл. 43, розраховують суму квадратів відхилень: загальну – C_Y, варіантів – C_V, помилок (залишку) – C_Z, а потім визначають фактичне значення критерію F.

Таблиця 42

План розміщення результатів вегетаційного досліду у розрахунковій таблиці

Варіант	Вихідні дані	Кількість спостережень	Суми за варіантами	Середнє за варіантами
1	$x_1(\mathrm{I}), x_1(\mathrm{II}),..., x_1(n)$	n_1	v_1	$\bar{x}_1$
2	$x_2(\mathrm{I}), x_2(\mathrm{II}),..., x_2(n)$	n_2	v_2	$\bar{x}_2$
–	–׀׀–׀׀–	–	–	–
–	–׀׀–׀׀–	–	–	–
$l-1$	$x_{l-1}(\mathrm{I}), x_{l-1}(\mathrm{II}),..., x_{l-1}(n)$	n_{l-1}	v_{l-1}	$\bar{x}_{l-1}$
l	$x_l(\mathrm{I}), x_l(\mathrm{II}),..., x_l(n)$	nl	v_l	$\bar{x}_1$
Загальна сума		$N = \sum n$	$\sum x = \sum v$	$\bar{x}$

Таблиця 43

Розрахункові формули для визначення сум квадратів відхилень, середнього квадрата дисперсій і критерію *F*

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{\text{факт.}}$	$F_{\text{теор.}}$	
					05	01
Загальна, C_Y	$\sum x^2 - C$	$N-1$	–	–	Знаходять за таблицею в додатку Б	
Варіантів, C_V	$\sum x_v^2/n - C$	$l-1$	sv^2	sv^2 / s^2		
Помилок, C_Z	$C_y - C_v$	$N-1$	s^2	–		

3. З'ясовують помилку досліду $s_{\bar{x}}$, помилку різниці середніх (s_d) та найменшу істотну різницю для порівняння середніх показників за варіантами.

Розглянемо конкретні приклади і техніку проведення статистичних розрахунків вегетаційного досліду з однаковою та різною кількістю повторень (відповідно приклади 1 і 2).

Корегуючий фактор *C* розраховують за формулою:

$$C = (\sum x)^2 / N \text{ або } C = \bar{x}_{\text{ср.}} \cdot \sum x.$$

Приклад 1. Провести обробку даних урожайності зерна тритикале ярого залежно від варіантів системи живлення у вегетаційному досліді з однаковою кількістю спостережень (повторень) і внести результати до табл. 44.

Проведення розрахунків. У таблиці результатів досліджень підраховують суми і середні показники за варіантами досліду. Далі

Таблиця 44

Урожай зерна тритикале ярого* (г/посудину)

Варіант досліду	Урожайність за повтореннями					Сума *V*	Середнє за *V*
1	61	98	63	63	67	322	64,4
2 (контроль)	64	58	67	61	64	314	62,8
3	57	62	55	59	60	293	58,6
4	70	74	76	71	78	369	73,8
5	89	86	91	83	90	439	87,8
6	53	51	51	57	54	266	53,2
7	87	93	93	89	95	457	91,4
Загальна сума						$\sum x = 2460$	$\bar{x} = 70,3$

* Кількість спостережень (повторень) кожного варіанта – 5.

визначають загальну кількість спостережень *N*, корегуючий чинник *C*, суму квадратів відхилень загальну C_y, варіантів C_v і помилок C_z.

$$N = \sum n = \sum x = 3;\ C \text{ (поправка)} = (\sum x)^2 / N = 2460^2 / 35 = 172902{,}9;$$
$$C_Y \text{ (загальна)} = \sum x^2 - C = (61^2 + 68^2 + ... + 95^2) - 172902{,}9 = 6687{,}1$$
$$\text{при } (N - 1) = 35 - 1 = 34 \text{ ступенях свободи;}$$
$$C_V \text{ (варіантів)} = \sum x_V^2 / n - C = (322^2 + 314^2 + ... + 457^2) / 5 - 172902{,}9 =$$
$$= 6420{,}3 \text{ при } (l - 1) = (7 - 1) = 6 \text{ ступенях свободи;}$$
$$C_Z = C_Y - C_V = 6687{,}1 - 6420{,}3 = 266{,}8$$
$$\text{при } (N - 1) - (l - 1) = 34 - 6 = 28 \text{ ступенях свободи.}$$

Після розрахунків сум квадратів відхилень для дисперсії варіантів і помилок складають таблицю дисперсійного аналізу (табл. 45) і розраховують фактичне значення критерію *F*.

Таблиця 45

Таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{\text{факт.}}$	$F_{\text{теор.}}$	
					05	01
Загальна, C_Y	6687,1	34	–	–	–	–
Варіантів, C_V	6420,3	6	1070,1	112,3	2,44	3,53
Помилок, C_Z	266,8	28	9,53	–	–	–

Теоретичні значення F_{05} – 2,44 і F_{01} – 3,53 знаходять за дод. В для 6 ступенів свободи дисперсії варіантів (чисельник) і 28 ступенів свободи дисперсії помилок (знаменник).

У цьому досліді фактичне значення критерію F значно перевищує теоретичне F_{05} і F_{01}, що свідчить про наявність істотної різниці між досліджуваними варіантами. Оскільки нульова гіпотеза спростовується, знаходимо помилку середньої ($s_{\overline{x}}$), помилку різниці середніх (s_d) і найменшу істотну різницю для потрібного рівня ймовірності.

$$s_x = \sqrt{s^2 / n} = \sqrt{9{,}53 / 5} = 1{,}38 \text{ г};$$
$$s_d = \sqrt{2 \cdot s^2 / n} = \sqrt{2 \cdot 9{,}53 / 5} = 1{,}95 \text{ г};$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}05 \cdot 1{,}95 = 4{,}0 \text{ г};$$
$$HIP_{05} = t_{05} \cdot s_d \cdot 100 / \overline{x} = 2{,}05 \cdot 1{,}95 \cdot 100 / 70{,}3 = 5{,}5\ \%;$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}76 \cdot 1{,}95 = 5{,}4 \text{ г};$$
$$HIP_{05} = t_{01} \cdot s_d \cdot 100 / \overline{x} = 2{,}76 \cdot 1{,}95 \cdot 100 / 70{,}3 = 7{,}7\ \%.$$

Теоретичне значення t-критерію знаходять за дод. Б для 28 ступенів свободи дисперсії помилок. Зведені результати досліджуваних показників урожайності тритикале ярого та результати статистичного аналізу доцільно подати у вигляді табл. 46.

Висновок. Досліджувані варіанти системи живлення – 4-й, 5-й і 7-й – істотно перевищували контроль (2) за рівнем урожайності

Таблиця 46

Підсумкова таблиця результатів досліджень на основі проведеного статистичного аналізу

Варіант досліду	Урожайність зерна, г	Відхилення від контролю		Групи для	
		г	%	HIP_{05}	HIP_{01}
1	64,4	1,6	2,5	II	II
2 (контроль)	62,8	–	–	–	–
3	58,6	–4,2	–6,7	III	II
4	73,8	11,0	17,5	I	I
5	87,8	25,0	39,8	I	I
6	53,2	–9,6	–15,3	III	III
7	91,4	28,6	45,5	I	I
HIP_{05}	–	4,0	5,5	–	–
HIP_{01}	–	5,4	7,7	–	–

за обох рівнів імовірності. Між 1-м варіантом і контролем істотної різниці не було. З імовірністю 95 % 3-й варіант істотно поступався за врожайністю контрольному. За більш точної оцінки (ймовірність 99 %) між цими варіантами істотної різниці не було.

Приклад 2. Провести оцінку показників сирої вегетативної маси 20-ти рослин тритикале ярого за різних варіантів мінерального живлення. Кількість повторень (спостережень) для варіантів 1–3, 5, 7 і 8 становить п'ять, для варіантів 4 і 6 – чотири (табл. 47).

Таблиця 47

Сира вегетативна маса 20-ти рослин тритикале ярого залежно від варіантів мінерального живлення, г

Варіант досліду	Сира вегетативна маса за повтореннями					Кількість визначень	Сума, V	Середнє за V
1 (контр.)	88	78	83	80	85	5	414	82,8
2	93	97	101	95	98	5	484	96,6
3	111	116	108	114	110	5	559	111,8
4 (контр.)	74	80	71	76	–	4	301	75,3
5	91	91	98	94	95	5	469	93,8
6	104	108	101	106	–	4	419	104,8
7	120	126	114	119	123	5	602	120,4
8	88	83	89	92	85	5	437	87,4
Загальна сума						$\sum n = 38$	$\sum x = 3685$	$\overline{x} = 97{,}0$

Проведення розрахунків. Характерною особливістю обробки результатів вегетаційного досліду з різною кількістю повторень за варіантам є необхідність визначення кількох показників найменшої істотної різниці, оскільки не всі середні рівнозначні. У цьому прикладі варіанти 4 і 6 спираються на чотири, а решта – на п'ять повторень (спостережень).

1-й етап. Визначають суми і середні за варіантами, загальну суму і середній показник вегетативної маси тритикале ярого в досліді.

2-й етап. Розраховують корегуючий фактор C, суму квадратів відхилень – загальну C_Y, варіантів C_V і помилок C_Z.

$$N = \sum n = 38;$$
$$C \text{ (поправка)} = (\sum x)^2 / N = 3685^2 / 38 = 357348;$$
$$C_Y \text{ (загальна)} = \sum x^2 - C = (88^2 + 78^2 + ... + 85^2) - 357348 = 7861$$
при $(N - 1) = 38 - 1 = 37$ ступенях свободи.

Під час розрахунків суми квадратів відхилень для варіантів слід мати на увазі, що до суми v входить різна кількість повторень n, через це розрахунки C_V проводять за формулою:

$$C_V = \sum(x_{v1}^2/n_1 + x_{v1}^2/n_2 + ... + x_{v1}^2/nl) - C = (414^2/5 + 484^2/5 + 559^2/5 + + 301^2/4 + 469^2/5 + 419^2/4 + 602^2/5 + 437^2/5) - 357348 = (34279{,}2 + + 46851{,}2 + ... + 38193{,}8) - 357348 = 7468;$$

$$C_Z = C_Y - C_V = 7861 - 7486 = 375$$

при $(N - 1) - (l - 1) = 30$ ступенях свободи.

3-й етап. Після розрахунку сум квадратів відхилень складають таблицю дисперсійного аналізу і визначають фактичний критерій F (табл. 48).

Таблиця 48

Результати дисперсійного аналізу однофакторного вегетаційного досліду з різною кількістю спостережень за варіантами

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна, C_Y	7861	37	–	–	–	–
Варіантів, C_V	7468	7	1067,0	85,4	2,34	3,30
Помилок, C_Z	375	30	12,5	–	–	–

Теоретичне значення $F_{05} = 2{,}34$ і $F_{01} = 3{,}30$ для 7 ступенів свободи дисперсії варіантів (чисельник) і 30 ступенів свободи дисперсії помилок (знаменник) знаходять за дод. В. Оскільки в нашому прикладі $F_{факт.} > F_{01}$, то нульова гіпотеза (H_0) спростовується і між варіантами досліду є істотна різниця.

4-й етап. Під час проведення оцінки істотності часткових розходжень у досліді з різною кількістю повторень досліджуваних варіантів необхідно враховувати нерівноточність порівняння середніх. Помилки середніх варіантів 1–3, 5, 7 і 8 спираються на $n_{1-3} = n_5 = n_7 = n_8 = 5$ повторень, а варіантів 4 і 6 – на $n_4 = n_6 = 4$ повторення. Тому помилку різниці між середніми, розрахованими на основі різної кількості повторень, слід визначити за формулою, яка враховує різну кількість повторень за досліджуваними варіантами:

$$s_d = \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}} = \sqrt{s^2 + (\frac{n_1 + n_2}{n_1 \cdot n_2})}\ ;$$

Помилка різниці середніх під час порівняння варіанта 4 з варіантом 6 ($n_4 = n_6 = 4$) становить:

$$s_d = \sqrt{2s^2 / n} = \sqrt{212{,}5 / 4} = 2{,}5 \text{ г.}$$

Під час порівняння середніх варіантів 1–3, 5, 7 і 8 ($n = 5$) з варіантами 4 і 6 ($n = 4$) помилка різниці середніх становить:

$$s'_d = \sqrt{s^2 \cdot ((n_1 + n_2) / (n_1 \cdot n_2))} = \sqrt{12{,}5 \cdot (4 + 5) / (4 \cdot 5)} = 2{,}37 \text{ г.}$$

Під час порівняння середніх варіантів 1–3, 5, 7 і 8 ($n_{1-3} = n_5 = n_7 = n_8 = 5$) помилка різниці середніх становить:

$$s''_d = \sqrt{2s^2 / n} = \sqrt{212{,}5 / 5} = 2{,}24 \text{ г.}$$

Відповідно до розрахованих помилок різниці середніх найменша істотна різниця становитиме:

$$HIP_{05} = t_{05} \cdot s_d = 2{,}04 \cdot 2{,}50 = 5{,}1 \text{ г;}$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}75 \cdot 2{,}50 = 6{,}9 \text{ г;}$$
$$HIP'_{05} = t_{05} \cdot s'_d = 2{,}04 \cdot 2{,}37 = 4{,}8 \text{ г;}$$
$$HIP'_{01} = t_{01} \cdot s'_d = 2{,}75 \cdot 2{,}37 = 6{,}5 \text{ г;}$$
$$HIP''_{05} = t_{05} \cdot s''_d = 2{,}04 \cdot 2{,}24 = 4{,}6 \text{ г;}$$
$$HIP''_{01} = t_{01} \cdot s''_d = 2{,}75 \cdot 2{,}24 = 6{,}2 \text{ г.}$$

Теоретичне значення критерію $t_{05} = 2{,}04$ і $t_{01} = 2{,}75$ знаходять за дод. Б для 30 ступенів свободи дисперсії помилок. Результати статистичної обробки та оцінку середніх показників досліджуваних варіантів доцільно подати у вигляді підсумкової табл. 49.

Висновок. Усі досліджувані варіанти системи живлення забезпечували достовірне підвищення вегетативної маси рослин тритикале ярого порівняно з другим контролем (варіант 4) за обох рівнів імовірності (95 і 99 %). Порівняно з першим контролем (варіант 2) усі варіанти, крім другого контролю (варіант 4), і варіанта 8 забезпечували істотне збільшення вегетативної маси 20-ти рослин для обох рівнів імовірності (І група). Варіант 8 був на одному рівні з контролем (різниця в межах HIP_{05}), а другий контроль (варіант 4) за вегетативною масою істотно поступався варіанту 1 за обох рівнів імовірності (ІІІ група).

Таблиця 49

Підсумкова таблиця результатів вегетаційного досліду на основі проведеного дисперсійного аналізу

Варіант досліду	Сира маса рослин, г	Порівняно з першим контролем (варіант 2)			Порівняно з другим контролем (варіант 4)		
		різниця	HIP_{05}	HIP_{01}	різниця	HIP_{05}	HIP_{01}
1	82,8	–	–	–	7,5	4,8	6,5
2 (контроль)	96,8	14,0	4,6	6,2	21,5	4,8	6,5
3	111,8	29,0	4,6	6,2	36,5	4,8	6,5
4 (контроль)	75,3	– 7,5	4,8	6,5	–	–	–
5	93,8	11,0	4,6	6,2	18,5	4,8	6,5
6	104,8	22,0	4,8	6,5	29,5	5,1	6,9
7	120,4	37,6	4,6	6,2	45,1	4,8	6,5
8	87,4	4,6	4,6	6,2	12,1	4,8	6,5

6.2. Багатофакторний дослід

Під час проведення дисперсійного аналізу багатофакторного вегетаційного досліду спочатку визначають суму квадратів відхилень – загальну C_Y, варіантів C_V і помилок C_Z. Далі суму квадратів відхилень варіантів (C_V) розкладають на складові компоненти – головні ефекти досліджуваних чинників та їхньої взаємодії. У двофакторному вегетаційному досліді $C_V = C_A + C_B + C_{AB}$; у трифакторному $C_V = C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}$.

Наведемо приклад статистичних розрахунків методом дисперсійного аналізу двофакторного вегетаційного досліду.

Приклад 3. У двофакторному вегетаційному досліді вивчено вплив трьох градацій чинника А (сорти пшениці твердої ярої) і трьох градацій чинника В – різні дози азоту у чотириразовому повторенні на врожай зерна 50-ти рослин (1 посудина) (табл. 50).

Проведення розрахунків. Дисперсійний аналіз двофакторного вегетаційного досліду ($l_a = 3$, $l_b = 3$, $n = 4$) проводять у кілька етапів.

1-й етап. Визначають загальну суму показників урожаю, середній урожай у досліді, суми і середні за варіантами досліду.

2-й етап. Розраховують корегуючий чинник C, суму квадратів відхилень – загальну C_Y, варіантів C_V і помилок C_Z.

Таблиця 50

Урожай зерна пшениці твердої ярої залежно від сортових особливостей (чинник А) і різних доз азоту (чинник В), г/50 рослин

Чинник A (сорт)	Чинник B (доза азоту)	Урожай, x				Сума, V	Середнє за V
A_1	B_1 (к)	41,4	44,1	42,6	43,3	171,4	42,9
	B_2	47,6	47,4	49,1	45,6	189,7	47,4
	B_3	53,2	50,9	54,4	52,7	211,2	52,8
A_2	B_1	32,6	36,0	34,3	33,8	136,7	34,2
	B_2	35,8	39,1	38,8	37,4	151,1	37,8
	B_3	41,0	41,7	43,3	39,6	165,6	41,4
A_3	B_1	43,7	44,9	42,2	41,0	171,8	43,0
	B_2	50,8	51,3	50,3	49,6	202,0	50,5
	B_3	47,0	48,2	45,4	47,5	188,1	47,0
Загальна сума						$\sum x = 1588$	$\bar{x} = 44,1$

$$N = l_a \cdot l_b \cdot n = 3 \cdot 3 \cdot 4 = 36;$$
$$C \text{ (поправка)} = (\sum x)^2 / N = 1587,6^2 / 36 = 70013,2;$$
$$C_Y \text{ (загальна)} = \sum x^2 - C = (41,4^2 + 44,1^2 + ... + 47,5^2) - 70013,2 =$$
$$= 1190,8 \text{ за } (N-1) = 36 - 1 = 35 \text{ ступенів свободи;}$$
$$C_V \text{ (варіантів)} = \sum x_v^2 / n - C = (171,4^2 + 189,7^2 + ... + 188,1^2) / 4 -$$
$$- 70013,2 = 1139,8 \text{ за } (l-1) = (l_a \cdot l_b - 1) = (9-1) =$$
$$= 8 \text{ ступенів свободи;}$$
$$C_Z = C_Y - C_V = 1190,8 - 1139,8 = 51,0$$
$$\text{за } (N-1) - (l-1) = 35 - 11 = 24 \text{ ступенів свободи.}$$

3-й етап. Для визначення сум квадратів відхилень чинників A і B та їхньої взаємодії AB складають допоміжну таблицю, до якої переносять суми показників за варіантами і розраховують суми показників за досліджуваними чинниками (суми врожаїв за рядами і колонками), та визначають суми квадратів відхилень для головних ефектів A, B та їхньої взаємодії.

Сума квадратів відхилень для чинника A (сорти):

$$C_A = \sum x_A^2 / l_b \cdot n - C = (572,3^2 + 453,4^2 + 561,9^2) / 4 \cdot 3 - 70013,2 =$$
$$= (327527,3 + 205571,6 + 315731,6) / 12 - 70013,2 = 722,7$$
$$\text{за } (l_a - 1) = (3-1) = 2 \text{ ступенів свободи.}$$

Сума квадратів відхилень для чинника *B* (дози азоту):

$$C_B = \sum x_B^2 / l_a \cdot n - C = (479{,}9^2 + 542{,}8^2 + 564{,}9^2) / 12 - 70013{,}2 =$$
$$= (230304 + 294632 + 319112) / 12 - 70013{,}2 = 324{,}1$$
за $(l_b - 1) = (3 - 1) = 2$ ступенів свободи.

Сума квадратів відхилень для взаємодії *AB*:

$$C_{AB} = C_V - C_A - C_B = 1139{,}8 - 722{,}7 - 324{,}1 = 93{,}0$$
за $(l_a - 1) \cdot (l_b - 1) = (3 - 1) \cdot (3 - 1) = 4$ ступенів свободи.

4-й етап. Складають таблицю дисперсійного аналізу та розраховують фактичний критерій *F* (табл. 51).

Таблиця 51

Результати дисперсійного аналізу двофакторного вегетаційного досліду ($l_a = 3$, $l_b = 3$, $n = 4$)

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна, C_Y	1190,8	35	–	–	–	–
Чинника *A*	722,7	2	361,4	191,2	3,37	5,53
Чинника *B*	324,1	2	162,1	85,8	3,37	5,53
Взаємодії *AB*	90,0	4	22,5	11,9	2,74	4,14
Помилок, C_Z	51,0	27	1,89	–	–	–

Теоретичне значення F_{05} і F_{01} знаходять за дод. В залежно від кількості ступенів свободи дисперсій головних ефектів чинників і їхньої взаємодії (чисельник) та 27 ступенів свободи дисперсії помилок (знаменник). Оскільки $F_{факт.}$ досліджуваних чинників і їхньої взаємодії перевищує $F_{теор.}$ за обох рівнів ймовірності, то в досліді є достовірна різниця між варіантами і нульова гіпотеза (H_0) спростовується.

5-й етап. Проводять оцінку істотності часткових розходжень. Для цього визначають помилку середньої ($s_{\bar{x}}$), помилку різниці середніх (s_d) і *НІР* для потрібного рівня ймовірності.

$$s_x = \sqrt{s^2 / n} = \sqrt{1{,}89 / 4} = 0{,}69 \text{ г};$$
$$s_d = \sqrt{2 \cdot s^2 / n} = \sqrt{2 \cdot 1{,}89 / 4} = 0{,}97 \text{ г}.$$
$$НІР_{05} = t_{05} \cdot s_d = 2{,}05 \cdot 0{,}97 = 2{,}0 \text{ г};$$
$$НІР_{01} = t_{01} \cdot s_d = 2{,}77 \cdot 0{,}97 = 2{,}7 \text{ г}.$$

Теоретичне значення $t_{05} = 2{,}05$ і $t_{01} = 2{,}77$ знаходять за дод. Б для 27 ступенів свободи дисперсії помилок. Результати досліджень і їхньої статистичної обробки доцільно подати у вигляді рис. 25.

Висновок. Різниця між будь-якими варіантами досліду, що перевищує 2,0 г, є істотною з імовірністю 95 %. Оскільки нас більше цікавить порівняння досліджуваних варіантів з контролем, детальніше зупинимося на їхній оцінці. Так, у дослідах за обох потрібних рівнів імовірності – 95 і 99 %, урожайність варіантів A_1B_2, A_1B_3, A_3B_2 і A_3B_3 істотно перевищувала контроль – A_1B_1 (I статистична група), варіанти A_2B_1 і A_3B_1 значно поступалися контролю (III група), а варіанти A_2B_3 і A_3B_1 були на одному рівні з контролем.

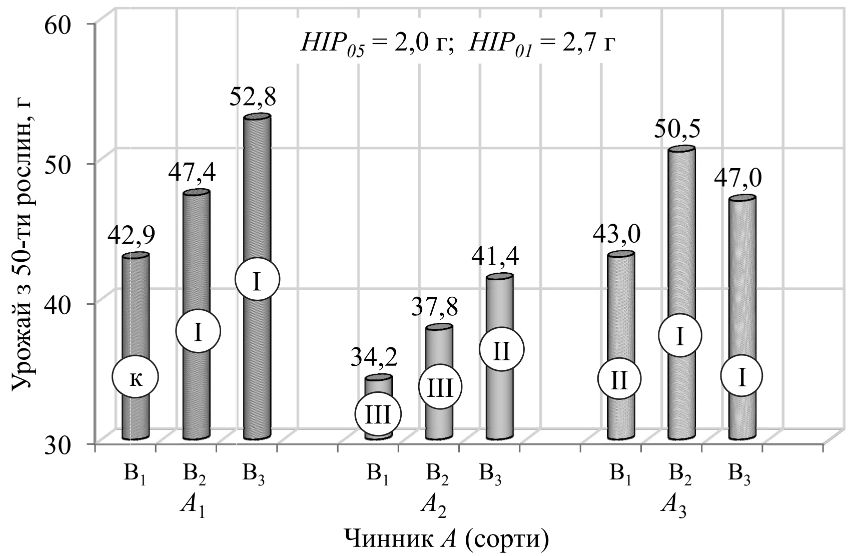


Рис. 25. Урожай 50-ти рослин пшениці твердої ярої трьох різних сортів (чинник *A*) залежно від рідних доз внесення азоту (чинник *B*)

7. Дисперсійний аналіз однофакторного польового досліду з однорічними і багаторічними культурами

Обробку даних польового досліду з однорічними культурами, поставленого методом рендомізованих повторень, проводять у кілька етапів:

– вихідні дані заносять до таблиці, визначають суми за варіантами і повтореннями досліду, розраховують середні за варіантами;
– далі розраховують суми квадратів відхилень для всіх джерел впливу на варіабельність ознаки: дисперсії загальної C_Y, варіантів C_V, помилок C_Z і повторень C_P;
– складають підсумкову таблицю дисперсійного аналізу і зіставляють отримані дані $F_{\text{факт.}}$ з $F_{\text{теор.}}$ для потрібних рівнів імовірності (перевіряють нульову гіпотезу). За $F_{\text{факт.}} > F_{\text{теор.}}$ проводять оцінку істотності порівнянь середніх показників за варіантами досліду і групують на основі показників *НІР*. Якщо $F_{\text{факт.}} < F_{\text{теор.}}$, нульова гіпотеза не спростовується і всі різниці між вибірковими середніми за варіантами розміщені в межах випадкових відхилень, у цьому випадку вираховують тільки помилки досліду – $s_{\overline{X}}$.

Суми квадратів відхилень для джерел варіювання дисперсії, $F_{\text{факт.}}$ для різних рівнів імовірності знаходять за формулами:

$$C \text{ (поправка)} = (\sum X)^2 / N;$$
$$C_Y \text{ (загальна дисперсія)} = \sum X^2 - C \text{ за } (N-1) \text{ ступенів свободи;}$$
$$C_P \text{ (дисперсія повторень)} = \sum X_P^2 / l - C \text{ за } (n-1) \text{ ступенів свободи;}$$
$$C_V \text{ (дисперсія варіантів)} = \sum X_V^2 / n - C \text{ за } (l-1) \text{ ступенів свободи;}$$
$$C_Z \text{ (дисперсія помилок)} = C_Y \cdot C_P \cdot C_V$$
$$\text{за } (l-1)\cdot(n-1) \text{ ступенів свободи;}$$
$$s_V^2 \text{ (середній квадрат дисперсії варіантів)} = C_v / (l-1);$$
$$s^2 \text{ (середній квадрат дисперсії помилок)} = C_z / (l-1)\cdot(n-1);$$
$$F_{\text{факт.}} = s_v^2 / s^2.$$

$F_{\text{теор}}$. (05 і 01) знаходять за дод. В для $(l-1)$ ступенів свободи дисперсії варіантів (чисельник) і $(l-1)\cdot(n-1)$ ступенів свободи дисперсії помилок (знаменник).

7.1. Обробка дослідів з однорічними культурами

На конкретному прикладі покажемо техніку проведення статистичних розрахунків даних польового досліду, поставленого методом рендомізованих повторень.

Приклад 1. Провести дисперсійний аналіз показників урожайності гібридів кукурудзи (табл. 52), визначити *НІР* з імовірністю 95 і 99 % та згрупувати показники врожайності за групами стосовно до контрольного варіанта.

Таблиця 52

Урожайність гібридів кукурудзи, т/га

Варіант (гібрид)	Урожайність за повтореннями				Сума, *V*	Середнє
	I	II	III	IV		
1 (контроль)	5,13	5,90	5,78	5,42	22,23	5,56
2	3,85	4,23	4,15	3,97	16,20	4,05
3	6,20	6,76	6,50	6,34	25,80	6,45
4	4,40	5,11	4,98	4,53	19,02	4,76
5	5,93	6,37	6,20	6,15	24,65	6,16
6	5,17	5,74	5,70	5,28	21,89	5,47
7	5,53	5,80	5,84	5,64	22,81	5,70
Сума, *P*	36,21	39,91	39,15	37,33	$\sum X = 152{,}6$	$\overline{x}_{\text{ср.}} = 5{,}45$

Проведення розрахунків. Визначають суми показників урожайності зерна кукурудзи за повтореннями, варіантами і в цілому по досліду, результати заносять до табл. 5.2. Правильність розрахунків перевіряють за співвідношенням: $\sum X_V = \sum X_P = \sum X = 152{,}6$.

Далі знаходять коригуючий фактор C, загальну суму квадратів відхилень C_Y, суму квадратів відхилень варіантів C_V, повторень C_P і помилок C_Z:

$$N = l \cdot n = 7 \cdot 4 = 28;$$
$$C = (\sum X)^2 / N = 152{,}6^2 / 28 = 831{,}67;$$
$$C_Y = \sum X^2 - C = (5{,}13^2 + 5{,}90^2 + ... + 5{,}64^2) - 831{,}67 =$$
$$= (26{,}32 + 34{,}81 + ... + 31{,}81) - 831{,}67 = 17{,}51$$
$$\text{за } (N-1) = (28-1) = 27 \text{ ступенів свободи;}$$
$$C_P = \sum X_P^2 / l - C = (36{,}21^2 + 39{,}91^2 + 39{,}15^2 + 37{,}33^2) / 7 - 831{,}67 =$$
$$= (1311{,}16 + 1592{,}81 + 1532{,}72 + 1393{,}53) / 7 - 831{,}67 = 1{,}22$$
$$\text{за } (n-1) = (4-1) = 3 \text{ ступенів свободи;}$$

$$C_V = \sum X_V^2 / n - C = (22{,}23^2 + 16{,}20^2 + ... + 22{,}81^2) / 4 - 831{,}67 = 16{,}11$$

за $(l - 1) = (7 - 1) = 6$ ступенів свободи;

$$C_Z = C_Y - C_P - C_V = 17{,}51 - 1{,}22 - 16{,}11 = 0{,}18$$

за $(l - 1) \cdot (n - 1) = (7 - 1) \cdot (4 - 1) = 18$ ступенів свободи.

На наступному етапі складають таблицю дисперсійного аналізу (табл. 53), визначають середні квадрати для дисперсії варіантів і помилок, розраховують $F_{\text{факт}}$.

Таблиця 53

Таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{\text{факт.}}$	$F_{\text{теор.}}$	
					05	01
Загальна, C_Y	17,51	27	–	–	–	–
Повторень, C_P	1,22	3	–	–	–	–
Варіантів, C_V	16,11	6	2,69	269,0	2,66	4,01
Помилок, C_Z	0,18	18	0,01	–	–	–

Теоретичні значення критерію F_{05} і F_{01} знаходять за дод. В для 6 ступенів свободи дисперсії варіантів (чисельник) і 18 ступенів свободи дисперсії помилок (знаменник). Оскільки $F_{\text{факт.}} > F_{\text{теор}}$. для обох рівнів імовірності (95 і 99 %), у досліді є істотна різниця між досліджуваними гібридами і нульова гіпотеза спростовується.

Далі для оцінки істотності різниці між варіантами досліду і їх групування стосовно до контролю розраховують помилку середньої $s_{\bar{x}}$, помилку різниці середніх s_d і HIP для 95 і 99 % рівнів ймовірності в абсолютних і відносних показниках:

$$s_{\bar{x}} = \sqrt{S^2 / n} = \sqrt{0{,}01 / 4} = 0{,}05 \text{ т/га};$$

$$s_d = \sqrt{2S^2 / n} = \sqrt{2 \cdot 0{,}01 / 4} = 0{,}07 \text{ т/га};$$

$$HIP_{05} = t_{05} \cdot s_d = 2{,}10 \cdot 0{,}07 = 0{,}15 \text{ т/га};$$

$$HIP_{01} = t_{01} \cdot s_d = 2{,}88 \cdot 0{,}07 = 0{,}20 \text{ т/га}.$$

$$HIP_{05} = t_{05} \cdot s_d \cdot 100 / \bar{x} = 0{,}15 \cdot 100 / 5{,}45 = 2{,}75 \text{ \%};$$

$$HIP_{01} = t_{01} \cdot s_d \cdot 100 / \bar{x} = 0{,}20 \cdot 100 / 0{,}07 = 3{,}67 \text{ \%}.$$

Теоретичне значення $t_{05} = 2{,}10$ і $t_{01} = 2{,}88$ знаходять за дод. Б для 18 ступенів свободи дисперсії помилок. Зведені результати дисперсійного аналізу доцільно подати у вигляді табл. 54.

Таблиця 54

Розподіл показників урожайності зерна гібридів кукурудзи за статистичними групами порівняно з контрольним варіантом досліду

Варіант (гібрид)	Урожайність, т/га	Відхилення від контролю		Група для	
		т/га	%	HIP_{05}	HIP_{01}
1 (контроль)	5,56	–	–	контроль	контроль
2	4,05	–1,51	–27,2	III	III
3	6,45	0,89	16,0	I	I
4	4,76	–0,80	–14,4	III	III
5	6,16	0,60	10,8	I	I
6	5,47	–0,09	–1,6	II	II
7	5,70	0,14	2,5	II	II
HIP_{05}	–	0,15	2,75	–	–
HIP_{01}	–	0,20	3,67	–	–

Висновок. Найбільшу, статистично достовірну врожайність зерна кукурудзи в досліді формували сорти 3 і 5. Ці сорти належали до II групи показників стосовно до контрольного варіанта. Сорти 2 і 3 були менш урожайні порівняно з контролем за обох рівнів імовірності (II група показників).

Для точнішої оцінки досліджуваних варіантів порівняно з контролем досліду, а також для можливості введення додаткових варіантів іноді збільшують кількість контролів (2 і більше) у кожній повторності.

На конкретному прикладі покажемо техніку проведення дисперсійного аналізу однофакторного польового досліду з подвійним набором контрольного варіанта в кожній повторності.

Приклад 2. Провести дисперсійний аналіз досліду з вивчення впливу різних сортів сої на збір білка з 1 га (табл. 55). Кількість варіантів (сортів) – 10. У кожній повторності контрольний варіант займає дві ділянки. Таким чином, за триразової повторності контроль має шестиразову повторність.

Проведення розрахунків. Спочатку визначають суми показників за повтореннями і варіантами, середні показники збору білка за варіантами, загальну суму показників у досліді, середній збір білка сої з гектара в середньому по досліду. Правильність розрахунків переві-

Таблиця 55

Збір білка з 1 га посівів різних сортів сої, кг/га

Варіант (сорт)	Збір білка за повтореннями			Сума, V	Середнє
	I	II	III		
1 (контроль)	710	696	721	2127	716,3
2	683	624	657	1964	654,7
3	714	701	710	2125	708,3
4	695	680	683	2058	686,0
5	886	839	855	2580	860,0
1 (контроль)	723	712	736	2171	–
6	752	706	725	2183	727,7
7	917	887	911	2715	905,0
8	826	814	820	2460	820,0
9	852	830	856	2538	846,0
10	726	682	708	2116	705,3
Суми P	8484	8171	8382	$\sum X = 25037$	$\bar{x}_{ср.} = 758{,}7$

ряють за співвідношенням $\sum X = \sum V = \sum P = 25037$. Відмінною рисою цього досліду є розміщення контрольного варіанта на двох ділянках у кожному повторенні. Середній показник контролю визначають за рівнянням: $\bar{X}_к = \sum X_к / 2n = (710 + ... + 736) / 2 \cdot 3 = 716{,}3$ кг/га.

На наступному етапі знаходять поправку C, загальну суму квадратів відхилень C_Y, суму квадратів відхилень повторень C_P, варіантів C_V і помилок C_Z. Оскільки у цьому досліді для визначення сум квадратів відхилень використовують великі числа, вихідні дані доцільно зменшити за співвідношенням $X_1 = X - A$, узявши за довільний початок число, близьке до середнього показника – 760. Перетворені показники переносять до табл. 56, визначають суми за повтореннями, варіантами і загальну суму. Правильність розрахунків перевіряють за співвідношенням: $\sum P_1 = \sum V_1 = \sum X_1 = -43$.

$$N = l \cdot n = 11 \cdot 3 = 33;$$
$$C = (\sum X_1)^2 / N = 43^2 / 33 = 56{,}03;$$
$$C_Y = \sum X_1^2 - C = (50^2 + 77^2 + ... + 52^2) - 56{,}03 = 214316{,}97$$
за $(N - 1) = (33 - 1) = 32$ ступенів свободи;
$$C_P = \sum X_{1P}^2 / l - C = (124^2 + 189^2 + 22^2) / 11 - 56{,}03 =$$
$$= (15376 + 35721 + 484) / 11 - 56{,}03 = 4633{,}15$$
за $(n - 1) = (3 - 1) = 2$ ступенів свободи;

Таблиця 56

Перетворені показники під довільний початок (A = 760)

Варіант (сорт)	Повторення ($X_1 = X - A$)			Сума, V
	I	II	III	
1 (контроль)	–50	–64	–39	–153
2	–77	–136	–103	–316
3	–46	–59	–50	–155
4	–65	–80	–77	–222
5	+126	+79	+95	+300
1 (контроль)	–37	–48	–24	–109
6	–8	–54	–35	–97
7	+157	+127	+151	+435
8	+66	+54	+60	+180
9	+92	+70	+96	+258
10	–34	–78	–52	–164
Суми P	+124	–189	22	$\sum X_1 = -43$

$$C_V = \sum X_{1V}^2 / n - C = (153^2 + 316^2 + ... + 164^2) / 3 - 56{,}03 = 207593{,}64$$
за $(l - 1) = (11 - 1) = 10$ ступенів свободи;
$$C_Z = C_Y - C_P - C_V = 214316{,}97 - 4633{,}15 - 207593{,}64 = 2090{,}18$$
за $(l - 1) \cdot (n - 1) = (11 - 1) \cdot (3 - 1) = 20$ ступенів свободи.

Далі складають таблицю дисперсійного аналізу (табл. 57) і розраховують $F_{факт.}$ (відношення середніх квадратів дисперсії варіантів до середнього квадрата дисперсії помилок – s_v^2 / s^2).

Таблиця 57

Таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна, C_Y	214317,0	32	–	–	–	–
Повторень, C_P	4633,15	2	–	–	–	–
Варіантів, C_V	207593,6	10	20759,4	198,7	2,35	3,37
Помилок, C_Z	2090,18	20	104,5	–	–	–

Теоретичне значення $F_{05} = 2{,}35$ і $F_{01} = 3{,}37$ знаходять за дод. В для 10 ступенів свободи дисперсії варіантів (чисельник) і 20 ступенів свободи дисперсії помилок (знаменник). $F_{факт.} > F_{теор.}$ для обох рівнів імовірності. Отже, нульова гіпотеза (H_0) спростовується і між варіантами є істотна різниця за обох рівнів імовірності.

Далі проводять оцінку істотності різниці між досліджуваними варіантами і контролем. Для цього спочатку розраховують помилку середньої $s_{\bar{x}}$ та помилку різниці середніх s_d.

$$s_{\bar{x}} = \sqrt{s^2 / n} = \sqrt{104{,}5 / 3} = 5{,}90 \text{ кг/га}.$$

Оскільки у цьому досліді кількість контрольних варіантів удвічі перевищує кожен з досліджуваних варіантів, під час порівняння досліджуваних показників з контролем помилку різниці середніх визначають за формулою:

$$s_d = \sqrt{s^2} \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}};$$

де n_1 – кількість контрольних варіантів у досліді; n_2 – кількість повторень у досліді.

$$s_d = \sqrt{104{,}5} \cdot \sqrt{\frac{6+3}{6 \cdot 3}} = 7{,}2 \text{ кг/га}.$$

Порівнюючи досліджувані варіанти між собою, помилку різниці середніх визначають за стандартною формулою:

$$s'_d = \sqrt{2 \cdot s^2 / n} = \sqrt{2 \cdot 104{,}5 / 3} = 8{,}3 \text{ кг/га}.$$

Найменша істотна різниця для 5 %-го і 1 %-го рівня ймовірності становить:

$$HIP_{05} = t_{05} \cdot s_d = 2{,}09 \cdot 7{,}2 = 15{,}05 \text{ кг/га};$$
$$HIP_{05} = t_{05} \cdot s_d \cdot 100 / x_{ср.} = 15{,}05 \cdot 100 / 758{,}7 = 1{,}98 \text{ \%};$$
$$HIP'_{05} = t_{05} \cdot s'_d = 2{,}09 \cdot 8{,}3 = 17{,}35 \text{ кг/га};$$
$$HIP'_{05} = t_{05} \cdot s'_d \cdot 100 / x_{ср.} = 17{,}35 \cdot 100 / 758{,}7 = 2{,}29 \text{ \%};$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}85 \cdot 7{,}2 = 20{,}5 \text{ кг/га};$$
$$HIP_{01} = t_{01} \cdot s_d \cdot 100 / x_{ср.} = 20{,}5 \cdot 100 / 758{,}7 = 2{,}70 \text{ \%};$$
$$HIP'_{01} = t_{01} \cdot s'_d = 2{,}85 \cdot 8{,}3 = 23{,}66 \text{ кг/га};$$
$$HIP'_{01} = t_{01} \cdot s'_d \cdot 100 / x_{ср.} = 23{,}66 \cdot 100 / 758{,}7 = 3{,}12 \text{ \%}.$$

Теоретичне значення $t_{05} = 2{,}09$ і $t_{01} = 2{,}85$ знаходять за дод. Б для 20 ступенів свободи дисперсії помилок. Показники HIP_{05} і HIP_{01} використовують для порівняння досліджуваних варіантів із контролем, а HIP'_{05} і HIP'_{01} – для порівняння досліджуваних варіантів між собою. Результати дисперсійного аналізу подають у вигляді табл. 58.

Таблиця 58

Розподіл показників збору білка з 1 га посівів різних сортів сої за статистичними групами порівняно з контрольним варіантом досліду

Варіант (сорт сої)	Збір білка, кг/га	Відхилення від контролю		Групи для	
		кг/га	%	HIP_{05}	HIP_{01}
1 (контроль)	716,3	–	–	контроль	контроль
2	654,7	–61,6	–8,60	III	III
3	708,3	–8,0	–1,12	II	II
4	686,0	–30,3	–4,23	III	III
5	860,0	143,7	20,06	I	I
6	727,7	–11,4	1,59	II	II
7	905,0	188,7	26,34	I	I
8	820,0	103,7	14,48	I	I
9	846,0	129,7	18,11	I	I
10	705,3	–11,0	–1,54	II	II
HIP_{05}	–	15,05	1,98	–	–
HIP_{01}	–	20,05	2,70	–	–

Висновок. Найбільший статистично достовірний збір білка з 1 га був у сортів 5, 7–9. Ці сорти належали до I групи показників порівняно з контролем за обох рівнів імовірності. Сорти 2 і 4 помітно поступалися контролю за показником збору білка з 1 га (III група показників), решта сортів – 3, 6 і 10 були на одному рівні з контролем (II група показників).

На практиці іноді трапляється вибраковувати результати досліду окремих ділянок за повтореннями. Якщо таких ділянок небагато і якщо будь-який з варіантів досліду, з урахуванням вибракуваних даних, має не менш ніж триразову повторність, можна провести дисперсійний аналіз такого досліду, попередньо відновивши вибракувані показники.

Теоретичний показник вибракуваних ділянок розраховують на основі пов’язаних показників за варіантами і повтореннями. При цьому теоретично розрахований показник не може нічого змінювати у фактичних ефектах варіантів. Він повинен показати таку ефективність дії досліджуваного чинника, яка б дорівнювала середній ефективності за рештою повторень, що мають дані цього варіанта.

Приклад 3. У досліді з вивчення впливу семи норм висіву на врожайність соняшнику були отримані такі показники (табл. 59). Унаслідок грубих помилок було вибраковано дві ділянки у другому повторенні (варіанти 4 і 6) і одну ділянку у третьому повторенні (варіант 6). Потрібно відновити вибракувані показники, розрахувати критерій $F_{факт.}$ і в разі спростування нульової гіпотези провести оцінку істотності між досліджуваними варіантами та контролем.

Таблиця 59

Урожайність гібридів соняшнику залежно від впливу норми висіву, ц/га

Досліджуваний варіант (гібрид соняшнику)	Урожайність за повтореннями					Кількість повторень
	I	II	III	IV	V	
1	26,8	25,7	25,3	26,3	25,0	5
2 (контроль)	23,6	21,4	21,8	22,5	22,9	5
3	28,1	25,6	27,2	26,9	27,5	5
4	25,4	–	23,0	22,3	24,3	4
5	22,7	20,9	21,3	20,8	22,7	5
6	24,4	–	–	23,1	23,6	3
7	23,8	23,0	22,0	22,4	22,7	5
8	20,6	19,4	19,6	20,3	20,5	5
Суми *P* з повним набором варіантів (1+2+3+5+7+8)	145,6	136,0	137,2	139,2	141,3	–
Середнє за шістьома варіантами	24,27	22,67	22,87	23,20	23,55	–

Проведення розрахунків. На першому етапі слід відновити вибракувані дані. Розрахунки проводять у такому порядку. Спочатку до табл. 59 заносять суми показників за повтореннями, але лише для тих варіантів які представлені у кожній повторності (у нашому прикладі це варіанти 1, 2, 3, 5, 7, 8), далі розраховують середній показник урожайності за повтореннями для варіантів, які представлені у кожному повторенні, тобто одержані суми показників ділять на 6. Зокрема, сума показників урожайності за першою повторністю для варіантів з повним набором повторень становить (26,8 + 23,6 + + 28,1 + 22,7 + 23,8 + 20,6) = 145,6 ц/га, отже, середній показник становить 145,6 / 6 = 24,27 ц/га.

Далі для «відновлення» вибракуваних даних складають допоміжну таблицю, до якої з табл. 59 переносять показники варіантів, що містять вибракувані ділянки та середні за повтореннями, визначені для варіантів з повним набором показників урожайності (табл. 60).

Таблиця 60

Допоміжна таблиця для відновлення вибракуваних даних

Варіант («неповний»)	Повторення					Сума	Середнє для варіанта	
	I	II	III	IV	V		4	6
4	25,4	–	23,0	22,3	24,3	95,0	23,75	–
6	24,4	–	–	23,1	23,6	71,1	–	23,70
Середнє за шістьма варіантам	24,27	22,67	22,87	23,20	23,55	–	23,47	23,67
Ефекти варіантів	–	–	–	–	–	–	+0,28	+0,03
Відновлені показники врожайності для «неповних» варіантів								
4	–	22,95	–	–	–	–	–	–
6	–	22,70	22,90	–	–	–	–	–

Середні показники за повтореннями, розраховані для варіантів, представлених в усіх повтореннях, порівнюють між собою, їхні відмінності зумовлені здебільшого рівнем родючості ґрунтів повторень. Для визначення середнього ефекту, наприклад, варіанта 4, у якому була вибракувана ділянка у другому повторенні, розраховують середній показник цього варіанта за рештою повторень ($\bar{x}_{4ср.}$ = 23,75) і середній показник за варіантами з повним набором ділянок у цих повтореннях ($\bar{x}_{ср}$ = 23,47). Зівставляючи ці числа (23,75 – 23,47 = 0,28), знаходять середній ефект варіанта 4 з вибракуваним показником. За умови, коли б ділянка четвертого варіанта другого повторення сформувала нормальний результат і не була вибракувана, він був би приблизно на 0,28 ц/га більший, ніж середній показник решти варіантів у цьому повторенні, а саме: 22,67 + 0,28 = 22,95 ц/га. За цим же принципом відновлюють інші вибракувані показники.

Науковець Дж. У. Снедекор (1961) запропонував відновлювати одиночно вибракувані показники за формулою:

$$X' = [l \cdot V + n \cdot P - \sum X] / (l - 1) \cdot (n - 1)$$

де l – кількість варіантів; n – кількість повторень; V – сума показників варіанта, де міститься вибракуваний показник; P – сума показників повторення, де міститься вибракуваний показник; $\sum X$ – загальна сума показників досліду.

Після відновлення вибракуваних даних складають розрахункову таблицю для проведення дисперсійного аналізу, до якої вносять відновлені показники. Далі підраховують суми показників по досліду та середні показники врожайності за варіантами (табл. 61). Правильність розрахунків перевіряють за відношенням: $\sum P = \sum V = \sum X = 934{,}0$.

Таблиця 61

Урожайність гібридів соняшнику залежно від впливу норми висіву, ц/га

Варіант	Урожайність за повтореннями					Сума, V	Середнє
	I	II	III	IV	V		
1	26,8	25,7	25,3	26,3	25,0	129,1	25,8
2 (контроль)	23,6	21,4	21,8	22,5	22,9	112,2	22,4
3	28,1	25,6	27,2	26,9	27,5	135,3	27,1
4	25,4	**23,0**	23,0	22,3	24,3	118,0	23,6
5	22,7	20,9	21,3	20,8	22,7	108,4	21,7
6	24,4	**22,7**	**22,9**	23,1	23,6	116,7	23,3
7	23,8	23,0	22,0	22,4	22,7	113,9	22,8
8	20,6	19,4	19,6	20,3	20,5	100,4	20,1
Суми P	195,4	181,7	183,1	184,6	189,2	$\sum X = 934$	$\bar{x}_{ср} = 23{,}4$

На наступному етапі визначають корегуючий фактор (C), суму квадратів відхилень – загальну (C_Y), повторень (C_P), варіантів (C_V) і помилок (C_Z):

$$N = l \cdot n = 5 \cdot 8 = 40;$$
$$C = (\sum X)^2 / N = 934^2 / 40 = 21809;$$
$$C_Y = \sum X^2 - C = (26{,}8^2 + 25{,}7^2 + ... + 20{,}5^2) - 21809 = 196$$
за $(N - 1) = (40 - 1) = 39$ ступенів свободи;
$$C_P = \sum X_P^2 / l - C = (195{,}4^2 + 181{,}7^2 + ... + 189{,}2^2) / 8 - 21809 = 15{,}4$$
за $(n - 1) = (5 - 1) = 4$ ступенів свободи;
$$C_V = \sum X_V^2 / n - C = (129{,}1^2 + 112{,}2^2 + ... + 100{,}4^2) / 5 - 21809 = 172{,}7$$
за $(l - 1) = (8 - 1) = 7$ ступенів свободи;

$$C_Z = C_Y - C_P - C_V = 196{,}0 - 172{,}7 - 15{,}4 = 7{,}9$$
за $[(l - 1)(n - 1)] - 3 = [(8 - 1)(5 - 1)] - 3 = 28$ ступенів свободи.

Отримані дані заносять до підсумкової таблиці дисперсійного аналізу (табл. 62) і визначають $F_{факт.}$ Під час визначення числа ступенів свободи дисперсії помилок слід від отриманого числа відняти кількість вибракуваних ділянок: $(l - 1)(n - 1) - 3 = (8 - 1)(5 - 1) - 3 = 7 \cdot 4 - 3 = 28 - 3 = 25$.

Таблиця 62

Таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна, C_Y	196,0	39	–	–	–	–
Повторень, C_P	15,4	4	–	–	–	–
Варіантів, C_V	172,7	7	24,67	77,09	2,41	3,46
Помилок, C_Z	7,9	28 – 3 = 25	0,32	–	–	–

Показник $F_{05} = 2{,}41$ і $F_{01} = 3{,}46$ знаходять за дод. В для 7 ступенів свободи дисперсії варіантів (чисельник) і 25 ступенів свободи дисперсії помилок (знаменник). $F_{факт.} > F_{теор.}$, отже, між варіантами доведено істотну різницю і нульова гіпотеза (H_0) спростовується.

Особливістю визначення найменшої істотної різниці у дослідах з відновленими показниками є необхідність урахування кількості фактичних показників, на основі яких вираховували статистичні показники. Так, помилку середньої розраховують за формулою:

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n_{1+}\, n_{2+\cdots+}\, n_l}} = \sqrt{\frac{0{,}32}{5+5+5+4+5+3+5+5}} = \sqrt{\frac{0{,}32}{37}} = 0{,}09.$$

Під час порівняння варіантів з повним набором ділянок (варіанти 1, 2, 3, 5, 7 і 8), помилка різниці середніх становитиме:

$$s_d = \sqrt{2s^2 / n} = \sqrt{2 \cdot 0{,}32 / 5} = 0{,}36 \text{ ц/га}.$$

Під час порівняння варіанта 4, який представлений у чотирьох повтореннях ($n = 4$), з варіантами, які наведені в кожному повторенні ($n = 5$), помилка різниці середніх становитиме:

$$s'_d = \sqrt{s^2 \cdot (n_4 + n)/(n_4 \cdot n)} = \sqrt{0{,}32 \cdot (4+5) / (4 \cdot 5)} = 0{,}38 \text{ ц/га}.$$

Під час порівняння варіанта 6 (кількість повторень – 3) з «повними» варіантами (n = 5) помилка різниці середніх становитиме:

$$s''_d = \sqrt{s^2 \cdot (n_6 + n)/(n_6 \cdot n)} = \sqrt{0{,}32 \cdot (3+5) / (3 \cdot 5)} = 0{,}41 \text{ ц/га}.$$

Далі шукають найменшу істотну різницю для порівняння відзначених груп варіантів:

$$HIP_{05} = t_{05} \cdot s_d = 2{,}06 \cdot 0{,}36 = 0{,}74 \text{ ц/га};$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}78 \cdot 0{,}36 = 1{,}00 \text{ ц/га};$$
$$HIP_{05} = t_{05} \cdot s'_d = 2{,}06 \cdot 0{,}38 = 0{,}78 \text{ ц/га};$$
$$HIP_{01} = t_{01} \cdot s'_d = 2{,}78 \cdot 0{,}38 = 1{,}06 \text{ ц/га};$$
$$HIP_{05} = t_{05} \cdot s''_d = 2{,}06 \cdot 0{,}41 = 0{,}84 \text{ ц/га};$$
$$HIP_{01} = t_{01} \cdot s''_d = 2{,}78 \cdot 0{,}41 = 1{,}14 \text{ ц/га}.$$

Показники t_{05} = 2,06 і t_{01} = 2,78 знаходять за дод. Б для 25 ступенів свободи дисперсії помилок. Результати досліду і його статистичного аналізу під час порівняння досліджуваних варіантів з контролем доцільно подати у вигляді табл. 63.

Таблиця 63

Розподіл показників урожайності соняшнику за статистичними групами стосовно до контрольного варіанта

Варіант	Урожайність, ц/га	Відхилення від контролю	*HIP*		Група для	
			05	01	HIP_{05}	HIP_{01}
1	25,8	3,4	0,74	1,00	I	I
2 (*St*)	22,4	–	–	–	(*St*)	(*St*)
3	27,1	4,7	0,74	1,00	I	I
4	23,6	1,2	0,78	1,06	I	I
5	21,7	–0,7	0,74	1,00	II	II
6	23,3	0,9	0,84	1,14	I	II
7	22,8	0,4	0,74	1,00	II	II
8	20,1	–2,3	0,74	1,00	III	III

Висновок. Урожайність гібридів 1, 3 і 4 була значно вищою порівняно з контролем досліду (гібрид 2) за обох рівнів імовірності. Урожайність гібрида 6 була вищою, ніж гібрида 2 з імовірністю 95 %, однак за більш точної оцінки (ймовірність 99 %) між цими варіантами різниці не було (II група показників). Гібрид 8 за рівнем зернової продуктивності значно поступався контрольному варіанту за обох рівнів імовірності (III група).

Таким чином, характерними особливостями статистичної обробки даних польового досліду з вибракуваними ділянками є:
- необхідність відновлення вибракуваних показників;
- зменшення кількості ступенів свободи дисперсії помилок на кількість вибракуваних показників;
- необхідність урахування кількості фактичних показників під час розрахунків помилки різниці середніх і найменшої істотної різниці для потрібних рівнів імовірності.

7.2. Обробка дослідів з багаторічними культурами

Під час проведення статистичних розрахунків дослідів з багаторічними культурами, які формують урожай кілька років поспіль, особливу увагу слід акцентувати на висновках обробки даних за весь період проведення досліду. Спочатку проводять дисперсійний аналіз одержаних даних конкретно за кожний рік, далі аналізують сумарний показник за період проведення досліду. Наведемо в конкретному прикладі техніку розрахунків даних досліду з багаторічними культурами.

Приклад. У дворічному польовому досліді вивчали вплив сортових особливостей на врожайність конюшини (табл. 64). Проведемо обробку показників урожайності методом дисперсійного аналізу.

Формуючи таблицю, слід передбачати додаткові рядки і стовпці для виділення сум показників за повтореннями і варіантами за кожний досліджуваний рік і в цілому по досліду.

Розв'язання. Спочатку розраховують суми показників за повтореннями і варіантами, окремо для кожного року і в сумі за роками. Правильність розрахунків перевіряють за відношенням $\sum X = \sum V = \sum P$. Далі для кожного року та в сумі за роками знаходять корегуючий фактор C, суму квадратів загального варіювання показників урожайності C_Y, варіювання повторень C_P, варіантів C_V і помилок C_Z.

Розрахунки показників для 2011 р.:

$$N = l \cdot n = 5 \cdot 4 = 20;\ C = (\sum X)^2 / N = 89{,}14^2 / 20 = 397{,}30;$$
$$C_Y = \sum X^2 - C = (4{,}10^2 + 4{,}33^2 + ... + 4{,}19^2) - 397{,}30 = 3{,}63$$
за $(N - 1) = (20 - 1) = 19$ ступенів свободи;

Таблиця 64

Урожайність рослинної біомаси конюшини, т/га

Рік	Варіант (сорт)	Урожайність за повтореннями				Сума, *V*	Середнє
		I	II	III	IV		
2011	1(контроль)	4,10	4,33	4,24	4,37	17,04	4,26
	2	4,49	4,62	4,30	4,59	18,00	4,50
	3	5,18	5,50	4,95	5,24	20,87	5,22
	4	4,06	4,11	3,96	4,16	16,29	4,07
	5	4,22	4,06	4,47	4,19	16,94	4,24
Суми *P*		22,05	22,62	21,92	22,55	$\sum x$ = 89,14	$\bar{x}_{ср}$ = 4,46
2012	1(контроль)	5,14	5,66	4,82	5,37	20,99	5,25
	2	5,43	5,88	5,47	5,54	22,32	5,58
	3	6,34	6,79	6,19	6,22	25,54	6,39
	4	4,78	4,92	4,64	4,80	19,14	4,79
	5	5,35	5,70	5,20	5,48	21,73	5,43
Суми *P*		27,04	28,95	26,32	27,41	$\sum x$ = 109,72	$\bar{x}_{ср}$ = 5,49
Разом за два роки	1(контроль)	9,24	9,99	9,06	9,74	38,03	9,51
	2	9,92	10,50	9,77	10,13	40,32	10,08
	3	11,52	12,29	11,14	11,46	46,41	11,60
	4	8,84	9,03	8,60	8,96	35,43	8,86
	5	9,57	9,76	9,67	9,67	38,67	9,67
Суми *P*		49,09	51,57	48,24	49,96	$\sum x$ = 198,86	$\bar{x}_{ср}$ = 9,94

$$C_P = \sum X_P^2 / l - C = (22{,}05^2 + 22{,}62^2 + ... + 22{,}55^2) / 5 - 397{,}30 =$$
$$= (486{,}2 + 511{,}7 + 480{,}5 + 508{,}5) / 5 - 397{,}30 = 0{,}08$$
за $(n - 1) = (4 - 1) = 3$ ступенів свободи;
$$C_V = \sum X_V^2 / n - C = (17{,}04^2 + 18{,}00^2 + ... + 16{,}94^2) / 4 - 397{,}30 = 3{,}27$$
за $(l - 1) = (5 - 1) = 4$ ступенів свободи;
$$C_Z = C_Y - C_P - C_V = 3{,}63 - 0{,}08 - 3{,}27 = 0{,}29$$
за $(N - 1) - (1 - 1) - (n - 1) = 19 - 4 - 3 = 12$ або
$(l - 1)(n - 1) = 3 \cdot 4 = 12$ ступенів свободи.

Розрахунки показників для 2012 р.:
$$N = l \cdot n = 5 \cdot 4 = 20;$$
$$C = (\sum X)^2 / N = 109{,}72^2 / 20 = 601{,}92;$$
$$C_Y = \sum X^2 - C = (5{,}14^2 + 5{,}66^2 + ... + 5{,}48^2) - 601{,}92 = 6{,}39$$
за $(N - 1) = (20 - 1) = 19$ ступенів свободи;

$$C_P = \sum X_P^2 / l - C = (27{,}04^2 + 28{,}95^2 + ... + 27{,}41^2) / 5 - 601{,}92 = 0{,}74$$
за $(n - 1) = (4 - 1) = 3$ ступенів свободи;
$$C_V = \sum X_V^2 / n - C = (20{,}99^2 + 22{,}32^2 + ... + 21{,}73^2) / 4 - 601{,}92 = 5{,}48$$
за $(l - 1) = (5 - 1) = 4$ ступенів свободи;
$$C_Z = C_Y - C_P - C_V = 6{,}39 - 0{,}74 - 5{,}48 = 0{,}17$$
за $(N - 1) - (l - 1) - (n - 1) = 19 - 4 - 3 = 12$ або
$(l - 1)(n - 1) = 3 \cdot 4 = 12$ ступенів свободи.

Разом за два роки:
$$N = l \cdot n = 5 \cdot 4 = 20;$$
$$C = (\sum X)^2 / N = 198{,}86^2 / 20 = 1977{,}26;$$
$$C_Y = \sum X^2 - C = (9{,}92^2 + 10{,}50^2 + ... + 9{,}67^2) - 1977{,}26 = 18{,}57$$
за $(N - 1) = (20 - 1) = 19$ ступенів свободи;
$$C_P = \sum X_P^2 / l - C = (49{,}09^2 + 51{,}57^2 + ... + 49{,}96^2) / 5 - 1977{,}26 = 1{,}22$$
за $(n - 1) = (4 - 1) = 3$ ступенів свободи;
$$C_V = \sum X_V^2 / n - C = (38{,}03^2 + 40{,}32^2 + ... + 38{,}67^2) / 4 - 1977{,}26 = 9{,}37$$
за $(l - 1) = (5 - 1) = 4$ ступенів свободи;
$$C_Z = C_Y - C_P - C_V = 18{,}57 - 1{,}22 - 9{,}37 = 7{,}98$$
за $(N - 1) - (l - 1) - (n - 1) = 19 - 4 - 3 = 12$ або
$(l - 1)(n - 1) = 3 \cdot 4 = 12$ ступенів свободи.

Далі складають таблицю дисперсійного аналізу (табл. 65), до якої заносять розраховані дані за кожний рік окремо та в сумі за два роки. Розраховують $F_{факт.}$ і перевіряють нульову гіпотезу H_0.

Теоретичне значення F_{05} і F_{01} знаходять за дод. В для 4 ступенів свободи дисперсії варіантів (чисельник) і 12 ступенів свободи дисперсії помилок (знаменник).

$F_{факт.}$ як за роками досліджень, так і в сумі перевищує $F_{теор.05}$, отже, нульова гіпотеза H_0 спростовується. За більш точного підходу (ймовірність 99 %) $F_{факт.}$, розрахований для суми врожайності за два роки, був нижчий ніж $F_{теор.01}$, що свідчить про відсутність істотної різниці між варіантами досліду ($F_{факт.}$ = 3,49; $F_{теор.01}$ = 5,41). На наступному етапі проводять оцінку істотності між середніми показниками врожайності за варіантами досліду.

Визначення для даних 2011 р.:
$$s_{\bar{x}} = \sqrt{s^2 / n} = 0{,}07 \text{ т/га}; \; s_d = \sqrt{2s^2 / n} = 0{,}10 \text{ т/га};$$
$$НІР_{05} = t_{05} \cdot s_d = 0{,}22 \text{ т/га}; \; НІР_{01} = t_{01} \cdot s_d = 0{,}31 \text{ т/га}.$$

Таблиця 65

Результати дисперсійного аналізу даних урожайності конюшини за 2011, 2012 рр., т/га

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$ 05	$F_{теор.}$ 01
Дані 2011 р.						
Загальна, C_Y	3,63	19	–	–	–	–
Повторень, C_P	0,08	3	–	–	–	–
Варіантів, C_V	3,27	4	0,82	41,0	3,26	5,41
Помилок, C_Z	0,29	12	0,02	–	–	–
Дані 2012 р.						
Загальна, C_Y	6,39	19	–	–	–	–
Повторень, C_P	0,74	3	–	–	–	–
Варіантів, C_V	5,48	4	1,37	97,9	3,26	5,41
Помилок, C_Z	0,17	12	0,014	–	–	–
Разом за два роки						
Загальна, C_Y	18,57	19	–	–	–	–
Повторень, C_P	1,22	3	–	–	–	–
Варіантів, C_V	9,37	4	2,34	3,49	3,26	5,41
Помилок, C_Z	7,98	12	0,67	–	–	–

Визначення для даних 2012 р.:

$$s_{\overline{x}} = \sqrt{s^2/n} = 0{,}06 \text{ т/га}; \; s_d = \sqrt{2s^2/n} = 0{,}08 \text{ т/га};$$
$$HIP_{05} = t_{05} \cdot s_d = 0{,}17 \text{ т/га}; \; HIP_{01} = t_{01} \cdot s_d = 0{,}25 \text{ т/га}.$$

Визначення суми показників за 2011–2012 рр.:

$$s_{\overline{x}} = \sqrt{s^2/n} = 0{,}41 \text{ т/га}; \; s_d = \sqrt{2s^2/n} = 0{,}58 \text{ т/га};$$
$$HIP_{05} = t_{05} \cdot s_d = 1{,}26 \text{ т/га}; \; HIP_{05} = t_{05} \cdot s_d = 1{,}77 \text{ т/га}.$$

Показники t_{05} = 2,18 і t_{01} = 3,06 знаходять за дод. Б для 12 ступенів свободи дисперсії помилок. Зведені дані статистичного аналізу в сумі за два роки та окремо за кожний рік подають у вигляді табл. 66.

Висновок. Найбільшу врожайність сирої біомаси як у сумі за два роки, так і безпосередньо за роками досліджень забезпечував варіант 3, решта сортів були на одному рівні зі стандартом або поступалися йому за врожайністю (варіант 4).

Таблиця 66

Розподіл показників урожайності біомаси конюшини за статистичними групами стосовно до контролю досліду

Рік	Сорт	Урожайність, т/га	Різниця з контролем	НІР 05	НІР 01	Група для НІР 05	Група для НІР 01
2011	1 (*St*)	4,26	–	0,22	0,31	–	–
	2	4,50	0,24			I	II
	3	5,22	0,96			I	I
	4	4,07	–0,19			II	II
	5	4,24	–0,02			II	II
2012	1 (*St*)	5,25	–	0,17	0,25	–	–
	2	5,58	0,33			I	I
	3	6,39	1,14			I	I
	4	4,79	–0,46			III	III
	5	5,43	0,18			I	II
Разом за два роки	1 (*St*)	9,51	–	1,26	1,77	–	–
	2	10,08	0,57			II	II
	3	11,60	2,09			I	I
	4	8,86	–0,65			II	II
	5	9,67	0,16			II	II

7.3. Обробка дослідів, проведених стандартними методами

Складання таблиці і розрахунок середніх показників для дослідів, поставлених стандартними методами, значно відрізняється від розрахунків середніх показників у дослідах, поставлених методом рендомізованих повторень. За стандартного методу є кілька шляхів приведення досліджуваних показників варіантів до середньої родючості грунту за показниками контрольних варіантів.

Перший спосіб. За цього способу показники досліджуваних варіантів порівнюють із середнім арифметичним значенням двох найближчих стандартів. Таким чином проводять порівняння стандарту розташованого через один або два досліджувані варіанти, – це, відповідно, ямб- і дактиль-методи.

Другий спосіб. Якщо стандартні варіанти розташовують через 2, 3 і більше дослідних варіантів, то варіанти досліду порівнюють з інтерпольованим контролем (стандартом).

Порівняння дослідних варіантів з найближче розташованим контролем, як правило, дає більшу помилку, ніж їхнє порівняння із середнім арифметичним показником двох найближчих стандартів або зі скалькульованим контролем, які об'єктивніше відображають вихідну родючість дослідних ділянок і придатніші для проведення порівнянь. Це пояснюється тим, що в основу розрахунків скалькульованого і середньоарифметичного контролю покладено не один, а два поділянкових показники.

Характерною особливістю статистичних розрахунків результатів дослідів, поставлених стандартними методами, є також те, що за цих методів не можна порівнювати між собою досліджувані варіанти, тому що нерідко вони досить віддалені один від одного і розміщені здебільшого на ділянках з різною родючістю ґрунту. Водночас порівняння ділянок між собою можна проводити через розрахований середній показник стандарту.

Наведемо приклад статистичних розрахунків дисперсійного аналізу досліду поставленого стандартним методом. У досліді із вивчення 14-ти сортів пшениці озимої, поставленому в один ярус стандартним дактиль-методом, одержано такі показники вмісту білка в зерні (табл. 67).

Розрахунки проводять у такому порядку.

1-й етап. Визначають різницю між вмістом білка досліджуваних сортів і середньоарифметичним показником двох сусідніх контролів. Так, різниця між вмістом білка у другому варіанті першого повторення і середньоарифметичним показником сусідніх контролів становила 14,8 – (13,0 + 13,4) / 2 = 1,60; третього варіанта першого повторення 15,3 – (13,0 – 13,4) / 2 = 2,1; четвертого варіанта першого повторення – – 12,8 – (13,4 – 13,2) / 2 = –0,5 тощо.

Під час визначення середнього показника вмісту білка для сортів, що розміщені на стиках повторень (у наведеному прикладі варіанти 2 і 3, розташовані у двох і трьох повтореннях), ураховують фактичне розміщення стандартних варіантів у досліді.

2-й етап. Розраховують середній показник вмісту білка стандартного варіанта у досліді $\bar{x}_{st}$ = (13,0 + 13,4 +...+ 13,3) / 22 = 13,61. Після цього визначають суми відхилень за варіантами V, повтореннями P, загальну суму всіх різниць d, середні різниці для кожного сорту. Правильність розрахунків перевіряють за відношенням:

Таблиця 67

Фактичні показники вмісту білка в зерні пшениці озимої та приведення їх до середнього стандарту в досліді, %
(дослідне поле ХНАУ ім. В. В. Докучаєва, 2012 р.)

Варіант досліду	Вміст білка за повтореннями			Середній вміст білка	Відхилення від середнього показника стандартів					Вміст білка, приведений до середнього стандарту
	I	II	III		повторення I	повторення II	повторення III	сума, V	середні	
St (1)	13,0	–	–	–	–	–	–	–	–	–
2	14,8	14,3	14,6	14,57	1,60	0,65	1,00	3,25	1,08	14,69
3	15,3	15,9	15,7	15,63	2,10	2,25	2,10	6,45	2,15	15,76
St (1)	13,4	14,1	13,7	13,73	–	–	–	–	–	–
4	12,8	12,8	12,5	12,70	–0,50	–1,10	–1,05	–2,65	–0,88	12,73
5	15,3	15,9	15,7	15,63	2,00	2,00	2,15	6,15	2,50	15,66
St (1)	13,2	13,7	13,4	13,43	–	–	–	–	–	–
6	13,3	13,1	13,6	13,33	–0,20	–0,80	0,05	–0,95	–0,32	13,29
7	13,5	13,8	13,6	13,63	0,00	–0,10	0,05	–0,05	–0,02	13,59
St (1)	13,8	14,1	13,7	13,87	–	–	–	–	–	–
8	4,2	14,7	14,6	14,50	0,60	0,70	0,90	2,20	0,73	14,34
9	16,1	16,5	16,6	16,40	2,50	2,50	2,90	7,90	2,63	16,24
St (1)	13,4	13,9	13,7	13,67	–	–	–	–	–	–
10	13,7	14,1	13,9	13,90	0,30	0,15	0,25	0,70	0,23	13,84
11	13,2	13,5	13,2	13,30	–0,20	–0,45	–0,45	–1,10	–0,37	13,24
St (1)	13,4	14,0	13,6	13,67	–	–	–	–	–	–
12	15,5	15,9	15,4	15,60	1,95	1,95	1,75	5,65	1,88	15,49
13	13,1	13,7	13,3	13,37	–0,45	–0,25	–0,35	–1,05	–0,35	13,26
St (1)	13,7	13,9	13,7	13,77	–	–	–	–	–	–
14	16,3	16,7	16,4	16,47	2,85	3,00	2,90	8,75	2,92	16,53
15	14,0	14,2	13,6	13,93	0,55	0,50	0,10	1,15	0,38	13,99
St (1)	13,2	13,5	13,3	13,33	–	–	–	–	–	–
Суми P					13,10	11,00	12,30	36,40	–	–

$$\sum P = \sum V = \sum d.$$

3-й етап. На наступному етапі фактичні показники вмісту білка підводять до середнього показника вмісту білка стандарту. Для цього до середнього показника стандартного варіанта (13,61 %) додають середню різницю кожного варіанта (враховуючи знак різниці).

Наприклад, приведений до стандартного варіанта показник вмісту білка другого сорту становить 13,61 + 1,08 = 14,69 %, третього і четвертого сортів – відповідно 13,61 + 2,15 = 15,76 % і 13,61 – 0,88 = = 12,73 % і под.

4-й етап. Для оцінки результатів за *F*-критерієм визначають суми квадратів відхилень. Для цього використовують відхилення від середнього стандарту. Розрахунки проводять таким чином.

Визначають загальну кількість ділянок варіантів досліду $N = 42$. Далі розраховують коригуючий фактор C, загальну суму квадратів відхилень різниць C_y, суму квадратів відхилень повторень C_p, варіантів C_v, і помилок C_z.

$$C = (\sum X_d)^2 / N = 36{,}4^2 / 42 = 31{,}55;$$
$$C_Y = \sum X_d^2 - C = (1{,}6^2 + 2{,}1^2 + ... + 0{,}1^2) - 31{,}6 = 62{,}5$$
за $(N - 1) = (42 - 1) = 41$ ступеня свободи;
$$C_P = \sum X_P^2 / l - C = (13{,}1^2 + 11{,}0^2 + ... + 13{,}3^2) / 14 - 31{,}55 = 0{,}16$$
за $(n - 1) = (3 - 1) = 2$ ступенів свободи;
$$C_V = \sum X_V^2 / n - C = (3{,}25^2 + 6{,}45^2 + ... + 1{,}15^2) / 4 - 31{,}55 = 37{,}89$$
за $(l - 1) = (14 - 1) = 13$ ступенів свободи;
$$C_Z = C_Y - C_P - C_V = 62{,}50 - 0{,}16 - 37{,}89 = 24{,}45$$
за $(N - 1) - (l - 1) - (n - 1) = 41 - 13 - 2 = 26$ або
$(l - 1)(n - 1) = 13 \cdot 2 = 26$ ступенів свободи.

Одержані результати переносять до таблиці дисперсійного аналізу (табл. 68). Показники $F_{05} = 2{,}15$ і $F_{01} = 2{,}96$ знаходять за дод. В для 13 ступенів свободи дисперсії варіантів (чисельник) і 26 ступенів свободи дисперсії помилок (знаменник).

5-й етап. Розраховують помилку середньої різниці та *НІР* для 05 або 01 %-го рівня ймовірності. Оскільки для проведення аналізу використовували не фактичні показники, а їхні відхилення від стан-

Таблиця 68

Таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна, C_Y	62,5	41	–	–	–	–
Повторень, C_P	0,16	2	–	–	–	–
Варіантів, C_V	37,89	13	2,92	3,11	2,15	2,96
Помилок, C_Z	24,45	26	0,94	–	–	–

дарту (різниця – d), то відразу визначають помилку середньої різниці s_d і далі розраховують $НІР_{05}$ та $НІР_{01}$.

$$s_d = \sqrt{2s^2/n} = \sqrt{2 \cdot 0{,}94 / 3} = 0{,}56\ \%;$$
$$НІР_{05} = t_{05} \cdot s_d = 2{,}06 \cdot 0{,}56 = 1{,}14\ \%;$$
$$НІР_{01} = t_{01} \cdot s_d = 2{,}78 \cdot 0{,}56 = 1{,}56\ \%.$$

Теоретичне значення t_{05} і t_{01} знаходять за дод. Б для 26 ступенів свободи дисперсії помилок. Результати аналізу доцільно подати у вигляді підсумкової табл. 69.

Висновок. У цьому досліді сорти 3, 5, 9, 12 і 14 за вмістом білка істотно перевищували показник контрольного варіанта за обох рівнів імовірності, решта варіантів були рівноцінні з контролем (II група).

Таблиця 69

Вміст білка у зерні різних сортів пшениці твердої ярої та їхній розподіл за статистичними групами стосовно до контролю досліду, %

Варіант досліду	Вміст білка, %	Відхилення від контролю	Група для	
			$НІР_{05}$	$НІР_{01}$
1 (контроль)	13,61	–	контроль	контроль
2	14,69	1,08	II	II
3	15,76	2,15	I	I
4	12,73	–0,88	II	II
5	15,66	2,05	I	I
6	13,29	–0,32	II	II
7	13,59	–0,02	II	II
8	14,34	0,73	II	II
9	16,24	2,63	I	I
10	13,84	0,23	II	II
11	13,24	–0,37	II	II
12	15,49	1,88	I	I
13	13,26	–0,35	II	II
14	16,53	2,92	I	I
15	13,99	0,38	II	II
$НІР_{05}$	–	1,14	–	–
$НІР_{01}$	–	1,56	–	–

7.4. Латинський квадрат і прямокутник

Згідно із цим методом, рендомізація додатково обмежується групуванням варіантів не лише за рядами, але й за стовпцями. Таке розміщення варіантів дозволяє визначити мінливість, пов'язану з ефектом стовпців і рядів, знизити помилку досліду. За цього методу кожний ряд і стовбець містить повний набір досліджуваних варіантів. Він забезпечує більшу точність порівняння варіантів, ніж метод рендомізованих повторень, лише у випадку значної мінливості за стовпцями. Також латинський квадрат може мати перевагу тоді, коли ділянки розміщені на одній лінії (одноярусне розташування).

Якщо в латинському квадраті вибраковується певна ділянка, то відновлений показник визначають за формулою:

$$x' = \frac{n \cdot (p + c + v) - 2\sum x}{(n-1)(n-2)},$$

де n – кількість рядів, стовпців і варіантів; p, c і v – суми даних ряду, стовпця і варіанта де розміщений вибракуваний показник; $\sum x$ – загальна сума показників досліду.

Наведемо приклад статистичних розрахунків польового досліду, поставленого методом латинського квадрата. У досліді вивчали вплив сортоособливостей на висоту рослин тритикале ярого. Загальна кількість ділянок у досліді – 36 (6 варіантів, 6 повторень) (табл. 70).

Таблиця 70

Висота різних сортів тритикале ярого у фазу повної стиглості, см

Ряди	Стовпці (повторення)						Суми за	
	1	2	3	4	5	6	рядами	стовпцями
A_1	A_3(63)	A_4(91)	A_3(67)	A_6(65)	A_4(90)	A_5(78)	454	A_1 – 438
A_2	A_1(73)	A_6(71)	A_5(76)	A_3(67)	A_2(75)	A_1(74)	436	A_2 – 477
A_3	A_4(87)	A_2(80)	A_1(69)	A_4(89)	A_5(80)	A_6(69)	474	A_3 – 385
A_4	A_5(76)	A_5(74)	A_6(71)	A_1(72)	A_6(74)	A_2(81)	448	A_4 – 538
A_5	A_6(69)	A_3(61)	A_2(77)	A_2(84)	A_3(64)	A_4(88)	443	A_5 – 458
A_6	A_2(80)	A_1(76)	A_4(93)	A_5(74)	A_1(74)	A_3(63)	460	A_6 – 419
Суми C	448	453	453	451	457	453	2715	$\sum x = 2715$

1-й етап. Спочатку визначають суми показників за стовпцями, рядами і варіантами. Правильність розрахунків перевіряють за відношенням: $\sum P = \sum C = \sum V = \sum X = 2715$.

Далі розраховують коригуючий чинник C, загальну суму квадратів відхилень C_y, суму квадратів відхилень – варіантів C_v, рядів C_p, стовпців C_c і помилок C_z:

$$N = l_p \cdot l_c = 6 \cdot 6 = 36;$$

$$C = (\sum X)^2 / N = 2715^2 / 36 = 204756{,}3;$$

$$C_{Y\ (\text{загальна})} = \sum X^2 - C = (63^2 + 73^2 + ... + 73^2) - 204756{,}3 = 2506{,}7$$

за $(N - 1) = (36 - 1) = 35$ ступенів свободи;

$$C_{C\ (\text{стовпців})} = \sum X_C^2 / l_p - C = (448^2 + 453^2 + ... + 453^2) / 6 - 204756{,}3 = 7{,}2$$

за $(l_C - 1) = (6 - 1) = 5$ ступенів свободи;

$$C_{P\ (\text{рядів})} = \sum X_P^2 / l_c - C = (454^2 + 436^2 + ... + 460^2) / 6 - 204756{,}3 = 150{,}5$$

за $(l_p - 1) = (6 - 1) = 5$ ступенів свободи;

$$C_{V\ (\text{варіантів})} = \sum X_V^2 / n - C = (438^2 + 477^2 + ... + 419^2) / 6 - 204756{,}3 = 2304{,}9$$

за $(n - 1) = (6 - 1) = 5$ ступенів свободи;

$$C_{Z\ (\text{помилок})} = C_Y - C_C - C_P - C_V = 2506{,}7 - 7{,}2 - 150{,}5 - 2304{,}9 = 44{,}1$$

за $35 - 5 \cdot 3 = 20$ ступенів свободи.

Одержані дані переносять до таблиці дисперсійного аналізу і визначають критерій $F_{\text{факт.}}$ (табл. 71).

Таблиця 71

Результати дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{\text{факт.}}$	$F_{\text{теор.}}$	
					05	01
Загальна, C_Y	2506,7	35	–	–	–	–
Стовпців, C_C	7,2	5	–	–	–	–
Рядів, C_P	150,5	5	–	–	–	–
Варіантів, C_V	2304,9	5	461,0	209,5	2,71	4,10
Помилок, C_Z	44,1	20	2,2	–	–	–

Для 5 ступенів свободи дисперсії варіантів (чисельник) і 20 ступенів свободи дисперсії помилок (знаменник) значення F_{05} і F_{01} відповідно становить 2,71 і 4,10 (див. дод. В). Таким чином, $F_{\text{факт.}} > F_{05}$ і F_{01}, тобто нульова гіпотеза спростовується за обох рівнів імовірності, що дає підставу стверджувати про істотну різницю між варіантами досліду.

2-й етап. Розраховують помилку вибіркової середньої ($s_{\overline{x}}$), помилку різниці середніх (s_d) і найменшу істотну різницю з рівнем імовірності 95 і 99 %:

$$s_{\overline{x}} = \sqrt{s^2 / n} = \sqrt{2{,}2 / 6} = 0{,}61 \text{ см};$$
$$s_d = \sqrt{2s^2 / n} = \sqrt{2 \cdot 2{,}2 / 4} = 0{,}86 \text{ см};$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}09 \cdot 0{,}86 = 1{,}80 \text{ см};$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}85 \cdot 0{,}86 = 2{,}46 \text{ см}.$$

Таким чином, висота рослин усіх сортів істотно відрізнялася від контрольного варіанта. Середні показники висоти і результати їхньої статистичної обробки доцільно подати у вигляді табл. 72.

Таблиця 72

Висота рослин тритикале ярого різних сортів та її статистичний аналіз

Сорт	Висота рослин, см	Порівняно зі стандартом	Група показників стосовно до контролю
A_1 контроль	73	–	контроль
A_2	80	7,0	I
A_3	64	–9,0	III
A_4	90	17,0	I
A_5	76	3,0	I
A_6	70	–3,0	III

Наведемо приклад статистичного аналізу результатів польового досліду, поставленого методом латинського прямокутника.

Приклад. В однофакторному польовому досліді вивчено вплив сортових особливостей соняшнику на його врожайність. Загальна кількість варіантів досліду – 12. Повторність – чотириразова. Площа облікової ділянки – 30 м². Результати врожайності наведено в табл. 73.

1-й етап. Підраховують суми показників за рядами, стовпцями, варіантами і визначають загальну суму показників досліду. Правильність розрахунків перевіряють за відношення: $\sum C = \sum V = \sum X = 93{,}22$. Наступним кроком є визначення коригуючого чинника C, суми квадратів відхилень загальної C_Y, стовпців C_C, рядів C_P, варіантів C_V і помилок C_Z:

Таблиця 73

Урожайність гібридів соняшнику, т/га

Ряди (Р)	Стовпці (C)				Суми за рядами, P	Суми за варіантами, V
	1	2	3	4		
1	A_4 (1,84)	A_2 (1,91)	A_{11}(1,96)	A_6 (1,67)	23,06	A_1 (7,35)
	A_{12}(2,03)	A_5(1,96)	A_1 (1,75)	A_3 (2,21)		A_2 (7,64)
	A_7(2,07)	A_9(1,83)	A_{10}(1,87)	A_8 (1,86)		A_3 (8,94)
2	A_1 (1,90)	A_3 (2,30)	A_5 (2,07)	A_2 (1,97)	23,84	A_4 (7,41)
	A_6(1,76)	A_8 (1,86)	A_4 (1,88)	A_7 (2,11)		A_5 (8,09)
	A_{11} (1,94)	A_{10} (1,92)	A_{12}(2,16)	A_9 (1,97)		A_6 (6,78)
3	A_8 (1,77)	A_7(2,01)	A_3 (2,30)	A_1 (1,83)	23,02	A_7 (8,33)
	A_{10} (1,90)	A_4 (1,83)	A_6 (1,64)	A_{12}(2,05)		A_8 (7,38)
	A_2 (1,94)	A_{11} (1,80)	A_9 (1,93)	A_5 (2,02)		A_9 (7,71)
4	A_9 (1,88)	A_{12} (2,11)	A_8 (1,89)	A_{10}(1,84)	23,30	A_{10} (7,53)
	A_3 (2,13)	A_1 (1,87)	A_2 (1,82)	A_4 (1,86)		A_{11} (7,71)
	A_5 (2,04)	A_6(1,71)	A_7(2,14)	A_{11}(2,01)		A_{12} (8,35)
Суми P	23,20	23,21	23,41	23,40	$\sum X = 93{,}22$	

$$N = l_p \cdot l_c \cdot l_g = 4 \cdot 4 \cdot 3 = 48;$$
$$C = (\sum X)^2 / N = 93{,}22^2 / 48 = 8689{,}97 / 48 = 181{,}04;$$
$$C_{Y(\text{загальна})} = \sum X^2 - C = (1{,}84^2 + 1{,}91^2 + ... + 2{,}01^2) - 181{,}04 = 1{,}033$$
за $(N - 1) = (48 - 1) = 47$ ступенів свободи;
$$C_{C(\text{стовпців})} = \sum X_C^2 / l - C = (23{,}2^2 + 23{,}21^2 + ... + 23{,}40^2) / 12 - 181{,}04 = 0{,}004$$
за $(l_P - 1) = (4 - 1) = 3$ ступенів свободи;
$$C_{P(\text{рядів})} = \sum X_P^2 / l - C = (23{,}06^2 + 23{,}84^2 + ... + 23{,}30^2) / 12 - 181{,}04 = 0{,}037$$
за $(l_C - 1) = (4 - 1) = 3$ ступенів свободи;
$$C_{V(\text{варіантів})} = \sum X_V^2 / n - C = (7{,}35^2 + 7{,}64^2 + ... + 8{,}35^2) / 4 - 181{,}04 = 0{,}912$$
за $(l - 1) = (12 - 1) = 11$ ступенів свободи;
$$C_{Z(\text{помилок})} = C_Y - C_C - C_P - C_V = 1{,}033 - 0{,}004 - 0{,}037 - 0{,}912 = 0{,}08$$
за $(N - 1) - (l - 1) - (l_p - 1) - (l_c - 1) = 30$ ступенів свободи.

2-й етап. Одержані результати переносять до підсумкової таблиці дисперсійного аналізу і розраховують фактичний критерій F (табл. 74). Критерій F визначають відношенням середнього квадрата дисперсії варіантів до середнього квадрата дисперсії помилок.

Критерій $F_{\text{теор.}}$ для обох рівнів імовірності знаходять за дод. В для 11 ступенів свободи дисперсії помилок. $F_{\text{факт.}}$ перевищує $F_{\text{теор}}$. за

Таблиця 74

Таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$ 05	$F_{теор.}$ 01
Загальна, C_Y	1,003	47	–	–	–	–
Рядів, C_P	0,037	3	–	–	–	–
Стовпців, C_C	0,004	3	–	–	–	–
Варіантів, C_V	0,912	11	0,083	27,67	2,12	2,90
Помилок, C_Z	0,080	30	0,003	–	–	–

обох рівнів імовірності, отже, нульова гіпотеза спростовується і між варіантами досліду є істотна різниця.

3-й етап. Далі проводять оцінку істотності часткових порівнянь досліджуваних варіантів. Для цього знаходять помилку середньої $s_{\overline{x}}$, помилку різниці середніх s_d і найменшу істотну різницю з різними рівнями ймовірності:

$$s_{\overline{x}} = \sqrt{s^2 / n} = \sqrt{0{,}003 / 4} = 0{,}027 \text{ т/га};$$

$$s_d = \sqrt{2s^2 / n} = \sqrt{2 \cdot 0{,}003 / 4} = 0{,}86 \text{ т/га};$$

$$HIP_{05} = t_{05} \cdot s_d = 2{,}04 \cdot 0{,}039 = 0{,}08 \text{ т/га};$$

$$HIP_{01} = t_{01} \cdot s_d = 2{,}75 \cdot 0{,}039 = 0{,}11 \text{ см}.$$

$$HIP_{05} = t_{05} \cdot s_d \cdot 100 / \overline{x}_{cp} = 0{,}08 \cdot 100 / 1{,}94 = 4{,}1\ \%;$$

$$HIP_{01} = t_{01} \cdot s_d \cdot 100 / \overline{x}_{cp} = 0{,}11 \cdot 100 / 1{,}94 = 5{,}7\ \%.$$

Для покращання оцінювання варіантів результати статистичної обробки доцільно представити у вигляді таблиці (табл. 75).

Висновок. У цьому досліді сорти A_3, A_5, A_7 і A_{12} істотно перевищували врожайність контрольного варіанта (A_1) за обох рівнів ймовірності. Сорти (гібриди) A_2, A_4, A_8 і A_{10} за рівнем врожайності були на рівні зі стандартом (II група) і лише врожайність сорту A_6 була істотно меншою, ніж на контрольному варіанті (III група).

Таблиця 75

Розподіл показників урожайності гібридів соняшнику за статистичними групами відносно контролю

Варіант (сорт)	Урожайність, т/га	Відхилення від контролю, т/га	Відхилення від контролю, %	Групи для HIP_{05}	Групи для HIP_{01}
A_1	1,84	–	–	контроль	контроль
A_2	1,91	0,07	3,8	II	II
A_3	2,24	0,40	21,7	I	I
A_4	1,85	0,01	0,5	II	II
A_5	2,02	0,18	9,8	I	I
A_6	1,70	– 0,14	– 7,6	III	III
A_7	2,08	0,24	13,0	I	I
A_8	1,85	0,01	0,5	II	II
A_9	1,93	0,09	4,9	I	II
A_{10}	1,88	0,04	2,2	II	II
A_{11}	1,93	0,09	4,9	I	II
A_{12}	2,09	0,25	13,6	I	I
HIP_{05}	–	0,08	4,1	–	–
HIP_{01}	–	0,11	5,7	–	–

8. Дисперсійний аналіз даних багатофакторного польового досліду

Багатофакторний дисперсійний комплекс – це сукупність конкретних результатів досліду, які дають змогу статистично оцінити дію та взаємодію двох і більше чинників на закономірність мінливості результатів досліджень.

Ефект взаємодії чинників у загальній варіабельності результативної ознаки показує ту її частку, яка викликана різною дією одного чинника за різних градацій (варіантів) іншого. У багатофакторних польових дослідах, закладених за повною факторіальною схемою, взаємодія чинників проявляється тоді, коли ефективність досліджуваних градацій одного чинника змінюється залежно від градацій іншого. Наприклад, сорти можуть по-різному реагувати на систему живлення, а ефективність норми висіву може значно змінюватися за різних способів сівби.

Взаємодія досліджуваних чинників може бути більшою (явище синергізму) або меншою (явище антагонізму) порівняно із сумою ефектів від роздільного застосування кожного з них. У першому випадку взаємодія чинників позитивна, у другому – негативна. Коли прибавка показника від сумісного застосування чинників рівноцінна сумі прибавок від їх роздільного застосування, взаємодія чинників відсутня (явище адитивізму).

8.1. Обробка даних польового досліду, закладеного методом рендомізованих повторень

Більшість двофакторних польових досліджень проводиться з використанням методу рендомізованих повторень (блоків). Наведемо приклад статистичних розрахунків двофакторного польового досліду, поставленого цим методом.

Приклад 1. У двофакторному польовому досліді вивчали вплив трьох попередників: пшениці озимої, кукурудзи на зерно і сої (чинник A) та п'яти варіантів системи живлення: контроль (без добрив); $N_{15}P_{15}K_{15}$; $N_{45}P_{45}K_{45}$; $N_{60}P_{60}K_{60}$ і $N_{90}P_{90}K_{90}$ (чинник B) на врожайність буряків цукрових (S – 50 м²). Розміщення варіантів у досліді та рівні врожайності, одержані на всіх ділянках, показані в табл. 76. Статистичну обробку показників проводять у кілька етапів.

1-й етап. Вихідні дані переносять до таблиці, розподіляючи показники окремо за факторами. Далі підраховують суми показників за варіантами та повторенням, визначають середні показники врожайності за варіантами досліду (табл. 77). Правильність розрахунків перевіряють за відношенням $\sum X = \sum P = \sum V = 1094{,}7$.

2-й етап. Розраховують коригуючий чинник C, загальну суму квадратів відхилень C_Y, суму квадратів відхилень дисперсії повторень C_P, варіантів C_V і помилок C_Z.

$$N = l_a \cdot l_b \cdot n = 3 \cdot 5 \cdot 4 = 60;$$

$$C = (\sum X^2) / N = 1094{,}7^2 / 60 = 19972{,}8;$$

$$C_Y = \sum x^2 - C = (13{,}6^2 + 14{,}1^2 + ...+ 22{,}2^2) - 19972{,}8 =$$
$$= (184{,}96 + 198{,}81 + ... + 492{,}84) - 19972{,}8 = 592{,}5$$

Таблиця 76

Схема розміщення варіантів досліду з вивчення впливу попередників і системи живлення на врожайність буряків цукрових, т/га (дослідне поле ХНАУ ім. В. В. Докучаєва, 2012 р.)*

Урожайність за повтореннями			
I	II	III	IV
A_1B_5 – 21,7	A_2B_4 – 20,1	A_2B_4 – 20,3	A_3B_1 – 15,5
A_3B_3 – 20,3	A_1B_3 – 17,9	A_3B_5 – 22,6	A_2B_4 – 19,4
A_2B_3 – 18,3	A_3B_1 – 14,6	A_1B_3 – 17,3	A_3B_3 – 20,7
A_3B_4 – 22,4	A_2B_3 – 17,9	A_3B_4 – 22,9	A_1B_5 – 21,6
A_3B_5 – 21,9	A_1B_4 – 20,8	A_2B_5 – 20,9	A_2B_5 – 21,7
A_1B_2 – 15,3	A_3B_4 – 21,7	A_1B_4 – 19,9	A_1B_1 – 15,4
A_1B_1 – 13,6	A_1B_1 – 14,1	A_2B_1 – 11,9	A_3B_2 – 17,8
A_2B_5 – 21,3	A_3B_5 – 21,8	A_1B_5 – 21,3	A_1B_3 – 17,6
A_3B_1 – 15,1	A_2B_2 – 15,8	A_3B_1 – 14,8	A_3B_5 – 22,2
A_1B_4 – 20,3	A_1B_5 – 21,5	A_2B_2 – 14,7	A_2B_2 – 15,5
A_2B_1 – 12,0	A_1B_2 – 15,8	A_3B_2 – 18,3	A_3B_4 – 22,6
A_3B_2 – 17,8	A_3B_3 – 20,3	A_1B_1 – 13,9	A_1B_2 – 15,4
A_2B_2 – 15,1	A_2B_1 – 12,6	A_3B_3 – 21,1	A_1B_4 – 20,7
A_1B_3 – 17,1	A_3B_2 – 18,1	A_2B_3 – 17,5	A_2B_1 – 12,3
A_2B_4 – 19,7	A_2B_5 – 21,0	A_1B_2 – 15,7	A_2B_3 – 18,3

* A_0–A_5 – градації чинника A, B_0–B_6 – градації чинника B.

за $(N - 1) = (60 - 1) = 59$ ступенів свободи;

$$C_P \text{ (повторень)} = \sum P^2 / l_a \cdot l_b - C =$$
$$= (271{,}9^2 + 274{,}0^2 + 273{,}1^2 + 275{,}7^2) / 5 \cdot 3 - 19972{,}8 =$$
$$= (73929{,}6 + 75076{,}0 + 74583{,}6 + 76010{,}49) / 15 - 19972{,}8 = 0{,}5$$

за $(n - 1) = (4 - 1) = 3$ ступенів свободи;

$$C_V \text{ (варіантів)} = \sum V^2 / n - C = (56{,}1^2 + 62{,}2^2 + ... + 88{,}5^2) / 4 - 19972{,}8 =$$
$$= (3147{,}2 + 3868{,}8 + ... + 7832{,}25) / 4 - 19972{,}8 = 586{,}4$$

за $(l - 1) = (15 - 1) = 14$ ступенів свободи;

$$C_Z = C_Y - C_P - C_V = 592{,}5 - 586{,}4 - 0{,}5 = 5{,}6$$

за $(l - 1)(n - 1) = 14 \cdot 3 = 42$ ступені свободи.

3-й етап. У досліді загальна сума квадратів варіантів складається із суми квадратів чинників А і В та їх взаємодії. На цьому етапі проведення дисперсійного аналізу двофакторного польового досліду, поставленого методом рендомізованих блоків, визначають

Таблиця 77

Вплив попередників і варіантів системи живлення на врожайність буряків цукрових, т/га

Чинник		Урожайність за повтореннями				Сума, *V*	Середнє
A	*B*	I	II	II	IV		
1	1	13,6	14,0	13,9	14,5	56,1	14,0
	2	15,3	15,8	15,7	15,4	62,2	15,6
	3	17,1	17,9	17,3	17,6	69,9	17,5
	4	20,3	20,8	19,9	20,7	81,7	20,4
	5	21,7	21,5	21,3	21,6	86,1	21,5
2	1	12,0	12,6	11,9	12,3	48,8	12,2
	2	15,1	15,8	14,7	15,5	61,1	15,3
	3	18,3	17,9	17,5	18,2	71,9	18,0
	4	19,7	20,1	20,3	19,4	79,5	19,9
	5	21,3	21,0	20,9	21,7	84,9	21,2
3	1	15,1	14,6	14,8	15,5	60,0	15,0
	2	17,8	18,1	18,3	17,8	72,0	18,0
	3	20,3	20,3	21,1	20,7	82,4	20,6
	4	22,4	21,7	22,9	22,6	89,6	22,4
	5	21,9	21,8	22,6	22,2	88,5	22,1
Сума *P*		271,9	274,0	273,1	275,7	$\sum x = 1094,7$	$\bar{x}_{ср.} = 18,2$

суми квадратів для чинників *A* і *B* та їх взаємодії. Для цього складають додаткову таблицю до якої переносять із табл. 77 показники врожайності кожного варіанта досліду у сумі за повтореннями і розраховують суми показників урожайності за кожним чинником (табл. 78).

Таблиця 78

Урожайність буряків цукрових залежно від впливу досліджуваних чинників у сумі за повтореннями, т/га

Чинник *A* (попередник)	Чинник B (варіанти внесення добрив)					Сума, *A*
	1	2	3	4	5	
1	56,1	62,2	69,9	81,7	86,1	356,0
2	48,8	61,1	71,9	79,5	84,9	346,2
3	60,0	72,0	82,4	89,6	88,5	392,5
Сума *B*	164,9	195,3	224,2	250,8	259,5	$\sum x = 1094,7$

$$C_A = \sum x_A^2 / l_b \cdot n - C = (356,0^2 + 346,2^2 + 392,5^2) / 5 \cdot 4 - 19972,8 = 59,5$$

за $(l_a - 1) = (3 - 1) = 2$ ступенів свободи;

$$C_B = \sum x_B^2 / l_a \cdot n - C = (164,9^2 + 195,3^2 + ... + 259,5^2) / 3 \cdot 4 - 19972,8 = 513,9$$

за $(l_b - 1) = (5 - 1) = 4$ ступенів свободи;

$$C_{AB} = C_Y - C_A - C_B = 586,4 - 59,5 - 513,9 = 13,0$$

за $(l_a - 1)(l_b - 1) = (3 - 1)(5 - 1) = 8$ ступенів свободи.

Далі складаємо таблицю дисперсійного аналізу і визначаємо істотність впливу головних ефектів досліджуваних чинників та їх взаємодії за критерієм *F* (табл. 79).

Таблиця 79

Результати дисперсійного аналізу двофакторного польового досліду ($l_a = 3$, $l_b = 5$), поставленого методом рендомізованих повторень (блоків)

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна, C_Y	592,5	59	–	–	–	–
Повторень, C_P	0,5	3	–	–	–	–
Чинника, *A*	59,5	2	29,8	229,2	3,23	5,18
Чинника, *B*	513,9	4	128,5	988,5	2,61	3,83
Взаємодії, *AB*	13,0	8	1,6	12,3	2,18	2,99
Похибки, C_Z	5,6	42	0,13	–	–	–

Значення F_{05} і F_{01} знаходять у додатках, за даними ступенів свободи для дисперсії головних ефектів *A* і *B* та взаємодії *AB* (чисельник) і 42 ступенів свободи дисперсії помилок (знаменник). У цьому досліді доведено істотну взаємодію досліджуваних чинників *A* і *B* з рівнем імовірності 95 і 99 % ($F_{факт.} > F_{теор.}$).

4-й етап. Проводять оцінку істотності часткових порівнянь ефектів досліджуваних чинників.

$$s_{\bar{x}} = \sqrt{s^2/n} = \sqrt{0,13/4} = 0,18;$$

$$s_d = \sqrt{2s^2/n} = \sqrt{2 \cdot 0,13/4} = \sqrt{0,065} = 0,25;$$

$$HIP_{05} = t_{05} \cdot s_d = 2,01 \cdot 0,25 = 0,50 \text{ т/га};$$

$$HIP_{01} = t_{01} \cdot s_d = 2,28 \cdot 0,25 = 0,57 \text{ т/га}.$$

5-й етап. Проводять оцінку істотності головних ефектів і їхньої взаємодії. У цьому прикладі часткові середні базуються на чотирьох повтореннях, а середні для головного ефекту чинників *A* і *B* відповідно на $n \cdot l_b = 4 \cdot 5 = 20$ і $n \cdot l_a = 4 \cdot 4 = 16$ спостереженнях. Розраховують s_d *і* HIP_{05} (HIP_{01}):

– для головного ефекту чинника *A*:

$$s_d = \sqrt{2\,s^2/n \cdot lb} = \sqrt{2 \cdot 0{,}13\ /\ 20} = 0{,}11 \text{ т/га};$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}01 \cdot 0{,}11 = 0{,}22 \text{ т/га};$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}28 \cdot 0{,}11 = 0{,}29 \text{ т/га};$$

– для головного ефекту чинника *B* і взаємодії *AB*:

$$s_d = \sqrt{2\,s^2/n \cdot \mathrm{la}} = \sqrt{2 \cdot 0{,}13\ /\ 12} = 0{,}11 \text{ т/га};$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}01 \cdot 0{,}15 = 0{,}30 \text{ т/га};$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}68 \cdot 0{,}15 = 0{,}40 \text{ т/га}.$$

Для покращання проведення оцінки, результати статистичної обробки доцільно представити у вигляді рис. 26.

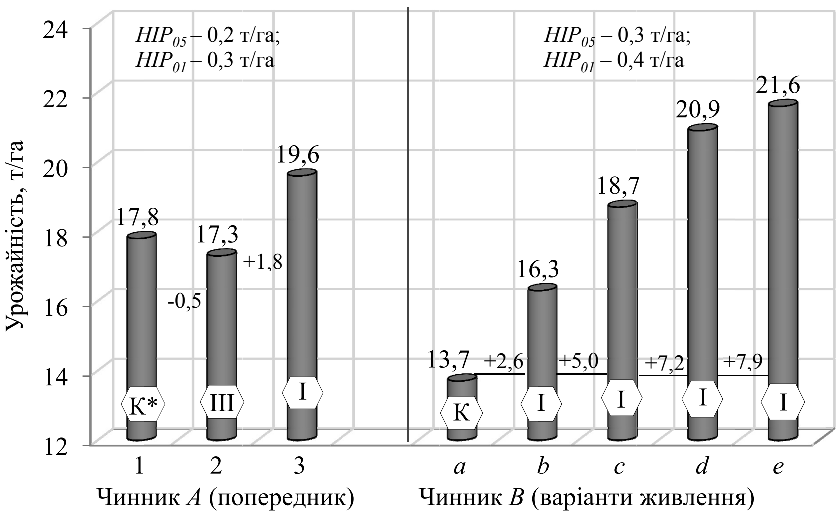


Рис. 26. Урожайність буряків цукрових у середньому за чинниками *A* (попередник) і *B* (система живлення), т/га.
Умовні позначення: * Групи показників: К – контроль; I – істотно вищі за контроль; III – істотно нижчі за контроль. Чинник А: 1 – пшениця озима; 2 – кукурудза на силос; 3 – соя. Варіанти системи живлення: *a* – контроль; *b* – $N_{15}P_{15}K_{15}$; *c* – $N_{45}P_{45}K_{45}$; *d* – $N_{60}P_{60}K_{60}$; *e* – $N_{90}P_{90}K_{90}$

Таким чином, усі варіанти системи живлення забезпечували істотне збільшення врожайності буряків цукрових з імовірністю 95 і 99 %. Після сої урожайність буряків істотно зростала – на 1,8 т/га (HIP_{05} – 0,2 т/га, HIP_{01} – 0,3 т/га), після кукурудзи на силос – навпаки, істотно зменшувалася – на 0,5 т/га порівняно з контролем.

Якщо різниця між будь-якими середніми показниками досліду перевищує 0,5 т/га (для 95 % ймовірності), то ці показники істотно відрізняються між собою.

8.2. Техніка обробки двофакторних польових дослідів, закладених методом розщеплених ділянок

Статистична обробка результатів польового досліду, поставленого методом розщеплених ділянок, принципово відрізняється від проведення статистичної обробки даних багатофакторних польових дослідів, поставлених методом рендомізованих повторень (блоків). Відмінною рисою у розрахунках багатофакторного досліду, проведеного методом розщеплених ділянок, є поділ помилки досліду C_z на складові компоненти, пов'язані з варіабельністю показників ділянок першого (помилка – Z_1), другого (помилка – Z_2) та наступних порядків. Таким чином, у багатофакторних дослідах, поставлених методом розщеплених ділянок, порівняння головних ефектів чинників та їх взаємодії проводиться за різними показниками.

Наведемо приклад проведення статистичних розрахунків результатів урожайності двофакторного польового досліду, поставленого методом розщеплених ділянок.

Приклад 1. У двофакторному досліді вивчали вплив трьох попередників (чинник *A*) – ділянки першого порядку (*S* – 300 м²) і шести норм висіву (чинник *B*) – ділянки другого порядку (*S* – 50 м²). Одержані результати врожайності зерна представлені в табл. 80. Проведемо обробку даних урожайності методом дисперсійного аналізу.

Проведення розрахунків. Дані цього двофакторного досліду з трьома варіантами чинника *A* і шістьма варіантами чинника *B*, поставленого у чотириразовій повторюваності, проводять у декілька етапів.

1-й етап. За аналогією з попередніми прикладами визначають середні показники, суми варіантів і повторень результатів урожай-

Таблиця 80

Урожайність зерна ячменю ярого залежно від впливу попередників та норм висіву, т/га

Чинник *A*	Чинник *B*, млн нас./га	Урожайність за повтореннями				Сума, *V*	Середнє
		I	II	III	IV		
Чистий пар (0)	2,5	3,21	3,13	3,32	3,27	12,93	3,23
	3,0	3,87	3,82	3,91	3,80	15,40	3,85
	3,5	4,22	4,18	4,13	4,25	16,78	4,20
	4,0	4,58	4,53	4,40	4,63	18,14	4,54
	4,5	4,73	4,88	4,65	4,81	19,07	4,77
	5,0	4,84	4,80	4,76	4,79	19,19	4,80
Кукурудза на силос (контроль) (1)	2,5	3,04	2,97	3,01	3,10	12,12	3,03
	3,0	3,56	3,48	3,52	3,60	14,16	3,54
	3,5	3,93	4,02	3,91	3,98	15,84	3,96
	4,0	4,13	4,26	4,19	4,23	16,81	4,20
	4,5	4,19	4,17	4,06	4,11	16,53	4,13
	5,0	4,03	4,12	3,93	3,98	16,06	4,02
Цукрові буряки (2)	2,5	2,85	2,73	2,81	2,88	11,27	2,82
	3,0	3,24	3,20	3,08	3,30	12,82	3,21
	3,5	3,54	3,66	3,60	3,59	14,39	3,60
	4,0	3,86	3,83	3,74	3,95	15,38	3,85
	4,5	3,90	3,81	3,90	3,84	15,45	3,86
	5,0	3,81	3,74	3,65	3,70	14,90	3,73
Сума за повтореннями *P*		69,53	69,33	68,57	69,81	$\sum x = 277{,}2$	$\bar{x} = 3{,}85$

ності представлених у таблиці вихідних даних. Правильність проведених розрахунків перевіряють за відношенням:

$$\sum x = \sum P = \sum V = 277{,}2.$$

2-й етап. Далі розраховують коригуючий чинник *C*, загальну суму квадратів C_Y, суму квадратів повторень C_P, варіантів C_V та помилок C_Z.

$$N = l_a \cdot l_b \cdot n = 3 \cdot 6 \cdot 4 = 72;$$
$$C = (\sum x^2) / N = 277{,}2^2 / 72 = 1067{,}53;$$
$$C_Y = \sum x^2 - C = (3{,}21^2 + 3{,}13^2 + ...+ 3{,}70^2) - 1067{,}53 =$$
$$= (10{,}304 + 9{,}797 + ... + 13{,}690) - 1067{,}53 = 21{,}381$$
за $(N - 1) = (72 - 1) = 71$ ступеня свободи;

$$C_{P\ (\text{повторень})} = \sum P^2 / l_a \cdot l_b - C =$$
$$= (69{,}53^2 + 69{,}33^2 + 68{,}57^2 + 69{,}81^2) / 3 \cdot 6 - 1067{,}53 =$$
$$= (4834{,}421 + 4806{,}649 + 4701{,}845 + 4873{,}436) / 18 - 1067{,}53 = 0{,}046$$
за $(n - 1) = (4 - 1) = 3$ ступенів свободи;
$$C_{V\ (\text{варіантів})} = \sum V^2 / n - C = (12{,}93^2 + 15{,}40^2 + ... + 14{,}90^2) / 4 - 1067{,}53 =$$
$$= (167{,}185 + 237{,}160 + ... + 222{,}010) / 4 - 1067{,}53 = 21{,}129$$
за $(l - 1) = (18 - 1) = 17$ ступенів свободи;
$$C_Z = C_Y - C_P - C_V = 21{,}381 - 0{,}046 - 21{,}129 = 0{,}206$$
за $(l - 1) \cdot (n - 1) = 17 \cdot 3 = 51$ ступеня свободи.

3-й етап. На цьому етапі, як і у двофакторному досліді поставленому методом рендомізованих повторень, визначають суми квадратів для чинників *A* і *B* та їх взаємодії. Для цього складають допоміжну таблицю, в яку переносять показники сум варіантів за повтореннями з табл. 80, тобто опускають дані кожного повторення. Далі розраховують суми та середні за чинниками *A* і *B* (табл. 81).

Таблиця 81

Суми показників за варіантами для визначення головних ефектів та їхньої взаємодії, т/га

Чинник *A*, попередник	Чинник *B*, норма висіву, млн нас./га						Сума, *A*	Середнє за *A*
	2,5	3,0	3,5	4,0	4,5	5,0		
0	12,93	15,40	16,78	18,14	19,07	19,19	101,51	4,23
1	12,12	14,16	15,84	16,81	16,53	16,06	91,52	3,81
2	11,27	12,82	14,39	15,38	15,45	14,90	84,21	3,51
Суми *B*	36,32	42,38	47,01	50,33	51,05	50,15	$\sum x = 277{,}2$	
Середнє за *B*	3,03	3,53	3,92	4,20	4,25	4,18	$\bar{x} = 3{,}85$	

Дисперсійний аналіз даних табл. 81 дає змогу поділити загальну дисперсію варіантів на окремі складові: варіювання чинників *A* і *B* та їхньої взаємодії. Розрахунки проводяться у такому порядку.

$$C_V = C_A + C_B + C_{AB} = 21{,}129;$$
$$C_A = \sum X_A^2 / l_b \cdot n - C\ (101{,}51^2 + 91{,}52^2 + 84{,}21^2) / 6 \cdot 4 - 1067{,}53 =$$
$$= (10304{,}280 + 8375{,}910 + 7091{,}324) / 24 - 1067{,}53 = 6{,}283$$
за $(l_a - 1) = (3 - 1) = 2$ ступенів свободи;
$$C_B = \sum X_b^2 / l_a \cdot n - C =$$
$$= (36{,}32^2 + 42{,}38^2 + ... + 50{,}15^2) / 3 \cdot 4 - 1067{,}53 = 14{,}085$$

за $(l_b - 1) = (6 - 1) = 5$ ступенів свободи;

$$C_{AB} = C_V - C_A - C_B = 21{,}129 - 6{,}283 - 14{,}085 = 0{,}761$$

за $(l_a - 1)(l_b - 1) = (3 - 1) \cdot (6 - 1) = 10$ ступенів свободи.

4-й етап. Характерною рисою двофакторних польових дослідів, поставлених методом розщеплених ділянок порівняно з двофакторними дослідами, поставленими методом рендомізованих повторень, є наявність двох помилок: перша для головних ділянок першого порядку (чинник *A*) і друга для ділянок другого порядку (чинник *B*) та взаємодії досліджуваних чинників. Для того щоб визначити ці помилки, потрібно розділити загальне варіювання помилок на складові компоненти: $C_Z = C_{ZI} + C_{ZI}$. Сума квадратів відхилень похибки I (C_{ZI}) дає можливість визначити істотність дії попередників (чинник *A*), а сума квадратів відхилень помилки II C_{ZII} дає змогу визначити ефект норми висіву (чинник *B*) та взаємодію досліджуваних чинників. Розрахунки проводять таким чином: спочатку визначають дисперсію помилок головних ділянок досліду C_{ZI}, після чого визначають дисперсію помилки другого порядку C_{ZII} як різницю між C_Z і C_{ZI}. Для визначення дисперсії помилки ділянок першого порядку складають допоміжну таблицю, до якої переносять суми показників досліджуваної ознаки ділянок першого порядку (табл. 82).

Таблиця 82

Урожайність зерна ячменю ярого залежно від попередників (чинник *A*) і норм висіву (чинник *B*) у сумі за повтореннями, т/га

Чинник *A*	Урожайність за повтореннями				Сума, *A*
	I	II	III	IV	
0	25,45	25,34	25,17	25,55	101,51
1	22,88	23,02	22,62	23,00	91,52
2	21,20	20,97	20,78	21,26	84,21
Сума *B*	69,53	69,33	68,57	69,81	$\sum x = 277{,}2$

Для цього визначають сумарні показники за ділянками першого порядку з табл. 80. Зокрема, для першої ділянки першого повторення сумарний показник становитиме: 3,21 + 3,87 + 4,22 + 4,58 + 4,73 + + 4,84 = 25,45; другої ділянки першого повторення – 3,04 + 3,56 + + 3,93 + 4,13 + + 4,19 + 4,03 = 22,88 і т. ін. Правильність розрахунків перевіряють за рівнянням: $\sum x = \sum P = \sum V = 277{,}2$.

Для визначення дисперсії помилок ділянок першого порядку C_{ZI}, знаходять загальну суму квадратів відхилень допоміжної табл. 82 C_{YI}, яка складається з дисперсії повторень, дисперсії ділянок першого порядку C_A і дисперсії помилок ділянок першого порядку C_{ZI}. Загальне варіювання помилок ділянок першого порядку розраховують за рівнянням: $C_{ZI} = C_{YI} - C_A - C_P$. Оскільки C_A і C_P відомі, визначають C_{YI}:

$$C_{YI} = \sum x_I^2 / l_b - C = (25{,}45^2 + 25{,}34^2 + ... + 21{,}26^2) / 6 - 1067{,}53 = 6{,}338;$$

$$C_{ZI} = C_{YI} - C_A - C_P = 6{,}338 - 6{,}283 - 0{,}046 = 0{,}009$$

за $(l_a - 1)\ (l_b - 1) = (3 - 1) \cdot (4 - 1) = 6$ ступенів свободи.

За різницею $C_Z - C_{ZI}$ визначають C_{ZII}:

$$C_{ZII} = C_Z - C_{ZI} = 0{,}206 - 0{,}009 = 0{,}197.$$

Після цього складають таблицю дисперсійного аналізу і визначають істотність впливу чинників *A* і *B* та їхньої взаємодії за *F*-критерієм (табл. 83).

Таблиця 83

Результати дисперсійного аналізу двофакторного польового досліду ($l_a = 3;\ l_b = 6$), поставленого методом розщеплених ділянок

Дисперсія	Сума квадратів	Ступені свободи	$F_{факт.}$	$F_{теор.}$	
				05	01
Загальна, C_Y	21,381	71	–	–	–
Повторень, C_P	0,046	3	–	–	–
Чинника, *A*	6,283	2	1571,00	5,14	10,92
Помилки, C_{ZI}	0,009	6	–	–	–
Чинника, *B*	14,085	5	704,25	2,40	3,41
Взаємодії, *AB*	0,761	10	19,00	2,02	2,70
Помилки, C_{ZII}	0,197	45	–	–	–

Теоретичні значення F_{05} і F_{01} знаходять у дод. В на підставі числа ступенів свободи для досліджуваних чинників *A* і *B* та їх взаємодії (чисельник) і відповідної їм помилки: C_{ZI} для чинника *A* і C_{ZII} – для чинника *B* і взаємодії *AB* (знаменник). У цьому досліді вплив досліджуваних чинників та їх взаємодії доведений для обох рівнів ймовірності ($F_{факт.} > F_{05}$ і $F_{факт.} > F_{01}$).

5-й етап. Проводять оцінку часткових різниць для ділянок першого порядку (попередник) і другого (норма висіву).

Для ділянок першого порядку:

$$s_{\bar{x}}' = \sqrt{s'^2/n} = \sqrt{0{,}002\,/4} = 0{,}022 \text{ т/га};$$
$$s_d' = \sqrt{2s'^2\,/\,n} = \sqrt{2\cdot 0{,}002/4} = \sqrt{0{,}065} = 0{,}03 \text{ т/га};$$
$$HIP_{05} = t_{05}\cdot s_d' = 2{,}45\cdot 0{,}03 = 0{,}07 \text{ т/га};$$
$$HIP_{01} = t_{01}\cdot s_d' = 3{,}71\cdot 0{,}03 = 0{,}11 \text{ т/га}.$$

Значення $t_{05} = 2{,}45$ і $t_{01} = 3{,}71$ беруть з дод. Б для 6 ступенів свободи помилки Z_I.

Для ділянок другого порядку (норми висіву):

$$s_{\bar{x}}'' = \sqrt{s''^2/n} = \sqrt{0{,}004\,/4} = 0{,}03 \text{ т/га};$$
$$s_d'' = \sqrt{2s''^2\,/\,n} = \sqrt{2\cdot 0{,}004/4} = 0{,}045 \text{ т/га};$$
$$HIP_{05} = t_{05}\cdot s_d'' = 2{,}01\cdot 0{,}045 = 0{,}09 \text{ т/га};$$
$$HIP_{01} = t_{01}\cdot s_d'' = 2{,}68\cdot 0{,}045 = 0{,}12 \text{ т/га}.$$

Значення $t_{05} = 2{,}01$ і $t_{01} = 2{,}68$ відповідають 45 ступеням свободи дисперсії помилки Z_{II} (див. дод. Б).

6-й етап. Проводять оцінку головних ефектів чинників A і B.

Для головного ефекту чинника A (попередник):

$$s_d = \sqrt{2\,s'^2/n\cdot lb} = \sqrt{2\cdot 0{,}002/24} = 0{,}013 \text{ т/га};$$
$$HIP_{05} = t_{05}\cdot s_d = 2{,}45\cdot 0{,}013 = 0{,}032 \text{ т/га};$$
$$HIP_{01} = t_{01}\cdot s_d = 3{,}71\cdot 0{,}013 = 0{,}048 \text{ т/га}.$$

Для головного ефекту чинника B (норма висіву):

$$s_d = \sqrt{2\,s''^2/n\cdot la} = \sqrt{2\cdot 0{,}004/12} = 0{,}03 \text{ т/га};$$
$$HIP_{05} = t_{05}\cdot s_d = 2{,}01\cdot 0{,}03 = 0{,}06 \text{ т/га};$$
$$HIP_{01} = t_{01}\cdot s_d = 2{,}68\cdot 0{,}03 = 0{,}08 \text{ т/га}.$$

Одержані дані $HIP_{05} = 0{,}07$ т/га і $HIP_{01} = 0{,}11$ т/га дають змогу оцінити ефективність попередника (чинник A) за різних варіантів норми висіву (чинник B).

Наприклад, за норми висіву 2,5 млн нас./га урожайність ячменю ярого після чистого пару була на 0,20 т/га вищою, ніж після кукурудзи на зерно і на 0,41 т/га, ніж після буряків цукрових, що значно більше за HIP_{05} і HIP_{01} (рис. 27). Так само проводимо оцінку ефективності попередників за рештою досліджуваних норм висіву.

Отже, врожайність зерна ячменю ярого після чистого пару за всіх норм висіву була істотно вищою, ніж після кукурудзи на зерно (група I), тоді як після буряків цукрових за всіх норм висіву врожай-

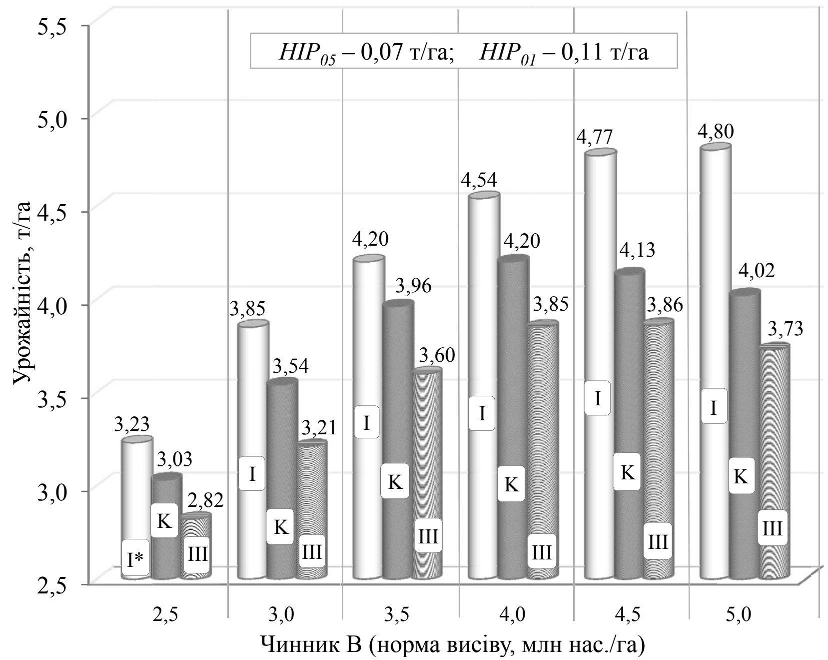


Рис. 27. Оцінка часткових порівнянь показників урожайності ячменю ярого після різних попередників за різних норм висіву.
Умовні позначення: * Групи показників стосовно до контролю чинника A (кукурудза на зерно).
Варіанти: – чистий пар; – кукурудза на зерно; – буряки цукрові

ність була істотно меншою, ніж на контролі – після кукурудзи на зерно (група III).

Найменша істотна різниця з імовірністю 95 і 99 % для часткових порівнянь ефекту норми висіву становить відповідно 0,01 і 0,12 т/га. Зокрема, після чорного пару врожайність зерна ячменю ярого за норми висву 2,5 млн нас./га була на 1,31 т/га меншою, ніж на контролі (норма висіву 4,0 млн нас./га), що значно менше за HIP_{05}. Аналогічні порівняння проводяться за іншими нормами висіву після різних попередників (рис. 28).

Отже, ефект норми висіву після різних попередників був неоднаковим. Зокрема, після чистого пару врожайність зерна ячменю ярого за норми висіву 5,0 млн нас./га була істотно вищою, ніж на

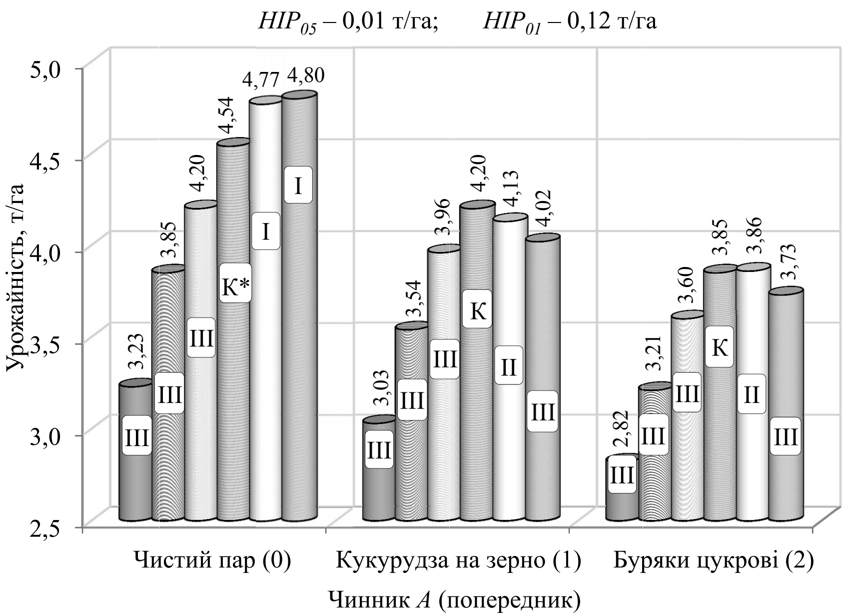


Рис. 28. Оцінка часткових порівнянь чинника *В* – показників урожайності за різних норм висіву після різних попередників.
Умовні позначення: * Групи показників стосовно до контролю чинника *В* (норма висіву – 4,0 млн нас./га). Норма висіву, млн нас./га:
– 2,5; – 3,0; – 3,5; – 4,0; – 4,5; – 5,0

контролі (норма висіву – 4,0 млн нас./га), а після кукурудзи на зерно, навпаки – істотно меншою (III група). Після всіх попередників урожайність зерна за норм висіву 2,5; 3,0 і 3,5 млн нас./га була істотно меншою, ніж на контрольному варіанті.

Аналіз головних ефектів показав істотну різницю між показниками обох досліджуваних чинників (рис. 29). Цей рисунок наглядно показує перевагу чистого пару при порівнянні показників головного ефекту чинника *А*. Максимальна достовірна врожайність зерна формувалася за норми висіву 4,5 млн нас./га. Як зменшення, так і збільшення норми висіву в середньому за чинником *А* спричиняло істотне зниження врожайності за обох рівнів ймовірності.

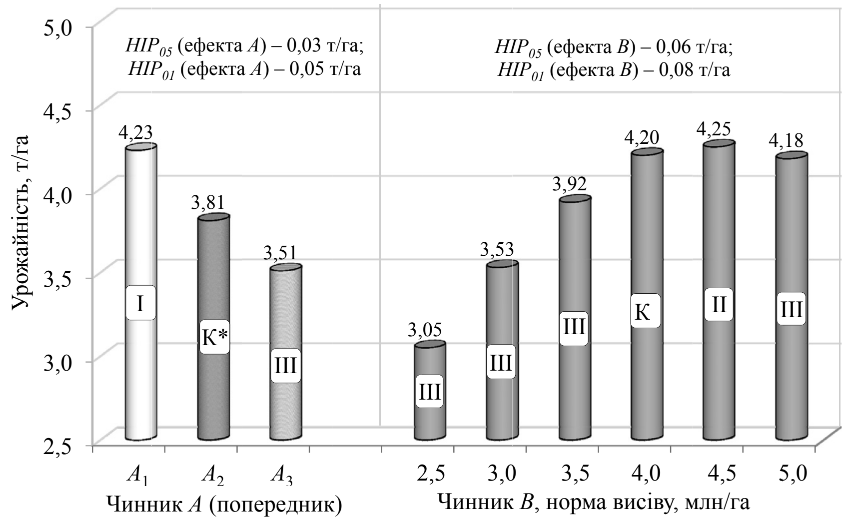


Рис. 29. Оцінка головних ефектів досліджуваних чинників.
Умовні позначення: * Групи показників відносно контролю чинників *А* і *В*

8.3. Техніка обробки трифакторного польового досліду, поставленого методом розщеплених ділянок

На практиці доволі часто виникає необхідність проведення складних трифакторних дослідів. Як правило, досліди, в яких передбачено вивчення впливу трьох і більше чинників за повною факторіальною схемою (вивчення усіх можливих комбінацій взаємодії досліджуваних чинників) проводяться методом розщеплених ділянок.

Дисперсійний аналіз дослідів за такої схеми значно ускладнюється, оскільки для оцінки впливу досліджуваних варіантів слід визначати показники трьох помилок (для трифакторного польового досліду). Також значно ускладнюється оцінка часткових порівнянь. Зокрема, якщо у двофакторному досліді, поставленому методом розщеплених ділянок визначаються два часткових порівняння ефектів, то у трифакторному досліді можлива кількість часткових порівнянь досліджуваних чинників зростає до вісімнадцяти. Водночас точність оцінки дії чинника, розміщеного на елементарних ділянках, а також ефектів взаємодії з цим чинником, значно зростає. Отже, у багатофакторному польовому досліді значно зростає ефективність оцінки чинника, розміщеного на ділянках найменшого порядку.

Значно доцільніше викладення теоретичного матеріалу методик розрахунків дисперсійного аналізу багатофакторних польових дослідів проводити на конкретних прикладах. Отже, наведемо приклад розрахунків дисперсійного аналізу трифакторного досліду, поставленого методом розщеплених ділянок.

Приклад 2. У трифакторному польовому досліді вивчали вплив двох попередників (чинник А) – ділянки першого порядку (S – 400 м^2), двох способів сівби (чинник В) – ділянки другого порядку (S_{II} – 200 м^2) та чотирьох варіантів системи удобрення (чинник С) – елементарні ділянки третього (найменшого) порядку (S_{III} – 50 м^2) (табл. 84).

Статистичну обробку даних трифакторного польового досліду, поставленого методом розщеплених ділянок, проводять за тим самим принципом, що і у двофакторному досліді – у декілька етапів.

1-й етап. У першу чергу перевіряють правильність розрахунків сум вихідних даних табл. 84 за рівнянням $\sum X = \sum V = \sum P$. Для полегшення подальших розрахунків визначають сумарні показники на головних ділянках і ділянках першого порядку. Для цього розраховують сумарний показник збору білка з ділянок першого і другого порядку кожного повторення. Так, сумарний збір білка після кукурудзи на зерно (A_0) за рядкового способу сівби (B_0) у першому повторенні становив 1265 кг/га, у другому – 1303 кг/га і т. ін.

2-й етап. Далі, як і у простих однофакторних дослідах, розраховують корегуючий фактор C, загальну суму квадратів C_Y, суму квадратів відхилень повторень C_P, варіантів C_V і помилок C_Z:

$$N = l_a \cdot l_b \cdot l_c \cdot n = 4 \cdot 2 \cdot 2 \cdot 4 = 64;$$
$$C_{(\text{корегуючий фактор})} = (\sum X^2) / N = 22556^2 / 64 = 7949580{,}3;$$
$$C_Y = \sum X^2 - C = (263^2 + 290^2 + ... + 428^2) - 7949580{,}3 = 122103{,}7$$
$$\text{за } (N - 1) = (64 - 1) = 63 \text{ ступенів свободи;}$$
$$C_{P(\text{повторень})} = \sum P^2 / l_a \cdot l_b \cdot l_c - C =$$
$$= (5630^2 + 5623^2 + 5704^2 + 5599^2) / 2 \cdot 2 \cdot 4 - 7949580{,}3 =$$
$$= 127199446 / 16 - 7949580{,}3 = 381{,}1$$
$$\text{за } (p - 1) = (4 - 1) = 3 \text{ ступенів свободи;}$$
$$C_{V(\text{варіантів})} = \sum V^2 / n - C = (1100^2 + 1275^2 + ... + 1702^2) / 4 - 7949580{,}3 =$$
$$= (1210000 + 1625625 + ... + 2896804) / 4 - 7949580{,}3 = 119107{,}7$$
$$\text{за } (l_a \cdot l_b \cdot l_c - 1) = 15 \text{ ступенів свободи.}$$

Таблиця 84

Збір білка з гектара пшениці твердої ярої за впливу технологічних чинників, кг/га

Чинники			Повторення				Сума, V	Середнє
A	B	C	I	II	III	IV		
0	0	0	263	290	278	269	1100	275
		1	310	327	322	316	1275	319
		2	342	331	356	340	1369	342
		3	350	355	363	361	1429	357
$\sum P$ для A_0B_0			1265	1303	1319	1286	$\sum X = 5173$	323
0	1	0	282	280	268	264	1094	274
		1	337	339	331	332	1339	335
		2	359	356	370	367	1452	363
		3	398	391	380	395	1564	391
$\sum P$ для A_0B_1			1376	1366	1349	1358	$\sum X = 5449$	341
$\sum P$ для A_0			2641	2669	2668	2644	$\sum X = 10622$	332
1	0	0	312	308	316	303	1239	310
		1	347	355	358	348	1408	352
		2	388	381	384	382	1535	384
		3	396	373	392	387	1548	387
$\sum P$ для A_1B_0			1443	1417	1450	1420	$\sum X = 5730$	358
1	1	0	332	324	338	321	1315	329
		1	376	389	402	381	1548	387
		2	416	409	409	405	1639	410
		3	422	415	437	428	1702	426
$\sum P$ для A_1B_1			1546	1537	1586	1535	$\sum X = 6204$	388
$\sum P$ для A_1			2989	2954	3036	2955	$\sum X = 11934$	373
Загальна сума Р			5630	5623	5704	5599	$\sum X = 22556$	$\bar{x}_{ср} = 352$

Примітка. Варіанти чинника A (попередник): 0 – кукурудза на зерно; 1 – соя. Варіанти чинника B: (спосіб сівби): 0 – рядковий; 1 – смуговий. Варіанти системи живлення (чинник C): 1 – контроль (без внесення добрив); 2, 3 і 4 – основне внесення мінеральних добрив у дозах відповідно $N_{15}P_{15}K_{15}$, $N_{30}P_{30}K_{30}$ і $N_{45}P_{45}K_{45}$.

За рівнянням $C_Y = C_V + C_P + C_Z$ визначаємо суму квадратів відхилень помилок: $C_Z = C_Y - C_P - C_V = 122103{,}7 - 119107{,}7 - 385{,}1 = 2610{,}9$ за $(n - 1)\ (l_a - 1)\ (l_b - 1)\ (l_c - 1) = 3 \cdot 1 \cdot 1 \cdot 3 = 9$ ступенів свободи.

3-й етап. У трифакторному польовому досліді значно зростає кількість складових загальної суми квадратів варіантів. Так, якщо у двофакторному польовому досліді сума квадратів варіантів є сумою трьох складових: чинника *А*, чинника *В* і їхньої взаємодії, то у трифакторному польовому досліді, поставленому методом розщеплених ділянок, їх кількість зростає до семи: сума квадратів відхилень чинників А, В і С, взаємодія двох чинників АВ, АС і ВС та взаємодія АВС. Загальна формула має вигляд:

$$C_V = C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}.$$

Спочатку визначають суми квадратів головних чинників, далі можливі комбінації двох досліджуваних чинників і на останньому етапі – суму квадратів взаємодії *АВС* (C_{ABC}), яка дорівнює різниці між загальною сумою квадратів відхилень варіантів (C_V) і сумою квадратів відхилень головних та подвійних ефектів чинників.

Формули для визначення сум квадратів відхилень досліджуваних чинників трифакторного польового досліду, поставленого методом розщеплених ділянок:

$$C_A = \sum X_A / l_b \cdot l_c \cdot n - C;$$
$$C_B = \sum X_B / l_a \cdot l_c \cdot n - C;$$
$$C_C = \sum X_C / l_a \cdot l_b \cdot n - C;$$
$$C_{AB} = \sum X_{AB} / l_c \cdot n - C - C_A - C_B;$$
$$C_{AC} = \sum X_{AC} / l_b \cdot n - C - C_A - C_C;$$
$$C_{BC} = \sum X_{BC} / l_a \cdot n - C - C_B - C_C;$$
$$C_{ABC} = C_V - C_A - C_B - C_C - C_{AB} - C_{AC} - C_{BC}.$$

Підставляючи дані з табл. 84, визначаємо суми квадратів відхилень складових варіантів досліду.

$$C_A = \sum X_A / l_b \cdot l_c \cdot n - C = (10622^2 + 11934^2) / 2 \cdot 4 \cdot 4 - 7949580{,}3 =$$
$$= (112826884 - 142420356) / 32 - 7949580{,}3 = 26896{,}0$$
за $(l_A - 1) = (2 - 1) = 1$ ступені свободи;
$$C_B = \sum X_B / l_a \cdot l_c \cdot n - C = (10903^2 + 11653^2) / 2 \cdot 4 \cdot 4 - 7949580{,}3 =$$
$= 254667818 / 32 - 7949580{,}3 = 8789{,}0$ за $(l_b - 1) = 1$ ступені свободи;
$$C_C = \sum X_C / l_a \cdot l_b \cdot n - C = (4748^2 + ... + 6243^2) / 2 \cdot 2 \cdot 4 - 7949580{,}3 =$$
$= 80637{,}1$ за $(l_c - 1) = (4 - 1) = 3$ ступенів свободи;
$$C_{AB} = \sum X_{AB} / l_c \cdot n - C - C_A - C_B = (5173^2 + 5449^2 +...+ 6204^2) / 4 \cdot 4 -$$
$$- 7949580{,}3 - 26896{,}0 - 8789{,}0 = 612{,}6$$

за $(l_a - 1)(l_b - 1) = (2 - 1) \cdot (2 - 1) = 1$ ступені свободи;
$$C_{AC} = \sum X_{AC} / l_b \cdot n - C - C_A - C_C = (2194^2 + 2614^2 +...+ 3250^2) / 2 \cdot 4 -$$
$$- 7949580{,}3 - 26896{,}0 - 80637{,}1 = 430{,}4$$
за $(l_a - 1)(l_c - 1) = (2 - 1)(4 - 1) = 3$ ступені свободи;
$$C_{BC} = \sum X_{BC} / l_a \cdot n - C - C_B - C_C =$$
$$(2339^2 + ... + 3266^2) / 2 \cdot 4 - 7949580{,}3 - 8789{,}0 - 80637{,}1 = 1523{,}9$$
за $(l_b - 1)(l_c - 1) = (2 - 1)(4 - 1) = 3$ ступенів свободи;
$$C_{ABC} = C_v - C_A - C_B - C_C - C_{AB} - C_{AC} - C_{BC} =$$
$$= 119107{,}7 - 26896{,}0 - 8789{,}0 - 80637{,}1 - 612{,}6 - 430{,}4 - 1523{,}9 = 218{,}7$$
за $(l_a - 1)(l_b - 1)(l_c - 1) = 3$ ступенів свободи.

4-й етап. Як уже згадувалося, у трифакторному польовому досліді, поставленому методом розщеплених ділянок, загальна сума квадратів відхилень помилки C_Z включає в себе суму квадратів відхилень помилок головних ділянок C_{ZI}; ділянок другого C_{ZII} і найменшого – третього порядку C_{ZIII}.

Для визначення суми квадратів відхилень помилок головних ділянок складають допоміжну табл. 85, до якої переносять сумарні показники збору білка з головних ділянок за повтореннями.

Таблиця 85

Збір білка з головних ділянок досліду за повтореннями

Чинник *A*	Повторення				Сума, *V*
	I	II	III	IV	
0	2641	2669	2668	2644	10622
1	2989	2954	3036	2955	11934
Суми *P*	5630	5623	5704	5599	22556

Правильність розрахунків перевіряють за рівнянням:

$$\sum X_A = \sum P = \sum V_A.$$

Загальну суму квадратів таблиці 85 визначають за формулою:

$$C_{YI} = \sum X_A^2 / l_b \cdot l_c - C = (2641^2 + 2669^2 +...+ 2955^2) / 2 \cdot 4 - 7949580 =$$
$= 27539{,}7$ за $n \cdot l_a - 1 = 4 \cdot 2 - 1 = 7$ ступенів свободи.
$$C_{ZI} = C_{YI} - C_P - C_A = 27539{,}7 - 385{,}1 - 26896{,}0 = 258{,}6$$
за $(n - 1)(l_a - 1) = (4 - 1)(2 - 1) = 3$ ступенів свободи.

5-й етап. Розраховують суму квадратів помилок другого порядку C_{ZII}, для чого складають другу допоміжну табл. 86, до якої з табл. 84

Таблиця 86

Сумарний збір білка з ділянок другого порядку

Чинник A	Чинник B	Повторення I	II	III	IV	Сума, V_{AB}
0	0	1265	1303	1319	1286	5173
	1	1376	1366	1349	1358	5449
1	0	1443	1417	1450	1420	5730
	1	1546	1537	1586	1535	6204
Суми P		5630	5623	5704	5599	22559

переносять сумарні показники досліджуваної ознаки для ділянок другого порядку за кожним повторенням.

Суму квадратів помилки ділянок другого порядку C_{ZII} визначають як різницю між загальною сумою квадратів C_{YII} та сумою квадратів табл. 85 (C_{YI}) і варіантів C_B і C_{VAB}: $C_{ZII} = C_{YII} - C_{YI} - C_B - C_{AB}$.

$$C_{VII} = X_{AB}^2 / l_c - C = (1265^2 + ... + 1535^2) / 4 - 7949580{,}3 = 37427{,}7$$
за $n \cdot l_a \cdot l_b - 1 = 4 \cdot 2 \cdot 2 - 1 = 15$ ступенів свободи;
$$C_{ZII} = C_{YII} - C_{YI} - C_B - C_{AB} = 27539{,}7 - 8789{,}0 - 612{,}6 = 486{,}4$$
за $(n-1)(l_b-1) + (n-1)(l_a-1)(l_b-1) = 6$ ступенів свободи.

6-й етап. На підставі проведених розрахунків визначаємо суму квадратів помилок елементарних ділянок C_{ZIII}:

$$C_{ZIII} = C_Y - C_{YII} - C_C - C_{AC} - C_{BC} - C_{ABC} =$$
$$= 122103{,}7 - 37427{,}7 - 80637{,}1 - 430{,}4 - 1523{,}9 - 218{,}7 = 1865{,}9$$
за $(n-1)(l_c-1) + (n-1)(l_a-1) \cdot (l_c-1) + (n-1)(l_b-1)(l_c-1) +$
$+ (n-1) \times (l_a-1)(l_b-1)(l_c-1) = (4-1)(4-1) +$
$+ (4-1)(2-1)(4-1) + (4-1) \times (2-1)(4-1) +$
$+ (4-1)(2-1)(2-1)(4-1) = 36$ ступенів свободи.

7-й етап. Для визначення достовірності впливу можливих комбінацій досліджуваних чинників та їхніх головних ефектів за критерієм F складають таблицю дисперсійного аналізу (табл. 87).

На підставі проведеного аналізу з використанням F-критерію можна сказати, що всі досліджувані чинники: попередники, способи сівби та варіанти системи живлення мали істотний вплив на збір білка з гектара за обох рівнів значущості ($F_{факт.} > F_{05}$ і $F_{факт.} > F_{01}$). Істотні зміни досліджуваного показника відбувалися залежно від взаємодії способів сівби та варіантів системи живлення (взаємодія BC) – $F_{факт.}$

Таблиця 87

Результати дисперсійного аналізу трифакторного польового досліду ($l_a = 2$; $l_b = 2$; $l_c = 4$; $n = 4$), поставленого методом розщеплених ділянок

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$ 05	01
Загальна, C_Y	122103,7	63	–	–	–	–
Головних ділянок, C_{YI}	27539,7	7	–	–	–	–
Ділянок другого порядку, C_{YII}	37427,7	15	–	–	–	–
Повторень, C_P	385,1	3	–	–	–	–
Чинника A, C_A	26896,0	1	26896,0	312,0	10,13	34,1
Помилки ділянок першого порядку, C_{ZI}	258,6	3	86,2	–	–	–
Чинника B, C_B	8789,0	1	8789,0	108,4	5,99	13,74
Взаємодії AB, C_{AB}	612,6	1	612,6	7,6	5,99	13,74
Помилки ділянок другого порядку, C_{ZII}	486,4	6	81,1	–	–	–
Чинника C, C_C	80637,1	3	26879,0	518,9	2,92	4,51
Взаємодії AC, C_{AC}	430,4	3	143,5	2,8	2,92	4,51
Взаємодії BC, C_{BC}	1523,9	3	508,0	9,8	2,92	4,51
Взаємодії ABC, C_{ACB}	218,7	3	72,9	1,4	2,92	4,51
Помилки ділянок третього порядку, C_{ZIII}	1865,9	36	51,8	–	–	–

були вищими за $F_{теор.}$ при обох рівнях ймовірності. Вплив взаємодії AB (попередник + спосіб сівби) був істотним з імовірністю 95 % ($F_{факт.} > F_{05}$). Істотність з імовірністю 99 % не доведена ($F_{факт.} < F_{01}$).

Достовірного впливу на зміну досліджуваного показника взаємодія AC і ABC не мала. Для цих комбінацій чинників $F_{факт.}$ становив 2,8 і 1,4 відповідно за F_{05} – 2,92.

8-й етап. Подальший аналіз представлених даних полягає у проведенні оцінки істотності можливих комбінацій досліджуваних чинників: їх головних ефектів і часткових взаємодій різних рівнів.

Порівняно з двофакторними дослідами оцінка впливу досліджуваних чинників трифакторного досліду значно ускладнюється через значне збільшення можливих комбінацій ефектів порівняння. Оскільки кількість часткових порівнянь досліджуваних чинників у трифакторному польовому досліді становить 18, то у двофакторному досліді, поставленому методом розщеплених ділянок, лише 2. Можливі комбінації оцінки головних ефектів та їхніх часткових порівнянь, стандартні помилки і t-значення для їхнього визначення наведено в табл. 88.

9-й етап. Для повноцінного оцінювання результатів проведеного експерименту проводиться оцінка головних і часткових порів-

Таблиця 88

Формули для розрахунку стандартних помилок і t-критерію для можливих комбінацій оцінки досліджуваних чинників трифакторного досліду, поставленого методом розщеплених ділянок

Комбінації оцінюваних чинників	Стандартні помилки	t-значення
Головний ефект чинника A	$\sqrt{2\, s_I^2/n \cdot l_b \cdot l_c}$	t_a
Головний ефект чинника B	$\sqrt{2s_{II}^2\,/\,n \cdot l_a \cdot l_c}$	t_b
Головний ефект чинника C	$\sqrt{2\, s_{III}^2\,/n \cdot l_a \cdot l_b}$	t_c
Часткових порівнянь B за різних C у середньому по A	$\sqrt{(l_c-1) \cdot s_{III}^2 + s_{II}^2\,/n \cdot l_a \cdot l_c}$	$t_{bc} = \frac{(l_c-1) \cdot s_{III}^2 \cdot t_c + s_{II}^2 \cdot t_b}{(l_c-1) \cdot s_{III}^2 + s_{II}^2}$
Часткових порівнянь A за різних C у середньому по B	$\sqrt{(l_c-1) \cdot s_{III}^2 + s_I^2/n \cdot l_a \cdot l_c}$	$t_{ac} = \frac{(l_c-1) \cdot s_{III}^2 \cdot t_c + s_I^2 \cdot t_a}{(l_c-1) \cdot s_{III}^2 + s_I^2}$
Часткових порівнянь C за різних A у середньому по B	$\sqrt{(l_a-1) \cdot s_I^2 + s_{III}^2\,/n \cdot l_a \cdot l_b}$	$t_{ca} = \frac{(l_a-1) \cdot s_I^2 \cdot t_a + s_{III}^2 \cdot t_c}{(l_a-1) \cdot s_I^2 + s_{III}^2}$
Часткових порівнянь C за різних B у середньому по A	$\sqrt{(l_b-1) \cdot s_{II}^2 + s_{III}^2\,/n \cdot l_a \cdot l_b}$	$t_{cb} = \frac{(l_b-1) \cdot s_{II}^2 \cdot t_b + s_{III}^2 \cdot t_c}{(l_b-1) \cdot s_{II}^2 + s_{III}^2}$
Часткових порівнянь B за різних A у середньому по C	$\sqrt{(l_a-1) \cdot s_I^2 + s_{II}^2\,/n \cdot l_c \cdot l_a}$	$t_{ba} = \frac{(l_a-1) \cdot s_I^2 \cdot t_a + s_{II}^2 \cdot t_b}{(l_a-1) \cdot s_I^2 + s_{II}^2}$
Часткових порівнянь A за різних B у середньому по C	$\sqrt{(l_b-1) \cdot s_{II}^2 + s_I^2\,/n \cdot l_b \cdot l_c}$	$t_{ab} = \frac{(l_b-1) \cdot s_{II}^2 \cdot t_b + s_I^2 \cdot t_a}{(l_b-1) \cdot s_{II}^2 + s_I^2}$

Продовження табл. 88

Комбінації оцінюваних чинників	Стандартні помилки	t-значення
Часткових порівнянь C за однакових A і B	$\sqrt{s_{III}^2\,/n}$	t_c
Часткових порівнянь A за однакових B і C	$\sqrt{s_I^2\,/n}$	t_a
Часткових порівнянь B за однакових A і C	$\sqrt{s_{II}^2\,/n}$	t_b
Часткових порівнянь B для одного A і різних C	$\sqrt{(l_c-1) \cdot s_{III}^2 + s_{II}^2\,/n \cdot l_c}$	$t_{ba} = \frac{(l_a-1) \cdot s_I^2 \cdot t_a + s_{II}^2 \cdot t_b}{(l_a-1) \cdot s_I^2 + s_{II}^2}$
Часткових порівнянь A для одного B і різних C	$\sqrt{(l_c-1) \cdot s_{III}^2 + s_I^2/n \cdot l_c}$	$t_{ab} = \frac{(l_b-1) \cdot s_{II}^2 \cdot t_b + s_I^2 \cdot t_a}{(l_b-1) \cdot s_{II}^2 + s_I^2}$
Часткових порівнянь B для одного C і різних A	$\sqrt{(l_a-1) \cdot s_I^2 + s_{II}^2\,/n \cdot l_a}$	$t_{bc} = \frac{(l_c-1) \cdot s_{III}^2 \cdot t_c + s_{II}^2 \cdot t_b}{(l_c-1) \cdot s_{III}^2 + s_{II}^2}$
Часткових порівнянь A для одного C і різних B	$\sqrt{(l_b-1) \cdot s_{II}^2 + s_I^2\,/n \cdot l_b}$	$t_{ac} = \frac{(l_c-1) \cdot s_{III}^2 \cdot t_c + s_I^2 \cdot t_a}{(l_c-1) \cdot s_{III}^2 + s_I^2}$
Часткових порівнянь C для одного A і різних B	$\sqrt{(l_b-1) \cdot s_{II}^2 + s_{III}^2\,/n \cdot l_b}$	$t_{ca} = \frac{(l_a-1) \cdot s_I^2 \cdot t_a + s_{III}^2 \cdot t_c}{(l_a-1) \cdot s_I^2 + s_{III}^2}$
Часткових порівнянь C для одного B і різних A	$\sqrt{(l_a-1) \cdot s_I^2 + s_{III}^2\,/n \cdot l_a}$	$t_{cb} = \frac{(l_b-1) \cdot s_{II}^2 \cdot t_b + s_{III}^2 \cdot t_c}{(l_b-1) \cdot s_{II}^2 + s_{III}^2}$
Часткових порівнянь A для одного або різних рівнів B і C	$s_{\overline{x}} = \sqrt{l_b \cdot (l_c-1) \cdot s_{III}^2 + (l_b-1) \cdot s_{II}^2 + s_I^2\,/n \cdot l_b \cdot l_c}$; $t_{abc} = \frac{l_b \cdot (l_c-1) \cdot t_c + (l_b-1) \cdot s_{II}^2 \cdot t_b + s_I^2 \cdot t_a}{l_b \cdot (l_c-1) + (l_b-1) \cdot s_{II}^2 + s_I^2}$	
Часткових порівнянь B для одного або різних рівнів A і C	$s_{\overline{x}} = \sqrt{l_a \cdot (l_c-1) \cdot s_{II}^2 + (l_a-1) \cdot s_I^2 + s_{II}^2\,/n \cdot l_a \cdot l_c}$; $t_{bac} = \frac{l_c \cdot (l_c-1) \cdot t_c + (l_a-1) \cdot s_I^2 \cdot t_a + s_{II}^2 \cdot t_b}{la \cdot (l_c-1) + (l_a-1) \cdot s_I^2 + s_{II}^2}$	
Часткових порівнянь C для одного або різних рівнів A і B	$s_{\overline{x}} = \sqrt{l_a \cdot (l_b-1) \cdot s_{II}^2 + (l_a-1) \cdot s_I^2 + s_{II}^2\,/\,n \cdot l_a \cdot l_b}$; $t_{cab} = \frac{l_a \cdot (l_c-1) \cdot t_c + (l_a-1) \cdot s_I^2 \cdot t_a + s_{III}^2 \cdot t_c}{l_a \cdot (l_c-1) + (l_a-1) \cdot s_I^2 + s_{III}^2}$	

нянь ефектів досліджуваних чинників за допомогою *НІР* або багаторангових критеріїв. Для цього визначають відповідні стандартні помилки, які використовують для цих оцінок. Спочатку проводиться оцінка істотності головних ефектів чинників:

– для головного ефекту чинника A:

$$s_d = \sqrt{2\,s_I^2/n \cdot l_b \cdot l_c} = \sqrt{2 \cdot 86{,}2\,/4 \cdot 2 \cdot 4} = 2{,}32;$$
$$HIP_{05} = t_{05} \cdot s_d = 3{,}18 \cdot 2{,}32 = 7{,}4 \text{ кг/га};$$
$$HIP_{01} = t_{01} \cdot s_d = 5{,}84 \cdot 2{,}32 = 13{,}5 \text{ кг/га}.$$

– для головного ефекту чинника B:

$$s_d = \sqrt{2\,s_{II}^2/n \cdot l_a \cdot l_c} = \sqrt{2 \cdot 81{,}1\,/4 \cdot 2 \cdot 4} = 2{,}25;$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}45 \cdot 2{,}25 = 5{,}5 \text{ кг/га};$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}71 \cdot 2{,}25 = 8{,}4 \text{ кг/га}.$$

– для головного ефекту чинника C:

$$s_d = \sqrt{2\,s_{III}^2/n \cdot l_a \cdot l_b} = \sqrt{2 \cdot 51{,}8\,/4 \cdot 2 \cdot 2} = 2{,}54;$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}04 \cdot 2{,}54 = 5{,}2 \text{ кг/га};$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}75 \cdot 2{,}54 = 7{,}0 \text{ кг/га}.$$

Результати оцінки впливу головних ефектів досліджуваних чинників представлені на рис. 30 і свідчать про те, що всі досліджувані чинники істотно впливали на мінливість збору білка з гектара за обох рівнів ймовірності (95 і 99 %) (рис. 30).

Усі досліджувані градації чинника C (варіанти системи живлення) забезпечували істотне збільшення збору білка з гектара порівняно з контрольним варіантом чинника.

У представленому досліді збір білка з гектара найбільших змін зазнавав за впливу досліджуваних варіантів системи живлення рослин. Цей показник мінімально зростав на 15 кг/га за HIP_{05} і HIP_{01} – відповідно – 2,5 і 7,0 кг/га.

10-й етап. На завершальному етапі статистичних розрахунків дисперсійного аналізу проводять оцінку часткових порівнянь ефектів досліджуваних чинників. Найвищою оцінка ефективності буде на елементарних ділянках нижчого порядку. У конкретному досліді – ділянки чинника C у можливих варіантах порівнянь з іншими чинниками. На конкретних прикладах покажемо механізм оцінки часткових порівнянь ефектів досліджуваних чинників:

– оцінка часткових порівнянь чинника C для однакових попередників (чинників A) і способів сівби (чинник B), тобто порівняння варіантів $A_0B_0C_0$–$A_0B_0C_1$ або $A_0B_0C_0$–$A_0B_0C_2$ і т. ін.

$$s_d = \sqrt{2\,s_{III}^2/n} = \sqrt{2 \cdot 51{,}8\,/4} = 5{,}1;$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}04 \cdot 5{,}1 = 10{,}4 \text{ кг/га};$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}75 \cdot 5{,}1 = 14{,}0 \text{ кг/га}.$$

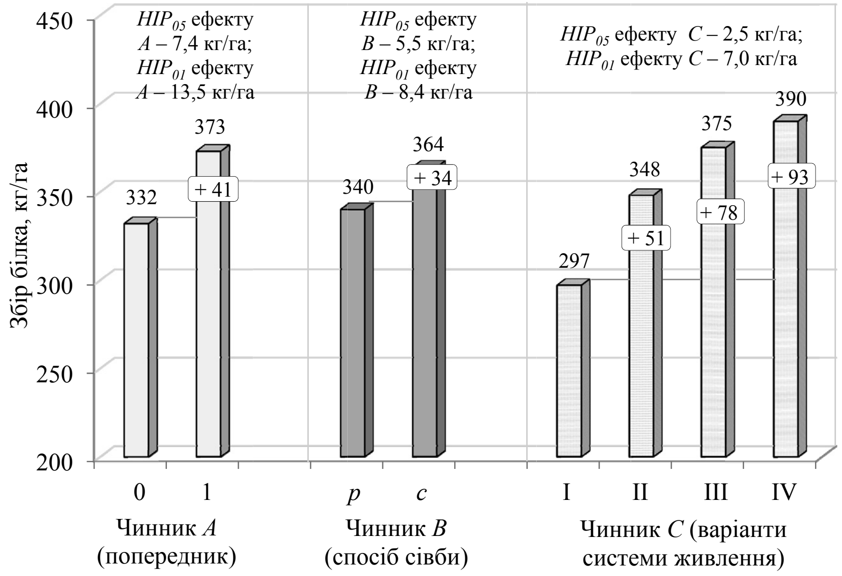


Рис. 30. Оцінка досліджуваних чинників як джерел впливу на мінливість збору білка з гектара.
Умовні позначення. Попередники: 0 – кукурудза на зерно; 1 – соя. Способи сівби: p – рядковий; c – смуговий. Варіанти системи живлення рослин: I – контроль; II – $N_{15}P_{15}K_{15}$; III – $N_{30}P_{30}K_{30}$; IV – $N_{45}P_{45}K_{45}$

– оцінка часткових порівнянь чинника B для однакових попередників (чинників A), та для однакових або різних варіантів системи живлення (чинник C), тобто порівняння $A_0B_0C_0$ $A_0B_1C_0$ $A_0B_0C_0$ $A_0B_1C_1$ тощо.

$$s_d = \sqrt{2 \cdot [(l_c - 1) \cdot s_{III}^2 + s_{II}^2\,/n \cdot l_c]} =$$
$$= \sqrt{2 \cdot [(4 - 1) \cdot 51{,}8 + 81{,}1\,/\,4 \cdot 4]} = 5{,}44.$$

$$t_{ba} = \frac{(l_a - 1) \cdot s_I^2 \cdot t_a + s_{II}^2 \cdot t_b}{(l_a - 1) \cdot s_I^2 + s_{II}^2} = \frac{(2 - 1) \cdot 86{,}2 \cdot 3{,}18 + 51{,}8 \cdot 2{,}45}{(2 - 1) \cdot 86{,}2 + 51{,}8} = 2{,}91.$$

$$HIP_{05} = t_{05ba} \cdot s_d = 2{,}91 \cdot 5{,}44 = 15{,}8 \text{ кг/га}.$$

Таким чином, на підставі проведених розрахунків встановлено, що між досліджуваними способами сівби і після кукурудзи на зерно

(A_0), і без застосування добрив (C_0) різниці не було ($d < HIP_{05}$), тоді як після сої на варіантах без внесення добрив проведення сівби смуговим способом забезпечувало істотне збільшення збору білка з гектара ($d = A_1B_1C_0 - A_1B_0C_0 = 329 - 310 = 19$ кг/га за $HIP_{05} - 16$ кг/га). Аналогічно оцінюємо вплив чинника B за різних варіантів чинника C і одного A. Далі проводимо оцінювання часткових порівнянь досліджуваних чинників у середньому за іншими чинниками. Так, HIP_{05} для часткових порівнянь чинника C у середньому за чинником A (тобто $B_0C_0 - B_0C_1$ чи $B_0C_0 - B_0C_2$ тощо) становить:

$$s_d = \sqrt{2\, s_{III}^2 / n \cdot l_a} = \sqrt{2 \cdot 51{,}8 / 4 \cdot 2} = 3{,}60;$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}04 \cdot 3{,}60 = 7{,}3 \text{ кг/га}.$$

Найменша істотна різниця для часткових порівнянь чинника B у середньому за чинником A при різних рівнях чинника C, тобто $B_0C_0 - B_1C_0$ або $B_0C_0 - B_1C_1$ і т. ін. становить:

$$s_d = \sqrt{2 \cdot [(l_c - 1) \cdot s_{III}^2 + s_{II}^2 / n \cdot l_a \cdot l_c]} = 3{,}84 \text{ кг/га}.$$
$$HIP_{05} = t_{ba} \cdot s_d = 2{,}18 \cdot 3{,}84 = 8{,}4 \text{ кг/га}.$$

На останньому етапі оцінюються часткові порівняння досліджуваних чинників за всіма можливими комбінаціями. Наприклад, HIP_{05} для часткових порівнянь чинника A за можливих комбінацій B і C, тобто $A_0B_0C_0 - A_1B_0C_0$ чи $A_0B_0C_0 - A_1B_1C_1$ тощо становитиме:

$$s_d = \sqrt{2 \cdot [l_b \cdot (l_c - 1) \cdot s_{III}^2 + (l_b - 1) \cdot s_{II}^2 + s_I^2] / n \cdot l_b \cdot l_c} =$$
$$= \sqrt{2 \cdot [2 \cdot (4 - 1) \cdot 51{,}8 + (2 - 1) \cdot 81{,}1 + 86{,}2] / 4 \cdot 2 \cdot 4} = 5{,}47.$$
$$t_{abc05} = \frac{l_b \cdot (l_c - 1) \cdot t_c + (l_b - 1) \cdot S_{II}^2 \cdot t_b + S_I^2 \cdot t_a}{l_b \cdot (l_c - 1) + (l_b - 1) \cdot S_{II}^2 + S_I^2} =$$
$$= \frac{2 \cdot (4 - 1) \cdot 2{,}04 + 81{,}1 \cdot 2{,}45 + 86{,}2 \cdot 3{,}18}{2 \cdot (4 - 1) + 81{,}1 + 86{,}2} = \frac{485{,}2}{169{,}3} = 2{,}86.$$
$$HIP_{05} = t_{abc05} \cdot s_d = 5{,}47 \cdot 2{,}86 = 15{,}6 \text{ кг/га}.$$

За аналогією з цими розрахунками визначають часткові порівняння чинників В і С за можливими комбінаціями решти чинників.

Таким чином, незважаючи на значні ускладнення під час проведення дисперсійного аналізу багатофакторного досліду порівняно з двофакторними, він дає змогу всебічно оцінити ефективність кожного досліджуваного чинника у взаємодії з різними комбінаціями інших чинників. Саме тому ці досліди дають більш об'єктивну оцінку досліджуваних чинників. Насамперед це стосується чинників, що розміщуються на елементарних ділянках нижчого порядку.

8.4. Техніка обробки дослідів, поставлених методом розщеплених блоків

Принциповою відмінністю цього методу є те, що елементарні ділянки чинника другого порядку розщеплюють ділянки першого порядку чинника A цілими смугами впоперек кожного повторення. Таке розміщення полегшує проведення польових робіт на ділянках другого порядку (чинник B), але знижує точність порівняння головних ефектів чинника, варіанти якого розміщені на ділянках другого порядку. Перевагою цієї схеми розміщення варіантів є поліпшення порівняння взаємодії досліджуваних чинників, насамперед порівнянь часткових показників одного чинника для заданих рівнів іншого. Цей метод розміщення варіантів ефективний для проведення попередніх (рекогносцирувальних) досліджень з метою встановлення тенденції впливу досліджуваних варіантів.

Для більш повного розуміння принципової різниці між розміщенням варіантів методами розщеплених ділянок і розщеплених блоків порівняємо їх. На рис. 31 наведено приклади розміщення од-

A_1		A_5		A_2		A_3		A_4	
B_1	B_2	B_6	B_1	B_4	B_2	B_6	B_3	B_2	B_5
B_6	B_3	B_2	B_4	B_1	B_5	B_2	B_5	B_4	B_1
B_4	B_5	B_3	B_5	B_6	B_3	B_1	B_4	B_6	B_3

Метод розщеплених ділянок

A_1	A_5	A_2	A_3	A_4
B_3	B_3	B_3	B_3	B_3
B_4	B_4	B_4	B_4	B_4
B_1	B_1	B_1	B_1	B_1
B_6	B_6	B_6	B_6	B_6
B_2	B_2	B_2	B_2	B_2
B_5	B_5	B_5	B_5	B_5

Метод розщеплених блоків

Рис. 31. Порівняння одного повторення двофакторних польових дослідів, поставлених методом розщеплених ділянок (зліва) і розщеплених блоків (справа). Кількість градацій (варіантів) чинника A – 5; чинника B – 6 шт.

накової кількості досліджуваних варіантів двофакторного польового досліду однієї повторності методами розщеплених ділянок і блоків.

На схемі, розміщеній праворуч, елементарні ділянки (чинника *B*) розташовані впоперек ділянок першого порядку (чинника *A*) і кожна ділянка другого порядку розщеплює блок. Іншою поширеною назвою методу розщеплення варіантів у досліді є смуговий метод, оскільки обидва досліджувані чинники накладаються смугами. Варіанти досліджуваних чинників (смуги) у кожному повторенні розміщують рендомізовано.

Характерною особливістю методу розщеплених блоків, порівняно з методом розщеплених ділянок, є наявність трьох помилок: головних ділянок, ділянок смуг і ділянок другого порядку (субділянок). Збільшення кількості помилок до трьох, зменшує кількість ступенів волі для оцінки головних ефектів варіантів чинника *B*, що розміщується смугами впоперек ділянок головного порядку. Розміщення чинника *B* смугами впоперек головних ділянок усуває частину варіації помилки ділянок другого порядку схеми розщеплених ділянок, оскільки з неї виділяється помилка ділянок смуг для схеми розщеплених блоків. Ця помилка, через менший показник, забезпечує більш точну *F*-оцінку ефектів взаємодії досліджуваних чинників.

Приклад 3. Одержані результати у двофакторному польовому досліді з вивчення впливу норм висіву (ділянки першого порядку – чинник *A*) і сортоособливостей (ділянки смуги – чинник *B*) (рис. 32). Загальна кількість ділянок у досліді – 64 (4 повторення, 4 варіанти чинників *A* і *B*). Ділянки другого порядку розташовані смугами по всій довжині головних ділянок – норм висіву.

Ділянки головного порядку (чинник *A*) за повтореннями, як і ділянки другого порядку (чинник *B*), за смугами розташовують випадково. Разом із тим таке розміщення ділянок у досліді вимагає визначення додаткової помилки для оцінки впливу головних ефектів чинника *B* (сортів).

Досліджувані сорти розміщені смугами впоперек рядів з цілим набором ділянок першого порядку. Будь-яка елементарна ділянка може позначатися символом $Xrcl_al_b$, де r – ряд (r = 4); c – стовбець (c = 4); l_a – норми висіву (l_a = 4) і l_b – сорти (l_b = 4). На представленому схематичному плані цифри у лапках показують суму досліджуваних показників ділянок першого порядку.

<table>
<tr><td colspan="16">Стовпці</td><td rowspan="5">Сума за рядами</td></tr>
<tr><td colspan="4">1</td><td colspan="4">2</td><td colspan="4">3</td><td colspan="4">4</td></tr>
<tr><td colspan="16">Чинник B (сорти)</td></tr>
<tr><td>3</td><td>0</td><td>2</td><td>1</td><td>3</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>3</td><td>0</td><td>1</td><td>3</td><td>2</td></tr>
<tr><td colspan="4">A_1 (126,8)</td><td colspan="4">A_2 (126,1)</td><td colspan="4">A_3 (120,1)</td><td colspan="4">A_0 (129,2)</td></tr>
<tr><td>31,5</td><td>34,9</td><td>25,8</td><td>34,6</td><td>31,3</td><td>34,7</td><td>25,3</td><td>34,8</td><td>34,3</td><td>24,3</td><td>33,5</td><td>28,0</td><td>35,1</td><td>31,9</td><td>34,1</td><td>28,1</td><td>502,2</td></tr>
<tr><td colspan="4">A_0 (128,1)</td><td colspan="4">A_3 (121,1)</td><td colspan="4">A_1 (126,7)</td><td colspan="4">A_2 (124,5)</td><td rowspan="2">500,4</td></tr>
<tr><td>33,7</td><td>35,8</td><td>27,4</td><td>31,2</td><td>29,6</td><td>33,9</td><td>24,9</td><td>32,7</td><td>33,7</td><td>27,3</td><td>32,9</td><td>32,8</td><td>33,9</td><td>34,2</td><td>31,7</td><td>24,7</td></tr>
<tr><td colspan="4">A_2 (123,4)</td><td colspan="4">A_1 (129,0)</td><td colspan="4">A_0 (129,7)</td><td colspan="4">A_3 (119,3)</td><td rowspan="2">501,4</td></tr>
<tr><td>30,8</td><td>33,2</td><td>24,1</td><td>35,3</td><td>31,3</td><td>34,9</td><td>27,6</td><td>35,2</td><td>32,6</td><td>28,8</td><td>34,4</td><td>33,9</td><td>33,1</td><td>33,1</td><td>28,6</td><td>24,5</td></tr>
<tr><td colspan="4">A_3 (117,2)</td><td colspan="4">A_0 (130,6)</td><td colspan="4">A_2 (126,7)</td><td colspan="4">A_1 (128,1)</td><td rowspan="2">502,6</td></tr>
<tr><td>28,3</td><td>32,4</td><td>23,7</td><td>32,8</td><td>34,7</td><td>33,0</td><td>29,2</td><td>33,7</td><td>35,0</td><td>25,7</td><td>34,1</td><td>31,9</td><td>34,5</td><td>34,4</td><td>32,3</td><td>26,9</td></tr>
<tr><td colspan="16">Сума показників за смугами</td><td></td></tr>
<tr><td>124,3</td><td>136,3</td><td>101,0</td><td>133,9</td><td>126,9</td><td>136,5</td><td>107,0</td><td>136,4</td><td>135,6</td><td>106,1</td><td>134,9</td><td>126,6</td><td>136,6</td><td>133,6</td><td>126,7</td><td>104,2</td><td rowspan="3">$\sum X = 2006,6$</td></tr>
<tr><td colspan="16">Сума показників за стовпцями</td></tr>
<tr><td colspan="4">495,5</td><td colspan="4">506,8</td><td colspan="4">503,2</td><td colspan="4">501,1</td></tr>
</table>

Рис. 32. Схематичний план досліду з вивчення сортоособливостей і норм висіву на площу верхнього листка пшениці твердої ярої (у см^2). Варіанти головних ділянок, передбачені для визначення впливу чотирьох норм висіву: 300; 350; 400 і 450 шт. нас./м^2. Варіанти ділянок другого порядку – чотири сорти пшениці ярої: Кучумовка, Чадо, Спадщина та Харківська 41

Площа прапорцевого листка для кожної ділянки другого порядку представлена на рис. 32 разом із сумами для ділянок першого порядку, рядів, стовпців і смуг, передбачених для досліджуваних сортів. Ці дані та суми варіантів по кожному чиннику, представлені в табл. 89, потрібні для проведення дисперсійного аналізу.

Розрахунки статистичних даних цього досліду проводять у кілька етапів.

1-й етап. Визначаємо суму показників за варіантами, стовпцями і рядами. Правильність розрахунків перевіряємо за рівнянням:

Таблиця 89

Сумарні показники площі листків рослин пшениці твердої ярої за варіантами для даних рис. 32

Чинник A (норма висіву)	Чинник B (сорти)				Сума, A
	0	1	2	3	
A_0	139,0	128,7	113,5	136,4	517,6
A_1	137,5	137,6	107,6	127,9	510,6
A_2	136,0	139,2	99,8	125,7	500,7
A_3	131,7	134,1	97,4	114,5	477,7
Суми B	544,2	539,6	418,3	504,5	2006,6

$$\sum x = \sum P = \sum V = 277{,}2.$$

2-й етап. Розраховуємо коригуючий чинник C, загальну суму квадратів відхилень C_Y, суму квадратів відхилень рядів C_P, варіантів чинника A (C_A), варіантів чинника B (C_B), смуг C_n, ділянок першого порядку C_{YI}, взаємодії чинників AB. Для розрахунку суми трьох помилок (C_{ZI}, C_{ZII}, C_{ZIII}) розраховуємо суму квадратів варіантів X_{AB}, представлених у табл. 89.

$$N = l_p \cdot l_c \cdot l_b = l_a \cdot l_c \cdot l_b = 4 \cdot 4 \cdot 4 = 64,$$

де l_p – кількість рядів; l_c – кількість стовпців; l_b – кількість варіантів чинника B і l_a – кількість варіантів чинника A.

$$C_{\text{(корегуючий чинник)}} = (\sum x^2) / N = 2006{,}6^2 / 64 = 62913{,}18;$$

$$C_Y = \sum x^2 - C = (31{,}5^2 + 34{,}9^2 + ...+ 26{,}9^2) - 62913{,}18 = 785{,}86;$$

$$C_{P\text{ (рядів)}} = \sum Xp^2 / l_c \cdot l_b - C = (502{,}2^2 + 500{,}4^2 + 501{,}4^2 + 502{,}6^2) / 16 - 62913{,}18 = 0{,}2$$

за $(l_p - 1) = (4 - 1) = 3$ ступенів свободи;

$$C_{C\text{(стовпців)}} = \sum Xc^2 / l_p \cdot l_b - C = (495{,}5^2 + ... + 501{,}1^2) / 16 - 62913{,}2 = 4{,}2$$

за $(l_c - 1) = (4 - 1) = 3$ ступенів свободи;

$$C_{VA\text{(чинника }A)} = \sum X_A^2 / l_b \cdot n - C = (517{,}6^2 +...+ 477{,}7^2) / 16 - 62913{,}2 = 56{,}8$$

за $(l_a - 1) = (4 - 1) = 3$ ступенів свободи;

$$C_{VB\text{(чинника }B)} = \sum X_B^2 / l_a \cdot n - C = (544{,}2^2 + 539{,}6^2 + 418{,}3^2 + 504{,}5^2) / 16 - 62913{,}2 = 637{,}9$$

за $(l_b - 1) = 3$ ступенів свободи;

$$C_{n\text{(смуг)}} = \sum X_n^2 / l_p - C = (124{,}3^2 + 136{,}3^2 + ... + 104{,}2^2) / 4 - 62913{,}18 = (15450{,}49 + 18577{,}69 + ... + 10857{,}64) / 4 - 62913{,}18 = 646{,}17$$

за $(l_p \cdot l_c - 1) = (4 \cdot 4 - 1) = 15$ ступенів свободи;

$$C_{YI\text{(головних ділянок)}} = \sum X_{An}^2 / l_b - C = (126{,}8^2 + 126{,}1^2 +...+ 128{,}1^2) / 4 - 62913{,}2 = (16078{,}24 + 15901{,}21 +...+ 16409{,}61) / 4 - 62913{,}2 = 63{,}89$$

за $(l_p \cdot l_c - 1) = (4 \cdot 4 - 1) = 15$ ступенів свободи.

Для розрахунку суми квадратів взаємодії чинників A і B, потрібно визначити загальну суму квадратів у табл. 89 (C_{YII}) за рівнянням:

$$C_{YII\text{(головних ділянок)}} = \sum X_{AB}^2 / n - C = (139{,}0^2 + 128{,}7^2 + ... + 114{,}5^2) / 4 - 62913{,}18 = (19321{,}00 + 16563{,}69 +...+ 13110{,}25) / 4 - 62913{,}18 = 763{,}3$$

за $(l_a \cdot l_b - 1) = (4 \cdot 4 - 1) = 15$ ступенів свободи.

За рівнянням $C_{YII} = C_{VA} + C_{VB} + C_{VAB}$ знаходимо C_{VAB}.

$$C_{VAB} = C_{YII} - C_{VA} - C_{VB} = 763{,}3 - 637{,}9 - 56{,}8 = 68{,}6$$

за $(l_a - 1)(l_b - 1) = (4 - 1)(4 - 1) = 3 \cdot 3 = 9$ ступенів свободи.

На кінцевому етапі визначаємо суми квадратів відхилень для трьох помилок досліду (C_{ZI}, C_{ZII} і C_{ZIII}).

$$C_{ZI\text{(головних ділянок)}} = C_{YI} - C_{VA} - C_C - C_P = 63{,}89 - 56{,}80 - 4{,}20 - 0{,}20 = 2{,}69$$

за $(l_p - 1)(l_c - 1) - (l_a - 1) = (4 - 1)(4 - 1) - (4 - 1) = 6$ ступенів свободи.

$$C_{ZII} = C_{YI} - C_C - C_{VB} = 646{,}17 - 4{,}2 - 637{,}9 = 4{,}07$$

за $(l_c - 1) - (l_b - 1) = (4 - 1)(4 - 1) = 3 \cdot 3 = 9$ ступенів свободи.

$$C_{ZIII} = C_Y - C_{YI} - C_{VB} - C_{ZII} - C_{VAB} = 785{,}86 - 63{,}89 - 637{,}9 - 4{,}1 - 68{,}6 = 11{,}37$$

за $(l_c - 1)(l_a - 1)(l_b - 1) = (4 - 1)(4 - 1)(4 - 1) = 27$ ступенів свободи.

2-й етап. Після визначення квадратів сум усіх необхідних показників досліду та визначення їх ступенів свободи складають таблицю дисперсійного аналізу (табл. 90).

Середні квадрати визначаємо шляхом ділення суми квадратів конкретних показників на відповідну для них кількість ступенів свободи. Наприклад, середній квадрат чинника A становитиме: $C_{VA}/(l_a - 1) = 56{,}8/3 = 18{,}9$; середній квадрат помилки $Z_{II} = C_{ZII}/(l_c - 1) \times (l_b - 1) = 0{,}45$ і т. п.

3-й етап. Фактичні значення показників F-критерію визначають шляхом поділу середніх квадратів на відповідні показники помилок: C_{ZI} – для норм висіву (чинник A); C_{ZII} – для сортів (чинник B); C_{ZIII} – для взаємодії AB.

Таблиця 90

Таблиця дисперсійного аналізу двофакторного польового досліду, поставленого методом розщеплених блоків

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна, C_Y	785,86	63	–	–	–	–
Головних ділянок, C_{YI}	63,89	15	–	–	–	–
Рядів, C_P	0,21	3	0,07	–	–	–
Стовпців, C_C	4,20	3	1,40	–	–	–
Чинника *А*	56,80	3	18,93	42,07	4,76	9,78
Помилки I, C_{ZI}	2,69	6	0,45	–	–	–
Чинника *В*	637,90	3	212,3	471,8	3,86	6,99
Помилки II, C_{ZII}	4,07	9	0,45	–	–	–
Взаємодії *АВ*	68,60	9	7,62	18,14	2,27	3,17
Помилки III, C_{ZIII}	11,37	27	0,42	–	–	–

Істотне перевищення показників $F_{факт.}$ порівняно з $F_{теор.}$ за обох рівнів ймовірності свідчить про істотний вплив досліджуваних чинників і їх взаємодії на мінливість площі верхнього листка пшениці твердої ярої.

Стандартні помилки і *t*-значення для оцінки головних ефектів та їх часткових ефектів наведені в табл. 91.

4-й етап. Розраховують HIP_{05} і HIP_{01} для оцінки головних ефектів і часткових порівнянь чинників:

– для головного ефекту чинника *А* (норма висіву):

$$s_d = \sqrt{2s_I^2 / n \cdot l_b} = \sqrt{2 \cdot 0{,}45 / 4 \cdot 4} = 0{,}24;$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}45 \cdot 0{,}24 = 0{,}59 \text{ см}^2;$$
$$HIP_{01} = t_{01} \cdot s_d = 3{,}71 \cdot 0{,}24 = 0{,}89 \text{ см}^2.$$

– для головного ефекту чинника *В* (сорт):

$$s_d = \sqrt{2s_{II}^2 / n \cdot l_a} = \sqrt{2 \cdot 0{,}45 / 4 \cdot 4} = 0{,}24;$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}26 \cdot 0{,}24 = 0{,}54 \text{ см}^2;$$
$$HIP_{01} = t_{01} \cdot s_d = 3{,}25 \cdot 0{,}24 = 0{,}78 \text{ см}^2.$$

Далі проводять оцінку істотності часткових різниць для головних ділянок (норми висіву) і ділянок другого порядку (сортів):

Таблиця 91

Формули для розрахунку стандартних помилок і *t*-критерію двофакторного досліду, поставленого методом розщеплених блоків

Комбінації оцінюваних чинників	Стандартні помилки	*t*-значення
Головний ефект чинника *А*	$\sqrt{s_I^2 / n \cdot l_b}$	t_a
Головний ефект чинника *В*	$\sqrt{s_{II}^2 / n \cdot l_a}$	t_b
Часткові порівняння чинника *А* за різних рівнів *В*	$\sqrt{[(l_b - 1) \cdot s_{III}^2 + s_I^2]/n \cdot l_b}$	$t_{AB} = \frac{(l_b - 1) \cdot s_{III}^2 \cdot t_C + s_I^2 \cdot t_a}{(l_b - 1) \cdot s_{III}^2 + s_I^2}$
Часткові порівняння чинника *В* за різних рівнів *А*	$\sqrt{[(l_a - 1) \cdot s_{III}^2 + s_{II}^2]/n \cdot l_a}$	$t_{BA} = \frac{(l_a - 1) \cdot s_{III}^2 \cdot t_C + s_{II}^2 \cdot t_b}{(l_a - 1) \cdot s_{III}^2 + s_{II}^2}$

Примітка. l_a, l_b і n – кількість варіантів головних ділянок, ділянок другого порядку і повторень відповідно; s_I^2, s_{II}^2 і s_{III}^2 – середні квадрати помилок; t_A, t_B і t_C – табличні показники для s_I^2, s_{II}^2 і s_{III}^2 відповідно.

– для головних ділянок:

$$s_d^I = \sqrt{2[(l_b - 1) \cdot s_{III}^2 + s_I^2]/n \cdot l_b} = \sqrt{2[(4 - 1) \cdot 0{,}42 + 0{,}45]/16} = 0{,}46.$$

$$t_{ab05} = \frac{(l_b - 1) \cdot s_{III}^2 \cdot t_c + s_I^2 \cdot t_a}{(l_b - 1) \cdot s_{III}^2 + s_I^2} = \frac{(4 - 1) \cdot 0{,}42 \cdot 2{,}05 + 0{,}45 \cdot 2{,}45}{(4 - 1) \cdot 0{,}42 + 0{,}45} = 2{,}15.$$

$$t_{ab01} = \frac{(l_b - 1) \cdot s_{III}^2 \cdot t_c + s_I^2 \cdot t_a}{(l_b - 1) \cdot s_{III}^2 + s_I^2} = \frac{(4 - 1) \cdot 0{,}42 \cdot 2{,}77 + 0{,}45 \cdot 3{,}71}{(4 - 1) \cdot 0{,}42 + 0{,}45} = 3{,}01.$$

$$HIP_{05} = t_{ab05} \cdot s_d^I = 2{,}15 \cdot 0{,}46 = 0{,}99 \text{ см}^2.$$
$$HIP_{01} = t_{ab01} \cdot s_d^I = 3{,}01 \cdot 0{,}46 = 1{,}38 \text{ см}^2.$$

– для ділянок другого порядку:

$$s_d^{II} = \sqrt{2[(l_a - 1) \cdot s_{III}^2 + s_{II}^2]/n \cdot l_a} = \sqrt{2[(4 - 1) \cdot 0{,}42 + 0{,}45]/16} = 0{,}46.$$

$$t_{ba05} = \frac{(l_b - 1) \cdot s_{III}^2 \cdot t_c + s_{II}^2 \cdot t_b}{(l_b - 1) \cdot s_{III}^2 + s_{II}^2} = \frac{(4 - 1) \cdot 0{,}42 \cdot 2{,}05 + 0{,}45 \cdot 2{,}26}{(4 - 1) \cdot 0{,}42 + 0{,}45} = 2{,}11.$$

$$t_{ba01} = \frac{(l_a - 1)\cdot S_{III}^2 \cdot t_c + S_{II}^2 \cdot t_b}{(l_a - 1)\cdot S_{III}^2 + S_{II}^2} = \frac{(4-1)\cdot 0{,}42 \cdot 2{,}77 + 0{,}45 \cdot 3{,}25}{(4-1)\cdot 0{,}42 + 0{,}45} = 2{,}89.$$

$$HIP_{05} = t_{ab05} \cdot s_d^{II} = 2{,}11 \cdot 0{,}46 = 0{,}97 \text{ см}^2.$$

$$HIP_{01} = t_{ab01} \cdot s_d^{II} = 2{,}89 \cdot 0{,}46 = 1{,}33 \text{ см}^2.$$

Оцінювання головних ефектів чинників за *НІР* краще проводити за допомогою рис. 33.

Наведений рисунок чітко демонструє ефективність кожної градації досліджуваних варіантів. Зокрема, збільшення норми висіву з 300 до 350 нас./м^2 не забезпечувало істотного зниження площі верхнього листка (різниця становила лише 0,5 см^2 за HIP_{05} і HIP_{01} відповідно – 1,0 і 1,4 см^2). Збільшення норми висіву з 300 до 400 нас./м^2 спричиняло достовірне зменшення площі верхнього листка (на 1,1 см^2) з імовірністю 95 %, тоді як за більш точного підходу (ймовірність 99 %), істотної різниці між цими варіантами не було. Норма висіву 450 нас./м^2 призводила до формування найменшої площі верхнього листка, яка за всіх рівнів ймовірності істотно відрізнялась від показників, одержаних за висіву 300 нас./м^2. За таким самим принципом проводять оцінку головних ефектів чинника *В* (сортів).

Оцінку часткових порівнянь досліджуваних чинників доцільно проводити, використовуючи допоміжну табл. 92, до якої переносять

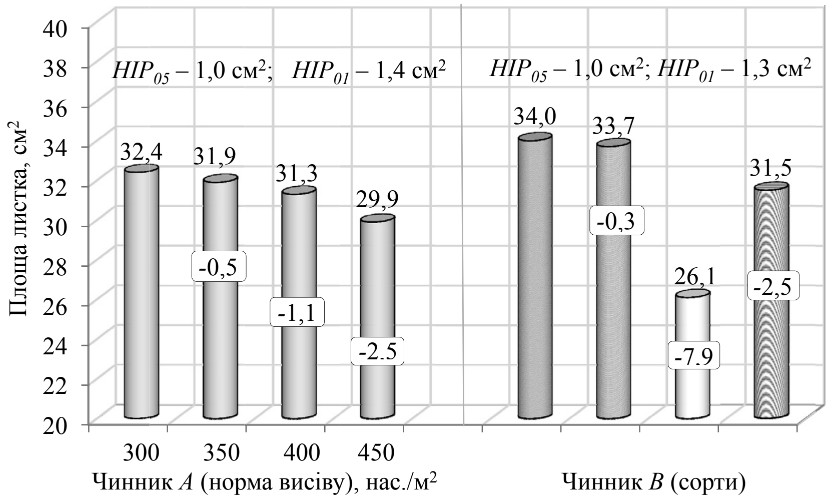


Рис. 33. Оцінка головних ефектів досліджуваних чинників

дані результатів досліду в середньому за повтореннями. Так, площа верхнього листка сорту Кучумовка за висіву 300 нас./м^2 становитиме: (35,8 + 33,7 + 34,4 + 35,1) / 4 = 34,8 см^2; за норми висіву 350 нас./м^2 – (34,9 + 35,2 + 32,9 + 34,5) / 4 = 34,4 см^2 і т. ін.

Таблиця 92

Площа верхнього листка рослин пшениці твердої ярої залежно від впливу норми висіву та сортоособливостей у середньому за повтореннями, см^2

Чинник *А*, норма висіву, нас./м^2	Чинник *В* (сорт)				HIP_{05} часткових порівнянь В
	Кучумовка	Чадо	Спадщина	Харків-ська 41	
300	34,8	32,2	28,3	34,1	0,97
350	34,7	34,4	26,9	32,0	
400	34,0	34,8	25,0	31,4	
450	32,9	33,5	24,4	28,6	
HIP_{05} часткових порівнянь *А*	0,99				

За всіх норм висіву досліджувані сорти формували різну площу верхнього листка, тобто часткові порівняння площі верхнього листка за різних градацій норми висіву показали істотну різницю між ними. Наприклад, часткові порівняння чинника *В* за одного рівня *А* становлять: $A_0B_0 - A_0B_2 = 34{,}8 - 28{,}3 = 6{,}5$ см^2, $A_1B_0 - A_1B_3 = 34{,}7 - 32 = 2{,}7$ см^2; $A_2B_1 - A_2B_3 = 34{,}8 - 31{,}4 = 3{,}4$ см^2 і т. ін. за HIP_{05} – 0,97 см^2. Аналогічно проводиться оцінка часткових порівнянь чинника *А* за одного чи різних рівнів *В*.

Отже, при проведенні досліду методом розщеплених блоків блоки головних ділянок чинника *А* розщеплюються таким чином, що кожний варіант чинника *В* смугами впоперек пересікає кожне повторення чинника *А*. У кожному блоці з повним набором ділянок чинника *А* проводиться незалежна рендомізація варіантів чинника *В*. До переваг цього методу слід віднести зменшення витрат на проведення технологічних операцій, пов'язаних з проведенням досліду, а також більш точну оцінку взаємодії досліджуваних чинників. Разом із тим цей метод має ряд недоліків, зокрема, меншу точність визначення оцінки ефектів чинника *В*, більшу комплексність розрахунків, ускладнення оцінки часткових порівнянь ефектів варіантів.

9. Дисперсійний аналіз однофакторного польового досліду з однорічними і багаторічними культурами

Більшість кількісних показників, які характеризують варіанти проведеного досліду, підпорядковуються закону нормального розподілу, і їх статистична обробка проводиться дисперсійним методом з урахуванням структури експерименту. Але частина кількісних показників досліду, зокрема кількість бур'янів або шкідників на обліковій площі, дегустаційна оцінка якості одержаної продукції, будь-яка оцінка стану посіву в балах та інші, зазвичай не підпорядковуються закону нормального розподілу, і результати експерименту потрібно перетворювати. Існує кілька можливих переходів від фактичних до перетворених показників, які можуть застосовуватися у тих чи інших випадках. Так, якщо між граничними показниками результатів досліджень існує велика різниця (наприклад, мінімальний показник – 10, а максимальний – 200, що має місце у дослідах із забур'яненістю, ураженістю шкідниками тощо), їх доцільно перетворювати за рівнянням $X' = \sqrt{X_{\text{факт}}}$.

Якщо у досліді окремі показники дорівнюють нулю, як, наприклад, бальна оцінка вилягання посівів, ураження шкідниками чи хворобами, вихідні дані доцільно перетворювати за рівнянням: $X' = \sqrt{1 + X_{\text{факт}}}$. Після переведення показників їх обробку проводять методом дисперсійного аналізу. Після оцінки перетворених показників у середньому за варіантами їх результати автоматично переходять на фактичні показники експерименту. Так, коли після проведення дисперсійного аналізу перетворених показників істотну прибавку показника забезпечує певний варіант, то він також значно перевищує контроль за фактичними показниками. Техніка розрахунків методом дисперсійного аналізу показана на прикладах 1, 2.

Приклад 1. У досліді з вивчення впливу гербіцидів на забур'яненість ячменю ярого, поставленому у чотирикратній повторності, визначена кількість бур'янів (табл. 93). Необхідно провести статистичну обробку одержаних даних методом дисперсійного аналізу.

1-й етап. Спочатку проводять оцінку одержаних результатів і визначають порядок перетворення фактичних результатів досліду. Оскільки діапазон варіювання показників значний ($X_{max.} - X_{min.} = 201 - 17 = 184$), фактичні показники доцільно перетворити за рівнянням $X_1 = \sqrt{X}$.

Таблиця 93

Забур'яненість посівів ячменю ярого, шт./м²

Варіант	Повторення					Сума, *V*	Середнє по *V*
	I	II	III	IV	V		
1	94	125	116	128	98	561	112
2	43	48	54	48	40	233	47
3 (контроль)	116	173	129	153	121	692	138
4	135	201	170	186	144	836	167
5	26	39	21	35	17	138	28
6	75	113	109	86	91	474	95

2-й етап. Перетворені показники використовують для проведення статистичного аналізу дисперсійним методом. Визначають суми показників за повтореннями, варіантами, загальну суму ($\sum X_1$) і середнє за варіантами перетворених показників (табл. 94). Правильність розрахунків перевіряють за рівнянням: $\sum P = \sum V = \sum X_1 = 284{,}5$.

Таблиця 94

Перетворені показники забур'яненості посівів ячменю

Варіант досліду	Повторення					Сума, *V*	Середнє	Середній фактичний показник
	I	II	III	IV	V			
1	9,7	11,2	10,8	11,3	9,9	52,9	10,6	112
2	6,6	6,9	7,3	6,9	6,3	34,0	6,8	46
3	10,8	13,2	11,4	12,4	11,0	58,8	11,8	139
4	11,6	14,2	13,0	13,6	12,0	64,4	12,9	166
5	5,1	6,2	4,6	5,9	4,1	25,9	5,2	27
6	8,7	10,6	10,4	9,3	9,5	48,5	9,7	94
Суми, *P*	52,5	62,3	57,5	59,4	52,8	$\sum X$=284,5	$x = 9{,}5$	$x_{ср.} = 9{,}5$

Після цього для перетворених показників визначають поправку C, загальну суму квадратів відхилень C_Y, суму квадратів відхилень повторень C_P, варіантів C_V і помилок C_Z.

$$N = l \cdot n = 6 \cdot 5 = 30;$$
$$C = (\sum X_1)^2 / N = 284{,}5^2 / 30 = 80940{,}25 / 30 = 2698{,}0;$$
$$C_Y = \sum X_1^2 - C = (9{,}7^2 + 11{,}2^2 + ... + 9{,}5^2) - 2698{,}0 = 235{,}6$$
за $(N - 1) = (30 - 1) = 29$ ступенів свободи;
$$C_P = \sum X_{1P}^2 / l - C = (52{,}5^2 + 62{,}3^2 + 57{,}5^2 + 59{,}4^2 + 52{,}8^2) / 6 - 2698{,}0 = 12{,}0$$
за $(n - 1) = (5 - 1) = 4$ ступенів свободи;
$$C_V = \sum X_{1V}^2 / l - C = (52{,}9^2 + 34{,}0^2 + ... + 48{,}5^2) / 5 - 2698{,}0 = 218{,}5$$
за $(l - 1) = (6 - 1) = 5$ ступенів свободи;
$$C_Z = C_Y - C_P - C_V = 235{,}6 - 12{,}0 - 218{,}5 = 5{,}1$$
за $(n - 1)(l - 1) = (5 - 1)(6 - 1) = 20$ ступенів свободи.

3-й етап. Складають таблицю результатів дисперсійного аналізу перетворених даних і розраховують критерій $F_{факт.}$ як співвідношення середнього квадрату дисперсії варіантів до дисперсії помилок (табл. 95).

Таблиця 95

Таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступінь свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна, C_Y	235,6	29	–	–	–	–
Повторень, C_P	12,0	4	–	–	–	–
Варіантів, C_V	218,5	5	43,7	168,0	2,71	4,10
Помилок, C_Z	5,1	20	0,26	–	–	–

$F_{05} = 2{,}71$ і $F_{01} = 4{,}10$ знаходять за дод. В для 5 ступенів свободи дисперсії варіантів і 20 ступенів свободи дисперсії помилок. За обох рівнів ймовірності $F_{факт.}$ вищий від $F_{теор.}$, отже, між варіантами досліду є істотна різниця і нульова гіпотеза спростовується.

4-й етап. Визначають помилку досліду $s_{\overline{x}1}$, помилку різниці середніх s_d і HIP для потрібних рівнів імовірності.

$$s_{\overline{x}1} = \sqrt{s^2/n} = \sqrt{0{,}26/5} = 0{,}23;$$
$$s_d = \sqrt{2s^2/n} = \sqrt{2 \cdot 0{,}26/5} = 0{,}32;$$
$$HIP_{05} = t_{05} \cdot s_d = 2{,}09 \cdot 0{,}32 = 0{,}67;$$
$$HIP_{01} = t_{01} \cdot s_d = 2{,}85 \cdot 0{,}32 = 0{,}91.$$

Теоретичне значення $t_{05} = 2{,}09$ і $t_{01} = 2{,}85$ знаходять за дод. Б для 20 ступенів свободи дисперсії помилок. Одержані результати дисперсійного аналізу представляють у вигляді табл. 96.

Таблиця 96

Розподіл показників забур'яненості посівів ячменю за статистичними групами відносно контролю для перетворених і фактичних показників

Варіант	Фактичний показник	Перетворений показник $X_1 = \sqrt{X}$	Відхилення від контролю для X_1	Групи для	
				HIP_{05}	HIP_{01}
1	112	10,6	– 1,2	III	III
2	46	6,8	– 5,0	III	III
3 (*st*)	139	11,8		(*st*)	(*st*)
4	166	12,9	1,1	I	I
5	27	5,2	– 6,6	III	III
6	94	9,7	– 2,1	III	III
HIP_{05}	–	–	0,7	–	–
HIP_{01}	–	–	0,9	–	–

Отже, забур'яненість варіантів 1, 2, 5 і 6 була значно нижчою, а на варіанті 4 істотно вищою, ніж на контролі за обох рівнів ймовірності (95 і 99 %). Як зазначалося вище, результати дисперсійного аналізу перетворених показників автоматично переносяться на фактичні середні дані досліду, тож у проведеному досліді після застосування гербіцидів у варіантах 1, 2, 5 і 6 забур'яненість істотно зменшувалася порівняно з контролем, а ефективність гербіциду, що вивчався у варіанті 4, була низькою (порівняно з контролем забур'яненість посівів істотно зростала за обох рівнів ймовірності).

Приклад 2. У досліді, поставленому методом рендомізованих повторень (блоків), проведено оцінку вилягання рослин ячменю ярого в балах (від 0 до 5) (табл. 97). Провести дисперсійний аналіз одержаних даних.

Порядок проведення розрахунків. 1-й етап. У першу чергу проводять оцінку одержаних результатів і визначають порядок перетворення фактичних результатів досліду. Оскільки у цьому досліді частина показників наближається до нуля, а два з них рівні нулю, фактичні показники доцільно перетворити за рівнянням: $X_1 = \sqrt{1 + X}$.

Таблиця 97

Ступінь вилягання рослин ячменю ярого у балах

Досліджувані варіанти	Повторення, X				Сума, V	Середнє
	I	II	III	IV		
1 (контроль)	4,2	3,5	4,7	3,1	15,5	3,88
2	1,5	2,2	1,7	1,3	6,7	1,68
3	0,5	0,8	1,1	0	2,4	0,60
4	3,4	3,1	3,8	2,7	13,0	3,25
5	0,7	0	0,3	0,3	1,3	0,33
6	1,3	1,9	1,3	1,7	6,2	1,55
7	2,1	1,5	2,4	1,1	7,1	1,78
8	3,6	3,0	4,1	2,9	13,6	3,40

2-й етап. Проводять статистичну обробку перетворених показників методом дисперсійного аналізу. Визначають суму показників за повтореннями, варіантами, загальну суму результатів, середні показники варіантів перетворених показників (табл. 98). Правильність розрахунків перевіряють за рівнянням:

$$\sum P = \sum V = \sum X_1 = 54{,}54.$$

3-й етап. Проводять статистичну обробку перетворених показників методом дисперсійного аналізу. Розраховують корегуючий

Таблиця 98

Таблиця перетворених показників $X = \sqrt{1+X}$

Варіант	Повторення, X_1				Сума, V	Середнє	Фактичне середнє
	I	II	III	IV			
1 (контроль)	2,28	2,12	2,39	2,02	8,81	2,20	3,88
2	1,58	1,79	1,64	1,52	6,53	1,63	1,68
3	1,22	1,34	1,45	1,00	5,01	1,25	0,60
4	2,10	2,02	2,19	1,92	8,23	2,06	3,25
5	1,30	1,00	1,14	1,14	4,58	1,15	0,33
6	1,52	1,70	1,52	1,64	6,38	1,60	1,55
7	1,76	1,58	1,84	1,45	6,63	1,66	1,78
8	2,14	2,00	2,26	1,97	8,37	2,09	3,40
Суми P	13,90	13,55	14,43	12,66	54,54	$\bar{x}_1 = 1{,}71$	$\bar{x}_{ср.} = 2{,}06$

фактор C, загальну суму квадратів відхилень, суму квадратів відхилень повторень C_P, варіантів C_V і помилок C_Z.

$$N = l \cdot n = 8 \cdot 4 = 32;$$
$$C_{(\text{корегуючий чинник})} = (\sum X_1)^2 / N = 54{,}54^2 / 32 = 92{,}96;$$
$$C_{Y(\text{загальне варіювання})} = (2{,}28^2 + 2{,}12^2 + ... + 1{,}97^2) - C = 4{,}74$$
$$\text{за } (N-1) = (32-1) = 31 \text{ ступеня свободи;}$$
$$C_P = \sum X_P^2 / l - C = (13{,}90^2 + 13{,}55^2 + 14{,}43^2 + 12{,}66^2) / 8 - 92{,}96 =$$
$$= (193{,}21 + 183{,}60 + 208{,}20 + 160{,}28) / 8 - 92{,}96 = 0{,}20$$
$$\text{за } (p-1) = (4-1) = 3 \text{ ступенів свободи;}$$
$$C_V = \sum X_V^2 / n - C = (8{,}81^2 + 6{,}53^2 + ... + 8{,}37^2) / 4 - 92{,}96 = 4{,}24$$
$$\text{за } (l-1) = (8-1) = 7 \text{ ступенів свободи;}$$
$$C_Z = C_Y - C_P - C_V = 4{,}74 - 0{,}20 - 4{,}24 = 0{,}30$$
$$\text{за } (l-1)(8-1) = 7 \cdot 3 = 21 \text{ ступеня свободи.}$$

3-й етап. Складають таблицю результатів дисперсійного аналізу перетворених даних, розраховують критерій $F_{факт.}$ і перевіряють гіпотезу приналежності середньої генеральній сукупності (табл. 99).

Таблиця 99

Таблиця дисперсійного аналізу

Дисперсія	Сума квадратів	Ступінь свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна, C_Y	4,74	31	–	–	–	–
Повторень, C_P	0,20	3	–	–	–	–
Варіантів, C_V	4,24	7	0,61	43,57	2,08	2,83
Помилок, C_Z	0,30	21	0,014	–	–	–

Теоретичні значення $F_{05} = 2{,}08$ і $F_{01} = 2{,}83$ знаходять за дод. В для 7 ступенів свободи дисперсії варіантів і 21 ступеня свободи дисперсії помилок. За обох рівнів імовірності $F_{факт.}$ перевищує $F_{теор.}$. Таким чином, між досліджуваними варіантами є достовірна різниця і нульова гіпотеза спростовується.

4-й етап. Розраховують помилку досліду $s_{\bar{x}_1}$, помилку різниці середніх s_d і НІР для потрібних рівнів ймовірності:

$$s_{\bar{x}_1} = \sqrt{s^2 / n} = \sqrt{0{,}014 / 4} = 0{,}06;$$
$$s_d = \sqrt{2s^2 / n} = \sqrt{2 \cdot 0{,}014 / 4} = 0{,}084;$$

$$НІР_{05} = t_{05} \cdot s_d = 2{,}49 \cdot 0{,}084 = 0{,}21;$$
$$НІР_{01} = t_{01} \cdot s_d = 3{,}65 \cdot 0{,}084 = 0{,}31.$$

Теоретичні значення $t_{05} = 2{,}49$ і $t_{01} = 3{,}65$ знаходять за дод. Б для 21 ступеня свободи дисперсії помилок. Результати статистичної обробки, проведеної методом дисперсійного аналізу, доцільно представити у вигляді підсумкової таблиці 100.

Таблиця 100

Бальна оцінка результатів вилягання посівів ячменю ярого за статистичними групами (для фактичних і перетворених показників)

Варіант досліду	Фактичний показники	Перетворений показник $X_1 = \sqrt{1+X}$	Відхилення від контролю для перетворених показників	Групи для	
				$НІР_{05}$	$НІР_{01}$
1 (*st*)	3,88	2,20	–	*st*	*st*
2	1,68	1,63	– 0,57	III	III
3	0,60	1,25	– 0,95	III	III
4	3,25	2,06	– 0,14	II	II
5	0,33	1,15	– 1,05	III	III
6	1,55	1,60	– 0,60	III	III
7	1,78	1,66	– 0,64	III	III
8	3,40	2,09	– 0,11	II	II
$НІР_{05}$	–	–	0,21	–	–
$НІР_{01}$	–	–	0,31	–	–

Висновок. Ступінь вилягання рослин ячменю ярого варіантів 2, 3, 5, 6 і 7 була значно меншою порівняно з контрольним варіантом (1) за обох рівнів ймовірності (95 і 99 %). Одержані результати статистичного аналізу і оцінка часткових порівнянь досліджуваних варіантів переносяться на фактичні результати показників бальної оцінки вилягання посівів.

При проведенні обліків і аналізів, частина показників може виражатися у відсотках або частках, зокрема ураженість рослин шкідниками, хворобами тощо. У цьому випадку вихідні показники досліду переводяться через кут, синус якого є квадратним коренем від частки або відсотка: X_1 = кут-арксинус $\sqrt{\text{відсоток}}$ (приклад 3). Для цього використовують перехідну таблицю дод. К.

У таблиці даних для дисперсійного аналізу зазвичай проставляють не індивідуальні спостереження (обліки), а середні по кожній ділянці досліджуваної ознаки. Облік діапазону змін, мінливості і варіабельності паралельних аналізів змішаного зразка збільшує обсяг розрахункових операцій і не приводить до помітної мінливості критерію істотності. Тому облік цієї мінливості має сенс лише у спеціальних методичних дослідженнях.

Приклад 3. У досліді, поставленому методом рендомізованих повторень, визначено частку рослин пшениці озимої, уражених летючою сажкою за різних варіантів захисту рослин (табл. 101). Провести статистичну обробку результатів представленого польового досліду методом дисперсійного аналізу.

Таблиця 101

Частка рослин пшениці озимої, уражених летючою сажкою, залежно від впливу варіантів захисту посівів, %

Варіант досліду	Повторення					Сума, *V*	Середнє
	I	II	III	IV	V		
1	17,6	20,4	15,7	19,3	18,0	91,0	18,2
2	5,8	7,0	5,2	6,4	6,0	30,4	6,1
3 (*st*)	69,3	73,8	68,8	70,6	72,2	354,7	70,9
4	53,5	57,2	55,3	57,4	60,2	283,6	56,7
5	0,9	1,7	0,5	1,9	1,3	6,3	1,3

1-й етап. Оскільки дані польового досліду з визначення ураженості рослин летючою сажкою представлені у відсотках, її доцільно перевести через кут, синус якого є квадратним коренем з відсотка – X_1 = кут-арксинус $\sqrt{\text{відсоток}}$ (таблиця 102). Для переведення вихідних результатів користуються дод. К.

2-й етап. Перетворені показники обробляють методом дисперсійного аналізу. Спочатку визначають суми перетворених показників за повтореннями, варіантами і загалом. Правильність розрахунків перевіряють за рівнянням $\sum X_1 = \sum V = \sum P = 760{,}1$. Далі розраховують поправку C, суму квадратів відхилень загальну C_Y, повторень C_P, варіантів C_V та помилок C_Z.

$$N = l \cdot n = 5 \cdot 5 = 25;$$
$$C_{\text{(корегуючий чинник)}} = (\sum X_1)^2 / N = 760{,}1^2 / 25 = 23110{,}0;$$

Таблиця 102

Таблиця перетворених показників

Варіант	Перетворені показники, кут-арксинус $\sqrt{\text{відсоток}}$ (X_1)					Сума, V	Середні показники	
							перетворені	фактичні
1	24,8	26,9	23,3	26,1	25,1	126,2	25,2	18,2
2	13,9	15,3	13,2	14,6	14,2	71,2	14,2	6,1
3 (*st*)	56,4	59,2	56,0	57,2	58,2	287,0	57,4	70,9
4	47,0	49,1	48,0	49,3	50,9	244,3	48,9	56,7
5	5,4	7,5	4,0	7,9	6,6	31,4	6,3	1,3
Суми P	147,5	158,0	144,5	155,1	155,0	760,1	$\bar{x} = 30{,}4$	$\bar{x} = 30{,}6$

$$C_{Y\ (\text{загальне варіювання})} = \sum X_I^2 - \text{C} = (24{,}8^2 + ... + 6{,}6^2) - 23110 = 9732{,}4$$
за $(N - 1) = (25 - 1) = 24$ ступенів свободи;
$$C_P = \sum X_{1P}^2 / l - C = (147{,}5^2 + 158{,}0^2 + ... + 155{,}0^2) / 5 - 23110 = 26{,}3$$
за $(p - 1) = (5 - 1) = 4$ ступенів свободи;
$$C_V = \sum X_{1V}^2 / n - C = (126{,}2^2 + 71{,}2^2 + ... + 31{,}4^2) / 5 - 23110 = 9696{,}4$$
за $(l - 1) = (5 - 1) = 4$ ступенів свободи;
$$C_Z = C_Y - C_P - C_V = 9732{,}4 - 9696{,}4 - 26{,}3 = 9{,}7$$
за $(l - 1)\ (n - 1) = 4 \cdot 4 = 16$ ступенів свободи.

3-й етап. Складають таблицю результатів дисперсійного аналізу переведених показників ураженості рослин пшениці озимої летючою сажкою, визначають критерій $F_{\text{факт.}}$ і перевіряють нульову гіпотезу (табл. 103).

Таблиця 103

Зведені результати дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{\text{факт.}}$	$F_{\text{теор.}}$	
					05	01
Загальна, C_Y	9732,4	24	–	–	–	–
Повторень, C_P	26,3	4	–	–	–	–
Варіантів, C_V	9696,4	4	2424,1	3973,9	3,01	4,77
Помилок, C_Z	9,7	16	0,61	–	–	–

Теоретичне значення $F_{05} = 3{,}01$ і $F_{01} = 4{,}77$ беруть з дод. В для 4 ступенів свободи дисперсії варіантів (чисельник) і 16 ступенів свободи дисперсії помилок (знаменник). У нашому випадку $F_{\text{факт.}}$ перевищує F_{05} і F_{01}, що свідчить про існування достовірної різниці між варіантами досліду.

4-й етап. Визначають помилку середньої $s_{\bar{x}_1}$, помилку різниці середніх s_d і найменшу істотну різницю для потрібного рівня ймовірності.

$$s_{\bar{x}_1} = \sqrt{s^2 / n} = \sqrt{0{,}61 / 5} = 0{,}35;$$
$$s_d = \sqrt{2s^2 / n} = \sqrt{2 \cdot 0{,}61 / 5} = 0{,}49;$$
$$НІР_{05} = t_{05} \cdot s_d = 2{,}12 \cdot 0{,}49 = 1{,}04;$$
$$НІР_{01} = t_{01} \cdot s_d = 2{,}92 \cdot 0{,}49 = 1{,}43.$$

Теоретичні значення $t_{05} = 2{,}12$ і $t_{01} = 2{,}92$ знаходять за дод. Б для 16-ти ступенів свободи дисперсії помилок. Зведені результати показників ураженості рослин пшениці озимої летючою сажкою та результати статистичного аналізу наведено в табл. 104.

Таблиця 104

Порівняльна таблиця результатів досліджень на основі статистичного аналізу

Варіант досліду	Фактичний показник	Переведений показник	Відхилення від контролю перетворених показників	Групи для	
				$НІР_{05}$	$НІР_{01}$
1	18,2	25,2	–32,2	III	III
2	6,1	14,2	–43,2	III	III
3	70,9	57,4	–	(*st*)	(*st*)
4	56,7	48,9	–8,5	III	III
5	1,3	6,3	–51,1	III	III
$НІР_{05}$	–	–	1,04	–	–
$НІР_{01}$	–	–	1,43	–	–

Висновок. Відсоток уражених рослин пшениці озимої летючою сажкою всіх досліджуваних варіантів був значно меншим порівняно з контролем з імовірністю 95 і 99 %. Це стосується також фактичних показників за правилом аналогії порівняння перетворених і фактичних показників.

10. Техніка проведення кореляційного і регресійного аналізу

10.1. Лінійна кореляція і регресія

Для дослідника важливим є визначення залежності продуктивності рослин від погодних умов та елементів технології вирощування. Під час вивчення кореляційної залежності між функціональною (y) та факторіальною (x) ознаками проводять певну кількість парних спостережень (n): (x_1; y_1); (x_2; y_2); (x_{n-1}; y_{n-1});... (x_n; y_n) і на основі отриманих показників визначають вибірковий коефіцієнт кореляції і регресії. Результати досліджень і статистичного аналізу відображають у вигляді теоретичної лінії регресії. Базові формули для визначення статистичних показників залежності між функціональною (y) і факторіальною (x) ознаками представлені в табл. 105.

На конкретних прикладах розглянемо послідовність розрахунків для маленьких ($n \leq 30$) (приклади 1 і 2) і великих ($n > 30$) вибірок.

Приклад 1. Провести кореляційний і регресійний аналізи зв'язку між довжиною верхнього міжвузля x і озерненістю колоса головного стебла пшениці м'якої ярої y (результати отримані на дослідному полі ХНАУ ім. В. В. Докучаєва у 2014 р.) (табл. 106).

Проведення розрахунків включає такі етапи.

1-й етап. Визначають середні показники $\bar{x}$ і $\bar{y}$, суми квадратів відхилень – $\sum(x-\bar{x})^2$ і $\sum(y-\bar{y})^2$ та загальну суму добутку відхилень обох ознак – $\sum(x-\bar{x})\,(y-\bar{y})$:

$$\bar{x} = (\textstyle\sum x) / n = 420{,}1 / 15 = 28{,}0 \text{ см};$$
$$\bar{y} = (\textstyle\sum y) / n = 507 / 15 = 33{,}8 \text{ шт.};$$
$$\textstyle\sum(x-\bar{x})^2 = \sum x^2 - (\sum x)^2 / n = 11926{,}45 - (420{,}1)^2 / 15 = 160{,}85;$$
$$\textstyle\sum(y-\bar{y})^2 = \sum y^2 - (\sum y)^2 / n = 17539 - (507)^2 / 15 = 402{,}4;$$
$$\textstyle\sum(x-\bar{x})\,(y-\bar{y}) = \sum xy - (\sum x \cdot \sum y) / n = 14453 - (420{,}1 \cdot 507) / 15 = 253{,}6.$$

2-й етап. Розраховують коефіцієнт кореляції, детермінації, регресії і лінійне рівняння регресії:

$$r = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \cdot \sum(y-\bar{y})^2}} = \frac{253{,}6}{\sqrt{160{,}85 \cdot 402{,}4}} = 0{,}997;$$

$$d_{yx} = r^2 = 0997^2 = 0{,}994;$$

Таблиця 105

Базові формули для проведення кореляційного і регресійного аналізу лінійної залежності

Показник	Формули для визначення
Коефіцієнт кореляції для малої вибірки ($n \leq 30$)	$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \cdot \sum(y-\bar{y})^2}} = \frac{\sum xy - (\sum x \cdot \sum y)/n}{\sqrt{(\sum x^2 - (\sum x)^2/n)\,(\sum y^2 - (\sum y)^2/n)}}$
Коефіцієнт кореляції для великої згрупованої вибірки ($n > 30$)	$r = \frac{\sum f(x-\bar{x})(y-\bar{y})}{\sqrt{\sum f(x-\bar{x})^2 \cdot \sum f(y-\bar{y})^2}} = \frac{\sum fxy - (\sum fx \cdot \sum fy)/n}{\sqrt{(\sum fx^2 - (\sum fx)^2/n)\,(\sum fy^2 - (\sum fy)^2/n)}}$
Коефіцієнт регресії для малої вибірки ($n \leq 30$)	$b_{yx} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{\sum xy - (\sum x \cdot \sum y)/n}{\sum x^2 - (\sum x)^2/n}$
Коефіцієнт регресії для великої згрупованої вибірки ($n > 30$)	$b_{yx} = \frac{\sum f(x-\bar{x})(y-\bar{y})}{\sum f(x-\bar{x})^2} = \frac{\sum fxy - (\sum fx \cdot \sum fy)/n}{\sum fx^2 - (\sum fx)^2/n}$
Помилка коефіцієнта кореляції і регресії	$s_r = \sqrt{\frac{1-r^2}{n-2}}$; $s_b = s_r \cdot \sqrt{\frac{\sum(y-\bar{y})^2}{\sum(x-\bar{x})^2}}$
Критерій істотності та ступені свободи	$t = r / s_r$; $\nu = n - 2$
Довірчі інтервали	$r \pm t_t \cdot s_r$; $b_{yx} \pm t_t \cdot s_b$
Рівняння регресії та помилка відхилення від регресії	$y = \bar{y} + b_{yx} \cdot (x - \bar{x})$; $s_{yx} = s_r \cdot \sqrt{\sum(y-\bar{y})^2}$

$$b_{yx} = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sum(x-\bar{x})^2} = 253{,}6 / 160{,}85 = 1{,}58 \text{ шт.};$$

$$y = \bar{y} + b_{yx} \cdot (x - \bar{x}) = 33{,}8 + 1{,}58 \cdot (x - 28{,}2) = 1{,}58x - 10{,}76.$$

Оскільки вільний член рівняння регресії негативний, то він не підлягає інтерпретації. Відповідно до коефіцієнта b_{yx} збільшення довжини міжвузля на 1,0 см приведе до збільшення озерненості колоса головного стебла у середньому на 1,6 шт. На основі оціночного рівняння регресії можна зіставити фактичні показники озерненості колоса головного стебла пшениці y з очікуваними $\hat{y}$ за рівнянням $\hat{y} = -10{,}76 + 1{,}58x$ (табл. 107).

Таблиця 106

Розрахунок допоміжних показників для визначення кореляції і регресії y за x

Номер пари	Ознаки		x^2	y^2	xy
	довжина міжвузля x, см	озерненість колоса y, шт.			
1	22,4	25	501,76	625	560,0
2	23,8	27	566,44	729	642,6
3	24,1	27	580,81	729	650,7
4	24,9	29	620,01	841	722,1
5	25,5	30	650,25	900	765,0
6	26,3	32	691,69	1024	841,6
7	27,4	33	750,76	1089	904,2
8	28,8	35	829,44	1225	1008,0
9	29,1	35	846,81	1225	1018,5
10	29,6	36	876,16	1296	1065,6
11	30,1	37	906,01	1369	1113,7
12	30,9	39	954,81	1521	1205,1
13	31,4	39	985,96	1521	1224,6
14	32,3	41	1043,29	1681	1324,3
15	33,5	42	1122,25	1764	1407,0
Сума	420,1	507	11926,45	17539	14453,0

Сума відхилень фактичних значень від розрахункових повинна дорівнювати нулю. У представленому прикладі $\sum d \neq 0$ ($\sum d = 4{,}5$), що зумовлено похибкою округлень.

3-й етап. На цьому етапі розраховують помилки, критерій значущості і довірчі інтервали для коефіцієнта кореляції r, регресії b_{yx}, і перевіряють нульову гіпотезу H_0:

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0{,}997^2}{15-2}} = 0{,}022;$$

$$s_b = s_r\sqrt{\frac{\sum(y-\overline{y})^2}{\sum(x-\overline{x})^2}} = 0{,}022 \cdot \sqrt{\frac{402{,}4}{160{,}85}} = 0{,}035 \text{ шт.};$$

$$s_{yx} = s_r\sqrt{\sum(y-\overline{y})^2} = 0{,}022 \cdot \sqrt{402{,}4} = 0{,}44 \text{ шт.};$$

$$t_{\text{факт.}} = r / s_r = 0{,}997 / 0{,}022 = 45{,}32;\ v = 15 - 2 = 13;\ t_{05} = 2{,}16;$$

$$r \pm t_{05} \cdot s_r = 0{,}997 \pm 2{,}16 \cdot 0{,}022 = 0{,}997 \pm 0{,}048 \text{ (від 0,949 до 1,000)};$$

Таблиця 107

Фактичні (y) і розраховані показники озерненості колоса залежно від довжини верхнього міжвузля (n = 15)

x, см	y, шт.	$\hat{y}$, шт.	$d = y - \hat{y}$	d^2
22,4	25	24,6	0,4	0,16
23,8	27	26,8	0,2	0,04
24,1	27	27,3	-0,3	0,09
24,9	29	28,6	0,4	0,16
25,5	30	29,5	0,5	0,25
26,3	32	30,8	1,2	1,44
27,4	33	32,5	0,5	0,25
28,8	35	34,7	0,3	0,09
29,1	35	35,2	-0,2	0,04
29,6	36	36,0	0,0	0,0
30,1	37	36,8	0,2	0,04
30,9	39	38,1	0,9	0,81
31,4	39	38,9	0,1	0,01
32,3	41	40,3	0,7	0,49
33,5	42	42,2	-0,2	0,04
Сума			$\sum d = 4{,}5$	$\sum d^2 = 3{,}91$

$$b_{yx} \pm t_{05} \cdot s_b = 1{,}58 \pm 2{,}16 \cdot 0{,}035 = 1{,}58 \pm 0{,}076 \text{ (від 1,50 до 1,66) шт.}$$

За t-критерієм ($t_{\text{факт.}} > t_{05}$) і довірчими інтервалами, які не включають нульового значення, кореляція та регресія значущі, отже, нульова гіпотеза на 5 %-му рівні спростовується.

4-й етап. За рівнянням регресії знаходять теоретичні значення y за мінімального і максимального значень x і будують теоретичну лінію регресії залежності y від x:

$$y_{(x-22{,}4)} = 1{,}58 \cdot 22{,}4 - 10{,}76 = 24{,}6 \text{ шт.};$$
$$y_{(x-33{,}5)} = 1{,}58 \cdot 33{,}5 - 10{,}76 = 42{,}2 \text{ шт.}$$

Знайдені координати (22,4; 24,6) і (33,5; 42,2) наносять на графік і, з'єднуючи їх прямою, отримують теоретичну лінію регресії y від x. Вона показує, що зі збільшенням довжини верхнього міжвузля на 1 см озерненість колоса збільшується на 1,58 шт. Відповідно до розрахованого коефіцієнта детермінації, близько 99 % змін озерненості колоса зумовлено довжиною верхнього міжвузля і лише на 1 %

змін пов'язані з іншими чинниками. На рисунку доцільно виділити довірчий інтервал для лінії регресії і вказати фактичні показники (рис. 34).

Для виділення довірчої зони необхідно вверх і вниз від теоретичної лінії регресії відкласти показник двох (95 %-ва зона) або трьох (99 %-ва зона) помилок відхилення від регресії, тобто $\pm 2s_{yx}$ або $\pm 3s_{yx}$, і з'єднати отримані точки граничними лініями інтервалів. Область, що знаходиться між цими лініями, зветься довірчою зоною регресії.

На рис. 34 пунктирними лініями обмежена 95 %-ва довірча зона розподілу показників відносно центральної лінії регресії, тобто зона у межах $y \pm 2s_{yx}$.

Загальна сума квадратів $\sum(y-\bar{y})^2$ може розкладатися на два компоненти: суму квадратів для регресії C_b і суму квадратів відхилень від регресії C_{byx}. Першу суму визначають за формулою:

$$C_b = [\sum(x-\bar{x})\cdot(y-\bar{y})]^2 / \sum(x-\bar{x})^2 = 253{,}6^2 / 160{,}85 = 399{,}8.$$

Другу суму квадратів знаходять за різницею:

$$C_{byx} = \sum(y-\bar{y})^2 - C_b = 402{,}4 - 399{,}8 = 2{,}60.$$

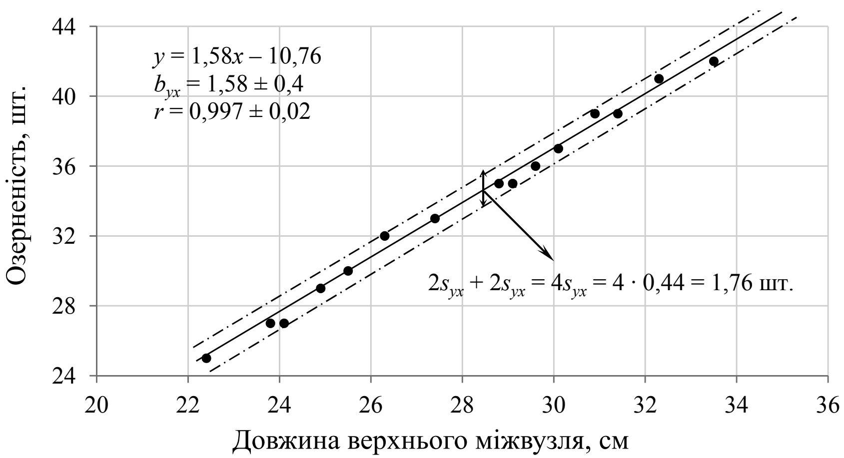


Рис. 34. Фактичний розподіл показників і теоретична лінія регресії при прямолінійній залежності між озерненістю колоса і довжиною міжвузля

Поділивши отримані суми квадратів на відповідні ступені свободи знаходять середні квадрати і визначають фактичний критерій F за яким перевіряють нульову гіпотезу щодо відсутності лінійного зв'язку між y і x. Розрахунки представляють у вигляді таблиці дисперсійного аналізу (табл. 108).

Таблиця 108

Дисперсійний аналіз *y*

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	$F_{теор.}$	
					05	01
Загальна	402,4	14	–	–	–	–
Регресії	399,8	1	399,8	1999	4,64	9,07
Відхилення від регресії	2,60	13	0,2	–	–	–

Відхилення від лінійності зумовлено вибірковим варіюванням, оскільки $F_{факт.} > F_{05}$, нульова гіпотеза H_0 про відсутність лінійного зв'язку між озерненістю колоса y і довжиною верхнього міжвузля x спростовується.

За середнім квадратом відхилення від регресії $s^2_{yx} = 0{,}2$ розраховують помилку відхилення від регресії s_{yx}. Вона дорівнює: $s_{yx} = \sqrt{s^2_{yx}} = \sqrt{0{,}2} = 0{,}45$ шт., тобто раніше розрахованому показнику.

Приклад 2. У фазу повної стиглості на 12-ти облікових ділянках визначили біологічну врожайність зерна y та довжину другого міжвузля x рослин тритикале ярого (табл. 109). Провести кореляційний та регресійний аналіз отриманих даних.

Проведення розрахунків включає кілька етапів.

1-й етап. Визначають середні показники $\bar{x}$ і $\bar{y}$, суми квадратів відхилень – $\sum(x-\bar{x})^2$ і $\sum(y-\bar{y})^2$ та загальну сума добутку відхилень обох ознак – $\sum(x-\bar{x})\,(y-\bar{y})$:

$$n = 12;$$
$$\bar{x} = (\sum x) / n = 46{,}1 / 12 = 3{,}84 \text{ см};$$
$$\bar{y} = (\sum y) / n = 38{,}43 / 12 = 3{,}20 \text{ т/га};$$
$$\sum(x-\bar{x})^2 = \sum x^2 - (\sum x)^2 / n = 182{,}25 - (46{,}1)^2 / 12 = 5{,}15;$$
$$\sum(y-\bar{y})^2 = \sum y^2 - (\sum y)^2 / n = 123{,}93 - (38{,}43)^2 / 12 = 0{,}86;$$
$$\sum(x-\bar{x})\,(y-\bar{y}) = \sum xy - (\sum x \cdot \sum y) / n = 145{,}57 - (46{,}1 \cdot 38{,}4) / 12 = -1{,}95.$$

2-й етап. Розраховують коефіцієнт кореляції r, детермінації d_{yx}, регресії b_{yx} і лінійне рівняння регресії:

Таблиця 109

Розрахунок допоміжних показників для визначення кореляції і регресії *y* за *x*

Номер пари	Ознаки		x^2	y^2	xy
	довжина другого міжвузля *x*, см	біологічна врожайність зерна *y*, т/га			
1	3,1	3,52	9,61	12,39	10,91
2	3,3	3,50	10,89	12,25	11,55
3	2,5	3,64	6,25	13,25	9,10
4	3,7	3,32	13,69	11,02	12,28
5	4,0	3,15	16,00	9,92	12,60
6	3,7	3,27	13,69	10,69	12,10
7	4,2	3,12	17,64	9,73	13,10
8	4,2	3,05	17,64	9,30	12,81
9	4,7	2,84	22,09	8,07	13,35
10	4,9	2,75	24,01	7,56	13,48
11	3,5	3,35	12,25	11,22	11,73
12	4,3	2,92	18,49	8,53	12,56
Сума	46,1	38,43	182,25	123,93	145,57

$$r = \frac{\sum(x-\overline{x})\,(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2\cdot\sum(y-\overline{y})^2}} = \frac{-1{,}95}{\sqrt{5{,}15\cdot 0{,}86}} = -0{,}929;$$

$$d_{yx} = r^2 = (-0{,}929)^2 = 0{,}863;$$

$$b_{yx} = \frac{\sum(x-\overline{x})\,(y-\overline{y})}{\sum(x-\overline{x})^2} = -\,1{,}95 \,/\, 5{,}15 = -0{,}38 \text{ т/га};$$

$$y = \overline{y} + b_{yx}\cdot(x-\overline{x}) = 3{,}20 + (-0{,}38)\cdot(x-3{,}84) = 4{,}66 - 0{,}38x.$$

3-й етап. На наступному етапі розраховують помилки, критерій значущості і довірчі інтервали для коефіцієнта кореляції r, регресії b_{yx}, і перевіряють нульову гіпотезу H_0:

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0{,}929^2}{12-2}} = 0{,}12;$$

$$s_b = s_r\sqrt{\frac{\sum(y-\overline{y})^2}{\sum(x-\overline{x})^2}} = 0{,}12\cdot\sqrt{\frac{0{,}86}{5{,}15}} = 0{,}05 \text{ т/га};$$

$$s_{yx} = s_r\sqrt{\sum(y-\overline{y})^2} = 0{,}12\cdot\sqrt{0{,}86} = 0{,}11 \text{ т/га};$$

$$t_{\text{факт.}} = r \,/\, s_r = 0{,}929 \,/\, 0{,}12 = 7{,}74;\ \nu = 12 - 2 = 10;\ t_{05} = 2{,}23;$$

$$r \pm t_{05}\cdot s_r = -0{,}929 \pm 2{,}23\cdot 0{,}12 = -0{,}929 \pm 0{,}268 \text{ (від } -1{,}0 \text{ до } -0{,}66);$$

$$b_{yx} \pm t_{05}\cdot s_b = -0{,}38 \pm 2{,}23\cdot 0{,}05 = -0{,}38 \pm 0{,}11 \text{ (від } -0{,}49 \text{ до } -0{,}27) \text{ шт.}$$

За t-критерієм ($t_{факт.} > t_{05}$) і довірчими інтервалами, які не включають нульового значення, кореляція та регресія значущі, отже, нульова гіпотеза на 5 %-му рівні спростовується.

4-й етап. За розрахованим рівнянням регресії $y = 4{,}66 - 0{,}38x$ знаходять теоретичні значення ознаки y за граничних показників ознаки x і будують лінію регресійної залежності y від x:

$$y_{(x-2{,}5)} = 4{,}66 - 0{,}38\cdot 2{,}5 = 3{,}71 \text{ т/га};$$
$$y_{(x-4{,}9)} = 4{,}66 - 0{,}38\cdot 4{,}9 = 2{,}80 \text{ т/га}.$$

Знайдені координати (2,5; 4,71) і (4,9; 2,80) наносять на графік і, з'єднуючи їх прямою, отримують теоретичну лінію регресії y від x (рис. 35).

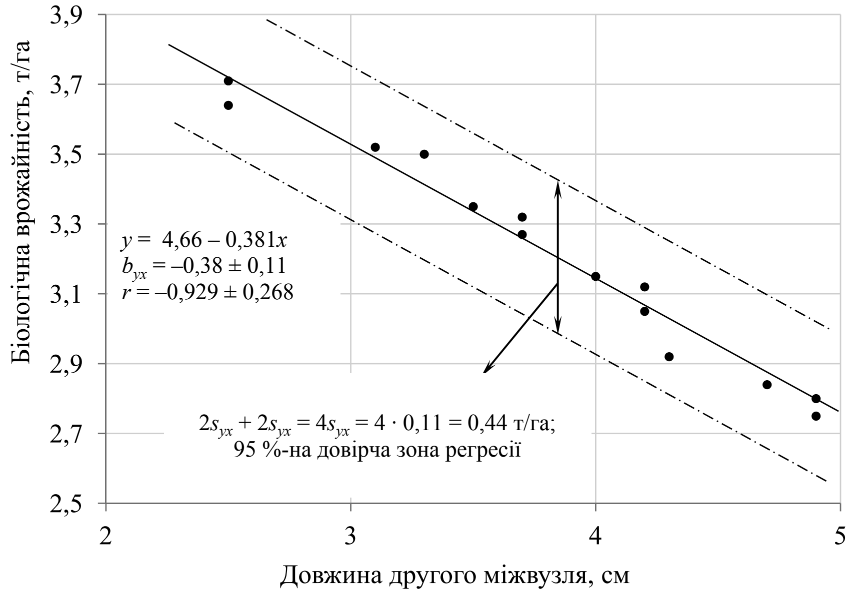


Рис. 35. Фактичний розподіл показників і теоретична лінія регресії при прямолінійній залежності між біологічною врожайністю зерна і довжиною другого міжвузля

Регресія вказує на зворотній зв'язок між досліджуваними ознаками. Вона показує, що зі збільшенням довжини другого міжвузля на 1 см біологічна врожайність зерна зменшиться на 0,38 т/га. Відповідно до розрахованого коефіцієнта детермінації, близько 86 % змін біологічної урожайності зерна зумовлено довжиною другого міжвузля і лише 14 % змін пов'язано з іншими чинниками. На рисунку доцільно виділити довірчий інтервал для лінії регресії, зазначити фактичні дані та показати основні статистичні показники. Фактичні дані, показані на рисунку крапками, достатньо компактно розосереджуються біля лінії прямолінійної регресії.

Приклад 3. Проведемо кореляційний і регресійний аналізи 54-х пар порівняння між висотою рослин *x* і довжиною колоса *y* головного стебла пшениці м'якої озимої (табл. 110).

Таблиця 110

Висота головного стебла *x* та довжина колоса *y* рослин пшениці озимої

Номер пари	*x*	*y*	Номер пари	*x*	*y*	Номер пари	*x*	*y*
1	**68**	**5,2**	19	91	7,4	37	88	7,3
2	101	9,8	20	104	10,1	38	90	7,6
3	83	6,6	21	96	8,5	39	109	10,4
4	75	6,0	22	73	5,7	40	85	7,1
5	82	6,4	23	91	7,7	41	97	8,4
6	95	8,1	24	94	7,8	42	107	10,3
7	86	7,4	25	76	6,0	43	80	6,3
8	88	7,4	26	87	6,9	44	97	8,9
9	90	7,8	27	96	9,2	45	108	**11,2**
10	74	5,7	28	101	9,7	46	88	7,1
11	92	8,0	29	78	6,2	47	79	6,1
12	99	8,2	30	**110**	10,6	48	101	9,8
13	85	7,5	31	74	5,7	49	81	6,3
14	80	6,1	32	91	8,0	50	90	7,2
15	93	8,4	33	74	5,7	51	106	10,4
16	82	6,6	34	102	10,2	52	69	5,2
17	94	8,0	35	80	6,2	53	71	5,4
18	99	8,8	36	97	8,4	54	96	8,2

Проведення розрахунків. 1-й етап. У цьому прикладі маємо справу з великою вибіркою ($n > 30$), тож доцільно згрупувати дані. За $n = 54$ виділяють 5–7 класів. Для рядів досліджуваних ознак визначають величину інтервалу і число класів:

$$i_x = (x_{max} - x_{min}) / 6 = (110 - 68) / 6 = 7{,}0 \text{ см};$$
$$i_y = (y_{max} - y_{min}) / 6 = (11{,}2 - 5{,}2) / 6 = 1{,}0 \text{ см}.$$

З табл. 110 вихідні дані переносять до табл. 111, визначають частоти показників у кожній групі та суми всіх частот за стовпцями f_x, строками f_y і загальне число результативних пар: $n = f_x = f_y = 54$. Щоб не помилитися під час визначення частот проявів ознак, використовують метод «штрихів» або «конвертиків».

Таблиця 111

Допоміжна таблиця для визначення частот прояву ознаки по кожній групі

x / *y*	Середні показники класів	68–74	75–81	82–88	89–95	96–102	103–110	Сума f_y
		Середні показники класів						
		71	78	85	92	99	106	
5,2–6,1	5,6	~~||||~~ ||	||||					11
6,2–7,1	6,6		||||	~~||||~~ |				10
7,2–8,1	7,6			||||	~~||||~~ ~~||||~~			14
8,2–9,1	8,6				|	~~||||~~ ||		8
9,2–10,1	9,6					||||	|	5
10,2–11,2	10,6					|	~~||||~~	6
Сума f_x		7	8	10	11	12	6	54

2-й етап. Складають розрахункову табл. 112, у якій замість границь класів проставляють їх модальні показники і перетворюють показники *x* і *y* за відношеннями:

$$x_1 = (x - A_x) / i_x = (x - 92{,}0) / 7{,}0;$$
$$y_1 = (y - A_y) / i_y = (y - 8{,}6) / 1{,}0.$$

За умовний початок A_x і A_y приймають близькі до модальних (середніх) значень ознак $\bar{x}$ і $\bar{y}$.

У розрахунковій табл. 112 визначають:

– добуток відхилень в одиницях інтервалу на їх частоту – $f_x x_1$ і $f_y y_1$ та їх суми $\sum f_x x_1 = -23$; $\sum f_y y_1 = -50$;

Таблиця 112

Розрахунок допоміжних показників для визначення кореляції і регресії *y* за *x*

$y_1 = (y - 89)/7$	$x_1 = (x - 7{,}7)/1$	–3	–2	–1	0	1	2	fy	$f_y y_1$	$f_y y_1^2$
	x / y	71	78	85	$A_x = 92$	99	106			
2	10,6	–	–	–	–	1	5	6	12	24
1	9,6	–	–	–	–	4	1	5	5	5
0	$A_y = 8{,}6$	–	–	–	1	7	–	8	0	0
–1	7,6	–	–	4	10	–	–	14	–14	14
–2	6,6	–	4	6	–	–	–	10	–20	40
–3	5,6	7	4	–	–	–	–	11	–33	99
fx		7	8	10	11	12	6	$\sum n = 54$	$\sum f_y y_1 = -50$	$\sum f_y y_1^2 = 182$
$f_x x_1$		–21	–16	–10	0	12	12	$\sum f_x x_1 = -23$		
$f_x x_1^2$		63	32	10	0	12	24	$\sum f_x x_1^2 = 141$		
$f x_1 y_1$		63	40	16	0	6	22	$\sum f x_1 y_1 = 147$		
$\bar{y}x$		5,6	6,1	7,0	7,7	9,1	10,4	–		

– добуток квадратів відхилень на їх частоти $f_x x_1^2$ і $f_y y_1^2$ та їх суми $\sum f_x x_1^2 = 141$ і $\sum f_y y_1^2 = 182$;

– суми добутків відхилень в інтервалах на їх частоту $f x_1 y_1$. Для цього частоту кожної групи f множать на відповідні показники y_1 і x_1, отримані дані складають. Наприклад, для першого стовпчика $f x_1 y_1 = = 7 \times (-3) \times (-3) = 63$; для другого стовпчика $f x_1 y_1 = 4 \cdot (-2) \cdot (-2) + 4 \times \times (-2) \cdot (-3) = 16 + 24 = 40$; для третього стовпчика $f x_1 y_1 = 4 \cdot (-1) \times \times (-1) + 6 \cdot (-1) \cdot (-2) = 4 + 12 = 16$ і т. ін. Потім знаходять загальну суму добутків $\sum fx\ i_x$ і i_y, $y_1 = (63 + 40 + 16 + 0 + 6 + 22) = 147$;

– середні значення $\bar{y}$ для кожної групи показників x. Наприклад, для першої групи показників x середнє значення $\bar{y}$ становить: $(7 \cdot 5{,}6) / 7 = 5{,}6$; для другої групи $\bar{y} = [(4 \cdot 6{,}6) + (4 \cdot 5{,}6)] / (4 + + 4) = = 6{,}1$; для третьої – $\bar{y} = [(4 \cdot 7{,}6) + (6 \cdot 6{,}6)] / (4 + 6) = 7{,}0$ і т. ін.;

– показники $\bar{x}$, $\bar{y}$, $\sum(x - \bar{x})$, $\sum(y - \bar{y})$ і $\sum(x - \bar{x})(y - \bar{y})$ у вихідних одиницях. Важливо враховувати, що якщо під час кодування

проводили ділення або множення на i_x і i_y, то суми квадратів у першому випадку слід помножити, а у другому поділити на i_x^2 і i_y^2; суму добутків відхилень у першому випадку слід помножити, а у другому поділити на $i_x \cdot i_y$.

$$n = 54;$$
$$\bar{x} = \sum x / n = (68 + 101 + ... + 96) /54 = 89{,}3 \text{ см};$$
$$\bar{y} = \sum y / n = (5{,}2 + 9{,}8 + ... + 8{,}2) /54 = 7{,}7 \text{ см};$$
$$\sum(x - \bar{x})^2 = i_x^2 \cdot [\sum f_x x_1^2 - (f_x x_1)^2 / n] = 7{,}0^2 \cdot (141 - (-23)^2 / 54) = 6428{,}8;$$
$$\sum(y - \bar{y})^2 = i_y^2 \cdot [\sum f_y y_1^2 - (f_y y_1)^2 / n] = 1{,}0^2 \cdot (182 - (-50)^2 / 54) = 135{,}7;$$
$$\sum(x - \bar{x})(y - \bar{y}) = i_x i_y \cdot [f x_1 y_1 - (\sum f_x x_1 \cdot \sum f_y y_1) / n] =$$
$$= 7{,}0 \cdot 1{,}0 \cdot (147 - (-23) \cdot (-50) / 54) = 879{,}9.$$

3-й етап. Розраховують вибірковий коефіцієнт кореляції, регресії і рівняння регресії ознаки y від x:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \cdot \sum(y - \bar{y})^2}} = \frac{879{,}9}{\sqrt{6428{,}8 \cdot 135{,}7}} = 879{,}9 / 934{,}0 = 0{,}942;$$
$$d_{yx} = r^2 = (0{,}942)^2 = 0{,}887;$$
$$b_{yx} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} = 879{,}9 / 6428{,}8 = 0{,}14 \text{ см};$$
$$y = \bar{y} + b_{yx} \cdot (x - \bar{x}) = 7{,}7 + 0{,}14 \cdot (x - 89{,}3) = -4{,}8 + 0{,}14x.$$

4-й етап. Знаходять помилки, критерій значущості, довірчі інтервали для коефіцієнта кореляції r і регресії b_{yx} та перевіряють нульову гіпотезу:

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - 0{,}942^2}{54 - 2}} = 0{,}047;$$
$$s_b = s_r \sqrt{\frac{\sum(y - \bar{y})^2}{\sum(x - \bar{x})^2}} = 0{,}047 \cdot \sqrt{\frac{135{,}7}{6428{,}8}} = 0{,}007 \text{ см};$$
$$s_{yx} = s_r \cdot \sqrt{\sum(y - \bar{y})^2} = 0{,}047 \cdot \sqrt{135{,}7} = 0{,}55 \text{ см};$$
$$t_{\text{факт.}} = r / s_r = 0{,}942 / 0{,}047 = 20{,}04;\ v = 54 - 2 = 52;\ t_{05} = 2{,}01;$$
$$r \pm t_{05} \cdot s_r = 0{,}942 \pm 2{,}01 \cdot 0{,}047 = 0{,}942 \pm 0{,}094 \text{ (від 0,848 до 1,000)};$$
$$b_{yx} \pm t_{05} \cdot s_b = 0{,}14 \pm 2{,}01 \cdot 0{,}007 = 0{,}14 \pm 0{,}014 \text{ (від 0,126 до 0,154) см}.$$

За t-критерієм ($t_{\text{факт.}} > t_{05}$) і довірчим інтервалам, які не включають нульового значення, кореляція та регресія значущі, отже, нульова гіпотеза на 5 %-му рівні спростовується.

5-й етап. За розрахованим рівнянням регресії $y = -4{,}8 + 0{,}14x$ знаходять теоретичні значення ознаки y за граничних показників ознаки x і будують лінію регресійної залежності y від x:

$$y_{(x-68)} = -4{,}8 + 0{,}14 \cdot 68 = 4{,}7 \text{ см};$$
$$y_{(x-110)} = -4{,}8 + 0{,}14 \cdot 110 = 10{,}6 \text{ см}.$$

Знайдені координати (68; 4,7) і (110; 10,6) наносять на графік і, з'єднуючи їх прямою, отримують теоретичну лінію регресії y від x (рис. 36). Регресія вказує на прямий зв'язок між досліджуваними ознаками. Вона показує, що зі збільшенням довжини рослин на 10 см довжина колоса головного стебла зросте на 1,4 см. Відповідно до розрахованого коефіцієнта детермінації, близько 88,7 % змін довжини колоса зумовлено висотою рослин і лише 11,3 % змін пов'язано з іншими чинниками. На рисунку пунктирними лініями показують 95 %-ву довірчу зону регресії.

Для перевірки гіпотези щодо лінійності зв'язку між y і x визначають суми квадратів для регресії C_b і відхилення від регресії C_{byx}:

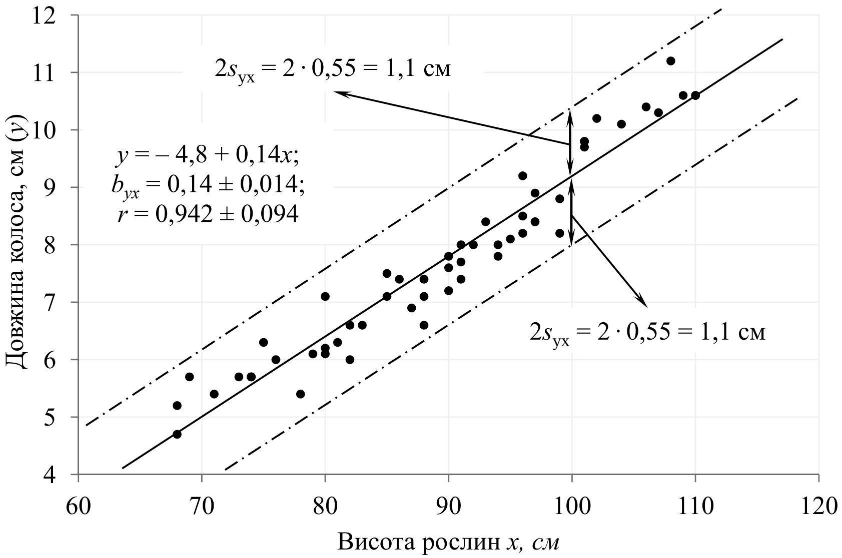


Рис. 36. Розподіл парних показників висоти рослин і довжини колоса рослин пшениці озимої навколо прямої лінії регресійної залежності між ними. Пунктирні лінії – 95 %-ва довірча зона регресії

$$C_b = [\sum(x-\bar{x})\,(y-\bar{y})]^2 / \sum(x-\bar{x})^2 = 879{,}9^2 / 6428{,}8 = 120{,}4.$$
$$C_{byx} = \sum(y-\bar{y})^2 - C_b = 135{,}7 - 120{,}4 = 15{,}3.$$

Поділивши отримані суми квадратів на відповідні ступені свободи, знаходять середні квадрати і визначають фактичний критерій F, за яким перевіряють нульову гіпотезу щодо відсутності лінійного зв'язку між y і x (табл. 113).

Таблиця 113

Дисперсійний аналіз *y*

Дисперсія	Сума квадратів	Ступінь свободи	Середній квадрат	$F_{факт.}$	F_{05}
Загальна	135,7	53	–	–	–
Регресії	120,4	1	120,40	415,1	4,03
Відхилення від регресії	15,3	52	0,29	–	–

Теоретичне значення $F_{05} = 4{,}03$ знаходять у дод. В для кількості ступенів свободи дисперсії чисельника, що дорівнює 1, і дисперсії знаменника – 52.

У цьому прикладі нульова гіпотеза щодо відсутності лінійного зв'язку між ознакою y і x спростовується ($F_{факт.} > F_{05}$), тож відхилення від лінійності зумовлено випадковим варіюванням показників.

Отримані статистичні результати свідчать, що між висотою рослин і довжиною колоса існує прямий тісний зв'язок, оскільки r для всієї генеральної сукупності знаходиться в межах інтервалу від 0,848 до 1. На підставі коефіцієнта детермінації ($d_{yx} = r^2 = 0{,}887$), можна зробити висновок про те, що близько 89 % змін довжини колоса зумовлено змінами висоти рослин і лише 11 % змін пов'язані з іншими чинниками.

За рівнянням $y = -4{,}8 + 0{,}14x$ для будь-яких значень x, включаючи ті, які відсутні у вихідних даних, теоретично можна визначити показник y. Разом із тим використовувати рівняння регресії для екстраполяції за межі даних таблиці не можна.

10.2. Множинна лінійна кореляція

Покажемо послідовність розрахунку коефіцієнта множинної лінійної кореляції на конкретному прикладі.

Приклад 4. На 12 паралельних ділянках досліду у фазу молочної стиглості визначили озерненість колосу головного стебла y, довжину верхнього міжвузля x і площу верхнього листка z пшениці м'якої ярої (табл. 114). Потрібно встановити залежність озерненості колоса від довжини верхнього міжвузля та площі верхнього листка $r_{y \cdot xz}$. Знаходити множинні коефіцієнти кореляції залежності довжини міжвузля від озерненості колоса та площі верхнього листка $r_{x \cdot yz}$, а також площі верхнього листка від озерненості колоса та довжини верхнього міжвузля $r_{z \cdot yx}$ недоцільно.

Проведення розрахунків. 1-й етап. Спочатку розраховують середні значення ознак, суми квадратів і суми добутків відхилень. За отриманими даними розраховують парні коефіцієнти кореляції – r_{yx}, r_{yz}, r_{xz} і коефіцієнт множинної кореляції $r_{y \cdot xz}$:

$$\bar{x} = 34{,}42;\ \bar{y} = 25{,}17;\ \bar{z} = 17{,}75;$$

Таблиця 114

Розрахунок допоміжних показників для визначення коефіцієнта множинної кореляції

Показники			x^2	y^2	z^2	xy	xz	yz
x, см	y, шт.	z, см^2						
31,3	24	17,1	979,6	576	292,4	751,2	535,2	410,4
27,6	20	15,4	761,8	400	237,2	552,0	425,0	308,0
42,8	31	20,5	1831,8	961	420,3	1326,8	877,4	635,5
32,2	23	17,7	1036,8	529	313,3	740,6	569,9	407,1
42,1	29	19,4	1772,4	841	376,4	1220,9	816,7	562,6
33,3	24	17,6	1108,9	576	309,8	799,2	586,1	422,4
29,4	21	16,1	864,3	441	259,2	617,4	473,3	338,1
37,2	27	18,7	1383,8	729	349,7	1004,4	695,6	504,9
29,7	22	16,4	882,1	484	269,0	653,4	487,1	360,8
39,0	29	18,8	1521,0	841	353,4	1131	733,2	545,2
36,1	25	17,8	1303,2	625	316,8	902,5	642,6	445,0
32,3	27	17,5	1043,3	729	306,3	872,1	565,3	472,5
$\sum x = 413{,}0$	$\sum y = 302$	$\sum z = 213{,}0$	$\sum x^2 = 14489{,}0$	$\sum y^2 = 7732$	$\sum z^2 = 3803{,}8$	$\sum xy = 10571{,}5$	$\sum xz = 7407{,}4$	$\sum yz = 5412{,}5$

$$\sum(x-\bar{x})^2 = \sum x^2 - (\sum x)^2 / n = 14489{,}0 - (413{,}0)^2 / 12 = 274{,}9;$$
$$\sum(y-\bar{y})^2 = \sum y^2 - (\sum y)^2 / n = 7732{,}0 - (302{,}0)^2 / 12 = 131{,}7;$$
$$\sum(z-\bar{z})^2 = \sum z^2 - (\sum z)^2 / n = 3803{,}8 - (213{,}0)^2 / 12 = 23{,}0;$$
$$\sum(x-\bar{x})(y-\bar{y}) = \sum xy - (\sum x \cdot \sum y) / n = 10571{,}5 - (413 \cdot 302) /12 = 177{,}7;$$
$$\sum(x-\bar{x})(z-\bar{z}) = \sum xz - (\sum x \cdot \sum z) / n = 7407{,}4 - (413 \cdot 213) /12 = 76{,}6;$$
$$\sum(y-\bar{y})(z-\bar{z}) = \sum yz - (\sum y \cdot \sum z) / n = 5412{,}5 - (302 \cdot 213) /12 = 52{,}0.$$

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \cdot \sum(y-\bar{y})^2}} = \frac{177{,}7}{\sqrt{274{,}9 \cdot 131{,}7}} = 177{,}7 / 190{,}3 = 0{,}934;$$

$$r_{xz} = \frac{\sum(x-\bar{x})\,(z-\bar{z})}{\sqrt{\sum(x-\bar{x})^2 \cdot \sum(z-\bar{z})^2}} = \frac{76{,}6}{\sqrt{274{,}9 \cdot 23{,}0}} = 76{,}6 / 79{,}5 = 0{,}964;$$

$$r_{yz} = \frac{\sum(y-y)\cdot(z-\bar{z})}{\sqrt{\sum(y-\bar{y})^2 \cdot \sum(z-\bar{z})^2}} = \frac{52{,}0}{\sqrt{131{,}7 \cdot 23{,}0}} = 52{,}0 / 55{,}0 = 0{,}945;$$

$$r_{y \cdot xz} = \sqrt{\left(r_{xy}^2 + r_{yz}^2 - 2r_{xy} \cdot r_{xz} \cdot r_{yz}\right) / (1 - r_{xz}^2)} =$$
$$= \sqrt{(\,0{,}934^2 + 0{,}945^2 - 2 \cdot 0{,}934 \cdot 0{,}964 \cdot 0{,}945) / (1 - 0{,}964^2)} = 0{,}942.$$

Критерій істотності F для коефіцієнта множинної кореляції визначається за рівнянням:

$$F = \frac{r^2}{1-r^2} \cdot \frac{(n-k)}{(k-1)} = \frac{0{,}942^2}{1-0{,}942^2} \cdot \frac{(12-3)}{(3-1)} = 35{,}3.$$

Теоретичне значення F_{05} за кількості ступенів свободи чисельника, що дорівнює 2 ($v_1 = k - 1 = 3 - 1 = 2$), і кількості ступенів свободи знаменника, що дорівнює 9 ($v_2 = n - k = 12 - 3 = 9$), становить 4,26. Оскільки $F_{факт.} > F_{05}$, нульова гіпотеза спростовується і множинний коефіцієнт кореляції істотно відрізняється від нуля.

Озерненість колоса тісно корелює з довжиною верхнього міжвузля і площею верхнього листка. Близько 89 % ($r_{y \cdot xz}^2 = 0{,}942^2 = 0{,}887$) варіативності озерненості колоса головного стебла зумовлено ознаками x і z.

2-й етап. Знаходять параметри a, b_1, b_2 і складають рівняння множинної регресії:

$$b_1 = \frac{\sum(z-\bar{z})^2 \cdot \sum(x-\bar{x})\,(y-\bar{y}) - \sum(x-\bar{x})\,(z-\bar{z}) \cdot \sum(y-\bar{y})\,(z-\bar{z})}{\sum(x-\bar{x})^2 \cdot \sum(z-\bar{z})^2 - [\sum(x-\bar{x}) \cdot (z-\bar{z})]^2} =$$
$$= \frac{23{,}0 \cdot 177{,}7 - 76{,}6 \cdot 52{,}0}{274{,}9 \cdot 23{,}0 - 76{,}6^2} = 0{,}23;$$

$$b_2 = \frac{\sum(x-x)^2 \cdot \sum(y-\overline{y})\,(z-\overline{z}) - \sum(x-\overline{x})\,(z-\overline{z}) \cdot \sum(x-\overline{x})\,(y-\overline{y})}{\sum(x-\overline{x})^2 \cdot \sum(z-\overline{z})^2 - [\sum(x-\overline{x})(z-\overline{z})]^2} =$$

$$= \frac{274{,}9 \cdot 52{,}0 - 76{,}6 \cdot 177{,}7}{274{,}9 \cdot 23{,}0 - 76{,}6^2} = 1{,}50;$$

$$a = \overline{y} - b_1 \cdot \overline{x} - b_2 \cdot \overline{z} = 25{,}17 - 0{,}23 \cdot 34{,}42 - 1{,}50 \cdot 17{,}75 = -9{,}4.$$

$$y = -9{,}4 + 0{,}23x + 1{,}5z.$$

3-й етап. Для фіксованих значень ознак x і z (табл. 115) за допомогою розрахованого рівняння множинної регресії знаходять значення ознаки y і переносять їх на рисунок та проводять через них лінії регресії (рис. 37). Розраховане рівняння об'єктивно показує тенденції зв'язків лише для показників x і z, які знаходяться в межах діапазону змін характерних для представлених у таблиці даних, тож у рівняння слід підставляти показники x і z, які не виходять за межі отриманих даних.

Таблиця 115

Озерненість колоса головного стебла пшениці м'якої ярої (шт. з колоса), розрахована за рівнянням $y = -9{,}4 + 0{,}23x + 1{,}5z$.

Довжина верхнього міжвузля x, см	Площа верхнього листка z, см²				
	16	17	18	19	20
30	21,5	23,0	24,5	26,0	27,5
34	22,4	23,9	25,4	26,9	28,4
38	23,3	24,8	26,3	27,8	29,3
42	24,3	25,8	27,3	28,8	30,3

10.3. Криволінійна кореляція і регресія

Універсальним показником тісноти зв'язку, незалежно від її форми, є кореляційне відношення η. Перед його розрахунком вибірку (парні спостереження) «вирівнюють» у порядку зростання за факторною ознакою x і розбивають на 4...8 груп. У кожній групі має бути не менше двох пар. За $n < 20$ користуються формулою:

$$\eta_{yx} = \sqrt{\frac{\sum(y-\overline{y})^2 - \sum(y-\overline{y}_x)^2}{\sum(y-\overline{y})^2}},$$

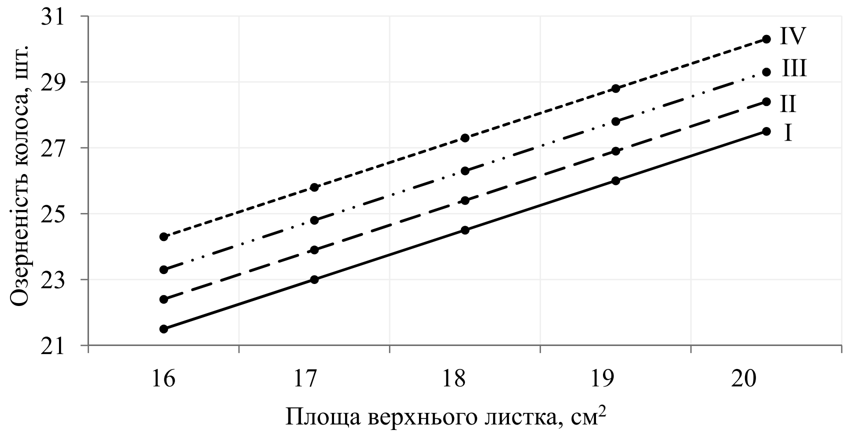


Рис. 37. Залежність між озерненістю колоса, довжиною верхнього міжвузля та площею верхнього листка рослин пшениці м'якої ярої. *Умовні позначення*: варіанти довжини міжвузля: I – 30 см, II – 34 см, III – 38 см, IV – 42 см

де $\sum(y-\overline{y})^2$ – сума квадратів відхилень окремих значень результативної ознаки y_i від середнього показника $\overline{y}$ (загальне варіювання); $\sum(y-\overline{y}_x)^2$ – сума квадратів відхилень внутрішньогрупових значень y від їх середньої $\overline{y}_x$ (варіювання в середині груп).

Для вибірок обсягом понад 20 ($n > 20$) доцільно підготувати кореляційну решітку, тоді формула буде мати вигляд:

$$\eta_{yx} = \sqrt{\sum f(\overline{y}_x - \overline{y})^2 / \sum f(y-\overline{y})^2},$$

де f – число груп (класів); $\sum(\overline{y}_x - \overline{y})^2$ – сума квадратів відхилень від загальної середньої (міжгрупове варіювання). Воно характеризує ту частку варіації результативної ознаки y, яка зумовлена мінливістю ознаки x; $\sum(y-\overline{y})^2$ – сума квадратів відхилень окремих значень від загальної середньої (загальне варіювання ознаки y).

Квадрат кореляційного відношення η_{yx}^2 найбільш точно показує частку мінливості результативної ознаки, зумовлену варіабельністю факторної ознаки для будь-якої форми зв'язку. Це відношення називають індексом кореляції і виражають у відсотках:

$$d_{yx}, \% = \eta_{yx}^2 \cdot 100.$$

Якщо прогнозується лінійний характер зв'язку між ознаками, то розрахунок проводять за формулою:

$$r = (\overline{xy} - \bar{x} \cdot \bar{y}) / (s_x \cdot s_y),$$

де $\bar{x}$ і $\bar{y}$ – середні вибіркові ознак; $\overline{xy}$ – середній добуток їх показників ($\sum xy/n$); s_x і s_y – стандартні відхилення ознак.

Дисперсії стандартних відхилень ознак визначають за формулами:

$$s_x^2 = \sum x^2 / n - (\bar{x})^2; \; s_y^2 = \sum y^2 / n - (\bar{y})^2.$$

Приклад 1. Під час аналізу залежності кількості зерен у колосі ячменю ярого (y, шт./колос), від глибини загортання насіння (x, см) були отримані такі пари спостережень (для зручності показники ознаки x вирівнюють):

y	18	17	20	25	29	31	27	27	26	25	23	21	18	16	12
x	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0	6,5	7,0	7,5	8,0	8,5	9,0	9,5

Для розрахунку показників лінійної, а далі криволінійної залежності складають табл. 116.

$$s_y^2 = \sum y^2 / n - (\bar{y})^2 = 526{,}2 - (22{,}3)^2 = 28{,}91; \; s_y = \sqrt{28{,}91} = 5{,}38;$$
$$s_x^2 = \sum x^2 / n - (\bar{x})^2 = 40{,}67 - (6{,}0)^2 = 4{,}67; \; s_x = \sqrt{4{,}67} = 2{,}16.$$
$$r = (\overline{xy} - \bar{x} \cdot \bar{y}) / (s_y \cdot s_x) = (130{,}5 - 6{,}0 \cdot 22{,}3) / (5{,}38 \cdot 2{,}16) =$$
$$= (130{,}5 - 133{,}8) / (11{,}62) = -0{,}28.$$

Коефіцієнт парної кореляції вказує на досить слабий зв'язок між досліджуваними ознаками, тож для характеристики залежності потрібно розрахувати кореляційне відношення. Для оцінки криволінійної залежності вибірку розбивають на групи з подальшим розрахунком середніх групових і відхилень від них. У нашому прикладі вибірку доцільно поділити на п'ять рівних за розміром груп, у кожній групі по три порівнювані пари (табл. 117).

Знаходимо індекс детермінації (d_{yx}) і кореляційне відношення (η_{yx}):

$$d_{yx} = \eta_{yx}^2 \cdot 100 = \frac{\sum(y-\bar{y})^2 - \sum(y-\bar{y}_x)^2}{\sum(y-\bar{y})^2} \cdot 100 =$$
$$= \frac{411{,}45 - 50{,}28}{411{,}45} \cdot 100 = 361{,}17 \cdot 100 / 411{,}45 = 87{,}8\ \%;$$
$$\eta_{yx} = \sqrt{\eta_{yx}^2} = \sqrt{0{,}8778} = 0{,}937.$$

Висновок. Між кількістю зерен у колосі ячменю ярого y і глибиною загортання насіння під час сівби x існує сильна залежність.

Таблиця 116

Робоча таблиця підготовки показників озерненості колоса ячменю і глибини загортання насіння для оцінки тісноти зв'язку між ними (n = 15)

Озерненість, колоса, шт.	Глибина загортання насіння, см	y^2	x^2	x^3	x^4	yx	yx^2
18	2,5	324	6,25	15,63	39,1	45,0	112,5
17	3,0	289	9,00	27,00	81,0	51,0	153,0
20	3,5	400	12,25	42,88	150,1	70,0	245,0
25	4,0	625	16,00	64,00	256,0	100,0	400,0
29	4,5	841	20,25	91,13	410,1	130,5	587,3
31	5,0	961	25,00	125,00	625,0	155,0	775,0
27	5,5	729	30,25	166,38	915,1	148,5	816,8
27	6,0	729	36,00	216,00	1296,0	162,0	972,0
26	6,5	676	42,25	274,63	1785,1	169,0	1098,5
25	7,0	625	49,00	343,00	2401,0	175,0	1225,0
23	7,5	529	56,25	421,88	3164,1	172,5	1293,8
21	8,0	441	64,00	512,00	4096,0	168,0	1344,0
18	8,5	324	72,25	614,13	5220,1	153,0	1300,5
16	9,0	256	81,00	729,00	6561,0	144,0	1296,0
12	9,5	144	90,25	857,38	8145,1	114,0	1083,0
$\sum y = 335$	$\sum x = 90{,}0$	$\sum y^2 = 7893$	$\sum x^2 = 610{,}0$	$\sum x^3 = 4500{,}04$	$\sum x^4 = 35144{,}8$	$\sum yx = 1957{,}5$	$\sum yx^2 = 12702{,}4$
$\bar{y} = 22{,}3$	$\bar{x} = 6{,}0$	$\bar{y}^2 = 526{,}2$	$\bar{x}^2 = 40{,}67$	$\bar{x}^3 = 300{,}0$	$\bar{x}^4 = 2343{,}0$	$\overline{yx} = 130{,}5$	$\overline{yx}^2 = 846{,}8$

Понад 90 % змін озерненості колоса зумовлено зміною глибини загортання насіння. Наскільки правильно характеристики цієї вибірки (η_{yx} і d_{yx}) відображають параметри сукупності, встановлюють за допомогою критерію істотності t_η, однак спочатку розраховують помилку вибірки (помилку кореляційного відношення):

Таблиця 117

Робоча таблиця підготовки вихідних даних для встановлення тісноти криволінійної залежності (n = 15, k = 5, $\bar{y}$ = 22,3)

x, см	y, шт.	$\bar{x}_y$, т/га	$\bar{y}_x$, см	$y-\bar{y}_x$	$y-\bar{y}$	$(y-\bar{y}_x)^2$	$(y-\bar{y})^2$
2,5	18			– 0,3	– 4,3	0,09	18,49
3,0	17	3,0	18,3	– 1,3	– 5,3	1,69	28,09
3,5	20			1,7	– 2,3	2,89	5,29
4,0	25			– 3,3	2,7	10,89	7,29
4,5	29	4,5	28,3	0,7	6,7	0,49	44,89
5,0	31			2,7	8,7	7,29	75,69
5,5	27			0,3	4,7	0,09	22,09
6,0	27	6,0	26,7	0,3	4,7	0,09	22,09
6,5	26			– 0,7	3,7	0,09	13,69
7,0	25			2,0	2,7	4,0	7,29
7,5	23	7,5	23,0	0,0	0,7	0,0	0,49
8,0	21			– 2,0	– 1,3	4,0	1,69
8,5	18			2,7	– 4,3	7,29	18,49
9,0	16	9,0	15,3	0,7	– 6,3	0,49	39,69
9,5	12			– 3,3	– 10,3	10,89	106,09
	$\sum y$ = = 335			$\sum(y-\bar{y}_x)$ = = 0,0	$\sum(y-\bar{y})$ = = 0,5	$\sum(y-\bar{y}_x)^2$ = = 50,28	$\sum(y-\bar{y})^2$ = = 411,45

$$s_\eta = \sqrt{\frac{1-\eta_{yx}^2}{n-2}} = \sqrt{(1-0{,}878)/(15\ -2)} = \sqrt{0{,}122/13}\ = 0{,}097;$$

$$t_{факт.} = \eta_{yx} / s_\eta = 0{,}937 / 0{,}097 = 9{,}6.$$

Теоретичне значення t_{01} для 13 ступенів свободи становить 3,01 (дод. Б). Оскільки $t_{факт.} > t_{01}$, достовірність зв'язку досліджуваних ознак встановлена з імовірністю 99 %. Висновок, зроблений на основі вибірки, поширюється на всю сукупність.

Реєстрація всіх пар спостережень (n = 15) на графіку дає діаграму розподілу показників. Емпіричну лінію розподілу будують за показниками середньогрупових показників ознак (у нашому випадку п'ять точок). Цим точкам на графіку відповідають такі координати x і y: 3,0 і 18,3; 4,5 і 28,3; 6,0 і 26,7; 7,5 і 23,0; 9,0 і 15,3 (рис. 38).

З урахуванням форми фактичної кривої для побудови прогнозованої лінії теоретично очікуваних показників x і y методом найменших квадратів вирішують рівняння параболи II порядку: $y = a + bx + cx^2$. Знаходження параметрів, або коефіцієнтів оціночного рівняння знач-

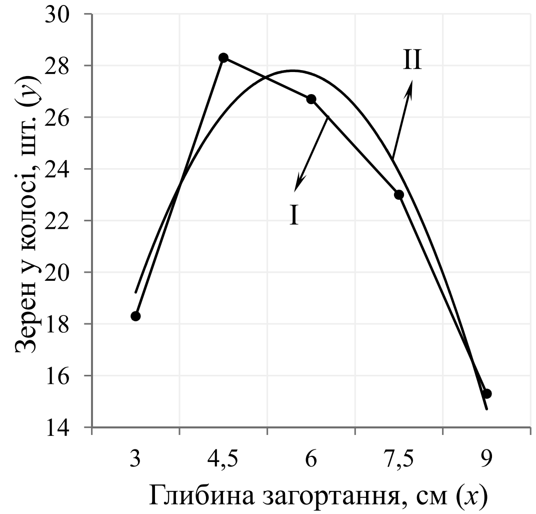


Рис. 38. Криволінійна залежність (η_{yx} = 0,937, α = 0,01, n = 15) між озерненістю колоса ячменю ярого (y) і глибиною загортання насіння (x). Емпірична крива (I) розрахована на основі п'яти пар середньо-групових показників x_y і y_x. Прогнозована крива (поліном другого порядку) – II, розрахована на основі рівняння: $y = -10{,}66 + 13{,}523x - 1{,}191x^2$

но спрощується при використанні відповідної комп'ютерної програми. У нашому прикладі оціночне рівняння має вигляд:

$$y = -10{,}66 + 13{,}532x - 1{,}191x^2.$$

Підставивши в це рівняння показники факторної ознаки – глибини загортання насіння x, отримуємо прогнозовані показники факторіальної ознаки – кількості зерен у колосі y:

x_i 2,5 3,0 3,5 4,0 4,5 5,0 5,5 6,0 6,5 7,0 7,5 8,0 8,5 9,0 9,5
y_i 15,7 19,2 22,1 24,4 26,1 27,2 27,7 27,6 26,8 25,7 23,8 21,7 18,3 14,6 10,4.

Таким чином, правильний вибір форми зв'язку дозволив визначити її тісноту з η = 0,936. Сильну криволінійну залежність озерненості колоса від глибини загортання насіння супроводжують, як правило, сильні лінійні залежності озерненості від погодних умов, рівня агротехніки і їх поєднання в різні роки. Розрахований рівень озерненості у визначеному діапазоні відображає типову варіацію цього показника для ячменю в умовах Лівобережного Лісостепу України.

Приклад 2. Потрібно розрахувати кореляційне відношення для великої вибірки (табл. 118), перевірити нульову гіпотезу H_0 та представити результати статистичної обробки у вигляді графіка.

Проведення розрахунків. Для проведення розрахунків вихідні дані групують і подають у вигляді кореляційної таблиці. При загаль-

Таблиця 118

Зв'язок між діаметром стебла (*x*, мм) та масою зерна (*y*, г) рослин кукурудзи

Номер пари	Ознака *x*	Ознака *y*	Номер пари	Ознака *x*	Ознака *y*	Номер пари	Ознака *x*	Ознака *y*
1	35	232	16	47	305	31	56	348
2	35	242	17	47	311	32	56	347
3	36	251	18	47	308	33	58	350
4	37	258	19	48	321	34	58	352
5	38	261	20	49	325	35	59	352
6	39	270	21	50	330	36	61	354
7	39	273	22	50	327	37	61	349
8	40	280	23	51	335	38	63	356
9	40	283	24	52	341	39	63	351
10	40	284	25	53	346	40	63	358
11	41	290	26	53	342	41	64	357
12	42	292	27	53	338	42	64	360
13	44	298	28	54	342	43	64	361
14	44	301	29	54	345	44	66	364
15	45	309	30	55	342	45	65	359

ній кількості порівнюваних пар, рівній 45 ($n = 45$), доцільно виділити 6–7 класів. Для обох рядів знаходять величину інтервалу групування і число груп:

$$i_x = (x_{\text{макс}} - x_{\text{мін}}) / 6\text{–}7 = (65 - 35) / 6\text{–}7 = 30 / 6 = 5 \text{ мм};$$
$$i_y = (y_{\text{макс}} - y_{\text{мін}}) / 6\text{–}7 = (364 - 232) / 6\text{–}7 = 132 / 6 = 22 \text{ г}.$$

Далі методом «рисок» або «конвертиків» результати табл. 118 переносять до кореляційної табл. 119. Правильність заповнення таблиці перевіряють за рівнянням $\sum f_x = \sum f_y = n = 45$.

Характер розподілу частот свідчить, що зв'язок між діаметром стебла та масою зерна з рослини носить криволінійний характер, тож у цьому прикладі коефіцієнт лінійної кореляції не можна використовувати для визначення тісноти зв'язку.

Для проведення подальших розрахунків складають розрахункову табл. 120, у якій замість меж класів проставляють їх середньо-групові значення та перетворюють показники *x* і *y* за відношеннями:

Таблиця 119

Розподіл показників таблиці 118 за групами

<table>
<tr><td rowspan="3">x / y</td><td rowspan="3">Середні показники у класах</td><td>35–39</td><td>40–44</td><td>45–49</td><td>50–54</td><td>55–59</td><td>60–65</td><td rowspan="3">Сума f_y</td></tr>
<tr><td colspan="6">Середні показники у класах</td></tr>
<tr><td>37</td><td>42</td><td>47</td><td>52</td><td>57</td><td>62</td></tr>
<tr><td>232–253</td><td>242,5</td><td>|||</td><td></td><td></td><td></td><td></td><td></td><td>3</td></tr>
<tr><td>254–275</td><td>264,5</td><td>||||</td><td></td><td></td><td></td><td></td><td></td><td>4</td></tr>
<tr><td>276–297</td><td>286,5</td><td></td><td>||||/</td><td></td><td></td><td></td><td></td><td>5</td></tr>
<tr><td>298–319</td><td>308,5</td><td></td><td>||</td><td>||||</td><td></td><td></td><td></td><td>6</td></tr>
<tr><td>320–341</td><td>330,5</td><td></td><td></td><td>||</td><td>||||/</td><td></td><td></td><td>7</td></tr>
<tr><td>342–364</td><td>353,0</td><td></td><td></td><td></td><td>||||</td><td>||||/ |</td><td>||||/ ||||/</td><td>20</td></tr>
<tr><td colspan="2">Сума f_x</td><td>7</td><td>7</td><td>6</td><td>9</td><td>6</td><td>10</td><td>$n = 45$</td></tr>
</table>

Таблиця 120

Розрахунок допоміжних показників для визначення кореляційного відношення

y_1	x_1	– 3	– 2	– 1	0	1	2	fy
	x / y	37	42	47	$A_x = 52$	57	62	
3	242,5	3	–	–	–	–	–	3
2	264,5	4		–	–	–	–	4
1	286,5	–	5	–	–	–	–	5
0	$A_y = 308{,}5$	–	2	4	–	–	–	6
–1	330,5	–	–	2	5	–	–	7
–2	353,0	–	–	–	4	6	10	20
f_x		7	7	6	9	6	10	$n = 45$
Сума f_{y1}		17	5	– 2	–13	–12	–20	$\sum f_{y1} = 25$
Групові середні $\bar{y}_{x1}$		2,43	0,71	–0,33	–1,44	–2,00	–2,00	$\bar{y}_1 = \sum f_{y1}/n = 0{,}56$
$(\bar{y}_{x1} - y)$		1,87	0,15	–0,89	–2,00	–2,56	–2,56	
$f_x(\bar{y}_{x1} - \bar{y}_1)^2$		24,48	0,16	4,75	36,00	39,32	65,50	$\sum f_x(\bar{y}_{x1} - \bar{y}_1)^2 = 170{,}21$
Сума $f(y_1 - \bar{y}_1)^2$		26,15	1,28	6,12	38,38	39,32	65,50	$\sum f(y_1 - \bar{y}_1)^2 = 176{,}75$
$\bar{y}_x$		255,1	292,8	315,8	340,5	353,0	353,0	$\bar{y} = \sum y / n = 14400/45=320$

$$x_1 = (x - A_x) / i_x = (x - 50) / 5;$$
$$y_1 = (y - A_y) / i_y = (y - 320) / 22.$$

За умовні початки A_x і Ay беруть значення, близькі до загальних середніх показників ознак. Так, при $\bar{x} = 50{,}6$ за умовний початок A_x доцільно взяти показник 52, а при $\bar{y} = 320{,}0$ за умовний початок A_y доцільно взяти показник 308,5.

У допоміжній таблиці визначають:

1) групові середні $\bar{y}_{x1}$ для кожного фіксованого значення x_1 і загальну середню $\bar{y}_{x1}$ в одиницях інтервалу. Наприклад, для $x_1 = 3$ показник $\bar{y}_{x1} = f_{y1} / f_x = 17 / 7 = 2{,}43$; для $x_1 = 2$ показник $\bar{y}_{x1} = f_{y1} / f_x = 5 / 7 = 0{,}71$ і т. п.;
2) суму квадратів відхилень середніх показників класів від загальної середньої. Для цього для кожної групи знаходять відхилення $(\bar{y}_{x1} - \bar{y}_1)$, квадрати відхилень $(\bar{y}_{x1} - \bar{y}_1)^2$ і суму квадратів $\sum f_x(\bar{y}_{x1} - \bar{y}_1)^2 = 24{,}48 + 0{,}16 + 4{,}75 + 36{,}0 + 39{,}32 + 65{,}50 = 170{,}21$.
3) суму квадратів відхилень загального варіювання $\sum f(y_1 - \bar{y}_1)^2$. Для цього у кожній клітині класу знаходять різницю між показником y_1 і $\bar{y}_1$, зводять її у квадрат, множать на частоту f даної клітини і складають по кожній групі. Наприклад, для першого класу за ознакою x сума $f(y_1 - \bar{y}_1)^2 = 3 \cdot (3 - 0{,}56)^2 + 4 \cdot (2 - 0{,}56)^2 = 17{,}86 + 8{,}29 = 26{,}15$; для другого класу – $f(y_1 - \bar{y}_1)^2 = 5 \cdot (1 - 0{,}56)^2 + 2 \cdot (0 - 0{,}56)^2 = 0{,}97 + + 0{,}31 = 1{,}28$ і т. ін. Далі знаходять загальну суму квадратів $\sum f(y_1 - \bar{y}_1)^2 = (26{,}15 + 1{,}28 + 6{,}12 + 38{,}38 + 39{,}32 + 65{,}5) = 176{,}75$;
4) середні групові значення $\bar{y}_x$ для кожного класу ознаки x за рівнянням: $\sum y_x / f_x$. Наприклад, для першого класу ознаки x середньо-групове значення $\bar{y}_x = \sum[(3 \cdot 242{,}5) + (4 \cdot 264{,}5)] / 7 = 255{,}1$, для другого класу – $\bar{y}_x = \sum[(5 \cdot 286{,}5) + (2 \cdot 308{,}5)] / 7 = 292{,}8$ і т. ін.

На наступному етапі розраховують вибіркове кореляційне відношення η_{yx}, його помилку s_η, критерій істотності і довірчий інтервал для кореляційного відношення всієї сукупності:

$$\eta_{yx}^2 = \frac{\sum f_x(\bar{y}_{x1} - \bar{y}_1)^2}{\sum f(y_1 - \bar{y}_1)^2} = 170{,}21 / 176{,}75 = 0{,}963\ (96{,}3\ \%);$$
$$\eta_{xy} = \sqrt{\eta_{xy}^2} = \sqrt{0{,}963} = 0{,}981;$$

$$s_\eta = \sqrt{\frac{1 - \eta_{xy}^2}{n - 2}} = \sqrt{\frac{1 - 0{,}981^2}{45 - 2}} = \sqrt{0{,}00088} = 0{,}03;$$
$$t_{\text{факт.}} = \eta_{xy} / s_\eta = 0{,}981 / 0{,}03 = 32{,}7.$$

Для представленої вибірки ($n = 45$) число її ступенів свободи (υ) становить: $n - 2 = 45 - 2 = 43$, тож теоретичний критерій $t_{05} = 2{,}01$ (дод. Б). Оскільки $t_{\text{факт.}} > t_{05}$, нульова гіпотеза ($H_0$) на 5 %-му рівні спростовується.

Довірчий інтервал для кореляційного відношення становить:

$$\eta_{xy} \pm t_{05} \cdot s_\eta = 0{,}981 \pm 2{,}01 \cdot 0{,}03 =$$
$$= 0{,}981 \pm 0{,}061\ (\text{або від } 0{,}920 \text{ до } 1{,}0).$$

Реєстрація всіх пар спостережень ($n = 45$) на графіку показує характер розподілу показників. Теоретичну лінію розподілу будують за показниками середньогрупових показників ознак (у нашому прикладі – шість точок, позначених хрестиком). Цим точкам на графіку відповідають такі координати пар x і y: 37 і 245,5; 42 і 264,5; 47 і 286,5; 52 і 308,5; 57 і 330,5; 62 і 353,5 (рис. 39).

Теоретична лінія показує, що маса зерна з однієї рослини зростає пропорційно збільшенню діаметра стебла рослин кукурудзи приблизно до 51–52 мм. Подальше збільшення діаметра стебла не сприяє збільшенню маси зерна з рослини.

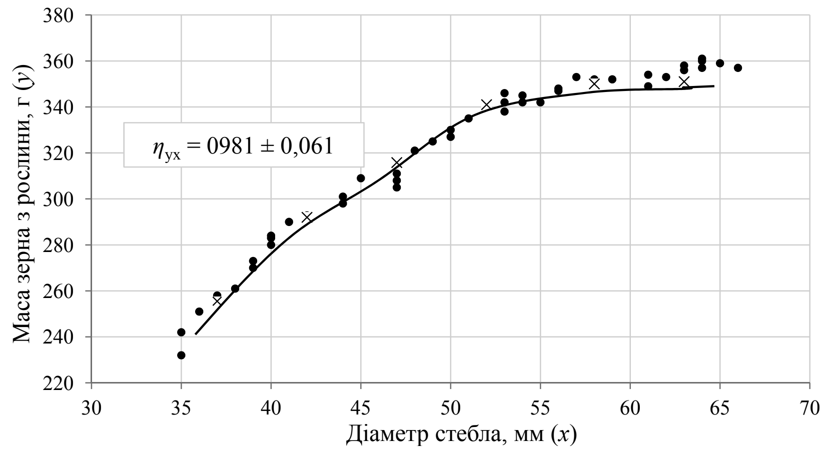


Рис. 39. Залежність маси зерна з однієї рослини кукурудзи (y) від діаметра її стебла (x)

11. Факторіальний аналіз

Основна мета факторіального аналізу полягає у визначенні частки впливу досліджуваних чинників та їх взаємодії на варіабельність певної досліджуваної ознаки, без урахування впливу повторень і помилок. Факторіальний аналіз дає чітке уявлення стосовно вкладу кожного з досліджуваних чинників у загальну варіабельність досліджуваного показника, виокремлюючи дію решти чинників. Разом із тим він не може дати відповідь на запитання – є чи немає істотної різниці між варіантами досліджуваних чинників. Тож факторіальний аналіз застосовують у комплексі з дисперсійним аналізом. Лише на підставі проведеного дисперсійного аналізу, навіть за великої частки певного чинника, можна стверджувати про його ефективність.

Типові формули для визначення ефекту чинників різних порядків та їх взаємодії представлені в табл. 121.

Розглянемо механізм проведення факторіального аналізу на прикладі двофакторного польового досліду.

Приклад. У двофакторному досліді з вивчення впливу норм висіву та способів сівби отримали такі показники врожайності насіння ріпака ярого (табл. 122). Проведемо факторіальний аналіз і визначимо вклади чинників у варіабельність досліджуваної ознаки.

Проведення розрахунків. Факторіальний аналіз проводять в кілька етапів.

1-й етап. Визначають середні показники врожайності насіння за варіантами чинників і загальну середню врожайність у досліді $\bar{y}$. Так, середня врожайність насіння за рядкового способу сівби становить $\bar{y}_p$ = (1,81 + 2,01 + 1,70) / 3 = 1,84 т/га; за норми висіву 1,0 млн /га – $\bar{y}_{1,0}$ = = (1,81 + 1,66 + 1,74) / 3 = 1,74 і т. ін. Результати заносять до табл. 122.

2-й етап. Визначають головні ефекти чинників *A* і *B*. Для цього від певного варіанта кожного чинника віднімають середню врожайність у досліді. Наприклад, головний ефект рядкового способу сівби (чинник *A*) становить – $Ey_{A1} = y_{A1} - \bar{y}$ = 1,84 – 1,70 = 0,14 т/га; широкорядного способу сівби (міжряддя 30 см) – $Ey_{A2} = y_{A2} - \bar{y}$ = 1,65 – 1,70 = = 0,05 т/га і т. п. У табл. 122 головні ефекти чинників виділені жирним шрифтом. Сума головних ефектів кожного чинника має спрямо-

Таблиця 121

Базові формули для знаходження ефектів чинників для дво- і трифакторного польового досліду

Досліджувані чинники та їх взаємодія		Формули для визначення ефекту
Для двофакторного досліду	Чинник *A*	$Ey_{Ai} = y_{Ai} - \bar{y}$
	Чинник *B*	$Ey_{Bi} = y_{Bi} - \bar{y}$
	Взаємодія *AB*	$Ey_{ABi} = y_{ABi} - \bar{y} - Ey_{Ai} - Ey_{Bi}$
Для трифакторного досліду	Чинник *A*	$Ey_{Ai} = y_{Ai} - \bar{y}$
	Чинник *B*	$Ey_{Bi} = y_{Bi} - \bar{y}$
	Чинник *C*	$Ey_{Ci} = y_{Ci} - \bar{y}$
	Взаємодія *AB*	$Ey_{AiBi} = y_{AiBi} - \bar{y} - Ey_{Ai} - Ey_{Bi}$
	Взаємодія *AC*	$Ey_{AiCi} = y_{AiCi} - \bar{y} - Ey_{Ai} - Ey_{Ci}$
	Взаємодія *BC*	$Ey_{BiCi} = y_{BiCi} - \bar{y} - Ey_{Bi} - Ey_{Ci}$
	Взаємодія *ABC*	$Ey_{AiBiCi} = y_{AiBiCi} - \bar{y} - Ey_{Ai} - Ey_{Bi} - Ey_{Ci} - Ey_{AiBi} - Ey_{AiCi} - Ey_{BiCi}$

Символами в таблиці позначені: $\bar{y}$ – середній показник досліджуваної ознаки в досліді; y_{Ai} , y_{Bi} , y_{Ci} – показник *i*-тої градації головного чинника *A, B* і *C* відповідно; y_{AiBi}, y_{AiCi}, y_{BiCi} – показники *i*-тої градації взаємодії чинників *AB*, *AC* і *BC* відповідно; y_{AiBiCi} – показник *i*-тої градації ефекту другого порядку (взаємодії чинників *ABC*).

Таблиця 122

Урожайність насіння ріпака ярого залежно від впливу способів сівби (чинник *A*) та норм висіву (чинник *B*), т/га

Чинник *A* (спосіб сівби)	Чинник *B* (норма висіву, млн нас./га)						Середнє за чинником *A*	
	1,0		2,0		3,0			
	*y**	*E***	*y*	*E*	*y*	*E*	*y*	*E*
Рядковий, 15 см	1,81	*–0,07*	2,01	*0,09*	1,70	*0,01*	1,84	**0,14**
Широкорядний, 30 см	1,66	*–0,03*	1,71	*–0,02*	1,59	*0,07*	1,65	**–0,05**
Широкорядний, 45 см	1,74	*0,10*	1,62	*–0,06*	1,43	*–0,04*	1,60	**–0,10**
Середнє за чинником *B*	1,74	**0,04**	1,78	**0,08**	1,57	**–0,13**	1,70	–

* *y* – урожайність; *E* – ефект першого та другого порядку.

вуватися до нуля. Так, сума головних ефектів чинників A і B становить відповідно – 0,14 + (–0,05) + (–0,10) = –0,01 т/га і 0,04 + 0,08 + + (–0,12) = = –0,01 т/га. У нашому випадку ця різниця відрізняється від нуля, що пояснюється округленням середніх показників.

3-й етап. Розраховують ефект першого порядку (взаємодії чинників) за базовою формулою табл. 121. Наприклад, ефект взаємодії чинників для рядкового способу сівби за висіву – 1,0 млн нас./га становить $Ey_{A1B1} = y_{A1B1} - \bar{y} - Ey_{A1} - Ey_{B1}$ = 1,81 – 1,70 – 0,14 – 0,04 = = –0,07 т/га, за висіву 2,0 млн/га – $Ey_{A1B2} = y_{A1B2} - \bar{y} - Ey_{A1} - Ey_{B2}$ = = 2,01 – 1,70 – 0,14 – 0,08 = 0,09 т/га і за висіву 3,0 млн/га – Ey_{A1B3} = = $y_{A1B3} - \bar{y} - Ey_{A1} - Ey_{B3}$ = 1,70 – 1,70 – 0,14 – (–0,13) = –0,01 т/га. Так само визначають інші значення ефекту першого порядку. У табл. 122 ефекти першого порядку показано курсивом.

4-й етап. Для визначення частки досліджуваних чинників у варіабельності врожайності насіння знаходять діапазон її мінливості за кожним чинником. Розрахунки доцільно занести у допоміжну табл. 123. Діапазон варіабельності ефекту чинника A коливається в межах від 0,10 до 0,14 т/га, тобто 0,24 т/га, чинника B – від 0,13 до 0,08 (0,21 т/га), взаємодії AB – від 0,07 до 0,10 (0,17 т/га).

Таблиця 123

Розмах мінливості показників урожайності насіння ріпака ярого залежно від дії досліджуваних чинників

Чинник	Ефект		Діапазон мінливості, т/га	Частка, %
	мінімальний	максимальний		
A	-0,10	0,14	0,24	38,7
B	-0,13	0,08	0,21	33,9
Взаємодія *AB*	-0,07	0,10	0,17	27,4
Сума	–	–	0,62	100,0

5-й етап. Діапазони варіабельності досліджуваних чинників становлять 0,24 + 0,21 + 0,17 = 0,62 т/га. Отриманий показник приймають за 100 % і визначають частки досліджуваних чинників у мінливості врожайності насіння ріпака ярого. Так, частка способу сівби (чинник A) становить 0,24 · 100 / 0,62 = 38,7 %, норми висіву (чинник B) – 0,21 × 100 / 0,62 = 33,9 % і взаємодії AB – 0,17 · 100 / 0,62 = 27,4 %.

6-й етап. Далі проводять дисперсійний аналіз цього досліду. У нашому прикладі $F_{факт.}$ головних ефектів чинників та їх взаємодії

значно перевищував $F_{теор.}$ за обох рівнів ймовірності (95 і 99 %), що свідчить про існування істотної різниці між варіантами досліду.

Висновок. У цьому прикладі досліджувані варіанти забезпечують істотні зміни врожайності насіння ріпака ярого. Це стосується як головних ефектів чинників, так і їх взаємодії. Серед досліджуваних чинників найбільші зміни врожайності насіння викликав чинник А – спосіб сівби (38,7 %). У досліді відзначена досить значна частка взаємодії досліджуваних чинників. Максимальна врожайність насіння ріпака на рядкових посівах – 2,01 т/га – була за норми висіву 2,0 млн нас./га, тоді як на широкорядних (міжряддя 45 см) – за норми висіву 1,0 млн нас./га – 1,74 т/га.

12. Коваріаційний аналіз

В агрономічних дослідженнях коваріаційний аналіз доцільно проводити для уточнення досліду в таких випадках:

– якщо на досліджувану ознаку може помітно впливати різноманітний вихідний стан умов експерименту (фізичний стан ґрунту, його родючість і т. ін.) які можуть бути кількісно виміряні на початку проведення досліджень;
– якщо на досліджувану ознаку в ході експерименту впливають незалежні від варіантів досліджень причини – ураженість рослин шкідниками, хворобами, вилягання рослин, забур’яненість посівів тощо.

Правильне застосування коваріаційного аналізу передбачає незалежний від варіантів досліду розподіл випадкової величини x. Якщо супутня ознака x має відношення до досліджуваних варіантів, то виключення її ефекту є неправомірним, оскільки це спричинить виключення певної частки ефекту варіанта. Наприклад, у дослідах із сортовивчення окремі сорти можуть мати меншу стійкість до деяких негативних чинників – шкідників, хвороб тощо, тож виключення впливу цих чинників є неправильним відносно до більш стійких сортів. У дослідах із впливу ценотичних чинників – норм висіву та способів сівби, коли густота рослин є результатом впливу цих факторів, не можна робити поправок на зрідженість.

На конкретних прикладах розглянемо механізм проведення коваріаційного аналізу.

Приклад 1. У досліді з вивчення впливу різних попередників отримали показники врожайності зерна ячменю ярого (ознака y) та забур'яненості посівів (шт./м^2), що наведені в табл. 124. Провести коваріаційний аналіз отриманих даних.

Таблиця 124

Урожайність зерна y (т/га) та забур'яненість посівів ячменю ярого x (шт./м^2) напередодні збирання врожаю

Варіант (попередник)		Урожайність за повтореннями			Суми V_x і V_y	Середнє значення
		I	II	III		
1	x	36,1	37,0	35,3	108,4	36,1
	y	4,21	4,37	4,10	12,68	4,23
2	x	31,2	33,0	28,7	92,9	31,0
	y	4,07	4,10	4,02	12,19	4,06
3	x	28,9	31,2	30,4	90,5	30,2
	y	3,96	4,17	3,92	12,05	4,02
4	x	37,7	39,1	36,8	113,6	37,9
	y	3,75	3,91	3,81	11,47	3,82
5	x	34,8	35,5	32,7	103,0	34,3
	y	4,48	4,60	4,41	13,49	4,50
6	x	21,6	23,3	22,4	67,3	22,4
	y	4,73	4,98	4,65	14,36	4,79
Суми P_x		190,3	199,1	186,3	$\sum x = 575{,}7$	$\bar{x} = 32{,}0$
Суми P_y		25,20	26,13	24,91	$\sum y = 76{,}24$	$\bar{y} = 4{,}24$

Проведення розрахунків. 1-й етап. У вихідній табл. 124 визначають суми та середні показники досліджуваних ознак (урожайність, т/га – y та забур'яненість шт./га – x) за варіантами та повтореннями. Правильність розрахунків перевіряють за рівняннями:

$$\sum x = \sum V_x = \sum P_x \text{ і } \sum y = \sum V_y = \sum P_y.$$

2-й етап. За формулами табл. 125 визначають суми квадратів за рядом x і y та суми добутків xy.

Суми рядів для ряду ознаки x:

$$N = l \cdot n = 6 \cdot 3 = 18;$$

Таблиця 125

Формули для визначення сум квадратів відхилень та добутків

Дисперсія	Суми квадратів і добутків		
	x^2	xy	y^2
Загальна, C_Y	$\sum x^2 - C$	$\sum xy - C$	$\sum y^2 - C$
Повторень, C_P	$\sum P^2_x / 1 - C$	$\sum P_x P_y / 1 - C$	$\sum P^2_y / 1 - C$
Варіантів, C_V	$\sum V^2_x / n - C$	$\sum V_x V_y / n - C$	$\sum V^2_y / n - C$
Залишок, C_Z	$C_Y - C_P - C_V$	$C_Y - C_P - C_V$	$C_Y - C_P - C_V$
–	$C = (\sum x)^2 / N$	$C = (\sum x)(\sum y) / N$	$C = (\sum y)^2 / N$

$$C = (\sum x)^2 / N = (575{,}7)^2 / 18 = 18412{,}8;$$
$$C_Y = \sum x^2 - C = (36{,}1^2 + 37{,}0^2 + ... + 22{,}4^2) - 18412{,}8 = 480{,}58;$$
$$C_P = \sum P^2_x / l - C = (190{,}3^2 + 199{,}1^2 + 186{,}3^2) / 6 - 18412{,}8 = 14{,}3;$$
$$C_V = \sum V^2_x / n - C = (108{,}4^2 + 92{,}9^2 +... + 67{,}3^2) / 3 - 18412{,}8 = 458{,}7;$$
$$C_Z = C_Y - C_P - C_Y = 480{,}58 - 14{,}3 - 458{,}7 = 7{,}58.$$

Суми рядів для ряду ознаки y:

$$C = (\sum y)^2 / N = (76{,}24)^2 / 18 = 322{,}9;$$
$$C_Y = \sum y^2 - C = (4{,}51^2 + 4{,}37^2 + ... + 4{,}65^2) - 322{,}9 = 2{,}045;$$
$$C_P = \sum P^2_y / l - C = (25{,}2^2 + 26{,}13^2 + 24{,}91^2) / 6 - 322{,}9 = 0{,}155;$$
$$C_V = \sum V^2_y / n - C = (12{,}68^2 + 12{,}19^2 + ... + 14{,}36^2) / 3 - 322{,}9 = 1{,}877;$$
$$C_Z = C_Y - C_P - C_V = 2{,}045 - 0{,}155 - 1{,}877 = 0{,}013.$$

Суми добутків xy:

$$C = (\sum x)(\sum y) / N = 575{,}7 \cdot 76{,}24 / 18 = 2438{,}409;$$
$$C_Y = \sum xy - C = (36{,}1 \cdot 4{,}21 + ... + 22{,}4 \cdot 4{,}65) - 2438{,}409 = 18{,}41;$$
$$C_P = \sum P_x \cdot P_y / l - C = (190{,}3 \cdot 25{,}2 + ... + 186{,}3 \cdot 24{,}91) / 6 - 2438{,}409 =$$
$$= 2439{,}79 - 2438{,}409 = 1{,}38;$$
$$C_V = \sum V_x \cdot V_y / n - C = (108{,}4 \cdot 12{,}68 +...+ 67{,}3 \cdot 14{,}36)/3 - 2438{,}41 =$$
$$= 2455{,}5 - 2438{,}41 = 16{,}76;$$
$$C_Z = C_Y - C_P - C_V = 18{,}41 - 1{,}38 - 16{,}76 = 0{,}27.$$

3-й етап. Суми квадратів і добутків заносять до таблиці коваріаційного аналізу (табл. 126) і знаходять коефіцієнт регресії ознаки y за ознакою x.

$$b_{yx} = \sum xy / \sum x^2 = 0{,}27 / 7{,}58 = 0{,}035 \text{ т/га}.$$

Суму квадратів для регресії визначають за співвідношенням:

Таблиця 126

Результати коваріаційного аналізу

Дисперсія	Суми квадратів і добутків			Ступінь свободи	b_{yx}	Середній квадрат	$F_{факт.}$	F_{05}
	x^2	y^2	xy					
Загальна	480,58	2,045	18,41	17	–	–	–	–
Повторень	14,3	0,155	1,38	2	–	–	–	–
Варіантів	458,7	1,877	16,76	5	–	3,35	$> 10^3$	3,48
Залишок I	7,58	0,013	0,27	10	–	0,027	–	–
Регресія C_b	–	0,010	–	1	0,035	0,010	33,0	5,12
Залишок II	–	0,003	–	9	–	0,0003	–	–

$$C_b = \sum(xy)^2 / \sum x^2 = 0{,}27^2 / 7{,}58 = 0{,}010.$$

Суму квадратів залишку після корегування дослідних даних знаходять за різницею: залишок II = залишок I – C_b = 0,013 – 0,010 = = 0,003 за 10 – 1 = 9 ступенів свободи.

Сума квадратів залишку I для ряду ознаки y, яка зазвичай використовується для визначення помилки досліду, складається з двох джерел варіювання: власне випадкове варіювання і варіювання, зумовлене залежністю між урожайністю та забур'яненістю посівів.

Суму квадратів для кореляційного зв'язку ознаки y з x (регресію) визначають шляхом поділу квадрата дисперсії залишку ряду xy на суму квадратів залишку для ряду x. Цей показник має один ступінь свободи, який вираховується із дисперсії залишку ряду ознаки y. У підсумку отримуємо суму квадратів для залишку II з 9 ступенями свободи (10 – 1 = 9). Середній квадрат другого залишку (0,003 / 9 ≈ 0,0003), характеризує помилку досліду після внесення поправки.

Критерій $F_{факт.}$ визначають шляхом поділу середнього квадрата дисперсії варіантів і регресії на дисперсію другого залишку. Теоретичний критерій F_{05} для дисперсії варіантів становить 3,48 (кількість ступенів свободи дисперсії чисельника – 5, знаменника – 9), а для регресії – 5,12 (кількість ступенів свободи дисперсії чисельника – 1, знаменника – 9) (див. дод. В).

Оскільки $F_{факт.} > F_{05}$, то зв'язок між ознаками y і x не випадковий і його можна застосовувати для корегування отриманих даних.

Якщо $F_{факт.} < F_{05,}$ то визначати поправки непотрібно, оскільки це не приведе до уточнення результатів експерименту.

4-й етап. У середні показники врожайності за варіантами досліду вводять поправки на регресію, тобто до показників урожайності, які нижчі від середнього показника, додають поправку, що дорівнює $b_{yx}(\bar{x} - x)$, а якщо їхня врожайність перевищує середній рівень, то поправку віднімають (табл. 127). Відкориговані середні врожаї за варіантами приведені до умов однакової забур'яненості всіх варіантів.

Таблиця 127

Внесення поправок для приведення середніх врожаїв (т/га) до однакової густоти рослин (тис./га)

Попередник	x	$\bar{x} - x$	$b_{yx}(\bar{x} - x) = 0{,}035\ (\bar{x} - x)$	Урожайність, т/га	
				фактична, y	відкоригована $y_1 = y + b_{yx}(\bar{x} - x)$
1	36,1	–4,1	–0,14	4,23	4,09
2	31,0	1,0	0,04	4,06	4,10
3	30,2	1,8	0,06	4,02	4,08
4	37,9	–5,9	–0,21	3,82	3,61
5	34,3	–2,3	–0,08	4,50	4,42
6	22,4	9,6	0,34	4,79	5,13
–	$\bar{x} = 32{,}0$	≈ 0,1	≈ 0,01	$\bar{y} = 4{,}24$	$\bar{y}_1 = 4{,}24$

5-й етап. Далі проводять оцінку істотності різниці між варіантами досліду (сортами):

$$s_{\bar{y}} = \sqrt{\frac{s_{II}^2}{n}} = \sqrt{\frac{0{,}0003}{3}} = 0{,}01 \text{ т/га};$$

$$s_d = \sqrt{\frac{2s_{II}^2}{n}} = \sqrt{\frac{2 \cdot 0{,}0003}{3}} = 0{,}014 \text{ т/га};$$

$$НІР_{05} = t_{05} \cdot s_d = 2{,}26 \cdot 0{,}014 = 0{,}032 \text{ т/га}.$$

Фактичне значення t-критерію для числа ступенів свободи другого залишку – 9 і рівня значущості – $\alpha = 0{,}05$ становить 2,26 (див. дод. Б).

Таким чином, усі різниці між середніми показниками врожайності зерна досліджуваних варіантів, які перевищують 0,03 т/га, істотні з 95 %-ю ймовірністю.

Приклад 2. У досліді з сортовивчення отримали показники врожайності та густоти рослин соняшнику напередодні збирання, які наведено в табл. 128. Провести коваріаційний аналіз отриманих даних.

Таблиця 128

Урожайність насіння y (т/га) та густота рослин x (тис./га) соняшнику напередодні збирання врожаю

Варіант (сорт)		Повторення I	II	III	IV	Суми V_x і V_y	Середні значення
1	x	70,6	68,6	69,8	72,0	281,0	70,25
	y	2,58	2,64	2,66	2,72	10,60	2,65
2	x	67,3	65,7	67,5	68,6	269,1	67,28
	y	2,65	2,68	2,88	2,76	10,97	2,74
3	x	73,4	76,2	74,6	75,4	299,6	74,90
	y	2,89	2,72	2,75	2,84	11,20	2,80
4	x	70,2	71,3	71,8	73,5	286,8	71,70
	y	2,40	2,28	2,39	2,42	9,49	2,37
5	x	67,2	67,7	70,3	68,0	273,2	68,30
	y	2,34	2,21	2,28	2,33	9,16	2,29
Суми P_x		348,7	349,5	354,0	357,5	$\sum x = 1409,7$	$\bar{x} = 70,49$
Суми P_y		12,86	12,53	12,96	13,07	$\sum y = 51,42$	$\bar{y} = 2,57$

Проведення розрахунків. Статистичну оброку проводять послідовно у кілька етапів.

1-й етап. У вихідній табл. 128 визначають суми та середні показники досліджуваних ознак за варіантами та повтореннями – урожайності y (т/га) і густоти рослин x (тис./м^2). Правильність розрахунків перевіряють за рівняннями: $\sum x = \sum V_x = \sum P_x$ і $\sum y = \sum V_y = \sum P_y$.

2-й етап. Визначають суми квадратів за рядами обох ознак (x і y) та суму добутків xy за формулами табл. 125.

Суми рядів для ряду ознаки x:

$$N = l \cdot n = 5 \cdot 4 = 20;$$
$$C = (\sum x)^2 / N = (1409,7)^2 / 20 = 99362,7;$$
$$C_Y = \sum x^2 - C = (70,6^2 + 68,6^2 + ... + 68,0^2) - 99362,7 = 170,6;$$
$$C_P = \sum P_x^2 / l - C = (348,7^2 + 349,5^2 + 354,0^2 + 357,5^2) / 5 - 99362,7 = 10,2;$$
$$C_V = \sum V_x^2 / n - C = (281,0^2 + 269,1^2 + 299,6^2 + 286,8^2 + 273,2^2) / 4 - 99362,7 = 99507,1 - 99362,7 = 144,4;$$

$$C_Z = C_Y - C_P - C_V = 170,6 - 10,2 - 144,4 = 16,0.$$

Суми рядів для ряду ознаки y:

$$C = (\sum y)^2 / N = (51,42)^2 / 20 = 132,2;$$
$$C_Y = \sum y^2 - C = (2,58^2 + 2,64^2 + ... + 2,33^2) - 132,2 = 0,909;$$
$$C_P = \sum P_y^2 / l - C = (12,86^2 + 12,53^2 + 12,96^2 + 13,07^2) / 5 - 132,2 = 0,034;$$
$$C_V = \sum V_y^2 / n - C = (10,6^2 + 10,97^2 + 11,2^2 + 9,49^2 + 9,16^2) / 4 - 132,2 = 133,027 - 132,2 = 0,827;$$
$$C_Z = C_Y - C_P - C_V = 0,909 - 0,034 - 0,827 = 0,048.$$

Суми добутків xy:

$$C = (\sum x) \cdot (\sum y) / N = 1409,7 \cdot 51,42 / 20 = 3624,339;$$
$$C_Y = \sum xy - C = (70,6 \cdot 2,58 + ... + 68,0 \cdot 2,33) - 3624,339 = 4,469;$$
$$C_P = \sum P_x \cdot P_y / l - C = (348,7 \cdot 12,86 + ... + 357,5 \cdot 13,07) / 5 - 3624,34 = 3624,776 - 3624,34 = 0,436;$$
$$C_V = \sum V_x \cdot V_y / n - C = (281,0 \cdot 10,6 + ... + 273,2 \cdot 9,16) / 4 - 3624,34 = 14510,391 / 4 - 3624,34 = 3627,598 - 3624,34 = 3,258;$$
$$C_Z = C_Y - C_P - C_V = 4,469 - 0,436 - 3,258 = 0,775.$$

3-й етап. Суми квадратів і добутків заносять до табл. 129 і знаходять коефіцієнт регресії ознаки y за ознакою x:

$$b_{yx} = \sum xy / \sum x^2 = 0,775 / 16,0 = 0,048 \text{ т/га}.$$

Відповідно до розрахованого коефіцієнта регресії зі зміною густоти рослин на 1 тис. шт./га врожайність насіння соняшнику в середньому збільшиться або зменшиться на 0,048 т/га.

Таблиця 129

Результати коваріаційного аналізу

Дисперсія	Суми квадратів і добутків x^2	y^2	xy	Ступінь свободи	b_{yx}	Середній квадрат	$F_{факт.}$	F_{05}
Загальна, C_Y	170,6	0,909	4,469	19	–	–	–	–
Повторень, C_P	10,2	0,034	0,436	3	–	–	–	–
Варіантів, C_V	144,4	0,827	3,258	4	–	0,207	207,0	3,36
Залишок I	16,0	0,048	0,775	12	–	0,004	–	–
Регресія, C_b	–	0,038	–	1	0,048	0,038	38,0	4,84
Залишок II	–	0,010	–	11	–	0,001	–	–

Суму квадратів для регресії знаходять за відношенням:

$$C_b = \sum(xy)^2 / \sum x^2 = 0{,}775^2 / 16{,}0 = 0{,}038.$$

Суму квадратів залишку після корегування дослідних даних знаходять за різницею: залишок II = залишок I – C_b = 0,048 – 0,038 = 0,1 за 12 – 1 = 11 ступенів свободи.

Аналіз даних табл. 129 показує, що середній квадрат залишку II (0,01/11 ≈ 0,001), який характеризує помилку досліду після внесення поправок, зменшився приблизно в чотири рази (0,001 проти 0,004).

Фактичний критерій F для дисперсії варіантів становить 3,36 (кількість ступенів свободи дисперсії чисельника – 4, знаменника – 11), а для регресії – 4,84 (кількість ступенів свободи дисперсії чисельника – 1, знаменника – 11) (див. дод. В).

Оскільки $F_{факт.} > F_{05}$, доцільно привести середні врожаї за сортами до однакової густоти рослин (табл. 130).

Таблиця 130

Внесення поправок для приведення середніх врожаїв у (т/га) до однакової густоти рослин (тис./га)

Варіант (сорт)	x	$\bar{x} - x$	$b_{yx}(\bar{x} - x) =$ $= 0{,}048\,(\bar{x} - x)$	Урожайність, т/га	
				фактична, y	відкоригована $y_1 = y + b_{yx} \cdot (\bar{x} - x)$
1	70,25	0,24	0,01	2,65	2,66
2	67,28	3,21	0,15	2,74	2,89
3	74,90	–4,41	–0,21	2,80	2,59
4	71,70	–1,21	–0,06	2,37	2,31
5	68,30	2,19	0,11	2,29	2,40
–	$\bar{x} = 70{,}49$	≈ 0	≈ 0	$\bar{y} = 2{,}57$	$\bar{y}_1 = 2{,}57$

4-й етап. Далі проводять оцінку істотності різниці між варіантами досліду (сортами):

$$s_{\bar{y}} = \sqrt{\frac{s_{II}^2}{n}} = \sqrt{\frac{0{,}001}{4}} = 0{,}016 \text{ т/га};$$

$$s_d = \sqrt{\frac{2s_{II}^2}{n}} = \sqrt{\frac{2 \cdot 0{,}001}{4}} = 0{,}022 \text{ т/га};$$

$$HIP_{05} = t_{05} \cdot s_d = 2{,}20 \cdot 0{,}022 = 0{,}048 \text{ т/га}.$$

Фактичне значення t-критерію для числа ступенів свободи другого залишку 8 і рівня значущості $\alpha = 0{,}05$ становить 2,2 (дод. Б).

Таким чином, усі різниці між середніми показниками врожайності насіння досліджуваних сортів, які перевищують 0,048 ≈ 0,05 т/га, істотні з 95 %-ю ймовірністю.

13. Розрахунок коефіцієнта спадковості

Статистичні показники ступенів спадковості селекційних ознак і комбінаційної здатності вихідних форм мають важливе значення у процесі підбору батьківських форм для схрещування і прогнозування ефективності селекції. Найбільше значення серед них має коефіцієнт спадковості – h^2. Він становить частку генетичної мінливості в загальній варіативності ознаки і розраховується за відношенням:

$$h^2 = s_r^2 / s_ф^2,$$

де s_r^2 – спадкова мінливість, $s_ф^2$ – загальна мінливість ознаки.

Крім генетичної мінливості, загальна фенотипічна мінливість включає модифікаційну (випадкову) мінливість s_z^2, зумовлену впливом екзогенних чинників: $s_ф^2 = s_r^2 + s_z^2$. Частку генетичної та модифікаційної мінливості в загальній фенотипічній мінливості ознак визначають з відношення дисперсій:

$$\frac{s_r^2}{s_ф^2} + \frac{s_z^2}{s_ф^2} = 1\ (100\ \%).$$

Зазвичай коефіцієнт спадковості визначають на основі коефіцієнта кореляції або регресії між фенотипами батьківських груп (приклад 1) та на основі класичного дисперсійного аналізу (прилад 2 – однофакторного, приклад 3 – двофакторного повного та приклад 4 – двофакторного ієрархічного комплексів).

Приклад 1. Розрахувати коефіцієнт спадковості h^2 при відборі сортів пшениці з високим вмістом білка в зерні (табл. 131).

За вихідними даними розраховують суми за графами, допоміжні величини ($\sum x$, $\sum y$, $\sum x^2$, $\sum y^2$ і $\sum xy$) (нижній рядок табл. 131). Далі вираховують коефіцієнт кореляції, регресії та фактичний t-критерій:

$$r = \frac{\sum xy - (\sum x \cdot \sum y)/n}{\sqrt{[\sum x^2 - (\sum x)^2/n]\,[\sum y^2 - (\sum y)^2/n\,]}} =$$

$$= \frac{3641{,}67 - (269{,}17 \cdot 270{,}36)/\,20}{\sqrt{[3630{,}30 - (269{,}17)^2/20]\,[3659{,}67 - (270{,}36)^2/20\,]}} =$$

$$= \frac{3641{,}67 - 3638{,}64}{\sqrt{[3630{,}30 - 3622{,}62]\,[3659{,}67 - 3654{,}73]}} = \frac{3641{,}67 - 3638{,}64}{\sqrt{7{,}68 \cdot 4{,}94}} =$$

$$= 0{,}492;$$

$$b_{yx} = \frac{\sum xy - (\sum x \cdot \sum y)/\,n}{\sum x^2 - (\sum x)^2/\,n} = \frac{3641{,}67 - (269{,}17 \cdot 270{,}36)/\,20}{3630{,}30 - (269{,}17)^2/\,20} = \frac{3{,}03}{7{,}68} =$$

$$= 0{,}395;$$

Таблиця 131

Дані щодо вмісту білка в зерні пшениці різних сортів та їхніх материнських форм, (%) і їх статистична обробка (*n* = 20)

Номер пари	Материнська форма, *x*	Сорт, *y*	x^2	y^2	xy
1	12,60	12,94	158,76	167,44	163,04
2	13,11	14,18	171,87	201,07	185,90
3	14,17	14,06	200,79	197,68	199,23
4	13,72	14,54	188,24	211,41	199,49
5	13,26	13,41	175,83	179,83	177,82
6	12,95	12,88	167,70	165,89	166,80
7	12,74	13,52	162,31	182,79	172,25
8	13,02	13,14	169,52	172,66	171,08
9	13,13	13,30	172,40	176,89	174,63
10	14,07	13,05	197,96	170,30	183,61
11	12,97	13,63	168,22	185,78	176,78
12	14,35	13,32	205,92	177,42	191,14
13	12,90	13,07	166,41	170,82	168,60
14	13,68	13,84	187,14	191,55	189,33
15	14,19	13,37	201,36	178,76	189,72
16	13,44	13,70	180,63	187,69	184,13
17	14,05	13,42	197,40	180,10	188,55
18	14,85	14,67	220,52	215,21	217,85
19	13,21	13,07	174,50	170,82	172,65
20	12,76	13,25	162,82	175,56	169,07
Сума	$\sum x = 269{,}17$	$\sum y = 270{,}36$	$\sum x^2 = 3630{,}3$	3659,67	3641,67

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - 0{,}492^2}{20 - 2}} = 0{,}205;$$

$$t_r = r\,/\,s_r = 0{,}492\,/\,0{,}205 = 2{,}40;$$

$t_{05} = 2{,}10$ (за $n - 2 = 20 - 2 = 18$ ступенів свободи).

Оскільки $t_{факт.} > t_{05}$ (2,40 > 2,10), кореляція та регресія значущі на 5 %-му рівні (нульова гіпотеза щодо незалежності *y* від *x* спростовується). Отже, вміст білка в зерні пшениці залежить від його вмісту в батьківських формах, тож розрахунок коефіцієнта спадковості h^2 на основі коефіцієнта кореляції та регресії буде обґрунтованим.

Коефіцієнт спадковості умовно прирівнюють до подвоєного коефіцієнта кореляції або регресії між фенотипами батьківських форм і потомством:

$$h^2 = 2r = 2 \cdot 0{,}492 = 0{,}984\ (98{,}4\ \%);$$

$$h^2 = 2b_{yx} = 2 \cdot 0{,}395 = 0{,}790\ (79{,}0\ \%).$$

Практика показує, що точніше зв’язок між батьківськими формами та потомством характеризує коефіцієнт регресії. На основі цього з 95 %-ю ймовірністю можна стверджувати, що майже 80 % загальної мінливості вмісту білка в зерні пшениці зумовлено спадковою мінливістю рослин. Таким чином, відбір батьківських форм за цією ознакою є доволі ефективним.

Приклад 2. Методом дисперсійного аналізу (однофакторний комплекс) необхідно розрахувати коефіцієнт спадковості озерненості колоса головного стебла пшениці м’якої ярої (табл. 132).

Для знаходження середніх квадратів дисперсії варіантів (батьківських форм) та помилок шукають поправку *C* (*Δ*), загальне варіювання C_Y, варіювання варіантів – форм C_V та помилок C_Z:

$$N = l \cdot n = 6 \cdot 4 = 24;$$

$$C = (\sum X)^2\,/\,N = 637{,}7^2\,/\,24 = 16944{,}2;$$

$$C_Y = \sum X^2 - C = (28{,}3^2 + 27{,}1^2 + ... + 27{,}4^2) - 16944{,}2 =$$

$$= (800{,}89 + 734{,}41 + ... + 750{,}76) - 16944{,}2 = 49{,}79$$

за $(N - 1) = (24 - 1) = 23$ ступенів свободи;

$$C_P = \sum X_P^2\,/\,l - C = (160{,}8^2 + 156{,}8^2 + 164{,}1^2 + 156{,}0^2)\,/\,6 - 16944{,}2 =$$

$$= (25856{,}64 + 24586{,}24 + 26928{,}81 + 24336{,}0)\,/\,6 - 16944{,}2 = 7{,}08$$

за $(n - 1) = (4 - 1) = 3$ ступенів свободи;

$$C_V = \sum X_V^2\,/\,n - C = (110{,}9^2 + 110{,}2^2 + ... + 110{,}9^2)\,/\,4 - 16944{,}2 = 41{,}49$$

за $(l - 1) = (6 - 1) = 5$ ступенів свободи;

Таблиця 132

Середня озерненість колоса головного стебла пшениці м'якої ярої, шт., суми і середні за варіантами та повтореннями (вихідна таблиця однофакторного дисперсійного комплексу)

Форма (варіант)		Повторення				Сума, V	Середнє
материнська	батьківська	I	II	III	IV		
Одна	1	28,3	27,1	28,8	26,7	110,9	27,7
	2	27,7	27,0	28,3	27,2	110,2	27,6
	3	24,2	23,8	25,4	24,0	97,4	24,4
	4	25,5	25,0	25,8	24,7	101,0	25,3
	5	27,0	26,6	27,7	26,0	107,3	26,8
	6	28,1	27,3	28,1	27,4	110,9	27,7
Суми P		160,8	156,8	164,1	156,0	$\sum x = 637,7$	$\bar{x} = 26,6$

$$C_Z = C_Y - C_P - C_V = 49,79 - 7,08 - 41,49 = 1,22$$
за $(l - 1)\ (n - 1) = (6 - 1)\ (4 - 1) = 15$ ступенів свободи.

Отримані результати заносять до таблиці дисперсійного аналізу (табл. 133) і визначають значущість дії генотипів на фенотипічну мінливість ознаки за F-критерієм.

Оскільки $F_{\text{факт.}} > F_{05}$, то батьківська форма має значний вплив на мінливість досліджуваної ознаки. Дисперсія варіантів визначається генотипом ознаки s_r^2 у сукупності з випадковою мінливістю – s_z^2. Таким чином, середній квадрат для варіантів s_v^2 (у нашому прикладі батьківських форм) складається з двох компонентів: $s_v^2 = s_z^2 + n \cdot s_r^2$. Звідси $s_r^2 = (s_v^2 - s_z^2) / n$, а $s_ф^2 = s_r^2 + s_z^2$ (показник n перед спадковою мінливістю приводить її до рівня вихідних дат).

Таблиця 133

Результати дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{\text{факт.}}$	F_{05}
Загальна, C_Y	49,79	23	–	–	–
Повторень, C_P	7,08	3	–	–	–
Варіантів – батьківська форма, C_V	41,49	5	8,30	103,8	2,90
Помилок, C_Z	1,22	15	0,08	–	–

За показниками дисперсійного аналізу розраховують спадкову і фенотипічну мінливість та коефіцієнт спадковості:

$$s_r^2 = (s_v^2 - s_z^2) / n = (8,30 - 0,08) / 4 = 2,06;$$
$$s_ф^2 = s_r^2 + s_z^2 = 2,06 + 0,08 = 2,14;$$
$$h^2 = s_r^2 / s_ф^2 = 2,06 / 2,14 = 0,963 \text{ або } 96,3\ \%.$$

Висновок. Коефіцієнт спадковості, який характеризує міру передачі досліджуваної ознаки (озерненість колоса головного стебла пшениці) від батьківської форми до сорту, становить 96,3 %, тож відбір за цією ознакою виправдовує себе повною мірою.

Приклад 3. За показниками вмісту олії в насінні гібридів соняшнику (табл. 134) визначити коефіцієнт спадковості на основі аналізу двофакторного дисперсійного комплексу.

Дисперсійний аналіз двофакторного комплексу проводять у кілька етапів.

1-й етап. Спочатку, як і під час проведення дисперсійного аналізу однофакторного польового комплексу, розраховують поправку,

Таблиця 134

Олійність насіння соняшнику, %, суми і середні за варіантами та повтореннями (вихідна таблиця двофакторного дисперсійного комплексу)

Чинник A	Чинник B	Повторення				Сума, V	Середнє
		I	II	III	IV		
A_1	B_1	46,7	44,8	46,5	47,0	185,0	46,3
	B_2	48,6	47,7	49,0	48,8	194,1	48,5
	B_3	45,5	44,9	45,1	46,5	182,0	45,5
A_2	B_1	56,7	55,5	58,3	58,7	229,2	57,3
	B_2	55,0	53,6	54,4	54,9	217,9	54,5
	B_3	58,1	57,0	57,4	58,6	231,1	57,8
A_3	B_1	53,1	51,8	52,7	53,2	210,8	52,7
	B_2	53,6	52,0	52,1	53,9	211,6	52,9
	B_3	55,3	54,4	55,8	52,8	218,3	54,6
A_4	B_1	52,2	50,9	51,1	53,0	207,2	51,8
	B_2	57,1	57,2	56,4	57,7	228,4	57,1
	B_3	53,5	53,0	54,2	54,8	215,5	53,9
Суми P		635,4	622,8	633,0	639,9	$\sum x = 2531,1$	52,7

Примітка. * Чинник A – материнська форма; чинник B – батьківська форма.

квадрати відхилень: загальний – C_y, повторень – C_p, варіантів – C_v та залишку – C_z.

$$N = l_a \cdot l_b \cdot n = 4 \cdot 3 \cdot 4 = 48;$$
$$C = (\sum X)^2 / N = 2531{,}1^2 / 48 = 133468{,}1;$$
$$C_Y = \sum X^2 - C = (46{,}7^2 + 44{,}8^2 + ... + 54{,}8^2) - 133468{,}1 =$$
$$= (2180{,}89 + 2007{,}04 + ... + 3003{,}04) - 133468{,}1 = 756{,}0$$
за $(N - 1) = (48 - 1) = 47$ ступенів свободи;
$$C_P = \sum X_P^2 / l - C = (635{,}4^2 + 622{,}8^2 + 633{,}0^2 + 639{,}9^2) / 12 - 133468{,}1 =$$
$$= (403733{,}2 + 387879{,}8 + 400689{,}0 + 409472{,}0) / 12 - 133468{,}1 = 13{,}1$$
за $(n - 1) = (4 - 1) = 3$ ступенів свободи;
$$C_V = \sum X_V^2 / n - C = (185{,}0^2 + 194{,}1^2 + ... + 215{,}5^2) / 4 - 133468{,}1 = 738{,}2$$
за $(l_a \cdot l_b - 1) = (12 - 1) = 11$ ступенів свободи;
$$C_Z = C_Y - C_P - C_V = 756{,}0 - 13{,}1 - 738{,}2 = 4{,}7.$$

2-й етап. Наступним кроком є оцінка істотності впливу материнських (чинник A) і батьківських (чинник B) форм та їхніх взаємодії на варіабельність досліджуваної ознаки. Для цього за схемою дисперсійного аналізу двофакторного досліду складають допоміжну таблицю (табл. 135), до якої заносять суми за всіма варіантами (сортами).

Таблиця 135

Вихідна таблиця дисперсійного аналізу двохфакторного комплексу для визначення головних ефектів чинників і їх взаємодії

Чинник A	Чинник B			Сума, A
	B_1	B_2	B_3	
A_1	185,0	194,1	182,0	561,1
A_2	229,2	217,9	231,1	678,2
A_3	210,8	211,6	218,3	640,7
A_4	207,2	228,4	215,5	651,1
Суми B	832,2	852,0	846,9	$\sum x = 2531{,}1$

Дисперсійний аналіз табл. 135 передбачає розчленування загальної суми квадратів варіантів – C_v, яке дорівнює 738,2, на складові компоненти: суму квадратів варіювання чинників A – C_A, B – C_B та їхню взаємодію – C_{AB}.

$$C_A = \sum X_A^2 / l_b \cdot n - C = (561{,}1^2 + ... + 651{,}1^2) / 3 \cdot 4 - 133468{,}1 =$$
$$= (314833{,}2 + 459955{,}2 + 410496{,}5 + 423931{,}2) / 12 - 133468{,}1 = 633{,}2$$

за $(l_a - 1) = (4 - 1) = 3$ ступенів свободи;
$$C_B = \sum X_B^2 / l_a \cdot n - C = (832{,}2^2 + 852{,}0^2 + 846{,}9^2) / 4 \cdot 4 - 133468{,}1 =$$
$$= (692556{,}8 + 725904{,}0 + 717239{,}6) / 16 - 133468{,}1 = 0{,}9$$
за $(l_b - 1) = (3 - 1) = 2$ ступенів свободи;
$$C_{AB} = C_V - C_A - C_B = 738{,}2 - 633{,}2 - 0{,}9 = 104{,}1$$
за $(l_a - 1)(l_b - 1) = (4 - 1)(3 - 1) = 6$ ступенів свободи.

3-й етап. Складають підсумкову таблицю дисперсійного аналізу (табл. 136) і проводять оцінку істотності дії та взаємодії чинників за F-критерієм.

Таблиця 136

Результати дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{факт.}$	F_{05}
Загальна, C_y	756,0	47	–	–	–
Повторень, C_p	13,1	3	–	–	–
Чинник A (материнська форма)	633,2	3	211,10	1508,0	2,90
Чинник B (батьківська форма)	0,9	2	0,45	3,2	3,30
Взаємодія AB	104,1	6	17,40	122,5	2,40
Помилок, C_z	4,70	33	0,14	–	–

Висновок. У результаті проведених розрахунків установлено істотний вплив материнських форм (чинник A) та взаємодії материнських і батьківських форм (AB) на результативність досліджуваної ознаки – вміст олії у насінні. У зв'язку із цим потрібно визначити коефіцієнти спадковості, які характеризують силу спадкового впливу материнських форм і їхню взаємодію з батьківськими формами.

У двофакторному комплексі дисперсія варіантів включає як спадкову (генетичну) мінливість, зумовлену генотипами материнських і батьківських форм та їхньою взаємодією, так і випадкову мінливість (помилку). Загальний коефіцієнт спадковості одержують шляхом підсумовування окремих коефіцієнтів спадковості:

$$h^2 = h_A^2 + h_B^2 + h_{AB}^2.$$

Фенотипічну мінливість генотипів материнських (s_A^2) і батьківських (s_B^2) форм, а також їхньої взаємодії будуть складати такі дисперсії двофакторного комплексу:

– материнські форми (чинник A) – $s_z^2 + n \cdot s_{AB}^2 + l_b \cdot s_A^2$;
– батьківські форми (чинник B) – $s_z^2 + n \cdot s_{AB}^2 + l_a \cdot s_B^2$;
– взаємодія (AB) – $s_z^2 + n \cdot s_{AB}^2$.

Оскільки в цьому прикладі істотними були дія материнських форм (чинник A) і взаємодія AB, має сенс визначити два коефіцієнти спадковості – h_A^2 (дії чинника A) і h_{AB}^2 (взаємодії AB).

Для цього визначають середньозважені дисперсії s_A^2, s_B^2, s_{AB}^2 за рівняннями:

$$s_A^2 = [(s_z^2 + n \cdot s_{AB}^2 + l_b \cdot s_A^2) - (s_z^2 + n \cdot s_{AB}^2)] / l_b \cdot n =$$
$$= [(0{,}14 + 4 \cdot 17{,}4 + 3 \cdot 211{,}1) - (0{,}14 + 4 \cdot 17{,}4)] / 3 \cdot 4 = 52{,}8;$$
$$s_B^2 = [(s_z^2 + n \cdot s_{AB}^2 + l_a \cdot s_B^2) - (s_z^2 + n \cdot s_{AB}^2)] / l_a \cdot n =$$
$$= [(0{,}14 + 4 \cdot 17{,}4 + 4 \cdot 0{,}45) - (0{,}14 + 4 \cdot 17{,}4)] / 4 \cdot 4 = 0{,}11;$$
$$s_{AB}^2 = [(s_z^2 + n \cdot s_{AB}^2) - s_z^2] / n = [(0{,}14 + 4 \cdot 17{,}4) - 0{,}14] / 4 = 17{,}4.$$

Фенотипічна мінливість становить:

$$s_ф^2 = s_A^2 + s_{AB}^2 + s_Z^2 = 52{,}8 + 17{,}4 + 0{,}14 = 70{,}3.$$

На підставі отриманих даних розраховують коефіцієнти спадковості:

$$h_A^2 = s_A^2 / s_ф^2 = 52{,}8 / 70{,}34 = 0{,}751\ (75{,}1\ \%);$$
$$h_{AB}^2 = s_{AB}^2 / s_ф^2 = 17{,}4 / 70{,}34 = 0{,}247\ (24{,}4\ \%).$$

Висновок. Під час підбору пар для схрещування слід ураховувати, що вміст олії у насінні гібридів соняшнику на 75,1 % залежить від материнської форми. Загальний коефіцієнт спадковості h^2 дорівнюватиме сумі двох істотних компонентів:

$$h^2 = h_A^2 + h_{AB}^2 = 75{,}1\ \% + 24{,}4\ \% = 99{,}5\ \%.$$

У двофакторних дослідах, поставлених за неповними факторіальними схемами, досліджують не всі можливі комбінації чинників (материнських і батьківських форм), унаслідок чого неможливо розрахувати взаємодію чинників. Двофакторну модель, поставлену за повною та неповною (пріоритетною) факторіальною схемами, можна відобразити так:

Форма та гібрид (сорт)	Двофакторна схема	
	повна факторіальна	неповна (пріоритетна)
Материнська (чинник A)	A_1 A_2	A_1 A_2
Батьківська (чинник B)	B_1 B_2 B_3 B_1 B_2 B_3	B_1 B_2 B_3 B_4 B_5 B_6
Гібриди (взаємодія AB)	A_1B_1 A_1B_2 A_1B_3 A_2B_1 A_2B_2 A_3B_3	A_1B_1 A_1B_2 A_1B_3 A_2B_4 A_2B_5 A_2B_6
Рівняння діючих чинників	$x_{ij} = \mu + A_i + B_j + A_iB_j + E_{ij}$	$x_{ij} = \mu + A_i + B_j + E_{ij}$

У пріоритетних (неповних) комплексах відсутнє вільне комбінування чинників, тож у них не визначають дисперсію взаємодії й інакше розраховують F-критерій.

Приклад 4. Проведено схрещування за неповною факторіальною схемою двох батьківських форм пшениці озимої (A_1 і A_2) з материнською формою (B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8), при цьому форму A_1 схрещували з B_1, B_2, B_3, B_4, а A_2 – з формами B_5, B_6, B_7, B_8. Ефективність гібридів F_1 оцінювали за довжиною верхнього міжвузля (табл. 137).

Таблиця 137

Довжина верхнього міжвузля рослин пшениці після схрещування батьківських форм проведеного за неповною (пріоритетною) схемою, см

Чинник A	Чинник B	Повторення				Суми V	
		I	II	III	IV	B	A
A_1	B_1	35	32	31	34	132	567
	B_2	38	41	37	40	156	
	B_3	36	38	38	40	152	
	B_4	32	32	30	33	127	
A_2	B_5	29	25	25	27	106	482
	B_6	31	31	29	33	124	
	B_7	37	34	32	36	139	
	B_8	30	25	30	28	113	
Суми P		268	258	252	271	$\sum A = \sum B = 1049$	

Примітка. * Чинник A – батьківська форма; чинник B – материнська форма.

Дисперсійний аналіз представленого двофакторного комплексу $AB = 2 \cdot 4$, де A – батьківська форма, а B – материнська форма, включає такі етапи:

1-й етап. Проводять дисперсійний аналіз за принципом однофакторного комплексу, тобто розраховують поправку $\varDelta$ (C), загальне варіювання C_y, варіювання повторень C_p, варіантів C_v і помилок C_z:

$$N = l_a \cdot l_b \cdot n = 2 \cdot 4 \cdot 4 = 32;$$
$$\varDelta \text{ (поправка)} = (\sum x)^2 / N = 1049^2 / 32 = 34387{,}5;$$
$$C_y = \sum x^2 - \varDelta = (35^2 + 32^2 + ... + 28^2) - 34387{,}5 = 619{,}5;$$
$$C_p = \sum x_p^2 / l_a \cdot l_b - \varDelta = (268^2 + 258^2 + 252^2 + 271^2) / 8 - 34387{,}5 = 29{,}1;$$
$$C_v = \sum x_v^2 / n - \varDelta = (132^2 + 156^2 + ... + 113^2) / 4 - 34387{,}5 = 536{,}3;$$
$$C_z = C_y - C_p - C_v = 619{,}5 - 29{,}1 - 536{,}3 = 54{,}1.$$
$$\sum x = \sum V = \sum A = \sum B = \sum P = 1049.$$

2-й етап. Оцінюють істотність дії батьківських і материнських форм на результативну ознаку (компонент взаємодії батьківських і материнських форм прихований всередині материнської форми).

$$C_A = \sum X_A^2 / l_b \cdot n - \varDelta = (567^2 + 482^2) / 4 \cdot 4 - 34387{,}5 =$$
$$= (321489 + 232324) / 16 - 34387{,}5 = 225{,}8;$$
$$C_B = C_v - C_A = 536{,}3 - 225{,}8 = 310{,}5.$$

Для рівномірних комплексів, поставлених за неповною факторіальною схемою, коли кожній градації чинника A відповідає однакова кількість градацій чинника B, число ступенів свободи чинника A становить $l_a - 1 = 2$, для $B - l_a (l_b - 1) = 6$ і для помилок $- (l_a \cdot l_b - 1) \times \times (n - 1) = 21$. У цьому прикладі $l_a = 2$, $l_b = 4$ і $v = 8$.

Під час статистичної обробки нерівномірних комплексів з неповною факторіальною схемою (l_b для $A_1 \neq l_b$ для A_2) вносять такі корективи: v для $B = l_b - l_a$ і v для помилок $= N - l_b - n - 1$.

3-й етап. Складають допоміжну таблицю дисперсійного аналізу і проводять оцінку істотності дії батьківських форм (табл. 138).

Фактичні значення F-критерію визначають за рівняннями:

$$F \text{ (для } A) = s_A^2 / s_Z^2 = 225{,}8 / 2{,}6 = 86{,}8;$$
$$F \text{ (для } A) = s_B^2 / s_Z^2 = 51{,}8 / 2{,}6 = 19{,}9.$$

Теоретичне значення F-критерію при заданих показниках ступенів свободи дисперсії чисельника та знаменника знаходять за дод. Б.

Фактичний F- критерій для обох батьківських форм перевищує теоретичний з 95 %-ю ймовірністю, тож доцільно провести компонентний аналіз двофакторного комплексу (табл. 139).

Таблиця 138

Результати дисперсійного аналізу

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{\text{факт.}}$	F_{05}
Загальна, C_Y	619,5	31	–	–	–
Повторень, C_P	29,1	3	–	–	–
Чинник A (батьківська форма)	225,8	1	225,8	86,8	4,32
Чинник B (материнська форма)	310,5	6	51,8	19,9	2,57
Помилок, C_Z	54,1	21	2,6	–	–

Таблиця 139

Компонентний аналіз двофакторної пріоритетної моделі

Джерело варіації	Оцінки дисперсії σ^2 за вибіркою	Очікувані дисперсії (параметри)
Чинник A	225,8	$\sigma_A^2 \cdot l_b \cdot n + \sigma_B^2 \cdot n + \sigma_Z^2$
Чинник B	51,8	$\sigma_B^2 \cdot n + \sigma_Z^2$
Залишок	2,6	σ_Z^2

У відповідності зі схемою компонентного аналізу точні значення дисперсії (генеральної) для рівномірного статистичного комплексу визначаються за формулами:

$$\sigma_A^2 = (s_A^2 - s_B^2) / l_b \cdot n = (225{,}8 - 51{,}8) / 4 \cdot 4 = 10{,}9;$$
$$\sigma_B^2 = (s_B^2 - s_Z^2) / n = (51{,}8 - 2{,}6) / 4 = 12{,}3; \ \sigma_Z^2 = s_Z^2 = 2{,}6.$$

На основі отриманих даних визначають фенотипічну дисперсію (мінливість) $\sigma_\Phi^2 = \sigma_A^2 + \sigma_B^2 + \sigma_Z^2 = 10{,}88 + 12{,}3 + 2{,}6 = 25{,}8$. Коефіцієнт спадковості розраховують для тих батьківських форм, вплив яких є істотним:

$$h_A^2 = \sigma_A^2 / \sigma_\Phi^2 = 10{,}9 / 25{,}8 = 0{,}42 \text{ (42 \%)};$$
$$h_B^2 = \sigma_B^2 / \sigma_\Phi^2 = 12{,}3 / 25{,}8 = 0{,}48 \text{ (48 \%)}.$$

Висновок. Оскільки вплив батьківських форм на досліджувану ознаку виявився істотним, то з показника коефіцієнта спадковості випливає, що довжина верхнього міжвузля пшениці озимої на 42 % визначається материнською формою та на 48 % – батьківською.

Визначити силу впливу батьківських і материнських форм, а також їхньої взаємодії на досліджувану ознаку можна також значно простішим методом, а саме – шляхом розрахунку коефіцієнта детермінації η^2, тобто відношення сум квадратів відхилень за даними таблиці дисперсійного аналізу. Наприклад, для третього прикладу сила впливу материнських форм η_A^2 та взаємодії материнських і батьківських форм η_{AB}^2 становитиме:

$$\eta_A^2 = C_A / C_Y = 633{,}2 / 756{,}0 = 0{,}838\ (83{,}8\ \%);$$
$$\eta_{AB}^2 = C_{AB} / C_Y = 104{,}1 / 756{,}0 = 0{,}138\ (13{,}8\ \%);$$
$$\eta^2_{(\text{загальний})} = \eta_A^2 + \eta_{AB}^2 = 0{,}838 + 0{,}138 = 0{,}976\ (97{,}6\ \%).$$

Легко помітити, що коефіцієнти спадковості, розраховані за цим принципом, мало відрізняються від показників h^2, отриманих у результаті компонентного аналізу, водночас їхній розрахунок значно простіший.

14. Пробіт-аналіз

Пробіт-аналіз (*probit analysis*) – вид регресійного аналізу, який застосовують для визначення впливу кількісної ознаки на бінарний відклик. Він належить до класу узагальнених лінійних моделей.

Ідея пробіт-аналізу вперше була представлена Ліссом у 1934 р. у статті, присвяченій впливу пестицидів на відсоток убитих шкідників. Він уперше запропонував для обліку відсотка вбитих шкідників використовувати ймовірнісний блок або пробіт.

Цей метод відомий також під назвою «дозаторний аналіз впливу кривих» і широко розповсюджений в області токсикології. Здебільшого йдеться про те, як на певну кількість індивідів впливають різні дози речовини (наприклад, певного інсектициду).

Пробіт-аналіз є специфічним методом статистичної оцінки впливу засобів захисту рослин на біологічні об’єкти, тобто результатів біологічних експериментів. Пробітом служить умовна випадкова змінна, яка виражає ймовірнісну стійкість шкідників до хімічного препарату. Суть аналізу зводиться до графічно-візуального пошуку лінії пробіт-регресії або залежності між дозою препарату та його ефектом, на основі якої визначають потрібну дозу летальності.

Чуттєвість певного виду шкідників до інсектициду може характеризуватися показником дози, яка викликає його повну загибель. Водночас загибель шкідників від дози препарату, навіть більшої за летальну, настає не відразу, а через певний проміжок часу. Недостатність дози препарату також одразу визначити неможливо. Для цього потрібно від кількох днів до кількох тижнів. Упродовж цього періоду в організмах шкідників відбуваються відновлювальні процеси. Відновлення не може бути повним, тож піддавати випробуванню одні і ті самі особини не можна. Крім того, задача визначення дози 100 %-ї смертності не виправдана з економічного погляду, адже це пов’язано зі значними економічними та біоенергетичними витратами. Цілком достатньою для характеристики певного інсектициду буде його доза, за якої загине 50 % шкідників. Ця доза летальності позначається LD_{50}.

Для проведення пробіт-аналізу всю досліджувану сукупність розподіляють на групи з подальшим випробуванням різних доз досліджуваного препарату в порядку зростання. У кожній наступній групі дозу препарату збільшують на певну величину. У першій досліджуваній групі, з найменшою (мінімальною) дозою препарату, гине найменше число слабких особин. У кожній наступній групі, зі збільшенням дози препарату, відсоток загиблих шкідників поступово зростатиме, досягаючи максимальної частки в останній групі. Таким чином, отримують зростаючий ряд загибелі шкідників.

Численні дослідження у фітопатології, токсикології, радіобіології та мікробіології показали, що розподіл частки шкідників, які негативно реагують на певний препарат, виражається кривою лінію, схожою на літеру *S* (рис. 40).

S-крива також характеризує процес росту та розвитку як рослин у цілому, так і їхніх окремих частин. Водночас у випадку впливу інсектицидів на шкідників *S*-крива несиметрична: крутизна вигину її верхньої і нижньої частини неоднакова. Статистична оцінка даних такого розподілу для визначення показника LD_{50} або інших доз препарату спрощується, якщо криву трансформувати в пряму лінію. Для вирівнювання графіка дози препарату переводять у логарифми, а відсотки загибелі шкідників – у пробіти і відкладають отримані значення відповідно на осі абсцис і на осі ординат. Значення пробітів представлені в дод. Т.

Приклад. У дослідах з вивчення ефективності інсектициду Карате отримали такі показники частки загибелі клопа-черепашки за

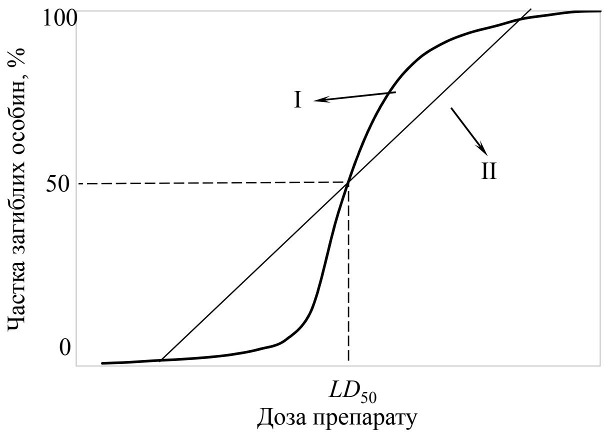


Рис. 40. Типова крива (I) та вирівняна лінія пробіт-регресії (II), що виражає залежність відсотка загиблих особин від дози препарату

різних доз препарату (табл. 140). Провести пробіт-аналіз наведених даних та визначити LD_{50} і LD_{95} (відповідно доза препаратів, що забезпечує 50 і 95 %-ву загибель шкідників).

Таблиця 140

Результати випробування інсектициду Карате

Доза препарату, мг/га (x)	Середня загибель, % (y)	Перетворення показників	
		$lg\ x$ (вісь x)	значення пр́обіт (вісь y)
100	9	2,000	3,59
120	12	2,079	3,82
140	14	2,146	3,92
160	22	2,204	4,23
180	36	2,255	4,64
200	51	2,301	5,03
220	70	2,342	5,52

У табл. 140 наведено дані стосовно середньої загибелі клопа-черепашки залежно від дози інсектициду і зроблені перетворення цих показників, необхідні для трансформації S-подібною кривої у пряму лінію шляхом переведення відсотків у пр́обіти (див. дод. Т), а доз препарату – в логарифми (за таблицею логарифмів або комп'ютерною програмою).

Визначені логарифми доз препарату відкладають на осі абсцис, а показники пр́обіт – на осі ординат. Через відмічені точки проводять

пряму лінію, яка повинна проходити якомога ближче до точок, насамперед тих, що відповідають летальності від 15 до 85 %. Летальну дозу інсектициду для LD_{50} та інших летальних доз визначають шляхом постановки перпендикуляра з відповідної точки – відсотка загибелі на осі ординат до перетину з лінією пр́обіт-регресії (рис. 41).

Наведений метод належить до найпростіших моделей системи пр́обітів. Він дозволяє лише приблизно визначити LD_{95} і LD_{99}, але за ним не можна встановити довірчі інтервали цих показників. Водночас цей метод порівняно з іншими точніше визначає LD_{50}. До того ж застосування складних модифікацій системи пр́обітів, представлених у спеціальній літературі, не завжди виправдовує себе, адже в більшості випадків висока точність не потрібна, а складні математичні розрахунки забирають багато часу.

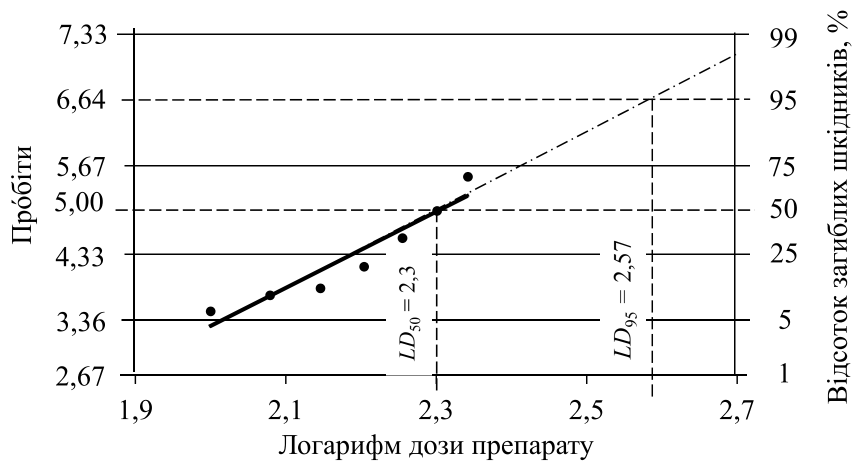


Рис. 41. Окомірна лінія пр́обіт-регресії (залежність доза – ефект) для визначення чуттєвості клопа черепашки до дії інсектициду Карате. З двох доз летальності LD_{50} визначено на основі інтерполяції, а LD_{95} – на основі екстраполяції (за межами досліджуваних доз препарату 2,00...2,34)

Контрольні запитання та завдання

1. Яка методика заокруглення чисел? Наведіть приклади.
2. Яка різниця між середньою простою та зваженою?
3. На основі чого бракують сумнівні дати та відновлюють вибракувані?
4. Який основний статистичний показник дисперсійного аналізу та як він використовується?
5. Які вибірки називають згрупованими, а які – незгрупованими? Наведіть приклади.
6. Якими способами доцільно користуватися для уникнення можливих помилок під час розподілу показників вибірки за групами?
7. Які методи слід використовувати для порівняння вибіркових часток за альтернативної мінливості?
8. З яких компонентів складається загальне варіювання досліджуваного показника в однофакторному вегетаційному досліді?
9. Наведіть послідовність проведення статистичних розрахунків однофакторного вегетаційного досліду.
10. З яких основних компонентів складається загальна сума варіювання показників у багатофакторному вегетаційному досліді?
11. Назвіть етапи проведення обробки даних польового досліду з однорічними культурами, поставленого методом рендомізованих повторень.
12. Які основні етапи проведення статистичної обробки дослідів, поставлених стандартними методами?
13. Як проводиться статистична обробка польового досліду, поставленого методом латинського квадрата або прямокутника?
14. Назвіть основні особливості проведення статистичної обробки результатів багатофакторних польових дослідів, поставлених методом розщеплених ділянок.
15. У чому полягає основна мета факторіального аналізу? У яких випадках його застосовують?
16. Для чого визначають коефіцієнт спадковості? Що він показує?
17. У чому полягає суть пробіт-аналізу? У яких випадках дослідницької справи його застосовують?

Перелік символів

s^2 – вибіркова дисперсія, середній квадрат;
s – стандартне відхилення;
$s\bar{y}$ – абсолютна помилка;
$\bar{x}$ – середня арифметична вибірки;
X – значення варіюючої ознаки;
HIP_{05}, HIP_{01} – найменші істотні різниці для 5 %-го та 1 %-го рівня значущості;
t_{05} *і* t_{01} – табличні значення критерію t для 5 %-го та 1 %-го рівня значущості;
$F_{ф}$ – емпіричне (фактичне) значення F-критерію Фішера;
F_{05}, F_{01} – табличні значення F-критерію для 5 %-го та 1 %-го рівня значущості;
e – абсолютна помилка експерименту;
E – відносна помилка експерименту;
μ – середня генеральної сукупності;
σ^2 – дисперсія генеральної сукупності;
σ – стандартне відхилення генеральної сукупності;
P – рівень імовірності;
P_1 – рівень значущості;
Me (медіана) – центральний показник вирівняного ряду показників;
Mo (мода) – показник, який найчастіше повторюється в сукупності;
E – відносна помилка вибіркової середньої;
V – коефіцієнт варіації, мінливості;
H_0 – нульова гіпотеза;
d – різниця між середніми вибірок;
s_d – помилка різниці між середніми вибірок;
l – кількість варіантів;
n – повторність вибірки;
N – загальна кількість спостережень у досліді;
v – число ступенів свободи;
C або Δ – корегуючий чинник (поправка) у дисперсійному аналізі;
C_Y, C_P, C_V, C_Z і т. п. – суми квадратів відхилень для різних джерел варіювання в дисперсійному аналізі;

$\chi^2_{ф}$ – фактичне значення критерія *хі*-квадрат Пірсона;
χ^2_{05}, χ^2_{01} – табличні значення критерія *хі*-квадрат для 5 %-го і 1 %-го рівня значущості;
r – коефіцієнт лінійної кореляції;
s_r – помилка коефіцієнта лінійної кореляції;
b_{yx} – коефіцієнт регресії y за x;
s_b – помилка коефіцієнта регресії;
s_{yx} – помилка відхилення від регресії;
$r_{xy \cdot z}$, $r_{xz \cdot y}$, $r_{zy \cdot x}$ – партикулярні лінійні коефіцієнти кореляції;
$R_{xy \cdot z}$, $R_{xz \cdot y}$, $R_{zy \cdot x}$ – множинні лінійні коефіцієнти кореляції;
η_{yx} – кореляційне відношення y до x;
s_η – помилка кореляційного відношення;
cov – коваріація;
h^2 – коефіцієнт спадковості.

Алфавітний покажчик

Словник термінів

• • • • •

Абсолютна помилка (*absolute error*) – різниця між істинними та фактичними показниками.

Абсолютне відхилення (*absolute deviation*) – середнє відхилення за модулем, відношення суми всіх відхилень за модулем до числа відхилень.

Автокореляція (*autocorrelation*) – стохастична варіація ознаки в часі.

Автоматична класифікація, або **кластерний** (**класифікаційний**) **аналіз** (*cluster analysis*), – автоматичне (комп'ютерне) групування даних за близькістю значень ознак.

Адитивність (*additivity*) – рівність сумісного ефекту досліджуваних чинників сумі окремих ефектів цих чинників.

Алгоритм (*algorithm)* – схема опису правильної послідовності розрахунків.

Арифметична або **вибіркова середня** (*arithmetic mean, average*) – середнє значення об'єктів вибірки за будь-якою ознакою.

Асиметрія, або **скошеність** (*skewness*) – несиметричність кривої розподілу, тобто лівий бік крутий, а правий – пологий або навпаки.

Атрибут (*attribute*) – ознака, властивість або якість.

Бартлет-тест (*Bartlett's test*) – статистичний тест на незалежність дисперсії (приналежність до однієї генеральної сукупності) або відповідність нормальному розподілу.

Біологічна модель (*biological model*) – відображення природи і характеру діючих у досліді чинників з урахуванням поставленої проблеми.

Біоессей (bioassay) – випробування активних речовин або препаратів на живих організмах.

Блок (*block*) – територіальна одиниця досліду, аналог повторення або його частини (*incomplete block*), який включає не менше двох варіантів.

Варіант (*variant or treatment*) – досліджуваний вид рослин, сорт, добриво, пестицид, інші чинники та умови життя рослин або їх дози, агротехнічні заходи, технології і т. ін.

Варіація (*variation*) – дисперсія або розсіяність, з одного боку, а з іншого – мінливість, або варіабельність (*variability*), властивість окремих об'єктів (особин) відрізнятися один від одного за тими чи іншими ознаками.

Вибірка (*sample*) – частина об'єктів сукупності (вибіркова сукупність), які підлягають безпосередньому обліку, вимірюванню.

Виключення (*correction on plot size*) – частина площі ділянки, яку виключають з обліку у зв'язку з пошкодженнями рослин, викликаними недосліджуваними чинниками.

Вирівнюваний посів, або **вирівнюваний дослід** (*blindfold experiment*), – суцільна сівба культури з метою оцінки ґрунтового покриву і його вирівнювання за 2–4 роки до закладання досліду.

Гіпотеза нульова – припущення про відсутність різниці в сукупності на підставі дослідження вибірки; гіпотеза альтернативна, або робоча, – протилежність нульовій (*hypothesis*: *nul. alternative*).

Гіпотеза статистична – припущення про закономірність розподілу показників ознаки у сукупності (*statistical hypothesis*).

Гістограма (*histogram or strip chart*) – стовбчасто-злитий графік, який характеризує чисельність (частоту прояви ознаки) окремих показників за групами великої вибірки; графічне представлення інтервального варіаційного ряду; сукупність зімкнутих прямокутників, де основою служить груповий інтервал, а висотою – чисельність класа.

Групування, або **класифікація** (*classification*), – розподіл великої вибірки на класи за певною ознакою.

Дані (*data*) – сукупність результатів наукового дослідження в цифрах.

Дата (*value*) – окремі значення досліджуваного показника.

Ділянка (*plot*) – елементарна територіальна одиниця польового досліду; дослідна (посівна) ділянка – ділянка, на якій проводять усі роботи, передбачені технологією вирощуваної культури та програмою досліджень; облікова ділянка (*yield plot*) – дослідна ділянка за виключенням захисних смуг, призначена для проведення обліків і відбору зразків.

Дисперсійний аналіз, або **аналіз варіації** (*variance analysis*), – метод статистичної оцінки дослідних даних, згрупованих за рядками та колонками з метою розкладання загальної варіації даних і ступенів свободи за джерелами варіації або за компонентами досліду (мінімум двох: досліджувані варіанти та випадковий чинник) і оцінки істотності різниць за F-критерієм – відношенням дисперсії варіантів до дисперсії помилок.

Дисперсія (*variance*) – розсіювання або розкид у загальновживаному сенсі і статистичний параметр (характеристика), який представляє середній квадрат (*mean square*) відхилень окремих показників сукупності (вибірки) від генеральної (вибіркової) середньої.

Довірча ймовірність (*confidence probability*), або статистична достовірність, дорівнює різниці: $P = 100 - \alpha$.

Довірчий інтервал (*confidence interval*, c. *limits*) – область знаходження окремих показників або середньої генеральної сукупности із заданою ймовірністю.

Дробовий облік, або **сліпий дослід** (*blindfold or dummy experiment*), – облік урожаю культури суцільної сівби на площадках перед закладанням досліду з метою рекогносцировки варіювання родючості ґрунту.

Експеримент, або **дослід** (*experiment*), – метод наукового дослідження, який проводиться шляхом активного вторгнення в природні процеси і явища або їх репрезентативного моделювання на підставі ретельно розробленого плану з метою отримання нових знань за чітко сформульованою проблемою.

Експериментальна помилка (*experimental error*) – сукупна помилка (*components of error*), обумовлена варіабельністю експериментальних одиниць (власне випадкова помилка – *random error*), недостатнім контролем побічних чинників (систематична помилка, або помилка зміщення – *systematic error or bias*), а також недбалістю в процесі проведення досліду (груба помилка – англійський аналог відсутній). Оскільки два останніх види помилок не мають розрахункових формул, то вони доповнюють випадкову помилку, яка і представляє, у підсумку, узагальнену помилку експерименту.

Експериментальна статистика (*statistical analysis*) – оцінка порівнюваних досліджень і експериментів на підставі контрольних критеріїв, або статистичних тестів.

Захисна смуга (*protective strips*), – бічні (за довжиною) та кінцеві або лобові (за шириною) частини дослідних ділянок для захисту облікової площі від крайового впливу (*marginal effects*) або випадкових чинників.

Значущість, або **істотність** (*significance*), – рівень значущості, ймовірність або ризик зробити помилковий висновок на підставі вибірки (наукового дослідження, досліду).

Квантилі (*quantile or fractile*) – рівні частини ранжованого ряда. Медіана, квартилі, децилі та центилі ділять ряд відповідно на 2, 4, 10 і 100 рівних частин. Числовими виразами служать граничні значення двох сусідніх частин.

Коваріаційний аналіз (*covariance analysis*) – потрійний дисперсійний аналіз (аналіз варіації показників x, y і їх добутків xy) плюс регресійний аналіз y на x з метою усунення впливу коваріанти та підвищення точності порівняння середніх.

Коваріація (*covariation*) – сумісна варіація двох ознак, розраховується як сума добутків відхилень (*covariance*) окремих значень x і y від своїх середніх.

Контроль (*control- or standard*) – варіант (варіанти) досліду, де відсутній досліджуваний чинник; абсолютний контроль – нульовий варіант (*zero treatment*); стандарт (*st*) – контрольний варіант – традиційний або широко поширений захід (сорт або гібрид) – база порівняння для нових заходів (сортів або гібридів).

Коригуючий чинник (*correction factor*) – коефіцієнт, або множник, для перетворення вихідних даних шляхом множення.

Коригуючий показник (*correction term*) – поправка зі знаком мінус, яка усуває відхилення фактичного результату від його справжньої величини.

Кореляційна матриця (*correlation matrix*) – впорядкована сукупність коефіцієнтів множинної кореляції.

Кореляційне відношення (*correlation ratio*) – показник тісноти (сили) зв'язку для лінійної та криволінійної залежності.

Кореляційний аналіз, або **аналіз кореляції** (*correlation analysis*), – методи вивчення сили (тісноти) стохастичного (імовірнісного) зв'язку між ознаками кількісної мінливості: коефіцієнт кореляції (лінійна залежність), кореляційне відношення (криволінійна залежність) і кореляційна матриця (множинна залежність).

Кореляція (*correlation*) – імовірнісний зв'язок однієї ознаки з іншою (простий) або однієї з кількома іншими (множинний), який визначається в середньому для достатньо великого числа спостережень. В загальнодоступному сенсі кореляція – це коли зміни однієї ознаки викликають зміни іншої. Чим більші ці зміни, тим тісніший кореляційний зв'язок. Асимптотичною межею кореляції є функціональний зв'язок або причинно-наслідкова залежність.

Коефіцієнт, або **індекс детермінації** (*coefficient of determination*), – квадрат коефіцієнта кореляції, частка (%) змін залежної змінної y, зумовлена змінами x.

Коефіцієнт кореляції (*correlation coefficient*) – показник тісноти (сили) лінійного зв'язку двох ознак. Він варіює в межах від –1 до +1: $-1 < r < 1$ (не округляти до 1, а брати третю значущу цифру після коми). Кореляція вважається сильною, якщо $r > 0,70$, середньою, якщо $0,30 < r < 0,70$ і слабкою, якщо $r < 0,30$.

Коефіцієнт регресії (*regression coefficient*) – кількісна міра зв'язку, яка показує, наскільки зміниться залежна змінна y (у своїх одиницях), якщо x зміниться на одиницю вимірювання.

Крива (*curve*) – графічна форма залежності ознак; теоретична крива строго підпорядковується математичній формулі, емпірична – зв'язує точки фактичних спостережень; розрахункова – крива, яку отримують за рівнянням регресії.

Критерій Крускаль-Уоліса, або *H*-критерій (*Kruskal-Wallis H-test*), – критерій для множинних порівнянь середніх у досліді (ранжовані *НІР*), відрізняється від критерію Тьюкі більш високими значеннями.

Критерій (метод) Тьюкі (*Tukey's test procedure*) – ранговий критерій для множинних порівнянь ранжованого ряду дослідних середніх однакової повторності. Розрахункову величину отримують з відношення різниці будь-якої пари середніх до загальної помилки середніх.

Критерій (метод) Шеффе (*Scheffe's procedure*) – ранговий критерій для множинних порівнянь як окремих середніх різної повторності, так і їх груп різної чисельності на основі контрастів. На відміну від критерію Тьюкі має вищий поріг істотності.

Критичний поріг Тьюкі, або *HSD*-критерій (*HSD-test honest significant difference*) – тест на істотність різниці між двома середніми, уточнює показник *НІР* для окремих пар ранжованого ряду середніх.

Латинський квадрат (*latin square*) – план досліду, де число варіантів дорівнює числу повторень ($v = n$). Подвійний набір повторень (за рядами та стовбцями) забезпечує контроль варіювання ґрунтової родючості в двох взаємно перпендикулярних напрямках.

Латинський прямокутник (*latin rectangle*) – латинський квадрат, де повторність кратна числу варіантів, тобто $v/n = k$, де $k \geq 2$.

Медіана (*median*) – центральне значення ознаки в ранжованому ряду за непарного обсягу вибірки або середнє з двох центральних значень за парного обсягу вибірки.

Метод блоків (*МБ*) (*CBD-block design or complete block design*) – те саме, що й метод організованих повторень (*МОП*).

Метод контрастів – спрощена версія теста Шеффе.

Метод організованих повторень (*МОП*) – план досліду, де варіанти скомпоновані (набір варіантів) на окремих частках дослідного поля з метою контролю одного напряму варіації ґрунтової родючості.

Метод перехресних ділянок (*МПД*) (*split block or criss-cross*) – план двофакторного досліду, де ділянки отримують шляхом розщеплення повторень у двох взаємно перпендикулярних напрямках.

Метод повної рендомізації (*МПР*) (*completely randomized design, CRD*) – план досліду, де варіанти розміщують за ділянками на підставі рендомізації (випадковості), тобто кожний варіант має рівний шанс потрапити на будь-яку ділянку.

Метод розщеплених ділянок (*МРД*) (*split-plot*) – план багатофакторного досліду, де ділянки для наступного чинника отримують шляхом розщеплення ділянок попереднього чинника: ділянки першого порядку (більші або головні ділянки) для чинника *А*, ділянки другого порядку для чинника *В* і т. д.

Мода (*mode or modal value*) – значення вибірки, яке найчастіше в ній трапляється.

НІР – найменша істотна різниця, межа випадкових відхилень або критична різниця (*LSD or CD, least significant difference or critical difference*).

Облік урожаю (*yield estimation*) – зважування товарної частки продукції (зерно, сіно, плоди, клубні): з усієї облікової ділянки (суцільний метод) або її незначної частини (метод пробних площадок) і зважування загальної продукції з розрахунком товарного врожаю на основі його частки у пробному снопі (метод пробного снопа).

Описова статистика (*descriptive statistics*) – розрахунок оцінок параметрів сукупності на основі вибірки або статистичних характеристик центральної тенденції, варіації, помилки (точності) вибірки, скошеності та плосковерхівковості (гостроверхівковості) емпіричної кривої.

Оцінка істотності різниці (різниць) (*significance test*) – точкова (*point*), за допомогою критеріїв, та інтервальна (*confidence interval estimation*) – оцінка на основі співставлення довірчих інтервалів (перекриваються чи ні).

Помилка вибірки (середньої) (*sample error or error of mean*) – відхилення середньої фактичних спостережень (середньої вибіркової) від середнього значення ознаки в досліджуваній сукупності (генеральної середньої).

Помилка першого роду (*error of first kind or type I error or α-error*) – відкидання правильної нульової гіпотези H_0 (прийняття неправильної H_A).

Помилка другого роду (*error of second kind or type II*) – відкидання правильної альтернативної гіпотези H_A (прийняття неправильної гіпотези H_0).

Парцели – ділянки (колишня назва) або смуги, які використовуються як повторення.

План досліду (*design of experiment*) – представлення основних параметрів досліду: варіантів, повторень, захисних смуг і т. ін. у вигляді рисунку.

Повторення (*replication or block*) – частина досліду з повним набором варіантів.

Повторність (*number of replications*) – число експериментальних одиниць (посудин, ділянок) для одного варіанта.

Поліном (*polynomial*) – рівняння поліноміальної (криволінійної) регресії загального вигляду: $y(x) = a + b_1x + b_2x^2 + ... + b_mx^m$, частіше другого ступеня, яке найбільш підходить для опису залежностей, форма яких ще невідома.

Поліноміальний тренд (*polynomial trend*) – нелінійний тренд для ймовірнісних процесів у часі, наприклад: $y(t) = a + b_1t + b_2t_2$.

Пробіт-аналіз (*probit analysis*) – статистична оцінка результатів випробування активних речовин на живих організмах.

Рендомізація (*randomization*) – принцип складання вибірки і розміщення варіантів випадково – на підставі жеребкування або таблиці випадкових чисел.

Рендомізований блок (*randomized complete block design*) – метод рендомізованих повторень (*МРП*) – план досліду з випадковим розміщенням варіантів за всіма повтореннями, різновид методу організованих повторень (*МОП*).

Рекогносцирувальний (розвідувальний) посів (*dummy experiment*) – суцільна сівба перед закладанням досліду (бажано дослідної культури) з обліком урожаю на площадках 10 м^2 і більше (дробовий облік) для виявлення виду і міри варіювання ґрунтової родючості.

Репліка (від ісп. *replica* – повторення, або реплікація), або неповний блок (*incomplete block*), – набір варіантів усередині повторення багатофакторного досліду.

Репрезентативність (*representability*) – відповідність вибірки генеральній сукупності і наявність достатнього обсягу.

Рівень значущості (*error of significance*) – ймовірність помилки (*probability of error*), помилка першого роду (*error rate*) = α, ризик зробити помилкове заключення про істотність досліджуваних ефектів.

Робастість (*robustness*) – чутливість тесту (експерименту).

Системний блок (англійська версія відсутня) – план досліду з однаковою послідовністю розміщення варіантів у всіх повтореннях, різновид *МОП*.

Скринінг (*screening*) – випробування, як правило, хімічних препаратів, суть якого полягає в поетапному відборі з великої кількості вихідних варіантів (препаратів).

Сліпий дослід (*dummy or blindfold experiment*) – розвідувальний посів дослідної культури з дробовим обліком урожаю.

Спостереження (*observation*, *measurement*) – простіше наукове дослідження, пасивна регістрація результатів вимірювання, аналізів і обліків сукупності за будь-якою ознакою.

Сукупність (*population*) – множина об'єктів, особин або чинників, які підлягають вивченню (генеральна, або вся сукупність).

Статистична достовірність, або **довірча ймовірність** (*confidence probability*), – рівень об'єктивності результатів проведеного дослідження $P = 100 - \alpha$.

Статистична модель досліду (*statistical model*) – математична формалізація агрономічної проблеми шляхом порівняння діючих у досліді чинників, представлених своїми символами.

Статистична помилка (*sample error*) – помилка вибірки або відхилення характеристик вибірки, насамперед середньої, від параметрів усієї сукупності. Вибірка, будучи частиною сукупності, не може повністю їй відповідати.

Статистична сукупність (*data*) – сукупність цифр або даних результатів досліджень.

Статистичний тест (H_0-*test*) – перевірка нульової гіпотези, прийняття H_0 або її відкидання (прийняття H_A або H_1).

Схема досліду (*scheme of experiment*) – сукупність варіантів досліду, яка представляється в певній послідовності з указанням їх цифрового коду та назви.

Схематичний план досліду (*plan of experiment arrangement*) – план досліду з указанням розмірів його складових, а також захисних смуг, географічної і топографічної прив'язки.

Техніка (методика) польового досвіду (*experimentation*) – планування і проведення досвіду плюс біометричні вимірювання та обліки, а також технологія польових робіт: обробки ґрунту, внесення добрив і пестицидів, посіву та збирання дослідної культури.

Точковий графік (*dot diagram*) – точки регістрації показників двох ознак парами у двомірній системі координат *X*, *Y*.

Точність досліду (*precision and accuracy*) – умовне поняття, яке визначається показником відносної узагальненої помилки дослідних середніх (*E*): до 3 % – висока, а при $E \geq 7$ % – неприйнятна точність.

Факторіальність схеми досліду – наявність усіх комбінацій (варіантів взаємодії) досліджуваних чинників.

Факторний дослід (*factor model*) – дослід, у якому досліджують не менше двох чинників, або дослід, де один чинник досліджуваний, а другий – контрольований.

Число ступенів свободи (*degrees of freedom*) – число незалежних відхилень (порівнянь).

Шахматний метод (*chess arrangement*) – план досліду, різновид систематичного блоку з ярусним зміщенням у послідовності розміщення варіантів.

Список рекомендованої літератури

1. Грицаєнко З. М. Методи біологічних та агрохімічних досліджень рослин і ґрунтів / З. М. Грицаєнко, А. О. Грицаєнко, В. П. Карпенко. – К.: ЗАТ «Нічлава», 2003. – 320 с.
2. Доспехов Б. А. Методика полевого опыта (с основами статистической обработки результатов исследований) / Б. А. Доспехов. – М.: Агропромиздат, 1985. – 351 с.
3. Кирюшин Б. Д. Основы научных исследований в агрономии / Б. Д. Кирюшин, Р. Р. Усманов, И. П. Васильев. – М.: Колос, 2009. – 398 с.
4. Лісовал А. П. Методи агрохімічних досліджень. – К.: НАУ, 2001. – 247 с.
5. Методика наукових досліджень в агрономії: навч. посібник / В. Г. Дідора, О. Ф. Смаглій, Е. Р. Ермантраут [та ін.]. – К.: «Центр навч. л-ри», 2013. – 264 с.
6. Основи наукових досліджень в агрономії: підручник / В. О. Єщенко, П. Г. Копитко, В. П. Опришко; за ред. В. О. Єщенка. – К.: Дія, 2005. – 288 с.
7. Основи наукових досліджень в агрономії: підручник / В. О. Єщенко, П. Г. Копитко, П. В. Костогриз, В. П. Опришко; за ред. В. О. Єщенка. – Вінниця: ПП«ТД«Едельвейс і К»», 2014. – 332 с.
8. Тимошенко І. І. Основи наукових досліджень в агрономії / І. І. Тимошенко, З. М. Майщук, Г. О. Косилович. – Львів: ЛДАУ, 2004. – 111 с.

Додатки

Додаток А

Таблиця випадкових чисел (*за Б. О. Доспеховим, 1985*)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	10	09	73	25	33	76	52	01	35	68	34	67	35	48	76	80	95	90	91	17
2	37	54	20	48	05	69	89	47	42	39	24	80	52	40	37	20	63	61	04	02
3	08	42	26	89	53	14	64	50	93	60	23	20	90	25	60	15	95	33	47	64
4	99	01	90	25	29	09	37	67	07	51	38	31	13	11	63	88	67	67	43	97
5	12	80	79	99	70	80	15	73	61	74	64	03	23	66	53	98	95	11	68	77
6	66	06	57	47	17	34	07	27	68	05	36	69	73	61	70	65	81	33	98	85
7	31	06	01	08	05	45	57	18	24	60	35	30	34	26	14	86	79	90	74	39
8	85	26	97	76	02	02	05	16	56	29	68	66	57	48	18	73	05	38	52	47
9	63	57	33	21	35	05	32	54	70	84	90	55	35	75	48	28	46	82	87	09
10	73	79	64	47	53	03	52	96	47	87	35	80	83	42	82	60	93	52	03	34
11	98	52	01	77	67	14	90	56	86	70	22	10	94	05	58	60	97	09	34	33
12	11	80	50	54	31	39	80	82	77	23	50	72	56	82	48	29	40	52	42	01
13	83	45	29	96	34	06	28	89	90	38	13	74	67	00	78	18	47	54	06	10
14	88	68	54	02	00	86	50	75	84	01	36	76	66	79	51	90	36	47	64	93
15	99	59	46	73	48	87	51	76	49	69	91	82	60	89	28	93	78	56	13	68
16	65	48	11	76	74	17	46	85	09	50	58	04	77	69	74	73	03	95	71	86
17	80	12	43	56	35	17	72	70	80	15	45	31	82	23	74	21	11	57	82	53
18	74	35	99	98	17	77	40	27	72	14	43	23	60	02	10	45	52	16	42	37
19	69	91	62	68	03	66	25	22	91	48	36	93	68	72	03	76	62	11	39	90
20	09	89	32	05	05	14	22	56	85	14	46	42	75	67	88	96	29	77	88	22
21	91	49	91	45	23	68	47	92	76	86	46	16	28	35	54	94	75	08	99	23
22	80	33	69	45	98	26	94	03	68	58	70	29	73	41	35	53	14	03	33	40
23	44	10	48	19	49	85	15	74	79	54	32	97	92	65	75	57	60	04	08	81
24	12	55	07	37	42	11	10	00	20	40	12	86	07	46	97	96	64	48	94	39
25	63	60	64	93	29	16	50	53	44	84	40	21	95	25	63	43	65	17	70	82
26	61	19	69	04	46	26	45	74	77	74	51	92	43	37	29	65	39	45	95	93
27	15	47	44	52	66	95	27	07	99	53	59	36	78	38	48	82	39	61	01	18
28	94	55	72	85	73	67	89	75	48	87	54	62	24	44	31	91	19	04	25	92
29	42	48	11	62	13	97	31	40	87	21	16	86	84	87	67	03	07	11	20	59
30	23	52	37	83	17	73	20	88	98	37	68	93	59	14	16	26	25	22	96	63

Додаток Б

Критичні значення одно- і двостороннього *t*-критерію, χ^2-критерію та мінімального коефіцієнта кореляції *r* для оцінки істотності різниць (кореляції) залежно від обсягу вибірки (ступенів свободи)

Число ступенів свободи	*t*-двосторонній при α, що дорівнює			χ^2		*r*	
	0,10	0,05	0,01	0,05	0,01	0,05	0,01
1	6,314	12,706	63,657	3,84	6,63	0,997	0,999
2	2,920	4,303	9,925	5,99	9,21	0,950	0,990
3	2,353	3,182	5,841	7,81	11,34	0,878	0,959
4	2,132	2,776	4,604	9,49	13,28	0,811	0,917
5	2,015	2,571	4,032	11,07	15,09	0,754	0,874
6	1,943	2,447	3,707	12,59	16,81	0,707	0,834
7	1,895	2,365	3,499	14,07	18,48	0,666	0,798
8	1,859	2,306	3,355	15,51	20,09	0,632	0,765
9	1,833	2,262	3,250	16,92	21,67	0,602	0,735
10	1,812	2,228	3,169	18,31	23,21	0,576	0,708
11	1,796	2,201	3,106	19,68	24,72	0,553	0,684
12	1,782	2,179	3,054	21,07	26,22	0,532	0,661
13	1,771	2,160	3,012	22,36	27,69	0,514	0,641
14	1,761	2,145	2,977	23,68	29,14	0,497	0,623
15	1,753	2,131	2,947	25,00	30,58	0,482	0,606
16	1,746	2,120	2,921	26,30	32,00	0,468	0,590
17	1,740	2,110	2,898	27,59	33,41	0,456	0,575
18	1,734	2,101	2,878	28,88	34,81	0,444	0,561
19	1,729	2,093	2,861	30,14	26,19	0,433	0,549
20	1,725	2,086	2,845	31,41	37,57	0,423	0,537
21	1,721	2,080	2,831	32,67	38,93	0,413	0,526
22	1,717	2,074	2,819	33,92	40,29	0,404	0,515
23	1,714	2,069	2,807	35,17	41,64	0,396	0,505
24	1,711	2,064	2,797	36,42	42,98	0,388	0,496
25	1,708	2,059	2,787	37,65	44,31	0,381	0,487
26	1,706	2,056	2,779	38,89	45,64	0,374	0,478
27	1,703	2,052	2,771	40,11	46,96	0,367	0,470
28	1,701	2,048	2,763	41,34	48,28	0,361	0,463
29	1,699	2,045	2,756	42,56	49,59	0,355	0,456
30	1,697	2,042	2,750	43,77	50,89	0,349	0,449
50	1,676	2,009	2,678	67,50	76,15	0,273	0,354
100	1,660	1,984	2,626	124,34	135,81	0,195	0,254
500	1,648	1,965	2,586	–	–	0,088	0,115
∞	1,645	1,960	2,576	–	–	–	–

Додаток В

Квантилі *F*-розподілу (теоретичні значення *F*-критерію)

Ступені свободи дисперсії знаменника	Ступені свободи дисперсії чисельника																
	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	100	∞
	Ймовірність помилки α = 0,05 (05 %)																
1	161	199	216	225	230	234	237	239	241	242	246	248	250	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66	8,62	8,60	8,58	8,56	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80	5,74	5,71	5,70	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,62	4,56	4,50	4,46	4,44	4,40	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,81	3,77	3,75	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44	3,38	3,34	3,32	3,28	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,08	3,05	3,03	2,98	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,86	2,82	2,80	2,76	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,70	2,67	2,64	2,59	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,72	2,65	2,57	2,53	2,50	2,45	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,65	2,54	2,46	2,42	2,40	2,35	2,30
13	4,64	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,54	2,46	2,38	2,34	2,32	2,26	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,49	2,39	2,31	2,27	2,24	2,19	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,25	2,21	2,18	2,12	2,07
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,40	2,35	2,20	2,12	2,04	1,99	1,96	1,90	1,84
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,25	2,24	2,10	2,00	1,92	1,87	1,84	1,77	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93	1,84	1,79	1,76	1,69	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	1,94	1,84	1,74	1,69	1,66	1,59	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,92	1,78	1,69	1,63	1,60	1,52	1,51
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,75	1,68	1,57	1,51	1,51	1,39	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,67	1,57	1,46	1,40	1,40	1,24	1,01

Продовження дод. В

Ступені свободи дисперсії знаменника	Ступені свободи дисперсії чисельника																
	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	100	∞
Ймовірність помилки α = 0,01 (01 %)																	
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6157	6208	6258	6286	6302	6334	6366
2	98,5	99,0	99,2	99,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,47	99,48	99,5	99,5	99,5
3	34,1	30,8	29,4	28,7	28,2	27,9	27,7	27,5	27,3	27,2	26,9	26,7	26,50	26,41	26,4	26,2	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,2	14,0	13,83	13,74	13,7	13,6	13,5
5	16,3	13,3	12,1	11,4	10,9	10,7	10,4	10,3	10,2	10,0	9,72	9,55	9,38	9,29	9,24	9,13	9,02
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,56	7,39	7,23	7,14	7,09	6,99	6,88
7	12,2	9,50	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,31	6,15	5,98	5,90	5,85	5,75	5,65
8	11,3	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,52	5,36	5,20	5,11	5,06	4,96	4,86
9	10,6	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	4,96	4,80	4,64	4,56	4,51	4,41	4,31
10	10,0	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,56	4,41	4,25	4,17	4,12	4,01	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,25	4,10	3,94	3,86	3,80	3,70	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,02	3,86	3,70	3,61	3,56	3,46	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	3,82	3,67	3,51	3,42	3,37	3,27	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,66	3,51	3,34	3,26	3,21	3,11	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,57	3,36	3,20	3,12	3,07	2,97	2,87
20	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,09	2,94	2,77	2,69	2,63	2,53	2,42
25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,21	3,13	2,84	2,70	2,54	2,45	2,40	2,29	2,17
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,70	2,55	2,38	2,29	2,24	2,13	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,80	2,44	2,37	2,20	2,11	2,05	1,94	1,81
50	7,17	5,06	4,20	3,72	3,41	3,18	3,02	2,88	2,78	2,70	2,42	2,26	2,10	2,00	1,94	1,82	2,68
100	6,90	4,82	3,98	3,51	3,20	2,99	2,82	2,69	2,59	2,51	2,22	2,06	1,89	1,79	1,73	1,59	1,43
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,04	1,87	1,69	1,59	1,52	1,36	1,00

Додаток Г

Значення χ^2-критерію

Число ступенів свободи	Рівень значущості							
	0,99	0,95	0,75	0,50	0,25	0,10	0,05	0,01
1	–	–	0,10	0,45	1,32	2,71	3,84	6,63
2	0,02	0,10	0,58	1,39	2,77	4,61	5,99	9,21
3	0,11	0,35	1,21	2,37	4,11	6,25	7,81	11,34
4	0,30	0,71	1,92	3,36	5,39	7,78	9,49	13,28
5	0,55	1,15	2,67	4,35	6,63	9,24	11,07	15,09
6	0,87	1,64	3,45	5,35	7,84	10,64	12,59	16,81
7	1,24	2,17	4,25	6,35	9,04	12,02	14,07	18,48
8	1,65	2,73	5,07	7,34	10,22	13,36	15,51	20,09
9	2,09	3,33	5,90	8,34	11,39	14,68	16,92	21,67
10	2,56	3,94	6,74	9,34	12,55	15,99	18,31	23,21
11	3,05	4,57	7,58	10,34	13,70	17,28	19,68	24,72
12	3,57	5,23	8,44	11,34	14,85	18,55	21,03	26,22
13	4,11	5,89	9,30	12,34	15,98	19,81	22,36	27,69
14	4,66	6,57	10,17	13,34	17,12	21,06	23,68	29,14
15	5,23	7,26	11,04	13,34	18,25	22,31	25,00	30,58
16	5,81	7,96	11,91	15,34	19,37	23,54	26,30	32,00
17	6,41	8,67	12,79	16,34	20,49	24,77	27,59	33,41
18	7,01	9,39	13,68	17,34	21,60	25,99	28,87	34,81
19	7,63	10,12	14,56	18,34	22,72	27,20	30,14	36,19
20	8,26	10,85	15,45	19,34	23,83	28,41	31,41	37,57
21	8,90	11,59	16,34	20,34	24,93	29,62	32,67	38,93
22	9,54	12,34	17,24	21,34	26,04	30,81	33,92	40,29
23	10,20	13,09	18,14	22,34	27,14	32,01	35,17	41,64
24	10,86	13,85	19,04	23,34	28,24	33,20	36,42	42,98
25	11,52	14,61	19,94	24,34	29,34	34,38	37,65	44,31
26	12,20	15,38	20,84	25,34	30,43	35,56	38,89	45,64
27	12,88	16,15	21,75	26,34	31,53	36,74	40,11	46,93
28	13,56	16,93	22,66	27,34	32,62	37,92	41,34	48,28
29	14,26	17,71	23,57	28,34	33,71	39,09	42,56	49,59
30	14,95	18,49	24,48	29,34	34,80	40,26	43,77	50,89
40	22,16	26,51	33,66	39,34	45,62	51,80	55,76	63,69
50	29,71	34,76	42,94	49,33	56,33	63,17	67,50	76,15
60	37,48	43,19	52,29	59,33	66,98	74,40	79,08	88,38
70	45,44	51,74	61,70	69,33	77,58	85,53	90,53	100,42
80	53,54	60,39	71,14	79,33	88,13	96,58	101,88	112,33
90	61,75	69,13	80,62	89,33	98,64	107,56	113,14	124,12
100	70,06	77,93	90,13	99,33	109,14	118,50	124,34	135,81

Додаток Д

Контрольні значення критерію *U* Манн-Уітні (T_{05} Уайта-Вілкоксона) для непов'язаних вибірок
(H_0: $T_ф > T_r$; $T_ф = R_{min}$: R_1 або R_2; n_1 і n_2 можна міняти місцями)

n_2	n_1: 4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	10																
5	11	17															
6	12	18	26														
7	13	20	27	36													
8	14	21	29	38	49												
9	14	22	31	40	51	52											
10	15	23	32	42	53	65	78										
11	16	24	34	44	55	68	81	96									
12	17	26	35	46	58	71	84	99	115								
13	18	27	37	48	60	73	88	103	119	136							
14	19	28	38	50	62	76	91	106	123	141	160						
15	20	29	40	52	65	79	94	110	127	145	164	184					
16	21	30	42	54	67	82	97	113	131	150	169	190	211				
17	21	32	43	56	70	84	100	117	135	154	174	195	217	240			
18	22	33	45	58	72	87	103	121	139	158	179	200	222	246	270		
19	23	34	46	60	74	90	107	124	143	163	183	205	228	252	277	303	
20	24	35	48	62	77	93	110	128	147	167	188	210	234	258	283	309	337
22	26	36	51	66	81	98	116	135	155	176	198	221	245	270	293	323	351
24	27	40	54	70	86	104	122	142	163	185	207	231	256	282	309	337	368

Додаток Е

Критичні значення *Q*-критерію Розенбаума для рівнів значущості $P_1 \leq 0{,}05$ і $P_1 \leq 0{,}01$

n	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
$P_1 = 0{,}05$																
11	6															
12	6	6														
13	6	6	6													
14	7	7	6	6												
15	7	7	6	6	6											
16	7	7	7	7	6	6										
17	7	7	7	7	7	7	7									
18	7	7	7	7	7	7	7	7								
19	7	7	7	7	7	7	7	7	7							
20	7	7	7	7	7	7	7	7	7	7						
21	8	7	7	7	7	7	7	7	7	7	7					
22	8	7	7	7	7	7	7	7	7	7	7	7				
23	8	8	7	7	7	7	7	7	7	7	7	7	7			
24	8	8	8	8	8	8	8	8	8	8	7	7	7	7		
25	8	8	8	8	8	8	8	8	8	8	8	7	7	7	7	
26	8	8	8	8	8	8	8	8	8	8	8	7	7	7	7	7
$P_1 = 0{,}01$																
11	9															
12	9	9														
13	9	9	9													
14	9	9	9	9												
15	9	9	9	9	9											
16	9	9	9	9	9	9										
17	10	9	9	9	9	9	9									
18	10	10	9	9	9	9	9	9								
19	10	10	10	9	9	9	9	9	9							
20	10	10	10	10	9	9	9	9	9	9						
21	11	10	10	10	9	9	9	9	9	9	9					
22	11	11	10	10	10	9	9	9	9	9	9	9				
23	11	11	10	10	10	10	9	9	9	9	9	9	9			
24	12	11	11	10	10	10	10	9	9	9	9	9	9	9		
25	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9	
26	12	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9

Додаток Ж

Показники критерію Ньюмана-Келса (*q*), Дуннетта (*D*) та максимального модуля (*m*) для оцінки істотності різниць ранжованих середніх* (α = 0,05)

Ступені свободи	Ранг середньої порівняно з базою порівняння, r**										
	1	2	3	4	5	6	7	8	9	10	15
***q*-критерій: $HIP_{05} = q \cdot s_y$; база порівняння $= \bar{y}_i$ (будь-яка середня)**											
5	2,57	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	7,00	7,72
6	2,45	3,46	4,34	4,90	5,31	5,63	5,90	6,12	6,32	6,49	7,14
7	2,36	3,34	4,17	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,76
8	2,31	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,48
10	2,23	3,15	3,88	4,33	4,65	4,91	5,12	5,31	5,46	5,60	6,11
15	2,13	3,01	3,67	4,08	4,37	4,60	4,78	4,94	5,08	5,20	5,65
20	2,09	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,99	5,01	5,43
30	2,04	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	5,21
60	2,00	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	5,00
***D*-критерій: $HIP_{05} = D \cdot s_d$; база порівняння $= \bar{y}_{st}$ (стандарт або контроль)**											
5	2,57	3,03	3,29	3,48	3,62	3,73	3,82	3,90	3,97	4,03	4,26
6	2,45	2,86	3,10	3,26	3,39	3,49	3,57	3,64	3,71	3,76	3,97
7	2,36	2,75	2,97	3,12	3,24	3,33	3,41	3,47	3,53	3,58	3,78
8	2,31	2,67	2,88	3,02	3,13	3,22	3,29	3,35	3,41	3,46	3,64
10	2,23	2,57	2,76	2,89	2,99	3,07	3,14	3,19	3,24	3,29	3,45
15	2,13	2,44	2,61	2,73	2,82	2,89	2,95	3,00	3,04	3,08	3,23
20	2,09	2,38	2,54	2,65	2,73	2,80	2,86	2,90	2,95	3,98	3,12
30	2,04	2,32	2,47	2,58	2,66	2,72	2,77	2,82	2,86	2,89	3,02
60	2,00	2,27	2,41	2,51	2,58	2,64	2,69	2,73	2,77	2,80	2,92
***m*-критерій: $HIP_{05} = m \cdot s_d$; база порівняння $= \bar{y}_0$ (середня за дослідом)**											
5	2,57	3,09	3,40	3,62	3,78	3,92	4,04	4,14	4,23	4,31	4,60
10	2,23	2,61	2,83	2,98	3,10	3,19	3,28	3,35	3,41	3,47	3,68
15	2,13	2,47	2,67	2,81	2,91	2,99	3,06	3,12	3,18	3,23	3,42
20	2,09	2,41	2,59	2,72	2,82	2,90	2,97	3,02	3,08	3,13	3,31
30	2,04	2,35	2,52	2,64	2,73	2,80	2,86	2,91	2,96	3,01	3,17
60	2,00	2,29	2,46	2,56	2,65	2,72	2,77	2,82	2,87	2,91	3,06

* Для проміжних значень кількості ступенів свободи застосовують інтерполяцію, тобто знаходять середньозбалансований показник. Наприклад, якщо $r = 4$, а число ступенів свободи – 9, то $q = (4{,}53+4{,}33)/2 = 4{,}43$; $D = (3{,}02+ + 2{,}89)/2 = 2{,}96$; $m = (3{,}62+4 \cdot 2{,}98)/5 = 3{,}11$.

** Для сусідніх середніх $r = 1$ (критерії *D*, *m*) і $r = 2$ (критерій *q*).

Додаток И

Значення корегуючого чинника *C* (критерію Дункана) для порівняння будь-яких середніх у вирівняному ряді: $HIP_\alpha = C(t_\alpha s_d)$

Ступені свободи для *z*	Ранг середньої, *r**										
	2	3	4	5	6	7	8	9	10	20	50
Рівень значущості α = 0,05											
4	1,00	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02
5	1,00	1,03	1,04	1,05	1,05	1,05	1,05	1,05	1,05	1,05	1,05
6	1,00	1,03	1,05	1,06	1,06	1,06	1,06	1,06	1,06	1,06	1,06
7	1,00	1,04	1,06	1,07	1,07	1,08	1,08	1,08	1,08	1,08	1,08
8	1,00	1,04	1,06	1,08	1,09	1,09	1,09	1,09	1,09	1,09	1,09
9	1,00	1,04	1,07	1,08	1,09	1,10	1,10	1,10	1,10	1,10	1,10
10	1,00	1,05	1,07	1,09	1,10	1,10	1,10	1,10	1,10	1,10	1,10
15	1,00	1,05	1,08	1,10	1,12	1,12	1,13	1,14	1,14	1,15	1,15
20	1,00	1,05	1,08	1,10	1,12	1,13	1,14	1,15	1,15	1,17	1,18
30	1,00	1,05	1,08	1,11	1,12	1,14	1,15	1,16	1,16	1,19	1,20
60	1,00	1,05	1,09	1,11	1,13	1,14	1,16	1,17	1,18	1,20	1,23
Рівень значущості α = 0,01											
3	1,00	1,03	1,04	1,05	1,07	1,08	1,08	1,09	1,09	1,13	1,13
4	1,00	1,04	1,06	1,08	1,09	1,09	1,11	1,11	1,12	1,15	1,15
5	1,00	1,05	1,07	1,08	1,10	1,11	1,12	1,13	1,14	1,19	1,19
6	1,00	1,05	1,08	1,09	1,11	1,12	1,14	1,15	1,15	1,20	1,20
7	1,00	1,05	1,08	1,10	1,12	1,13	1,15	1,16	1,17	1,21	1,21
8	1,00	1,05	1,08	1,10	1,12	1,14	1,15	1,16	1,17	1,22	1,22
9	1,00	1,06	1,08	1,10	1,12	1,14	1,16	1,17	1,17	1,24	1,24
10	1,00	1,06	1,09	1,11	1,13	1,15	1,16	1,17	1,18	1,24	1,24

* Для сусідніх середніх вирівняного ряду $r = 2$.

Додаток К

Кути, які відповідають відсоткам: кут арксинус $\sqrt{\text{відсоток}}$

%	Десяті частки відсотка									
	0	1	2	3	4	5	6	7	8	9
0	0,0	1,8	2,6	3,1	3,6	4,0	4,4	4,8	5,1	5,4
1	5,7	6,0	6,3	6,6	6,8	7,0	7,3	7,5	7,7	7,9
2	8,1	8,3	8,5	8,7	8,9	9,1	9,3	9,5	9,6	9,8
3	10,0	10,1	10,3	10,5	10,6	10,8	10,9	11,1	11,2	11,4
4	11,5	11,7	11,8	12,0	12,1	12,2	12,4	12,5	12,7	12,8
5	12,9	13,0	13,2	13,3	13,4	13,6	13,7	13,8	13,9	14,1
6	14,2	14,3	14,4	14,5	14,6	14,8	14,9	15,0	15,1	15,2
7	15,3	15,4	15,6	15,7	15,8	15,9	16,0	16,1	16,2	16,3
8	16,4	16,5	16,6	16,7	16,8	17,0	17,1	17,2	17,3	17,4
9	17,5	17,6	17,7	17,8	17,8	18,0	18,0	18,2	18,2	18,3
10	18,4	18,5	18,6	18,7	18,8	18,9	19,0	19,1	19,2	19,3
11	19,4	19,5	19,6	19,6	19,7	19,8	19,9	20,0	20,1	20,2
12	20,3	20,4	20,4	20,5	20,6	20,7	20,8	20,9	21,0	21,0
13	21,1	21,2	21,3	21,4	21,5	21,6	21,6	21,7	21,8	22,0
14	22,1	22,1	22,1	22,2	22,3	22,4	22,5	22,6	22,6	22,7
15	22,8	22,9	23,0	23,0	23,1	23,2	23,3	23,3	23,4	23,5
16	23,6	23,7	23,7	23,8	29,3	24,0	24,0	24,1	24,2	24,3
17	24,4	24,4	24,5	24,6	24,6	24,7	24,8	24,9	25,0	25,0
18	25,1	25,2	25,2	25,3	25,4	25,5	25,6	25,6	25,7	25,8
19	25,8	25,9	26,0	26,1	26,1	26,2	26,3	26,4	26,4	26,5
20	26,6	26,6	26,7	26,8	26,9	26,9	27,0	27,1	27,1	27,2
21	27,3	27,4	27,4	27,5	27,6	27,6	27,7	27,8	27,8	27,9
22	28,0	28,0	28,1	28,2	28,2	28,3	28,4	28,4	28,5	28,6
23	28,7	28,7	28,8	28,9	28,9	29,0	29,1	29,1	29,2	29,3
24	29,3	29,4	29,5	29,5	29,6	29,7	29,7	29,8	29,9	29,9
25	30,0	30,1	30,1	30,2	30,3	30,3	30,4	30,5	30,5	30,6
26	30,7	30,7	30,8	30,9	30,9	31,0	31,0	31,0	31,2	31,2
27	31,2	31,3	31,4	31,5	31,6	31,6	31,7	31,8	31,8	31,9
28	32,0	32,0	32,1	32,1	32,2	32,3	32,3	32,4	32,5	32,5
29	32,6	32,6	32,7	32,8	32,8	32,9	33,0	33,0	33,1	33,2
30	33,2	33,3	33,3	33,4	33,5	33,5	33,6	33,6	33,7	33,8
31	33,8	33,9	34,0	34,0	34,1	34,1	34,2	34,3	34,3	34,3
32	34,4	34,5	34,6	34,6	34,7	34,8	34,8	34,9	35,0	35,0
33	35,1	35,1	35,2	35,2	35,3	35,4	35,4	35,5	35,6	35,6

Продовження дод. К

%	Десяті частки відсотка									
	0	1	2	3	4	5	6	7	8	9
34	35,7	35,7	35,8	35,9	35,9	36,0	36,0	36,1	36,2	36,2
35	36,1	36,3	36,4	36,5	36,5	36,6	36,6	36,7	36,8	36,8
36	36,9	36,9	37,0	37,0	37,1	37,2	37,2	37,3	37,4	37,4
37	37,5	37,5	37,6	37,6	37,7	37,8	37,8	37,9	37,9	38,0
38	38,1	38,1	38,2	38,2	38,3	38,4	38,4	38,5	38,5	38,6
39	38,6	38,7	38,8	38,8	38,9	38,9	39,40	39,1	39,1	39,2
40	39,2	39,3	39,4	39,4	39,5	39,5	39,6	39,6	39,7	39,8
41	39,8	39,9	39,9	40,0	40,0	40,1	40,2	40,2	40,3	40,3
42	40,4	40,5	40,5	40,6	40,6	40,7	40,7	40,8	40,9	40,9
43	41,0	41,0	41,1	41,2	41,2	41,3	41,3	41,4	41,4	41,5
44	41,6	41,6	41,7	41,7	41,8	41,8	41,9	42,0	42,0	42,1
45	42,1	42,2	42,2	42,3	42,4	42,4	42,5	42,5	42,6	42,6
46	42,7	42,8	42,8	42,9	42,9	43,0	43,1	43,1	43,2	43,2
47	43,3	43,3	43,4	43,4	43,5	43,6	43,6	43,7	43,7	43,8
48	43,8	43,9	44,0	44,0	44,1	44,1	44,2	44,3	44,3	44,4
49	44,4	44,5	44,5	44,6	44,7	44,7	44,8	44,8	44,9	44,9
50	45,0	45,0	45,1	45,2	45,2	45,3	45,3	45,4	45,5	45,5
51	45,6	45,6	45,7	45,8	45,8	45,9	45,9	46,0	46,0	46,1
52	46,2	46,2	46,3	46,3	46,4	46,4	46,5	46,6	46,6	46,7
53	46,7	46,8	46,8	46,9	47,0	47,0	47,1	47,1	47,2	47,2
54	47,3	47,4	47,4	47,5	47,5	47,6	47,6	47,7	47,8	47,8
55	47,9	47,9	48,0	48,0	48,1	48,2	48,2	48,3	48,3	48,4
56	48,4	48,5	48,6	48,7	48,7	48,7	48,8	48,8	48,9	49,0
57	49,0	49,1	49,1	49,2	49,3	49,3	49,4	49,4	49,5	49,5
58	49,6	49,7	49,7	49,8	49,8	49,9	49,9	50,0	50,1	50,1
59	50,2	50,2	50,3	50,4	50,4	50,5	50,5	50,6	50,6	50,7
60	50,8	50,8	50,9	50,9	51,0	51,1	51,1	51,2	51,2	51,3
61	51,4	51,4	51,5	51,5	51,6	51,6	51,7	51,8	51,8	51,9
62	51,9	52,0	52,1	52,1	52,2	52,2	52,3	52,3	52,4	52,5
63	52,5	52,6	52,6	52,7	52,8	52,8	52,9	53,0	53,0	53,1
64	53,1	53,2	53,3	53,3	53,4	53,4	53,5	53,6	53,6	53,7
65	53,7	53,8	53,8	53,9	54,0	54,0	54,1	54,2	54,2	54,3
66	54,3	54,4	54,4	54,5	54,6	54,6	54,7	54,8	54,8	54,9
67	54,9	55,0	55,1	55,1	55,2	55,2	55,3	55,4	55,4	55,5
68	55,6	55,6	55,7	55,7	55,8	55,9	55,9	56,0	56,0	56,1

Продовження дод. К

%	Десяті частки відсотка									
	0	1	2	3	4	5	6	7	8	9
69	56,2	56,2	56,3	56,4	56,4	56,5	56,5	56,6	56,7	56,7
70	56,8	56,8	56,9	57,0	57,0	57,1	57,2	57,2	57,3	57,4
71	57,4	57,5	57,5	57,6	57,7	57,7	57,8	57,9	57,9	58,0
72	58,0	58,1	58,2	58,2	58,3	58,4	58,4	58,5	58,6	58,6
73	58,7	58,8	58,8	58,9	59,0	59,0	59,1	59,2	59,2	59,3
74	59,3	59,4	59,5	59,5	59,6	59,7	59,7	59,8	59,9	59,9
75	60,0	60,1	60,1	60,2	60,3	60,3	60,4	60,5	60,5	60,6
76	60,7	60,7	60,8	60,9	60,9	61,0	61,1	61,1	61,2	61,3
77	61,3	61,4	61,5	61,6	61,6	61,7	61,8	61,8	61,9	62,0
78	62,0	62,1	62,2	62,2	62,3	62,4	62,4	62,5	62,6	62,6
79	62,7	62,8	62,9	62,9	63,0	63,1	63,2	63,2	63,3	63,4
80	63,4	63,5	63,6	63,6	63,7	63,8	63,9	63,9	64,0	64,1
81	64,2	64,2	64,3	64,4	64,4	64,5	64,6	64,7	64,8	64,8
82	64,9	65,0	65,0	65,1	65,2	65,3	65,4	65,4	65,5	65,6
83	65,6	65,7	65,8	65,9	66,0	66,0	66,1	66,2	66,3	66,3
84	66,4	66,5	66,6	66,7	66,7	66,8	66,9	67,0	67,0	67,1
85	67,2	67,3	67,4	67,4	67,5	67,6	67,6	67,8	67,9	67,9
86	68,0	68,1	68,2	68,3	68,4	68,4	68,5	68,6	68,7	68,8
87	68,9	69,0	69,0	69,1	69,2	69,3	69,4	69,5	69,6	69,6
88	69,7	69,8	69,9	70,0	70,1	70,2	70,3	70,4	70,4	70,5
89	70,6	70,7	70,8	70,9	71,0	71,1	71,2	71,3	71,4	71,5
90	71,6	71,7	71,8	71,8	72,0	72,0	72,2	72,2	72,3	72,4
91	72,5	72,6	72,7	72,8	73,0	73,0	73,2	73,3	73,4	73,5
92	73,6	73,7	73,8	73,9	74,0	74,1	74,2	74,3	74,4	74,6
93	74,7	74,8	74,9	75,0	75,1	75,2	75,4	75,5	75,6	75,7
94	75,8	75,9	76,1	76,2	76,3	76,4	76,6	76,7	76,8	77,0
95	77,1	77,2	77,3	77,5	77,6	77,8	77,9	78,0	78,2	78,3
96	78,5	78,6	78,8	78,9	79,1	79,2	79,4	79,5	79,7	79,9
97	80,0	80,2	80,4	80,5	80,7	80,9	81,1	81,3	81,5	81,7
98	81,9	82,1	82,3	82,5	82,7	83,0	83,2	83,4	83,7	84,0
99	84,3	84,6	84,9	85,2	85,6	86,0	86,4	86,9	87,4	88,2
100	90,0	–	–		–	–	–	–	–	–

Додаток Л

Співвідношення між показниками *r* і *z*

Десяті частки, *r*	Соті частки, *r*									
	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
	значення *z*									
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
0,2	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
0,3	0,309	0,321	0,332	0,343	0,354	0,365	0,377	0,388	0,400	0,412
0,4	0,424	0,436	0,448	0,460	0,472	0,485	0,498	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,663	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,776	0,793	0,811	0,829	0,848
0,7	0,867	0,887	0,908	0,929	0,951	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
0,9	1,472	1,527	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647

Додаток М

Щільність імовірності нормального розподілу $\Phi(t)$

(імовірність повторюваності показника X за різних значень t)

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}$$

t	Соті частки									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3949	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3122	3101	3078	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2505	2541	2516	2492	2468	2444
1,0	0,2420	2395	2371	2347	2323	2298	2275	2265	2251	2202
1,1	2178	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1692	1669	1647	1626	1604	1582	1561	1540	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0658	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0296	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0162	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0094	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0054	0053	0051	0050	0049	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0003	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Додаток Н

Таблиця квадратних коренів і квадратів

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
1	1,0000	1	26	5,0990	676	51	7,1414	2601
2	1,4142	4	27	5,1962	729	52	7,2111	2704
3	1,7321	9	28	5,2915	784	53	7,2801	2809
4	2,0000	16	29	5,3852	841	54	7,3485	2916
5	2,2361	25	30	5,4772	900	55	7,4162	3025
6	2,4495	36	31	5,5678	961	56	7,4833	3136
7	2,6458	49	32	5,6569	1024	57	7,5498	3249
8	2,8284	64	33	5,7446	1089	58	7,6158	3364
9	3,0000	81	34	5,8310	1156	59	7,6811	3481
10	3,1623	100	35	5,9161	1225	60	7,7460	3600
11	3,3166	121	36	6,0000	1296	61	7,8102	3721
12	3,4641	144	37	6,0828	1369	62	7,8740	3844
13	3,6056	169	38	6,1644	1444	63	7,9373	3969
14	3,7417	196	39	6,2450	1521	64	8,0000	4096
15	3,8730	225	40	6,3246	1600	65	8,0623	4225
16	4,0000	256	41	6,4031	1681	66	8,1240	4356
17	4,1231	289	42	6,4807	1764	67	8,1854	4489
18	4,2426	324	43	6,5574	1849	68	8,2462	4624
19	4,3589	361	44	6,6332	1936	69	8,3066	4761
20	4,4721	400	45	6,7082	2025	70	8,3666	4900
21	4,5826	441	46	6,7823	2116	71	8,4261	5041
22	4,6904	484	47	6,8557	2209	72	8,4853	5184
23	4,7958	529	48	6,9282	2304	73	8,5440	5329
24	4,8990	576	49	7,0000	2401	74	8,6023	5476
25	5,0000	625	50	7,0711	2500	75	8,6603	5625

Продовження дод. Н

N	$\sqrt{N}$	N^2	*N*	$\sqrt{N}$	N^2	*N*	$\sqrt{N}$	N^2
76	8,7178	5776	101	10,0499	10201	126	11,2250	15876
77	8,7750	5929	102	10,0995	10404	127	11,2694	16129
78	8,8318	6084	103	10,1489	10609	128	11,3137	16384
79	8,8882	6241	104	10,1980	10816	129	11,3578	16641
80	8,9443	6400	105	10,2470	11025	130	11,4018	16900
81	9,0000	6561	106	10,2956	11236	131	11,4455	17161
82	9,0554	6724	107	10,3441	11449	132	11,4891	17424
83	9,1104	6889	108	10,3923	11664	133	11,5326	17689
84	9,1652	7056	109	10,4403	11881	134	11,5758	17956
85	9,2195	7225	110	10,4881	12100	135	11,6190	18225
86	9,2736	7396	111	10,5357	12321	136	11,6619	184496
87	9,3274	7569	112	10,5830	12544	137	11,7047	18769
88	9,3808	7744	113	10,6301	12769	138	11,7473	19044
89	9,4340	7921	114	10,6771	12996	139	11,7898	19321
90	9,4868	8100	115	10,7238	13225	140	11,8322	19600
91	9,5394	8281	116	10,7703	13456	141	11,8743	19881
92	9,5917	8464	117	10,8167	13689	142	11,9164	20164
93	9,6487	8649	118	10,8628	13924	143	11,9583	20449
94	9,6954	8836	119	10,9087	14161	144	12,0000	20736
95	9,7468	9025	120	10,9545	14400	145	12,0416	21025
96	9,7980	9216	121	11,0000	14641	146	12,0830	21316
97	9,8489	9406	122	11,0454	14884	147	12,1244	21609
98	9,8995	9604	123	11,0905	15129	148	12,1655	21904
99	9,9499	9801	124	11,1355	15376	149	12,2066	22201
100	10,0000	10000	125	11,1803	15625	150	12,2474	22500

Продовження дод. Н

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
151	12,2882	22801	176	13,2665	30976	201	14,1774	40401
152	12,3288	23104	177	13,3041	31329	202	14,2127	40804
153	12,3693	23409	178	13,3417	31684	203	14,2478	41209
154	12,4097	23716	179	13,3791	32041	204	14,2829	41616
155	12,4499	24025	180	13,4164	32400	205	14,3178	42025
156	12,4900	24336	181	13,4536	32761	206	14,3527	42436
157	12,5300	24649	182	13,4907	33124	207	14,3875	42849
158	12,5698	24964	183	13,5277	33489	208	14,4222	43264
159	12,6095	25281	184	13,5647	33856	209	14,4568	43681
160	12,6491	25600	185	13,6015	34225	210	14,4914	44100
161	12,6886	25921	186	13,6382	34596	211	14,5258	44521
162	12,7279	26244	187	13,6748	34969	212	14,5602	44944
163	12,7671	26569	188	13,7113	35344	213	14,5945	45369
164	12,8062	26896	189	13,7477	35721	214	14,6287	45796
165	12,8452	27225	190	13,7840	36100	215	14,6629	46225
166	12,8841	27556	191	13,8203	36481	216	14,6969	46656
167	12,9228	27889	192	13,8564	36864	217	14,7309	47089
168	12,9615	28224	193	13,8924	37249	218	14,7648	47524
169	13,0000	28561	194	13,9284	37636	219	14,7986	47961
170	13,0384	28900	195	13,9642	38025	220	14,8324	48400
171	13,0767	29241	196	14,0000	38416	221	14,8661	48841
172	13,1149	29584	197	14,0357	38809	222	14,8997	49284
173	13,1529	29929	198	14,0712	39204	223	14,9382	49729
174	13,1909	30276	199	14,1067	39601	224	14,9666	50176
175	13,2288	30625	200	14,1421	40000	225	15,0000	50625

Продовження дод. Н

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
226	15,0333	51067	251	15,8430	63001	276	16,6132	76176
227	15,0665	51529	252	15,8745	63504	277	16,6433	76729
228	15,0997	51984	253	15,9060	64009	278	16,6733	77284
229	15,1327	52441	254	15,9374	64516	279	16,7033	77841
230	15,1658	52900	255	15,9687	65025	280	16,7332	78400
231	15,1987	53361	256	16,0000	65536	281	16,7631	78961
232	15,2315	53824	257	16,0312	66049	282	16,7929	79524
233	15,2643	54289	258	16,0624	66564	283	16,8226	80089
234	15,2971	54756	259	16,0935	67081	284	16,8523	80656
235	15,3297	55225	260	16,1245	67600	285	16,8819	81225
236	15,3623	55696	261	16,1555	68121	286	16,9115	81796
237	15,3948	56169	262	16,1864	68644	287	16,9411	82369
238	15,4272	56644	263	16,2173	69169	288	16,9706	82944
239	15,4596	57121	264	16,2481	69696	289	17,0000	83521
240	15,4919	57600	265	16,2788	70225	290	17,0294	84100
241	15,5242	58081	266	16,3095	70756	291	17,0587	84681
242	15,5563	58564	267	16,3401	71289	292	17,0880	85264
243	15,5885	59049	268	16,3707	71824	293	17,1172	85849
244	15,6205	59536	269	16,4012	72361	294	17,1464	86436
245	15,6525	60025	270	16,4317	72900	295	17,1756	87025
246	15,6844	60516	271	16,4621	73441	296	17,2047	87616
247	15,7162	61009	272	16,4924	73984	297	17,2337	88209
248	15,7480	61504	273	16,5227	74529	298	17,2627	88804
249	15,7797	62001	274	16,5529	75076	299	17,2916	89401
250	15,8114	62500	275	16,5831	75625	300	17,3205	90000

Продовження дод. Н

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
301	17,3494	90601	326	18,0555	106276	351	18,7350	123201
302	17,3781	91204	327	18,0831	106929	352	18,7617	123904
303	17,4069	91809	328	18,1108	107584	353	18,7883	124609
304	17,4356	92416	329	18,1384	108241	354	18,8149	125316
305	17,4642	93025	330	18,1659	108900	355	18,8414	126025
306	17,4929	93636	331	18,1934	109561	356	18,8680	126736
307	17,5214	94249	332	18,2209	110224	357	18,8944	127449
308	17,5499	94864	333	18,2483	110889	358	18,9209	128164
309	17,5784	95481	334	18,2757	111556	359	18,9473	128881
310	17,6068	96100	335	18,3030	112225	360	18,9737	129600
311	17,6352	96721	336	18,3303	112896	361	19,0000	130321
312	17,6635	97344	337	18,3576	113569	362	19,0263	131044
313	17,6918	97969	338	18,3848	114244	363	19,0526	131769
314	17,7200	98596	339	18,4120	114921	364	19,0788	132496
315	17,7482	99225	340	18,4391	115600	365	19,1050	133225
316	17,7764	99856	341	18,4662	116281	366	19,1311	133956
317	17,8045	100489	342	18,4932	116964	367	19,1572	134689
318	17,8326	101124	343	18,5203	117649	368	19,1833	135424
319	17,8606	101761	344	18,5472	118336	369	19,2094	136161
320	17,8885	102400	345	18,5742	119025	370	19,2354	136900
321	17,9165	103041	346	18,6011	119716	371	19,2614	137641
322	17,9444	103684	347	18,6279	120409	372	19,2873	138384
323	17,9722	104329	348	18,6548	121104	373	19,3132	139129
324	18,0000	104976	349	18,6815	121801	374	19,3391	139876
325	18,0278	105625	350	18,7083	122500	375	19,3649	140625

Продовження дод. Н

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
376	19,3907	141376	401	20,0250	160801	426	20,6398	181476
377	19,4165	142129	402	20,0499	161604	427	20,6640	182329
378	19,4422	142884	403	20,0749	162409	428	20,6882	183184
379	19,4679	143641	404	20,0998	163216	429	20,7123	184041
380	19,4936	144400	405	20,1246	164025	430	20,7364	184900
381	19,5192	145161	406	20,1494	164836	431	20,7605	185761
382	19,5448	145924	407	20,1742	165649	432	20,7846	186624
383	19,5704	146689	408	20,1990	166464	433	20,8087	187489
384	19,5959	147456	409	20,2237	167281	434	20,8327	188356
385	19,6214	148225	410	20,2485	168100	435	20,8567	189225
386	19,6469	148996	411	20,2731	168921	436	20,8806	190096
387	19,6723	149769	412	20,2978	169744	437	20,9045	190969
388	19,6977	150544	413	20,3224	170569	438	20,9284	191844
389	19,7231	151321	414	20,3470	171396	439	20,9523	192721
390	19,7484	152100	415	20,3715	172225	440	20,9762	193600
391	19,7737	152881	416	20,3961	173056	441	21,0000	194481
392	19,7990	153664	417	20,4206	174889	442	21,0238	195364
393	19,8242	154449	418	20,4450	174724	443	21,0476	196249
394	19,8494	155236	419	20,4695	175561	444	21,0713	197136
395	19,8746	156025	420	20,4939	176400	445	21,0950	198025
396	19,8997	156816	421	20,5183	177241	446	21,1187	198916
397	19,9249	157609	422	20,5426	178084	447	21,1424	199809
398	19,9499	158404	423	20,5670	178929	448	21,1660	200704
399	19,9750	159201	424	20,5913	179776	449	21,1896	201601
400	20,0000	160000	425	20,6155	180625	450	21,2132	202500

Продовження дод. Н

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
451	21,2368	203401	476	21,8174	226576	501	22,3830	251001
452	21,2603	204304	477	21,8403	227529	502	22,4054	252004
453	21,2838	205209	478	21,8632	228484	503	22,4277	253009
454	21,3073	206116	479	21,8861	229441	504	22,4499	254016
455	21,3307	207025	480	21,9089	230400	505	22,4722	255025
456	21,3542	207936	481	21,9317	231361	506	22,4944	256036
457	21,3776	208849	482	21,9545	232324	507	22,5167	257049
458	21,4009	209764	483	21,9773	233289	508	22,5389	258064
459	21,4243	210681	484	22,0000	234256	509	22,5610	259081
460	21,4476	211600	485	22,0227	235225	510	22,5832	260100
461	21,4709	212521	486	22,0454	236196	511	22,6053	261121
462	21,4942	213444	487	22,0681	237169	512	22,6274	262144
463	21,5174	214369	488	22,0907	238144	513	22,6495	263169
464	21,5407	215296	489	22,1133	239121	514	22,6716	264196
465	21,5639	216225	490	22,1359	240100	515	22,6936	265225
466	21,5870	217156	491	22,1585	241081	516	22,7156	266256
467	21,6102	218089	492	22,1811	242064	517	22,7376	267289
468	21,6333	219024	493	22,2036	243049	518	22,7596	268324
469	21,6565	219961	494	22,2261	244036	519	22,7816	269361
470	21,6795	220900	495	22,2486	245025	520	22,8035	270400
471	21,7025	221841	496	22,2711	246016	521	22,8254	271441
472	21,7256	222784	497	22,2935	247009	522	22,8473	272484
473	21,7486	223729	498	22,3159	248004	523	22,8692	273529
474	21,7715	224676	499	22,3383	249001	524	22,8910	274576
475	21,7945	225625	500	22,3607	250000	525	22,9129	275625

Продовження дод. Н

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
526	22,9347	276676	551	23,4734	303601	576	24,0000	331776
527	22,9565	277729	552	23,4947	304704	577	24,0208	332929
528	22,9783	278784	553	23,5160	305809	578	24,0416	334084
529	23,0000	279841	554	23,5372	306916	579	24,0624	335241
530	23,0217	280900	555	23,5584	308025	580	24,0832	336400
531	23,0434	281961	556	23,5797	309136	581	24,1039	337556
532	23,0651	283024	557	23,6008	310249	582	24,1247	338724
533	23,0868	284089	558	23,6220	311364	583	24,1454	339889
534	23,1084	285156	559	23,6432	312481	584	24,1661	341056
535	23,1301	286225	560	23,6643	313600	585	24,1868	342225
536	23,1517	287296	561	23,6854	314721	586	24,2074	343396
537	23,1733	288369	562	23,7065	315844	587	24,2281	344569
538	23,1948	289444	563	23,7276	316969	588	24,2487	345744
539	23,2164	290521	564	23,7487	318096	589	24,2693	346921
540	22,2379	291600	565	23,7697	319225	590	24,2899	348100
541	23,2594	292681	566	23,7908	320356	591	24,3105	349281
542	23,2809	293764	567	23,8118	321489	592	24,3311	350464
543	23,3024	294849	568	23,8328	322624	593	24,3516	351649
544	23,3238	295936	569	23,8537	323761	594	24,3721	352836
545	23,3452	297025	570	23,8747	324900	595	24,3926	354025
546	23,3666	298116	571	23,8956	326041	596	24,4131	355216
547	23,3880	299209	572	23,9165	327184	597	24,4336	356409
548	23,4094	300304	573	23,9374	328329	598	24,4540	357604
549	23,4307	301401	574	23,9583	329476	599	24,4745	358801
550	23,4521	302500	575	23,9792	330625	600	24,4949	360000

Продовження дод. Н

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
601	24,5153	361201	626	25,0200	391876	651	25,5147	423801
602	24,5357	362404	627	25,0400	393129	652	25,5343	425104
603	24,5561	363609	628	25,0599	394384	653	25,5539	426409
604	24,5764	364816	629	25,0799	395641	654	25,5734	427716
605	24,5967	366025	630	25,0998	396900	655	25,5930	429025
606	24,6171	367236	631	25,1197	398161	656	25,6125	430336
607	24,6374	368449	632	25,1396	399424	657	25,6320	431649
608	24,6571	369664	633	25,1595	400689	658	25,6515	432964
609	24,6779	370881	634	25,1794	401956	659	25,6710	434281
610	24,6982	372100	635	25,1992	403225	660	25,6905	435600
611	24,7184	373321	636	25,2190	404496	661	25,7099	436921
612	24,7386	374544	637	25,2389	405769	662	25,7294	438244
613	24,7588	375769	638	25,2587	407044	663	25,7488	439569
614	24,7790	376996	639	25,2784	408321	664	25,7682	440896
615	24,7992	378225	640	25,2982	409600	665	25,7876	442225
616	24,8193	379456	641	25,3180	410881	666	25,8070	443556
617	24,8395	380689	642	25,3377	412164	667	25,8263	444889
618	24,8596	381924	643	25,3574	413449	668	25,8457	446224
619	24,8797	383161	644	25,7772	414736	669	25,8650	447561
620	24,8998	384400	645	25,3969	416025	670	25,8844	448900
621	24,9199	385641	646	25,4165	417316	671	25,9037	450241
622	24,9399	386884	647	25,4362	418600	672	25,9230	451584
623	24,9600	388129	648	25,4558	419904	673	25,9422	452929
624	24,9800	389376	649	25,4755	421201	674	25,9615	454276
625	25,0000	390625	650	25,4951	422500	675	25,9808	456625

Продовження дод. Н

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
676	26,0000	456976	701	26,4764	491401	726	26,9444	527076
677	26,0192	458329	702	26,4953	492804	727	26,9629	528529
678	26,0384	459684	703	26,5141	494209	728	26,9815	529984
679	26,0576	461041	704	26,5330	495616	729	27,0000	531441
680	26,0768	462400	705	26,5518	497025	730	27,0185	532900
681	26,0960	463761	706	26,5707	498436	731	27,0370	534361
682	26,1151	465124	707	26,5895	499849	732	27,0555	535824
683	26,1343	466489	708	26,6083	501264	733	27,0740	537289
684	26,1534	497856	709	26,6271	502681	734	27,0924	538756
685	26,1725	469225	710	26,6458	504100	735	27,1109	540225
686	26,1916	470596	711	26,6646	505521	736	27,1293	541696
687	26,2107	471969	712	26,6833	506944	737	27,1477	543169
688	26,2298	473344	713	26,7021	508369	738	27,1662	544644
689	26,2488	474721	714	26,7208	509796	739	27,1846	546121
690	26,2679	476100	715	26,7395	511225	740	27,2029	547600
691	26,2869	477481	716	26,7582	512656	741	27,2213	549081
692	26,3059	478864	717	26,7769	514089	724	27,2397	550564
693	26,3249	480249	718	26,7955	515524	743	27,2580	552049
694	26,3439	481636	719	26,8142	516961	744	27,2764	553536
695	26,3629	483025	720	26,8328	518400	745	27,2947	555025
696	26,3818	484416	721	26,8514	519841	746	27,3130	555516
697	26,4008	485809	722	26,8701	521284	747	27,3313	558009
698	26,4197	487204	723	26,8887	522729	748	27,3496	559504
699	26,4386	488601	724	26,9072	524176	749	27,3679	561001
700	26,4575	490000	725	26,6258	525625	750	27,3861	562500

Продовження дод. Н

N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2	N	$\sqrt{N}$	N^2
751	27,4044	564001	776	27,8568	602176	801	28,3019	641601
752	27,4226	565504	777	27,8747	603729	802	28,3196	643204
753	27,4408	567009	778	27,8927	605284	803	28,3373	644809
754	27,4591	568516	779	27,9106	606841	804	28,3549	646416
755	27,4773	570025	780	27,9285	608400	805	28,3725	648025
756	27,4955	571536	781	27,9464	609961	806	28,3901	649636
757	27,5136	573049	782	27,9643	611524	807	28,4077	651249
758	27,5318	574564	783	27,9821	613089	808	28,4253	652864
759	27,5500	576081	784	28,0000	614656	809	28,4429	654481
760	27,5681	577600	785	28,0179	616225	810	28,4605	656100
761	27,5862	579121	786	28,0357	617796	811	28,4781	657721
762	27,6043	580644	787	28,0535	619369	812	28,4956	659344
763	27,6225	582169	788	28,0713	620944	813	28,5132	660969
764	27,6405	583696	789	28,0891	622521	814	28,5307	662596
765	27,6586	585225	790	28,1069	624100	815	28,5482	664225
766	27,6767	586756	791	28,1247	625681	816	28,5657	665856
767	27,6948	588289	792	28,1425	627264	817	28,5832	667489
768	27,7128	589824	793	28,1603	628849	818	28,6007	669124
769	27,7308	591361	794	28,1780	630436	819	28,6182	670761
770	27,7489	592900	795	28,1957	632025	820	28,6356	672400
771	27,7669	594441	796	28,2135	633616	821	28,6531	674041
772	27,7849	595984	797	28,2312	635209	822	28,6705	675684
773	27,8029	597529	798	28,2489	636804	823	28,6880	677329
774	27,8209	599076	799	28,2666	638401	824	28,7054	678976
775	27,8388	600625	800	28,2843	640000	825	28,7228	680625

Продовження дод. Н

N	$\sqrt{N}$	*N²*	*N*	$\sqrt{N}$	*N²*	*N*	$\sqrt{N}$	*N²*
826	28,7402	682276	856	29,2575	732736	886	29,7658	784996
827	28,7576	683929	857	29,2746	734449	887	29,7825	786769
828	28,7750	685584	858	29,2916	736164	888	29,7993	788544
829	28,7924	687241	859	29,3087	737881	889	29,8161	790321
830	28,8097	688900	860	29,3258	739600	890	29,8329	792100
831	28,8271	690561	861	29,3428	741321	891	29,8496	793881
832	28,8444	692224	862	29,3598	743044	892	29,8664	795664
833	28,8617	693889	863	29,3769	744769	893	29,8831	797449
834	28,8791	695556	864	29,3939	746496	894	29,8998	799236
835	28,8964	697225	865	29,4109	748225	895	29,9166	801025
836	28,9137	698896	866	29,4279	749956	896	29,9333	802816
837	28,9310	700569	867	29,4449	751689	897	29,9500	804609
838	28,9482	702244	868	29,4618	753424	898	29,9666	806404
839	28,9655	703921	869	29,4788	755161	899	29,9833	808201
840	28,9828	705600	870	29,4958	756900	900	30,0000	810000
841	29,0000	707281	871	29,5127	758641	901	30,0167	811801
842	29,0172	708964	872	29,5296	760384	902	30,0333	813604
843	29,0345	710649	873	29,5466	762129	903	30,0500	815409
844	29,0517	712336	874	29,5635	763876	904	30,0666	817216
845	29,0639	714025	875	29,5804	765625	905	30,0832	819025
846	29,0861	715716	876	29,5973	767376	906	30,0998	820836
847	29,1033	717409	877	29,6142	769129	907	30,1164	822649
848	29,1204	719104	878	29,6311	770884	908	30,1330	824464
849	29,1376	720801	879	29,6479	772641	909	30,1496	826281
850	29,1548	722500	880	29,66,48	774400	910	30,1662	828100
851	29,1719	724201	881	29,6816	776161	911	30,1828	829921
852	29,1890	725904	882	29,6985	777924	912	30,1993	831744
853	29,2062	727609	883	29,7153	779689	913	30,2159	833569
854	29,2233	729316	884	29,7321	781456	914	30,2324	835396
855	29,2404	731025	885	29,7489	783225	915	30,2490	837225

Закінчення дод. Н

N	$\sqrt{N}$	N²	N	$\sqrt{N}$	N²	N	$\sqrt{N}$	N²
916	30,2655	839056	945	30,7409	893025	974	31,2090	948676
917	30,2820	840889	946	30,7571	894916	975	31,2250	950665
918	30,2985	842724	947	30,7734	896809	976	31,2410	952576
919	30,3150	844561	948	30,7896	898704	977	31,2570	954529
920	30,3315	846400	949	30,8058	900601	978	31,2730	956484
921	30,3480	848241	950	30,8221	602500	979	31,2890	958441
922	30,3645	850084	951	30,8383	904401	980	31,3050	960400
923	30,3809	851929	952	30,8545	906304	981	31,3209	962361
924	30,3974	853776	953	30,8707	908209	982	31,3369	964324
925	30,4138	855625	954	30,8869	910116	983	31,3528	966289
926	30,4302	857476	955	30,9031	612025	984	31,3688	968256
927	30,4467	859329	956	30,9192	913936	985	31,3847	970225
928	30,4631	861184	957	30,9354	915849	986	31,4006	972196
929	30,4795	863041	958	30,9516	917764	987	31,4166	974169
930	30,4959	864900	959	30,9677	919681	988	31,4325	976144
931	30,5123	866761	960	30,9839	921600	989	31,4484	978121
932	30,5287	868624	961	31,0000	923521	990	31,4643	980100
933	30,5450	870489	962	31,0161	925444	991	31,4802	982081
934	30,5614	872356	963	31,0322	927369	992	31,4960	984064
935	30,5778	874225	964	31,0483	929296	993	31,5119	986049
936	30,5941	876096	965	31,0644	931225	994	31,5278	988036
937	30,6105	877969	966	31,0805	933156	995	31,5436	990025
938	30,6268	879844	967	31,0966	935089	996	31,5595	992016
939	30,6434	881721	968	31,1127	937024	997	31,5753	994009
940	30,6594	883600	969	31,1288	938961	998	31,5911	996004
941	30,6757	885481	970	31,1448	940900	999	31,6070	998001
942	30,6920	887364	971	31,1609	942841	1000	31,6228	1000000
943	30,7083	889249	972	31,1769	944784			
944	30,7246	891136	973	31,1929	946729			

Додаток П

Кут-арксинус (x_1) від вихідних дат (x, %)

x, %	Десяті частки відсотка									
	0	1	2	3	4	5	6	7	8	9
0	0	5,7	8,1	10,0	11,5	12,9	14,2	15,3	16,4	17,5
10	18,4	19,4	20,3	21,1	22,0	22,8	23,6	24,4	25,1	25,8
20	26,6	27,3	28,0	28,7	29,3	30,0	30,7	31,3	31,9	32,6
30	33,2	33,8	34,4	35,1	35,7	36,3	36,9	37,5	38,1	38,6
40	39,2	39,8	40,4	41,0	41,6	42,1	42,7	43,3	43,9	44,4
50	45,0	45,6	46,1	46,7	47,3	47,9	48,4	49,0	49,6	50,2
60	50,8	51,4	51,9	52,5	53,1	53,7	54,3	54,9	55,6	56,2
70	56,8	57,4	58,1	58,7	59,3	60,0	60,7	61,3	62,0	62,7
80	63,4	64,2	64,9	65,6	66,4	67,2	68,0	68,9	69,7	70,6
90	71,6	72,5	73,6	74,7	75,8	77,1	78,5	80,0	81,9	84,3
100	90,0	–	–	–	–	–	–	–	–	–

Додаток Р

Значення функції ψ [r / (n + 1)]

r / (n + 1)	ψ	r / (n + 1)	ψ	r / (n + 1)	ψ
0,000	∞	0,25	– 0,67	0,50	0,00
0,02	– 2,05	0,29	– 0,56	0,62	0,31
0,05	– 1,64	0,30	– 0,53	0,70	0,52
0,10	– 1,28	0,33	– 0,44	0,80	0,84
0,15	– 1,04	0,40	– 0,25	0,90	1,28
0,20	– 0,84	0,43	– 0,17	0,99	2,33

Додаток С

Критичні значення X-критерію Ван-дер-Вардена H_0: $X < X_{05}$ (X_{01})

$n = n_1 + n_2$	$\|n_1 - n_2\| = 0$ або 1		$\|n_1 - n_2\| = 2$ або 3	
	α = 0,05	α = 0,01	α = 0,05	α = 0,01
8	2,40	–	2,30	–
10	2,60	3,20	2,49	3,10
12	2,86	3,60	2,79	3,58
14	3,11	3,94	3,06	3,88
16	3,39	4,26	3,36	4,25
18	3,63	4,60	3,60	4,58
20	3,86	4,94	3,84	4,92
30	4,88	6,35	4,87	6,34
50	6,50	8,51	6,51	8,50

Додаток Т

Пробіт-значення від відсотку загибелі шкідників при випробуванні пестицидів

Загибель шкідників, %	0	1	2	3	4	5	6	7	8
0	–	2,67	2,95	3,12	3,25	3,36	3,45	3,52	3,39
10	3,72	3,77	3,38	3,87	3,92	3,96	4,01	4,05	4,08
20	4,16	4,19	4,23	4,26	4,29	4,33	4,36	3,39	4,42
30	4,48	4,50	4,53	4,56	4,59	4,61	4,64	4,67	4,69
40	4,75	4,77	4,80	4,82	4,85	4,87	4,90	4,92	4,95
50	5,00	5,03	5,05	5,08	5,10	5,13	5,15	5,18	5,20
60	5,25	5,28	5,31	5,33	5,36	5,39	5,41	5,44	5,47
70	5,52	5,55	5,58	5,61	5,64	5,67	5,71	5,74	5,77
80	5,84	5,88	5,92	5,95	5,99	6,04	6,08	6,13	6,18
90	6,28	6,34	6,41	6,48	6,55	6,64	6,75	6,88	7,05

Додаток У

Критичні значення коефіцієнта кореляції на 5 %-вому і 1%-вому рівні значущості

Ступені свободи (n – 2)	0,05	0,01	Ступені свободи (n – 2)	0,05	0,01	Ступені свободи (n – 2)	0,05	0,01
1	0,997	1,000	16	0,468	0,590	35	0,325	0,418
2	0,950	0,990	17	0,456	0,575	40	0,304	0,393
3	0,878	0,959	18	0,444	0,561	45	0,288	0,372
4	0,811	0,917	19	0,433	0,549	50	0,273	0,354
5	0,754	0,874	20	0,423	0,537	60	0,250	0,325
6	0,707	0,834	21	0,413	0,526	70	0,232	0,302
7	0,666	0,798	22	0,404	0,515	80	0,217	0,283
8	0,632	0,765	23	0,396	0,505	90	0,205	0,267
9	0,602	0,735	24	0,388	0,496	100	0,195	0,254
10	0,576	0,708	25	0,381	0,487	150	0,159	0,208
11	0,553	0,684	26	0,374	0,478	200	0,138	0,181
12	0,532	0,661	27	0,367	0,470	300	0,113	0,148
13	0,514	0,641	28	0,361	0,463	400	0,098	0,128
14	0,497	0,623	29	0,355	0,456	500	0,088	0,115
15	0,482	0,606	30	0,349	0,449	1000	0,062	0,081

Навчальне видання

РОЖКОВ Артур Олександрович
ПУЗІК Володимир Кузьмич
КАЛЕНСЬКА Світлана Михайлівна
ПУЗІК Людмила Михайлівна
ПОПОВ Сергій Іванович
МУЗАФАРОВ Наіль Мініярович
БУХАЛО Василь Якович
КРИШТОП Євген Анатолійович

ДОСЛІДНА СПРАВА В АГРОНОМІЇ

Книга друга
Статистична обробка результатів агрономічних досліджень

Навчальний посібник

Редактори *О. В. Васильєва*, *А. М. Чорна*
Технічний редактор *Є. В. Онишко*
Коректори *І. О. Бутильська*, *М. А. Захарченко*
Комп'ютерний набір і верстка *А. О. Рожков*

Підп. до друку 19.02.2016. Формат 60×84/16.
Папір офсетний. Гарнітура Таймс. Друк офсетний.
Ум.-друк. арк. 18,8. Обл.-вид. арк. 20,0.
Наклад 300 прим. Зам. № 15-12

Видання і друк ТОВ «Майдан»
61002, Харків, вул. Чернишевська, 59
Тел.: (0572) 700-37-30
Свідоцтво про внесення суб'єкта видавничої справи
до Державного реєстру видавців і розповсюджувачів
видавничої продукції ДК №1002 від 31.07.2002 р.