

СТАТИСТИЧНИЙ АНАЛІЗ РЕЗУЛЬТАТІВ В РОСЛИННИЦТВІ

ЗАПОРІЖЖЯ, 2026



**УДК 633:519.2 (075)
С78**

*Рекомендовано Вченою радою Таврійського державного
агротехнологічного університету імені Дмитра Моторного
(Протокол № 3 від 27.10.2026 р.)*

Рецензенти:

В. В. Дідур – доктор технічних наук, професор, завідувач кафедри агроінженерії Уманського національного університету.

А. В. Шепель – кандидат сільськогосподарських наук, доцент кафедри землеробства Херсонського державного аграрно-економічного університету.

Н. А. Дьоміна – кандидат технічних наук, доцент, завідувачка кафедри вищої математики і фізики Таврійського державного агротехнологічного університету імені Дмитра Моторного.

Аксьонов І. В. Статистичний аналіз результатів в рослинництві : навчальний посібник. / І. В. Аксьонов, М. О. Колесніков, А. М. Шкіндер-Барміна; Таврійський державний агротехнологічний університет імені Дмитра Моторного. Запоріжжя: 2026. 207 с.

Навчальний посібник присвячено теоретичним і практичним аспектам статистичного аналізу результатів досліджень у рослинництві. У виданні висвітлено основи математичної статистики, вибіркового методу, описової статистики, перевірки статистичних гіпотез, дисперсійного, кореляційного, регресійного та багатовимірного аналізу, а також сучасних методів обробки й інтерпретації експериментальних даних. Особливу увагу приділено застосуванню статистичних методів у польових дослідках, селекції та оцінюванні продуктивності сільськогосподарських культур. Матеріали посібника спрямовані на формування у здобувачів вищої освіти навичок володіння методами статистичних та біометричних аналізів, що застосовуються у рослинництві. Навчальний посібник розраховано на здобувачів ступеня вищої освіти «Доктор філософії», а також може бути рекомендовано здобувачам ступенів вищої освіти «Бакалавр», «Магістр», працівникам в галузі рослинництва та науковцям у закладах вищої освіти III-IV рівня акредитації.

Електронна версія підручника розміщена в репозитарії ТДАТУ за посиланням <http://elar.tsatu.edu.ua/handle/123456789/2791>

УДК 633:519.2 (075)

© Аксьонов І.В., Колесніков М.О.,
Шкіндер-Барміна А.М., 2026
© ТДАТУ, 2026

ЗМІСТ

	Назва розділу	
ВСТУП		5
Розділ 1	ПІДГОТОВКА ДАНИХ ДОСЛІДІВ ДО СТАТИСТИЧНОЇ ОБРОБКИ	9
1.1. Роль математичної статистики в аналізі результатів польових досліджень		9
1.2. Основні статистичні терміни та показники		11
1.3. Типи виборок		18
1.4. Округлення дослідних даних		21
1.5. Вибракування сумнівних даних		21
1.6. Відновлення даних, що випали під час проведення дослідів		22
1.7. Точність дослідів (ассурасу)		25
Контрольні питання. Розрахункові задачі. Тестові завдання до Розділу 1		26
Розділ 2	СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ МІНЛИВОСТІ	30
2.1. Статистичні характеристики кількісної мінливості		30
2.2. Статистичні характеристики якісної мінливості		36
Контрольні питання. Розрахункові задачі. Тестові завдання до Розділу 2		42
Розділ 3	СТАТИСТИЧНІ ГІПОТЕЗИ І ТЕСТИ	46
3.1. Нульова та альтернативна гіпотези		46
3.2. Алгоритм статистичного теста		47
3.3. Вибір способу статистичної обробки даних дослідів		49
3.4. Оцінка суттєвих відмінностей між значеннями вибірок дослідів		53
Контрольні питання. Розрахункові задачі. Тестові завдання до Розділу 3		55
Розділ 4	ДИСПЕРСІЙНИЙ АНАЛІЗ	60
4.1. Сутність і алгоритм дисперсійного аналізу		60
4.2. Дисперсійний аналіз даних однофакторного дослідів		66
4.3. Дисперсійний аналіз даних багатфакторного дослідів		72
Контрольні питання. Розрахункові задачі. Тестові завдання до Розділу 4		82
Розділ 5	КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ	85
5.1. Застосування та використання кореляційного аналізу		86
5.2. Аналіз парної лінійної кореляції		94
5.3. Парна лінійна кореляція між якісними (альтернативними) ознаками		100
5.4. Аналіз множинної кореляції		101
5.5. Аналіз криволінійної кореляції		106
5.6. Застосування та використання регресійного аналізу		109
5.7. Парна лінійна залежність		112
5.8. Множинна лінійна залежність		120

	5.9. Криволінійна залежність	129
	Контрольні питання. Розрахункові задачі. Тестові завдання до Розділу 5	134
Розділ 6	КОВАРІАЦІЙНИЙ АНАЛІЗ В СТАТИСТИЧНОМУ ОБРОБІТКУ РЕЗУЛЬТАТІВ ДОСЛІДІВ	139
	6.1. Загальне поняття про коваріаційний аналіз	139
	6.2. Основні формули для розрахунків при коваріаційному аналізі	142
	6.3. Алгоритм виконання коваріаційного аналізу	143
	6.4. Приклад коваріаційного аналізу	146
	Контрольні питання. Розрахункові задачі. Тестові завдання до Розділу 6	155
Розділ 7	ВИЗНАЧЕННЯ КОЕФІЦІЄНТА УСПАДКОВУВАНOSTI	159
	7.1. Поняття про успадкування кількісних ознак та коефіцієнт успадковуваності	159
	7.2. Проведення аналізу мінливості кількісних ознак	165
	7.3. Алгоритм обчислення коефіцієнта успадковуваності за допомогою коефіцієнтів кореляції та регресії	166
	7.4. Використання коефіцієнту успадковуваності для генетичного покращення селекційного матеріалу	171
	7.5. Алгоритм обчислення коефіцієнта успадковуваності у широкому сенсі (H^2) за допомогою дисперсійного аналізу	172
	7.6. Алгоритм обчислення коефіцієнта успадковуваності у широкому сенсі (H^2) у двофакторному досліді за допомогою дисперсійного аналізу	175
	Контрольні питання. Розрахункові задачі. Тестові завдання до Розділу 7	183
Розділ 8	ПЕРЕВІРКА ГІПОТЕЗ ПРО ЗАКОНИ РОЗПОДІЛУ	187
	8.1. Поняття про критерій відповідності χ^2 (Хі-квадрат)	187
	8.2. Використання χ^2 критерію в генетичному та селекційному аналізах	188
	8.3. Приклади розрахунків критерію χ^2	190
	Контрольні питання. Розрахункові задачі. Тестові завдання до Розділу 8	193
	Короткий вказівник символів	198
	Глосарій	199
	Додатки	202
	Список літератури	207

ВСТУП

Розвиток сучасної аграрної науки нерозривно пов'язаний із застосуванням кількісних методів оцінювання результатів експериментальних досліджень. Впровадження нових сортів і гібридів сільськогосподарських культур, оптимізація технологій вирощування, удосконалення селекційних програм, вивчення генетичної мінливості, адаптивності рослин до змін клімату, оцінювання ефективності агротехнологічних заходів потребують не лише накопичення експериментального матеріалу, але й його об'єктивної статистичної інтерпретації. Математична статистика забезпечує наукову достовірність висновків, дозволяє відокремити закономірності від випадкових коливань і є необхідною складовою сучасної методології наукових досліджень.

За останні десятиліття роль статистичного аналізу в аграрній науці значно зросла. Якщо раніше статистичні методи застосовувалися переважно для обробки результатів польових дослідів, то сьогодні вони є обов'язковим інструментом практично всіх напрямів біологічних і аграрних досліджень – від рослинництва, класичної селекції та генетики до молекулярної біології, біотехнології, цифрового землеробства та систем точного землеробства. Розвиток інформаційних технологій, автоматизація збору експериментальних даних, використання дистанційного моніторингу, безпілотних літальних апаратів, геоінформаційних систем і технологій штучного інтелекту призводять до постійного збільшення обсягів інформації, що потребує використання сучасних статистичних методів її аналізу.

Основи математичної статистики формувалися протягом понад століття. Вагомий внесок в розвиток математичної статистики в рослинництві та сільськогосподарській науці зробили видатні вчені Карл Пірсон, Рональд Фішер та Вільям Госсет (Стьюдент). Вони заклали підґрунтя біометрії та аналізу польових досліджень.

Так, Карл Пірсон (Karl Pearson) є засновником математичної статистики, який розробив концепції кореляційного та регресійного аналізу, що застосовуються для вивчення зв'язків між умовами вирощування (вологою, температурою) та кількісними показниками врожаю.

Рональд Фішер (Ronald Fisher), працюючи на Ротамстедській дослідній станції (найстарішій агрономічній станції у світі), розробив методологію дисперсійного аналізу, принципи рандомізації та теорію планування сільськогосподарських експериментів, що дозволило точно оцінювати врожайність різних сортів рослин за умов природної мінливості ґрунту.

Вільям Госсет (William Sealy Gosset) публікував свої роботи під псевдонімом «Стьюдент» і працював над завданнями практичної статистики для перевірки малих вибірок, створивши знаменитий t-критерій Стьюдента, яким досі користуються агрономи для оцінювання достовірності відмінностей між сортами, гібридами, агротехнічними заходами та системами удобрення.

Фундаментальну роль у розвитку математичної статистики в рослинництві відіграв Джордж Снедекор (George Snedecor). Якщо Фішер був першовідкривачем теорії, то Снедекор став головним популяризатором та

практиком, який адаптував складні математичні методи для звичайних агрономів, селекціонерів та біологів. Значною мірою завдяки його працям набув поширення F-критерій Фішера-Снедекора (буква «F» – від Fisher), який використовується для оцінювання варіації між дослідними варіантами та визначення статистичної значущості отриманих результатів. Розроблені ним статистичні таблиці суттєво спростили практичне застосування методів дисперсійного аналізу в експериментальних дослідженнях.

Визначний внесок у розвиток методики польового експерименту зробив також Борис Доспехов, фундаментальна праця якого (вперше видана у 1965 році) стала класичним підручником для багатьох поколінь агрономів, селекціонерів та аспірантів в Україні та інших країнах.

Все більше особливого значення статистичний аналіз набуває у сортознавстві та селекції рослин. Створення нових сортів і гібридів ґрунтується на оцінюванні великої кількості генотипів за комплексом кількісних і якісних ознак, які істотно змінюються під впливом генетичних особливостей та умов вирощування. Урожайність, продуктивність, зимостійкість, посухостійкість, стійкість до хвороб, біохімічний склад продукції, якість плодів чи насіння є полігенними ознаками, що характеризуються значною мінливістю. Статистичні методи дозволяють об'єктивно оцінити генотипову та фенотипову варіацію, визначити достовірність відмінностей між сортами, оцінити силу взаємозв'язків між ознаками та ефективність селекційного добору.

Не менш важливою є роль статистичного аналізу у плодівництві, овочівництві, виноградарстві та декоративному садівництві. Багатофакторний характер досліджень багаторічних насаджень, значний вплив погодних умов, особливості росту та плодоношення рослин обумовлюють необхідність використання спеціальних методів статистичної обробки експериментальних даних. Коректне застосування критеріїв достовірності дозволяє уникнути помилкових висновків щодо ефективності сортів, підщеп, способів формування крон, систем удобрення, режимів зрошення чи інших елементів технології вирощування.

Водночас сучасна аграрна освіта висуває нові вимоги до підготовки фахівців. Майбутній агроном, селекціонер чи науковець повинен не лише володіти методиками проведення експериментів, а й уміти правильно планувати дослід, здійснювати статистичну обробку результатів, інтерпретувати отримані дані та представляти їх відповідно до міжнародних вимог. За сучасних умов статистична грамотність належить до ключових професійних компетентностей фахівця аграрної галузі.

Особливої актуальності це набуває у зв'язку з інтеграцією української науки до Європейського дослідницького простору. Сучасні міжнародні журнали, що індексуються в наукометричних базах Scopus та Web of Science, приділяють значну увагу обґрунтованості статистичних методів, використаних під час аналізу експериментальних результатів. Некоректний вибір статистичних критеріїв, недостатня кількість повторень або неправильна інтерпретація результатів нерідко стають причиною відхилення наукових статей. Тому

оволодіння сучасними методами статистичного аналізу є необхідною умовою підготовки конкурентоспроможних наукових кадрів.

Паралельно зі стрімким розвитком методів статистичного аналізу відбувається активне впровадження цифрових технологій у наукові дослідження. Поряд із класичними ручними розрахунками широко використовуються спеціалізовані програмні продукти, в тому числі Microsoft Excel, Statistica, SPSS, SAS, Jamovi, JASP та інші. Використання таких програм значно скорочує час обробки великих масивів інформації, мінімізує ймовірність арифметичних помилок і відкриває можливості застосування складних багатофакторних моделей аналізу даних. Проте, ефективне використання програмного забезпечення можливе лише за умови глибокого розуміння теоретичних основ математичної статистики та біометрії.

Практика свідчить, що під час виконання курсових, кваліфікаційних та дисертаційних робіт, підготовки наукових статей, звітів за результатами науково-дослідних робіт і інших проєктів студенти, аспіранти та молоді науковці часто стикаються з труднощами вибору відповідного статистичного методу, правильного виконання розрахунків та інтерпретації отриманих результатів. Незважаючи на значну кількість наукової літератури з математичної статистики, більшість видань має універсальний характер і недостатньо враховує специфіку агробіологічних досліджень. Водночас аграрна галузь має свої особливості, пов'язані з високою мінливістю біологічних об'єктів, впливом природно-кліматичних факторів, необхідністю аналізу багаторічних спостережень, багатофакторних польових дослідів та селекційних експериментів.

Саме тому виникла потреба у створенні навчального посібника, який поєднує класичні положення математичної статистики із практичними прикладами, адаптованими до потреб аграрної науки. Посібник орієнтований на формування практичних навичок статистичної обробки експериментальних даних, що отримуються під час проведення польових, лабораторних та виробничих досліджень. Значна увага приділена алгоритмам виконання розрахунків, поясненню сутності статистичних критеріїв, аналізу типових помилок під час обробки результатів та інтерпретації статистичних показників.

Основним завданням даного навчального посібника – є формування у здобувачів освіти, викладачів та наукових працівників системних знань щодо сучасних методів статистичної обробки експериментальних даних, які застосовуються в аграрній науці. У посібнику послідовно викладено основні положення описової статистики, теорії вибіркового методу, оцінювання варіації, перевірки статистичних гіпотез, кореляційного, регресійного та дисперсійного аналізу, наведено приклади їх практичного застосування при обробці результатів досліджень у рослинництві.

Навчальний посібник призначений для здобувачів вищої освіти освітньо-наукового рівня, а також ступенів бакалавра і магістра, викладачів закладів вищої освіти аграрного профілю, працівників науково-дослідних установ. Він буде корисним під час вивчення таких дисциплін, як «Біометрія в рослинництві», «Математична статистика в агрономії», «Рослинництво», «Генетика та селекція сільськогосподарських культур», «Методика наукових досліджень з модулем по

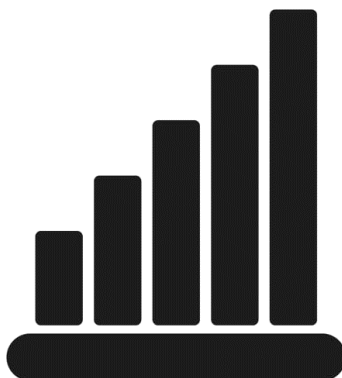
підготовці та написанню магістерської роботи (проєкту)», а також інших дисциплін, навчальними програмами яких передбачено проведення експериментальних досліджень та статистичну обробку їх результатів.

Особливістю видання є його практична спрямованість. Усі приклади максимально адаптовано до специфіки агробіологічних досліджень. Розглянуто типові задачі, що виникають під час аналізу результатів польових, лабораторних і вегетаційних дослідів, селекційних випробувань, сортовивчення, оцінювання продуктивності та якості продукції рослинництва. Значну увагу приділено покроковому виконанню розрахунків, інтерпретації отриманих статистичних показників і правильному оформленню результатів досліджень. Такий підхід дозволяє читачам не лише засвоїти теоретичні основи статистичного аналізу, але й навчитися застосовувати їх у власній науковій та професійній діяльності.

Матеріали посібника можуть бути використані під час виконання курсових і кваліфікаційних робіт, підготовки магістерських досліджень, дисертацій, наукових статей, звітів за результатами науково-дослідних робіт, а також у практичній діяльності наукових установ і підприємств агропромислового комплексу.

Автори сподіваються, що навчальний посібник сприятиме формуванню сучасної статистичної культури дослідника, підвищенню якості експериментальних досліджень та широкому впровадженню сучасних методів аналізу даних у практику аграрної науки й освіти.

РОЗДІЛ 1. ПІДГОТОВКА ДАНИХ ДОСЛІДІВ ДО СТАТИСТИЧНОЇ ОБРОБКИ



- 1.1. Роль математичної статистики в аналізі результатів польових досліджень*
- 1.2. Основні статистичні терміни та показники*
- 1.3. Типи виборок*
- 1.4. Округлення дослідних даних.*
- 1.5. Вибракування сумнівних даних*
- 1.6. Відновлення даних, що випали під час проведення дослідю*
- 1.7. Точність дослідю (accuracy)*

1.1. Роль математичної статистики в аналізі результатів польових досліджень

Результати досліджень в рослинництві аналізують методами математичної статистики, тобто систематизують, обробляють і роблять обґрунтовані висновки та пропозиції. При цьому оперують певними поняттями, термінами та символами.

Науковим співробітникам, які мають справу з вимірами та обробкою даних наукових дослідів, постійно доводиться звертатися до елементарних основ математичної статистики, щоб отримати максимально корисну інформацію з результатів вимірів.

Необхідність статистичної обробки та подання експериментальних даних у дослідників виникає відразу після закінчення проведення дослідів та початку проведення аналізу експериментальних результатів досліджень.

Для проведення аналізу отриманих та висновків за результатами проведених досліджень обов'язково виконують первинну математичну обробку отриманих даних у дослідях. Математичному обробітку даних дослідів підлягають кількісні ознаки, які вивчаються в дослідях.

З 60-х років ХХ століття математична статистика стала обов'язковою умовою аналізу експериментальних даних у галузі сільського господарства.

Одним із факторів, що вплинув на необхідність використання статистичних

розрахунків експериментальних даних є те, що всі біологічні об'єкти в рослинництві мають протилежні властивості.

Широка амплітуда мінливості ознак у різних об'єктів змушує експериментаторів вдаватися як до усереднення даних, і до оцінки меж мінливості та сили зв'язку між ознаками.

Іншою важливою обставиною, що вплинула на процес впровадження статистичних методів для аналізу ознак, факторів, біологічних явищ та властивостей, з'явилося те, що практично всі ознаки, біологічні явища та властивості підкоряються статистичним закономірностям, характерним не окремим об'єктам, а цілим сукупностям об'єктів.

Виявилося, що якщо згрупувати дані, отримані шляхом вимірювання будь-яких морфологічних ознак або біологічних ознак, в єдину сукупність, то ця сукупність матиме вигляд суто статистичної сукупності.

Істотне завдання математичної статистики – відібрати в спланованому і закладеному досліді об'єкти для досліджень, які будуть об'єктивно відображати вплив факторів, що вивчаються.

При плануванні та проведенні науково-дослідних експериментів істотну роль відіграють статистичні методи, які дають можливість:

- ✓ встановлювати ступінь достовірності подібності та відмінності досліджуваних об'єктів на підставі результатів вимірювань їх показників;
- ✓ аналізувати наявність чи відсутності залежності між різними показниками;
- ✓ компактно та інформативно описати результати експериментів;
- ✓ виявляти інформативність показників.

У дослідженнях з рослинництва об'єктом вивчення є найчастіше сукупність предметів та явищ.

Тому будь-яке множинне явище, наприклад, рослин, їх частин, ділянок, у польовому досліді називають статистичною сукупністю.

Предмети та явища у кожній сукупності суворо індивідуальні. Вони відрізняються один від одного рядом ознак, наприклад, масою, обсягом і т.д. Кожна ознака має різний ступінь виразності, тобто ця ознака змінюється, варіює в певних межах ознаки, що вивчається.

Звідси, властивість предметів (рослин, ознак продуктивності рослин, врожайності на ділянках польового досвіду) відрізнятися один від одного навіть у однорідних сукупностях називається мінливістю або варіюванням, випадковістю.

Мінливість ознак – властивість, яка характерна всім предметам природи. Так наприклад, варіюючою ознакою рослин є висота, кількість і маса насіння однієї рослини, маса 1000 насінин, вміст протеїну, жиру в насінні і т. д.

Варіювання виникає внаслідок того, що рослини одного самого сорту, лінії завжди відрізняються своєю спадковістю. Крім того, формування ознак продуктивності рослин часто відбувається за різними погодними умовами вегетаційного періоду. Отже, при будь-якому проведенні досліджень дані дослідів завжди варіюватимуть у тих чи інших межах. Можливі значення варіюючої ознаки X називають у математичній статистиці варіантами і

позначають x_1, x_2, x_3 .

Мінливість, варіювання ознак створює труднощі в тих випадках, коли потрібно дати загальну характеристику певній групі рослин, ґрунтів і так далі за окремими ознаками або порівняти дві такі групи та знайти різницю між ними.

Дослідникам не завжди вдається досліджувати за тією чи іншою ознакою всі рослини, всю сукупність дослідів з двох причин:

- 1) сукупність дуже велика за обсягом;
- 2) щодо тієї чи іншої ознаки генеральної сукупності доводиться знищувати всю цю групу, сукупність.

Тому в науково-дослідній роботі запобігають до вивчення частини сукупності, через яку роблять узагальнення про всю генеральну сукупність.

При застосуванні статистичної обробки результатів дослідів науковим співробітникам необхідно чітко виділяти та представляти, що собою уявляю генеральна сукупність.

Генеральна сукупність – це вся сукупність особин, рослин, варіантів дослідів або вся група об'єктів, що підлягають вивченню у запланованих дослідженнях.

Вибіркова сукупність, вибірка – частина об'єктів сукупності, або вибірка з генеральної сукупності, або частина об'єктів дослідів, яка поставлена дослідником під контроль, під перевірку.

В більш широкому аспекті розуміння з боку статистики, вибірка – це деяка кількість об'єктів (особин), вибраних з даного угруповання з метою оцінки параметрів (середнє, дисперсія і т.д.) цього угруповання.

Наприклад, рослини окремого сорту, гібрида або варіанта дослідів, які відбираються для визначення їх продуктивності або інших показників, утворюють вибірку.

Характерна особливість більшості агрономічних досліджень у тому, що вибірка містить лише малу частину досліджуваного угруповання, оскільки всю сукупність досліджувати неможливо.

Вибірка є "представником" того угруповання, звідки воно відібрано для аналізу.

Взагалі ж у статистиці поняття «генеральна сукупність» ми можемо розуміти – як безліч будь-яких однорідних елементів, з якого за певним правилом виділяється деяка підмножина, яка називається вибіркою.

Наведені характеристики генеральної сукупності та вибірки показують нерозривну їх пов'язаність між собою, яку повинен урахувати дослідник при організації та проведенні досліджень.

Найважливішим властивістю вибірки у статистиці є її випадковість. Принцип випадковості означає рівну ймовірність попадання кожного елемента досліджуваної генеральної сукупності у вибірку. Порушення принципу випадковості може призвести до отримання не вірогідних даних та серйозних помилок у висновках. Так, наприклад, у деяких посібниках можна зустріти методики відбору "типових представників" генеральної сукупності на основі візуальних даних (наприклад, "модельні варіанти"). Така методика сама по собі може бути необхідною, наприклад, для підбору особин в експериментальну

групу, де важлива вирівняність об'єктів, що досліджуються, по ряду показників. Проте, для характеристики генеральної сукупності такий добір неприйнятний, оскільки значною мірою він є суб'єктивним і не може бути основою для отримання статистично обґрунтованих висновків. З такої вибірки, наприклад, свідомо виключаються нетипові, але досить численні особини. Це веде до невірних оцінок: заниження показників різноманітності, зміщення міжгрупових відмінностей, спотворення вкладів факторів, що досліджуються, в загальну мінливість ознаки. Принцип випадковості дозволяє теоретично описати розподіл параметра, що оцінюється (наприклад, середнє арифметичне) в генеральній сукупності, що складається з усіх мисливих випадкових вибірок даного обсягу, які можна було б витягти з даної сукупності. Найголовніше те, що реальна вибірка підкоряється цьому розподілу лише тоді, коли вона випадкова. Це становище є основою статистичних методів.

Цілком очевидно, що не завжди можливо досліджувати за тією чи іншою ознакою всі особливості рослин. У цих випадках вдаються до вивчення частини її, через яку роблять загальні висновки. Такий метод називають вибіркоvim, і вважається основним під час статистичного вивчення сукупності.

Сутність вибіркового методу в дослідях з агрономії полягає в тому, щоб за малою спробою отримати відомості і зробити висновок про всю генеральну сукупність. Тут дуже важливо правильно вибрати обсяг вибірки.

Обсяг вибірки – це кількість елементів у генеральній сукупності та вибірці, її позначають буквою n . Правильний обсяг вибірки – це застосування методу повторності дослідів. Повторність дослідів у просторі виключає помилку. Отже, помилки нівелюються у процесі повторення.

В результаті спостережень за рахунок збільшення повторень отримуємо відомості про числовій величині досліджуваної ознаки цієї сукупності.

Об'єкти досліджень в рослинництві – це окремі рослини, групи рослин, середовище та умови вирощування рослин сортів або гібридів. Усім їм властиве явище мінливості, чи варіювання. Ступінь варіювання, виражену математично, називають варіацією. Якщо тисячі насіння однієї й тієї ж культури, одного сорту посіяти і вирощувати в однакових умовах, рослини завжди будуть відрізнятися за зростанням, масою, зовнішнім виглядом, врожаєм, його якістю і т. д.

Таким чином, головне завдання статистичної обробки експериментальних даних – знайти такі показники, які характеризують особливості сукупностей, що вивчаються, порівняти їх між собою, встановити вірогідність отриманих результатів та зробити висновки стосовно ефективності проведення досліджень.

1.2. Основні статистичні терміни та показники

Метою більшості досліджень є збір даних, які підлягають статистичній обробці, аналізу і згодом допомагають отримати вірогідну інформацію про результати проведеного дослідів та зробити висновки стосовно ефективності досліджуваних факторів.

Зазвичай, дослідники отримують дані дослідів із вибірки індивідуумів.

Мета статистичної обробки даних дослідів полягає в тому, щоб згрупувати

ці дані та витягти з них достовірну інформацію.

Будь-яка безліч ідентифікованих об'єктів, рослин, факторів, явищ, що відрізняються за конкретною ознакою один від одного незначно, але зберігають подібність за деякими суттєвими характеристикам називається сукупністю.

Сукупністю можна називати ділянку, варіант досліда, поле рослин, сорт або гібрид сільськогосподарської культури та інше.

Сукупності представлені окремими членами або одиницями. Сукупність членів чи одиниць називається обсягом сукупності. Саме виміряне значення кожної одиниці сукупності називається варіантом сукупності.

Відмінності між варіантами сукупності називаються варіацією чи дисперсією.

Якщо обсяг сукупності будь-якої вибірки позначити через n , а кожен варіант сукупності через x , то в такому випадку ряд варіантів сукупності можна записати таким чином: $x_1, x_2, x_3, \dots, x_n$.

Отже, якщо проведеними дослідом встановлено досліджено вплив будь-якого агротехнічного прийому на врожайність сортів чи гібридів, то кількість використаних для дослідів сортів або гібридів можна назвати сукупністю, обсяг якої, наприклад, дорівнює $n=11$.

Дія агротехнічного прийому на формування врожаності одним сортом чи гібридом є варіант сукупності.

Якщо вимірний матеріал є лише частиною якогось загального матеріалу, то така сукупність буде називатися вибірковою сукупністю, а весь величезний матеріал, з якого взято вибірка сукупність називається генеральною сукупністю.

Наприклад, якщо виміряні 100 зерен якогось сорту пшениці являють собою вибірку сукупність з обсягом $n = 100$, тобто вага зерна всього сорту пшениці можна вважати генеральною сукупністю обсягом, рівним $N \rightarrow \infty$.

Обробка даних та інтерпретація отриманих розрахунків залежить від того, які ознаки досліджувались і були виміряні в експерименті.

Ознаки об'єктів в рослинництві в основному представлені трьома категоріями.

- ознаки сукупності, що відрізняються певною якістю. Такі ознаки відносяться до якісних або альтернативних. Наприклад, сукупність сортів гороху представлена червоними та білими квітками;

- ознаки, що мають кількісний захід, але відмінності між окремими варіантами сукупності виражаються цілими числами. Наприклад, кількість, маса насіння однієї рослини пшениці. Така кількісна мінливість та такі кількісні ознаки називаються дискретними;

- ознаки, що мають кількісний захід, але між окремими випадками сукупності немає чітко виражених (дискретних) кордонів. Наприклад, маса зерна пшениці з $1,0 \text{ м}^2$ виражена в грамах, але не в кілограмах. Така мінливість та такі ознаки називаються кількісними ознаками з безперервною мінливістю (мірні ознаки, континуальні).

Методи статистики, що застосовують для аналізу отриманих результатів та

встановлення ефективності досліджувань, ґрунтуються на вірогідності та рівні значущості.

Вірогідність отриманих даних дослідів (reliability) безпосередньо пов'язана з репрезентативністю певної вибірки, на основі якої дослідником робляться висновки. Репрезентативність – це показник того, наскільки результати вибірки відповідають генеральним параметрам досліду. Іншими словами, вірогідність показує наскільки ймовірно те, що залежність, яка встановлена для даної вибірки, буде знову виявлена (підтверджена) на даних іншої вибірки, відібраної з тієї ж самої генеральної сукупності.

Рівень значущості (p-level) є мірою довіри до отриманого результату досліду. Її вищий рівень відповідає нижчому рівню довіри до знайденої у статистичному аналізі закономірності. Рівень значущості показує можливість помилки, пов'язаної з поширенням спостережуваного результату на всю генеральну сукупність, тобто можливість того, що знайдена в досліджуваній вибірці закономірність (зв'язок між змінними параметрами) є лише випадковою особливістю цієї вибірки.

Так, наприклад. Рівень значущості t-критерію, що оцінює ступінь відмінностей між двома вибірками, становить 0,032. В цьому випадку ймовірність випадковості отриманих відмінностей між вибірками становитиме дещо більше ніж 3,0%. Вибір певного рівня значущості, вище якого результати відкидаються як хибні, є досить довільним. Зазвичай результат $p=0,05$ є прийнятною межею статистичної значущості. Тобто величина рівня значущості, що дорівнює або менша, ніж 0,05, говорить про достовірність отриманого результату.

Проте слід пам'ятати, що цей рівень все ще включає досить велику ймовірність помилки (5%). Результати, отримані лише на рівні $p=0,01$, зазвичай розглядаються як статистично значущі, а на рівні $p=0,005$ чи $p=0,001$ – як високо значущі. Однак слід пам'ятати, що дана класифікація рівнів значущості досить довільна і є лише неформальною угодою, прийнятою на основі практичного досвіду в тій чи іншій області дослідження.

Якщо вимірювання ознак у об'єктів, що вивчаються в досліді, зроблено, вони записуються в журналах у порядку виміру об'єктів, варіантів, тобто. матеріал являє собою набір цифр часто в хаотичному порядку.

У практичних завданнях статистики є сукупність спостережень $x_1, x_2, \dots, x_n$, на основі яких необхідно зробити ті чи інші висновки стосовно отриманих результатів дослідження.

Для компактного опису сукупності використовуються методи описової статистики.

Показниками, що описують вибірку, є параметри та положення статистики. При статистичній обробці результатів досліджень застосовуються математичні величини. При підготовці до статистичного аналізу дослідник повинен мати на увазі для чого він ці величини застосовує і як доцільно їх застосовувати у різних випадках.

Показниками, що описують вибірку, є параметри положення (середнє арифметичне значення, мода, медіана, сума квадратів відхилень, мінімум,

максимум, верхній та нижній квартилі та інш.) та розкиду (дисперсія, стандартне відхилення, коефіцієнт варіації, та інш.).

Середнє арифметичне значення (mean) показує "центральне становище" змінної.

Середня арифметична величина є дуже важливим параметром, що характеризує вибіркoву сукупність. Вона використовується для характеристики будь-яких сукупностей, ознак, варіантів.

Середня арифметична величина є узагальненою характеристикою сукупності, ознаки, варіанту.

Середнє арифметичне, яке часто називають просто «середнє значення», одержують шляхом складання всіх значень ознаки, варіанту та поділу отриманої суми на кількість значень у вибірці.

Середня арифметична величина виражається за допомогою алгебраїчної формули. Набір кількості (n) спостережень змінної x відображають за формулою:

$$\bar{x} = (x_1 + x_2 + x_3 + \dots + x_n)/n$$

Середня арифметична величина характеризується рядом властивостей:

✓ якщо до кожного значення вибірки додати або відібрати одну і ту ж величину, або помножити і розділити на ту саму величину, то середня арифметична збільшиться або зменшиться на цю величину;

✓ алгебраїчна сума відхилень окремих варіантів сукупності від середньої арифметичної цієї сукупності дорівнює нулю;

✓ сума квадратів відхилень варіант сукупності від середньої арифметичної $\bar{x}$ менше суми квадратів відхилень від будь-якої іншої величини.

Помилка середньої арифметичної.

Для представлення своїх результатів варіантів вибірки вказують значення середньої арифметичної і середньої помилки.

Чим більшою є різноманітність ознаки, тим більша помилка, і чим більша чисельність вибірки, тим менша помилка

Якби всі об'єкти були однакові, тобто різноманітність була б одно нулю, то й помилка дорівнювала нулю. У цьому випадку навіть один екземпляр точно характеризував би всю генеральну сукупність.

Середнє квадратичне значення.

Точнішим показником, що характеризує варіацію або розсіювання варіант навколо середнього арифметичного значення, є середнє квадратичне значення. Середнє квадратичне значення ґрунтується на розгляді відхилень значень ознаки окремих одиниць сукупності від середньої арифметичної.

При цьому використовується спосіб усереднення відхилень варіантів від середньої арифметичної, що дозволяє обійти труднощі, обумовлену рівністю нулю їхньої суми алгебри.

Цей спосіб зводиться до розрахунку квадратів відхилень варіантів від середньої з них подальшим усередненням.

Мода (mode) вибірки – значення статистики, що найчастіше зустрічається у вибірці (максимум функції щільності розподілу). Мода випадкової величини x

позначається як $\text{mod}(x)$. За необхідністю мода розраховується у MS Excel за допомогою функції МОДА.

Медіана (median) – це значення, яке розбиває вибірку на дві рівні частини. Половина спостережень лежить нижче за медіану і половина – вище за неї. Якщо кількість спостережень у вибірці непарна, то медіана розраховується як середнє двох середніх значень. Медіана випадкової величини x позначається $\text{med}(x)$. За необхідністю медіана розраховується у MS Excel за допомогою функції МЕДІАНА.

Сума квадратів відхилень (sum of squares) – це сума зведених у квадрат відхилень випадкової величини від її середнього значення.

Квартилі - це набір квартилів для $p = 0.25, 0.5, 0.75$. Їх оцінками (квартилями емпіричного розподілу) є величини, що ділять вибірку даних на чотири групи (точки), що містять (по можливості) однакові кількості спостережень.

Нижній квартиль (Q_1) – це значення, нижче якого в упорядкованому множині даних знаходиться чверть даних вибірки. Нижній квартиль, (відомий як 25-й емпіричний квартиль) визначається як середнє число між найменшим числом та медіаною набору даних і визначає, де 25% даних вибірки знаходяться нижче або ліворуч від нього (якщо дані впорядковані на часовій шкалі від найменшого до найбільшого).

Другий квартиль (Q_2) – дорівнює медіані і це є медіана набору даних. 50% даних вибірки знаходиться нижче цієї точки.

Верхній квартиль (Q_3) – це середнє значення між медіаною і найвищим значенням набору даних. Верхній квартиль показує значення, вище якого у впорядкованому безлічі даних є чверть значень. Квартиль відомий як 75-й емпіричний квартиль. 75% даних знаходяться вище цієї точки.

Різниця $Q_3 - Q_1$ – це інтерквартильна широта і є робастною мірою розсіювання. У зв'язку з тим, що для обчислення квартилів необхідно впорядкувати дані від найменшого до найбільшого, квартилі є формою статистики порядку.

Дисперсія (variance) – це усереднена для вибірки сума квадратів відхилень. Дисперсія є однією з основних характеристик мінливості величини, яка вивчається в досліді. Найважливішим дійством дисперсії є те, що вона може бути розчленована на частини, кожна з яких обумовлена дією будь-якого фактора. Сума цих частин є повною дисперсією випадкової величини.

Стандартне (середньоквадратичне) відхилення (Standard deviation) також є мірою різноманітності об'єктів, які входять у вибірку. Стандартне відхилення розраховується як квадратний корінь з дисперсії, отже, має ту ж розмірність, що і ознака, яка вивчається дослідником. Стандартне відхилення показує, як у середньому відхиляється значення ознаки кожного об'єкта досліді від середнього арифметичного. Чим сильніший розкид значень, тим більше стандартне відхилення. Оскільки стандартне відхилення – величина з певною одиницею виміру (метри, кілограми, штуки тощо), то по стандартному відхиленню можна порівнювати мінливість лише одних і тих самих ознак, що виражені в одних одиницях і мають порівняні середні величини. Середнє

квадратичне відхилення показує, наскільки її середнє коливається в досліджуваній сукупності, і виражається у тих самих одиницях вимірювання, як і варіанти. Це обмежує застосування даного параметра. Для порівняння рівня мінливості ознак різної розмірності використовується коефіцієнт варіації.

Коефіцієнт варіації (coefficient of variation) кількісної ознаки уявляє собою відношення стандартного відхилення до середнього арифметичного значення і виражається у процентах

Коваріація (covariance) є мірою сполученою мінливості двох ознак. Коваріація відбиває дисперсію, обумовлену спільним варіюванням ознак. За необхідністю коваріація розраховується у MS Excel із застосуванням команди КОВАР. Під час розгляду цифрових значень результатів досліджень неможливо виявити будь-які закономірності. Тому, існують деякі прості способи упорядкування матеріалу:

Перший спосіб. Можна переписати отримані дані у порядку наростання їх величини чи порядку зменшення величин. Такий спосіб упорядкування даних називається ранжуванням.

При ранжуванні дані досліджувані упорядковуються від меншого показника ознаки до більшого (або навпаки) та привласнюють їм ранги.

Труднощі у призначенні рангів виникають у тому випадку, якщо зустрічаються збігаються значення. У цьому випадку обчислюються середні ранги, які є середнім арифметичним значенням рангів, призначених для чисел у разі, якби вони відрізнялися на деяку невелику величину. У будь-якому випадку в результаті розрахунку найбільший ранг повинен дорівнювати загальній кількості елементів.

Ранжування даних дає досліднику вже деяке уявлення про характер експериментального матеріалу. При цьому позначають мінімальне та максимальне значення варіантів, яким надають назву ліміти мінливості. Найменше значення варіанти називається мінімальним значенням (min), а найбільше – максимальним значенням (max).

У центрі такого впорядкованого (ранжованого) ряду цифр зосереджено основна кількість варіантів із "середнім значенням" ознаки.

Наприклад, дослідник виконує дослідження по встановленню продуктивності створених сортосразків соняшнику за ознакою маса насіння одного кошику (в грамах) у 15 рослин. Отримано дані: 65, 69, 59, 62, 81, 77, 64, 79, 85, 57, 59, 68, 65, 78, 66.

Ранжуємо варіанти сукупності у порядку наростання маси насіння кошику (в грамах) рослин соняшнику: 57, 59, 59, 62, 64, 65, 65, 66, 68, 69, 77, 78, 79, 81, 85.

Такий ранжований ряд дає уявлення про те, що ліміти мінливості ознаки маса насіння ознаки, як частота серцевих скорочень, є у межах min = 57г; max=85 г. Середнє значення ознаки маса насіння одного кошика можливо оцінити на рівні 68,9 г.

Другий спосіб. Він уявляє собою найбільш наочний спосіб упорядкування варіантів сукупності. Відповідно до вимог цього способу досліднику необхідно виписати в стовпчик усі величини кількісною ознак, варіантів, які зустрілися

разом.

Цифри можуть бути записані як у порядку наростання від найменшого значення до більшого, наприклад від 1 до 11, так і порядок зменшення, тобто від 11 до 1.

За ознакою маса насіння одного кошику рослини соняшнику запишемо отримані цифри в порядку наростання, та визначимо скільки разів кожна цифра зустрічається у розподілі ознаки. Різні величини ознаки, варіатив, які зустрілися у розподілі, будуть називатися класами та позначатися x_i , а числа, що відповідають кожному класу – чисельностями чи частотами і позначатимуться літерою f . Створюється таблиця і дані заносяться в робочу таблицю:

Ознака: маса насіння кошику соняшнику, г (x_i)	Скільки разів зустрічається значення ознаки (f)
57	1
59	2
62	1
64	1
65	2
66	1
68	1
69	1
77	1
78	1
79	1
81	1
85	1
Сума	15

Такий ряд чисел, що складається з класів та частот, називається **варіаційним рядом**. Варіаційний ряд дає чіткішу картину закономірностей розподілу проти простого ранжирования.

1.3. Типи вибірок

Для формування вибірки можна використовувати такі прийоми:

- ✓ проста повторна вибірка;
- ✓ проста випадкова вибірка;
- ✓ проста безповторна вибірка;
- ✓ систематична імовірнісна вибірка;
- ✓ серійна (гніздова) вибірка;
- ✓ районована вибірка.

Проста повторна вибірка.

Використання такої вибірки ґрунтується на припущенні, що кожен респондент з рівною часткою ймовірності може потрапити у вибірку. На основі

генерального списку сукупності складаються картки із номерами респондентів. Вони поміщаються в колоду, перемішуються і їх навмання виймається картка, записується номер, потім повертається назад. Далі процедура повторюється стільки разів, який обсяг вибірки нам необхідний. Недоліки: повторення одиниць відбору.

Процедура побудови простої випадкової вибірки включає у собі наступні етапи.

Перший етап: необхідно отримати повний список членів генеральної сукупності та пронумерувати цей список. Такий список, нагадаємо, називається основою вибірки;

Другий етап: визначити очікуваний обсяг вибірки, тобто очікуване число рослин, ознак, варіантів, що досліджуються в досліді.

Третій етап: витягти з таблиці випадкових чисел стільки чисел, скільки досліднику потрібно вибірових одиниць. Якщо у вибірці має опинитися 100 рослин, із таблиці беруть 100 випадкових чисел. Ці випадкові числа можуть генеруватись комп'ютерною програмою.

Четвертий етап: вибрати зі списку-основи спостереження, номери яких відповідають виписаним випадковим числам.

Проста випадкова вибірка має очевидні переваги.

Цей метод простий для розуміння. Результати дослідження можна поширювати на досліджувану сукупність. Більшість підходів до отримання статистичних висновків передбачають збір інформації за допомогою простої випадкової вибірки.

Однак метод має суттєві обмеження:

1. Часто складно створити основу вибірового спостереження, яка дозволила б провести просту випадкову вибірку.
2. Результатом застосування простої випадкової вибірки може стати велика сукупність, або сукупність, розподілена по великій географічній території, що значно збільшує час та вартість збирання даних.
3. Результати застосування простої випадкової вибірки часто характеризуються низькою точністю та більшою стандартною помилкою, ніж результати застосування інших імовірнісних методів

Проста безповторна вибірка.

Процедура побудови вибірки така ж, тільки картки з номерами респондентів не повертаються назад до колоди.

Систематична імовірнісна вибірка.

Вибірка є спрощеним варіантом простої імовірнісної вибірки. На основі генерального списку сукупності через певний інтервал (K) відбираються респонденти. Розмір інтервалу (K) визначається випадково. Найбільш достовірний результат досягається при однорідній генеральній сукупності, інакше можливі збіг величини кроку та якихось внутрішніх циклічних закономірностей вибірки (змішування вибірки).

Серійна (гніздова) вибірка.

Одиниці відбору є статистичні серії (родина, вид, сорт, сортозразок тощо). Відібрані елементи піддаються суцільному обстеженню. Відбір статистичних

одиниць може бути організований як при випадковій так і систематичній вибірці. Недоліки: можливість більшої однорідності, ніж у генеральній сукупності.

Районована вибірка.

У разі неоднорідної генеральної сукупності, перш ніж використовувати ймовірнісну вибірку з будь-якою технікою відбору, рекомендується розділити генеральну сукупність на однорідні частини, така вибірка називається районованою. Групами районування можуть виступати як сорт, сортозразок, так і будь-яка ознака рослин, яка закладена в основу проведення дослідження. Ознака, на основі якої здійснюється розподіл, називається ознакою розшарування та районування.

При проведенні дослідження дослідника цікавить властивість або кількісна ознака, які вивчаються на у двох або у більший кількості вибірках з метою їх подальшого порівняння. Ці вибірки можуть знаходитися у різних співвідношеннях. Незалежні вибірки характеризуються тим, що можливість відбору будь-якого показника ознаки однієї вибірки залежить від відбору будь-якого з випробуваних іншої вибірки.

Залежні вибірки характеризуються тим, що кожному випробуванню однієї вибірки поставлено у відповідність до певного критерію випробування з іншої вибірки.

Результати лабораторних досліджень – один і об'єктів для статистичної обробки в наукових дослідженнях.

При проведенні статистичного аналізу отриманих даних не завжди вдається вичленувати найважливіші результати з величезної кількості матеріалу.

Така ситуація пов'язана з не правильним використанням різних критеріїв. Насамперед, по перше досліднику необхідна ідея: що буде досліджувати і які показники буде порівнюватиме в досліді. По-друге, дослідник повинен правильно вибрати критерій визначення достовірності і тільки потім грамотно використати математичні формули в статистичній обробки даних.

Найчастіше для визначення та отримання достовірних результатів статистичного аналізу в науковій літературі рекомендують застосовувати *t*-критерій Стьюдента.

***t*-критерій Стьюдента** – загальна назва для класу методів статистичної перевірки гіпотез (статистичних критеріїв), заснованих на розподіл Стьюдента. Найчастіші випадки застосування *t*-критерію пов'язані з перевіркою рівності середніх значень у двох вибірках досліді.

Цей критерій був розроблений Вільямом Госсетом для оцінки якості пива в компанії Гіннес. У зв'язку з зобов'язаннями перед компанією з нерозголошення комерційної таємниці (керівництво Гіннеса вважало такою використання статистичного аналізу у своїй роботі), стаття Госсета вийшла 1908 року в журналі «Біометрика» під псевдонімом Студент (Студент). Цей критерій найчастіше використовується для перевірки гіпотези: «Середні дві вибірки відносяться до однієї і тієї ж сукупності». В цьому разі можна застосувати критерій Стьюдента (за умови достатньо більших обсягів вибірок ($n \geq 30$), або переконавшись, що статистичні ряди близькі до нормального закону розподілу).

При використанні критерію можна виділити два випадки.

У першому випадку його застосовують для перевірки гіпотези про рівність генеральних середніх двох незалежних, незв'язаних вибірок (так званий двовибірковий t -критерій). У цьому випадку є контрольна група та експериментальна (дослідна) група. Кількість виконаних випробувань у групах може бути різна.

У другому випадку, коли та сама група об'єктів передбачає застосування числового матеріалу для перевірки гіпотез про середні значення ознак досліду, використовується так званий парний t -критерій вибірки при цьому називають залежними, пов'язаними.

Перед статистичною обробкою дані необхідно відповідним чином підготувати: округлити, обчислити середні арифметичні за кожною дослідною ділянкою і варіантом, вибракувати сумнівні і відновити випали дані, перетворити їх.

1.4. Округлення дослідних даних

У дослідженнях користуються наступним правилом: для отримання достатньо точних чисел необхідно мати досвідчені дані з трьома цифрами. Наприклад, врожайність з дослідної ділянки слід записувати 0,225 т/га; 2,25 т/га; 22,5 т/га.

Для більш ретельного округлення використовують зменшене у 4 рази стандартне відхилення певного варіаційного ряду. Якщо першою цифрою для $s/4$ виявиться ціле число, то дані округляють до цілого числа.

При розрахунку суми квадратів беруть додаткову цифру, тобто. якщо вихідні дані мають десяті частки, то квадрати обчислюють до сотих. Якщо цифра за останньою цифрою більше 5 або після 5 слідує цифра більше нуля, то останню значну цифру збільшують на одиницю. Так, числа 95,78 та 95,767 округляють до 95,8. Якщо за останньою цифрою коштує 5, а потім нулі, то останню значну непарну цифру збільшують на одиницю: $94,550 = 94,6$, а парна цифра залишається незмінною: $74,450 = 74,4$.

1.5. Вибракування сумнівних даних

Сумнівні дані, які значно відрізняються від решти даних якого-небудь варіанту, визначають тільки за допомогою математичної статистики. Суб'єктивний брак даних неприпустимо.

Розглянемо це на конкретному прикладі.

В досліді, де була 6-кратна повторність у варіанті з вивченням впливу інокуляції насіння сої без внесення добрив, врахували масу 1000 насінин. За підрахунками даних досліджень було отримано наступні результати за ознакою маса 1000 насінин по кожному із повторень: 152,1 г; 152,9 г; 153,7 г; 150,4 г; 151,8 г; 150,5 г; 149,9 г; 140,4 г; 151,5 г; 150,9 г.

Щоб переконатися, що всі дані належать до певного варіаційного ряду, виконують наступні операції.

1. Ранжують варіаційний ряд у порядку, в якому спостерігається зростання маси 1000 насінин (г) сої: 140,4 г; 149,9 г; 150,4 г; 150,5 г; 150,9 г; 151,5 г; 151,8 г; 152,1 г; 152,9 г; 153,7 г.

Двом першим і двом останнім даним маси 1000 насінин відповідно надають позначення: X_1, X_2, X_{n-1}, X_n . $X_1 = 140,4$ г; $X_2 = 149,9$ г; $X_{n-1} = 152,9$ г; $X_n = 153,7$ г.

Найбільш сумнівними будуть крайні чисельні показники: 142,1 г і 153,7 г. Сумнівність цих показників перевіряють шляхом розрахунку критерію τ показників X_1 і X_n для 5-ти та 1-го рівня значущості.

Критерій τ_1 для показника X_1 розраховують за формулою:

$$\tau_1 = (X_2 - X_1) : (X_{n-1} - X_1) = (149,9 - 140,4) : (152,9 - 140,4) = 9,5 : 12,5 = 0,76.$$

$$\tau_1 = 0,76.$$

Критерій τ_2 для показника X_n розраховують за формулою:

$$\tau_2 = (X_n - X_{n-1}) : (X_n - X_2) = (153,7 - 152,9) : (153,7 - 149,9) = 0,8 : 3,8 = 0,210.$$

$$\tau_2 = 0,210.$$

Далі знаходять теоретичні значення критеріїв τ відповідно до даних додатку 1.

При кількості аналізованих показників $n = 10$, критерії τ при 5-ти та 1-го рівня значущості рівняються: $t_{0,95} = 0,477$; $t_{0,99} = 0,597$ (дані теоретичних критеріїв τ для 5-ти та 1-го рівня значущості беруть з додатку 1).

Отримані розрахункові дані критеріїв τ порівнюють з теоретичними критеріями τ , що наведено в додатку 1. Якщо розрахункові критерії τ більше або рівні теоретичним критеріям τ , то дані, які перевіряються, відносять до сумнівних.

Розрахунковий критерій τ_1 дорівнюється 0,760 і він більший за значенням теоретичних критеріїв $t_{0,95} = 0,477$ та $t_{0,99} = 0,597$. Це підтверджується як на рівні довірчої ймовірності $t_{0,95}$, так і на рівні $t_{0,99}$, коли згідно теорії ймовірностей можливість помилитися становить лише один випадок на 100. Таким чином, показник маси 1000 насінин 142,1 г, який має критерій $\tau_1 = 0,879$, сумнівний і підлягає обов'язковій вибраковки.

Розрахунковий критерій τ_2 дорівнюється 0,210 і він за значенням менше теоретичних критеріїв $t_{0,95} = 0,477$ та $t_{0,99} = 0,597$. Таким чином, максимальний показник маси 1000 насінин 153,7 г, який має критерій $\tau_2 = 0,210$, не сумнівний і не підлягає вибраковки.

Середню арифметичну масу 1000 насінин сої слід обчислювати не з 10 (десяти) даних, а з 9 (дев'яти) даних, які залишилися після вибраковки сумнівного значення маси 1000 насінин 142,1 (г): $149,9 + 150,4 + 150,5 + 150,9 + 151,5 + 151,8 + 152,1 + 152,9 + 153,7 = 151,5$ г.

Без виконання вибраковки сумнівного значення (142,1 г) середня арифметична маса 1000 сої була б заниженою – 150,6 г.

Слід брати до уваги, що виконувати бракування сумнівних даних дослідів за наведеними формулами можна лише за загальної чисельності отриманих даних не менше як $n = 4$.

1.6. Відновлення даних, що випали під час проведення досліду

Обробка даних часто ускладнюється випаданням даних на деяких ділянках дослідів (пошкодження рослин птахами; шкідниками, хворобами, пошкодження рослин під час виконання механічних агроприйомів по догляду за посівами, затоплення ділянки після злив, проїзд транспорту і т. д.).

Через випадання даних середні арифметичні у разі можуть бути або завищені, якщо випав результат з дуже низьким числовим значенням, або занижені, якщо випав результат із максимальним числовим значенням. Через це виникають помилки, які можна усунути, якщо відновити дані, які випали, тобто обчислити їх найбільш ймовірні значення. При випаданні одного результату користуються наступною формулою:

$$X_{ls} = [(L \times V) + (n \times P) - \sum X] : [(L - 1) \times (n - 1)],$$

де:

L – кількість варіантів в досліді;

V – сума даних у варіанті, де випав результат; n – кількість повторностей у досліді;

P – сума даних у повторності, де випав результат;

$\sum X$ – сума даних у всьому досліді, за винятком результату, який випав.

З метою закріплення знань з відновлення даних, які випали під час проведення дослідів розглянемо приклад. В досліді закладено 4 (чотири) варіантів з вивчення густоти стояння соняшнику у 3-кратному повторенні. Результати отриманих даних з урожайності дослідів наведено в таблиці 1.1.

Таблиця 1.1

Вплив густоти стояння рослин на формування врожайності соняшнику

Варіант (густина стояння рослин, тис./га)	Врожайність у повтореннях, ц/га		
	1 повторення	2 повторення	3 повторення
40 (контроль)	27,9	26,7	28,2
30	24,2	X_{ls}	24,7
50	34,1	33,2	33,4
60	31,7	30,7	29,9

Дані таблиці 1.1 показують, що у другому повторенні другого варіанту відсутній результат. Результат повинно відновити.

Визначаємо суму врожайності 1-го та 3-го повторень другого варіанту: $V = 24,2 + 24,7 = 48,9$.

Визначаємо суму врожайності всіх варіантів другого повторення, де випав результат другого варіанту: $P = 26,7 + 33,2 + 30,7 = 90,6$.

Розраховуємо суму даних врожайності за повтореннями всього дослідів, за винятком результату, який випав у другому варіанті другого повторення:

$\sum X = 27,9 + 26,7 + 28,2 + 24,2 + 24,7 + 34,1 + 33,2 + 33,4 + 31,7 + 30,7 + 29,9 = 324,7$.

Розраховані дані вставляємо у формулу і визначаємо результат, який випав в процесі проведення дослідів:

$$X_{ls} = [(L \times V) + (n \times P) - \sum X] : [(L - 1) \times (n - 1)] = [(4 \times 48,9) + (3 \times 90,6) - 324,7] : [(4 - 1) \times (3 - 1)] = [195,6 + 271,8 - 324,7] : [3 \times 2] = 142,7 : 6 = 23,8 \text{ ц/га.}$$

Результат, що випав в досліді у другому повторенні другого варіанту за проведеними розрахунками становить 23,8 ц/га.

Відновлений результат ставлять у друге повторення, другого варіанту, і виконують відповідний аналіз – розраховують середню арифметичну врожайність другого варіанту за всіма повтореннями цього варіанту.

При випаданні одночасно кількох даних в одному досліді можна використовувати метод статистичної обробки для дослідів з неповним числом даних.

Для підтвердження отримання вірогідних результатів дослідів в статистичній обробки даних застосовують довірчий інтервал та точність дослідів.

Довірчий інтервал (confidence interval) – інтервал значень навколо оцінки параметра (який обчислюється за вибірковими даними), в якому знаходиться "справжнє" значення цієї статистики. При цьому ймовірність того, що "справжнє" (генеральне) значення не потрапляє до цього інтервалу, дорівнює рівню значущості.

Наприклад, якщо середня вибірки за параметром довжина качана кукурудзи дорівнює 21 см, а нижня та верхня межі довірчого інтервалу з рівнем значущості $p = 0,05$ дорівнюють 18 см і 24 см відповідно, то можна укласти з 95%-ної ймовірністю, що інтервал з межами 18 см та 24 см "накриває" середню популяцію. Відповідно, ймовірність того, що "справжнє" середнє не потрапило в цей інтервал, становить 5,0%.

Якщо встановити більший рівень довіри (тобто рівень значущості зробити менше), то інтервал стане ширшим і ймовірність, з якою він "накриває" невідоме середнє вибірки рослин, зростає.

Довірчий інтервал розраховується на основі помилки параметра та значення t -критерію Стюдента. Нижня межа довірчого інтервалу дорівнює

$\bar{x} - tm$, верхня межа довірчого інтервалу дорівнює $\bar{x} + tm$, де t – значення t -критерію для вибраного рівня значущості (кількість спостережень береться для випадку нескінченної великої сукупності), m – помилка параметра, яка вирахована за вибірковими даними.

Допустимо, обчислений за вибіркою середня висота рослин соняшнику в варіанті дослідів становить 143,4 см, стандартна помилка дорівнює 2,39. Середнє значення висоти рослин соняшнику записується як $143,4 \pm 2,39$. При найбільш поширеному у дослідженнях з рослинництва рівні 0,05 і при нескінченно великій сукупності значення t -критерію буде становити 1,96. Отже, довірчий інтервал для середнього значення висоти рослин соняшнику рівняється $\bar{x} \pm tm = 143,4 \pm 1,96 \times 2,39 = 143,4 \pm 4,68$ см. Таким чином, з ймовірністю 95% генеральне середнє висоти рослин соняшнику у варіанті дослідів знаходиться в інтервалі від 138,72 см ($143,4 - 4,68$) до 148,08 см ($143,4 + 4,68$).

1.7. Точність дослід (accuracy)

Стандартна помилка є іменованою величиною (має ту ж розмірність, що і досліджуваний параметр). Тому її використання для порівняльної оцінки точності дослід, з якою визначені вибіркові значення параметрів по відношенню до їх генеральних параметрів, незручно.

Для цієї мети застосовують показник точності дослід. Під точністю дослід розуміють відношення стандартної помилки середнього до середнього значення випадкової величини. Точність дослід виражається в процентах і визначає ступінь надійності отриманих даних.

В дослідженнях з рослинництва задовільною прийнято вважати точність, яка не перевищує 5%. Виходячи з цієї точності, визначають і мінімальну кількість спостережень.

Отримані у досліді дані, що виражені у великих чи малих кількісних одиницях, вважаються неоднорідними, а розподіл їх значень не відповідає нормальному. Це можливо при обліках та підрахунках у досліді кількості бур'янів, шкідників, ураженості рослин хворобами тощо. Такі результати досліджень не підпорядковуються законам нормального розподілу в межах статистичного аналізу і призводять до виникання неоднорідності дисперсій за вибірками, збільшення варіювання даних варіантів дослід.

Обробка таких даних дослід класичним способом, тобто методами параметричної статистики, вимагає попереднього перетворення вихідних даних з метою зниження їх абсолютних величин, а також ступеня варіації шляхом вилучення квадратного кореня, логарифмування і так далі.

Для перетворення вихідних даних в більшості випадках застосовують спосіб вилучення квадратного кореня із отриманого результату.

Перетворення вихідних даних цим способом передбачає застосування наступної формули:

$$x_1 = \sqrt{X},$$

де: x_1 – перетворена дана; X – вихідна дана.

Якщо результати досліджень виражені числами, які за значенням близькі до нуля, такі вихідні дані перетворюють за формулою:

$$x_1 = \sqrt{X + 1}.$$

За підрахунками чисельностей рідкісних подій, якщо більшість значень не перевищує 10, перевагу віддають формулі:

$$x_1 = \sqrt{X + 0,5}.$$

Перетворення вихідних даних через квадратний корінь дозволяють

уточнити суттєвість відмінностей між малими середніми. Це особливо цінно у дослідженнях з прийомами хімічної боротьби зі бур'янами, хворобами, шкідливими комахами, де відмінності серед відносно неефективних варіантів є менший інтерес, ніж різницю між препаратами, що забезпечують гарний захист посівів рослин.

Коли стандартні відхилення вибірок (варіантів) приблизно пропорційні середнім (відношення дають близькі величини) і коли множені головні ефекти (збільшення ефекту в рази) переважають над складовими, перетворення вихідних даних дослідів виконують за способом логарифмування.

За наявності нульових значень до всіх вихідних дат додають одиницю, оскільки логарифм нуля дає мінус нескінченність.

При великій кількості нульових значень слід вибирати інший спосіб перетворення. Перетворену дану позначають через x_1 , вихідну дану позначають через X . Перетворену дану визначають за формулою:

$$x_1 = \lg X \text{ або } x_1 = \lg X + 1.$$

При $X < 1$ вихідну дану множать на ціле число, наприклад на число 10 або на число 100; тоді $x_1 = \lg (100X)$.

Логарифмічне перетворення дає максимальний ефект при близькості стандартних відхилень до вибірових середніх, тобто коли $V \rightarrow 100\%$. Отримані дані рекомендується перетворювати, коли ступінь їхньої варіації перевищує 40 % або коли фактичний розподіл відповідає експоненційній кривій ($y = e^x$), наприклад, дані водопроникності ґрунту та мікроструктури, вміст обмінного Са в ґрунті, гребінність ґрунту, діаметр елементарних частинок ґрунту і т.д.

Оцінку суттєвості відмінностей між середніми слід проводити на основі перетворених даних, а потім зробити зворотний перехід до вихідних середніх, які називаються званих уточненими або наведеними.

Контрольні питання

1. Що називають генеральною сукупністю?
2. Що таке вибіркова сукупність?
3. Які види ознак використовують у статистичних дослідженнях?
4. Що характеризує середнє арифметичне?
5. У чому полягає відмінність між медіаною та модою?
6. Які показники характеризують розсіювання даних?
7. Що таке коефіцієнт варіації і для чого його використовують?
8. Які способи формування вибірки застосовують у наукових дослідженнях?
9. Для чого виконують ранжування експериментальних даних?
10. Які статистичні критерії застосовують для перевірки достовірності результатів досліджень?

Розрахункові задачі

Задача 1. Розрахунок основних статистичних показників

У досліді з вивчення впливу біостимулятора на висоту рослин гороху отримано такі результати (см): 61; 64; 60; 65; 63; 62; 66; 64; 61; 64.

Необхідно визначити:

1. Обсяг вибірки (n).
2. Середнє арифметичне ($\bar{x}$).
3. Мінімальне та максимальне значення.
4. Розмах варіації.
5. Медіану.
6. Моду.

Задача 2. Розрахунок дисперсії, стандартного відхилення та коефіцієнта варіації

Отримано дані маси 1000 насінин сої (г): 148,6; 149,8; 151,2; 150,3; 149,9; 150,5; 151,0; 149,7.

Необхідно визначити:

1. Середнє арифметичне.
2. Дисперсію вибірки.
3. Стандартне відхилення.
4. Коефіцієнт варіації (%).
5. Зробити висновок щодо мінливості ознаки.

Задача 3. Розрахунок довірчого інтервалу

За результатами польового досліді отримано: середня урожайність пшениці $\bar{x} = 6,25$ т/га; середня помилка $m = 0,18$ т/га; рівень значущості $p = 0,05$; $t = 1,96$.

Необхідно:

1. Обчислити довірчий інтервал середньої урожайності.
2. Записати результат у вигляді: $\bar{x} \pm tm$
3. Зробити висновок щодо меж, у яких із довірчою ймовірністю 95 % знаходиться генеральне середнє.

Задача 4. Відновлення втраченого результату досліді

У польовому досліді з трьома повтореннями вивчали вплив густоти стояння рослин на урожайність соняшнику.

Варіант	I повторення	II повторення	III повторення
Контроль	31,5	30,9	31,2
Варіант 2	33,1	X	32,8
Варіант 3	35,6	35,1	34,9
Варіант 4	36,2	35,8	36,0

Необхідно:

1. Визначити: L ; n ; V ; P ; ΣX .
2. За наведеною у посібнику формулою обчислити відсутнє значення X_{ls} .
3. Обчислити середню урожайність другого варіанта після відновлення результату.

Задача 5. Вибракування сумнівного результату (τ -критерій)

У досліді отримано такі значення маси 1000 насінин пшениці (г): 39,8; 40,2; 40,5; 40,8; 41,0; 41,2; 41,4; 41,7; 42,0; 46,5. Відомо: $n = 10$; $\tau_{0.05} = 0,477$; $\tau_{0.01} = 0,597$.

Необхідно:

1. Побудувати ранжований ряд.
2. Визначити: X_1 ; X_2 ; X_{n-1} ; X_n .
3. Обчислити: τ_1 та τ_2
4. Порівняти отримані значення з табличними.
5. Встановити, чи необхідно вибракувати крайні значення.
6. За потреби обчислити нове середнє арифметичне після вибракуння.

Тестові завдання

1. Сукупність, що містить усі можливі об'єкти дослідження, називається:
 - А. Вибірковою
 - Б. Генеральною
 - В. Альтернативною
 - Г. Дискретною
2. Середнє арифметичне характеризує:
 - А. Розмах варіації
 - Б. Частоту появи значень
 - В. Центральне положення вибірки
 - Г. Довірчий інтервал
3. Яка ознака є дискретною?
 - А. Урожайність, т/га
 - Б. Висота рослин
 - В. Кількість бобів на рослині
 - Г. Вміст білка, %
4. Який показник найчастіше характеризує мінливість даних?
 - А. Медіана
 - Б. Мода
 - В. Стандартне відхилення

Г. Квартиль

5. Медіана — це:

А. Найчастіше значення

Б. Найбільше значення

В. Значення, яке ділить впорядковану вибірку навпіл

Г. Середнє квадратичне

6. При якому рівні значущості результат зазвичай вважають статистично достовірним?

А. $p > 0,10$

Б. $p > 0,50$

В. $p \leq 0,05$

Г. $p = 1,00$

7. Який метод використовують для порівняння середніх двох незалежних вибірок?

А. Кореляційний аналіз

Б. Регресійний аналіз

В. t-критерій Стьюдента

Г. Коефіцієнт варіації

8. Ранжування даних означає:

А. Видалення сумнівних результатів

Б. Упорядкування даних за зростанням або спаданням

В. Побудову гістограми

Г. Обчислення дисперсії

9. Який показник виражають у відсотках?

А. Дисперсію

Б. Стандартне відхилення

В. Середню помилку

Г. Коефіцієнт варіації

10. Для визначення зв'язку між двома кількісними ознаками застосовують:

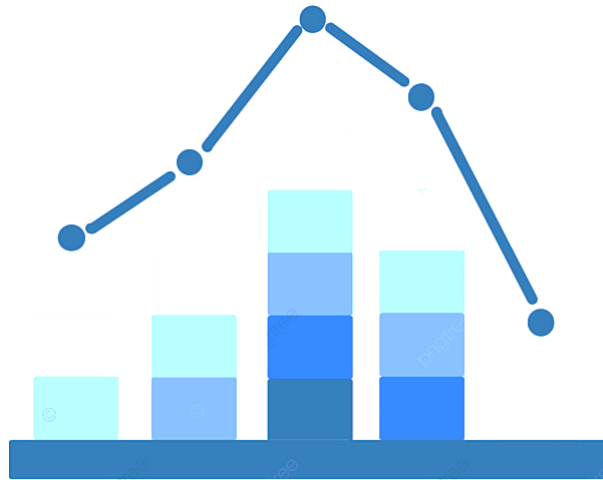
А. t-критерій Стьюдента

Б. χ^2 -критерій

В. Кореляційний аналіз

Г. Ранжування

РОЗДІЛ 2. СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ МІНЛИВОСТІ



2.1. Статистичні характеристики кількісної мінливості

2.2. Статистичні характеристики якісної мінливості

2.1. Статистичні характеристики кількісної мінливості.

При проведенні досліджень та аналізу отриманих результатів розрізняють мінливість кількісну та якісну.

Кількісно змінюватися можуть: елементи продуктивності, врожайність, вміст цукру, жиру, кислот, вітамінів, крохмалю або білка в урожаї, висота рослин, вміст поживних елементів у ґрунті, тобто все, що має масу, розмір, обсяг і т. п.

Розрізняють два види кількісної мінливості: безперервну та уривчасту або дискретну.

Безперервна мінливість — значення ознаки виражені мірами обсягу, довжини, маси і т.д..

Уривчаста або дискретна мінливість — відмінності між одиницями спостереження за ознаками рослин виражаються цілими числами, між якими немає і не може бути переходів, наприклад, кількість зерен у колосі пшениці і т. д.

Хід аналізу варіаційних рядів кількісної мінливості залежить від обсягу вибірки.

Малий обсяг вибірки становить менш 30 одиниць. Великий обсяг вибірки становить більше 30 одиниць.

Основними статистичними характеристиками кількісної мінливості при аналізу як малих, так і великих вибірок є:

$\bar{x}$ — середня арифметична; s^2 — дисперсія;

s — стандартне відхилення (середнє квадратичне відхилення);

s_x — помилка середньої арифметичної (помилка вибіркової середньої);

V — коефіцієнт варіації;

$s_{x\%}$ – відносна помилка вибіркової середньої; B – коефіцієнт вирівняності.

Наприкінці аналізу дають інтервальну оцінку середньої арифметичної. **Проста середньоарифметична величина** являє собою середнє значення ознаки, при визначенні якої загальний обсяг даної ознаки в сукупності даних залишається незмінним і порівну розподіляється між усіма одиницями, що входять до цієї сукупності.

Середня арифметична величина характеризується рядом властивостей:

✓ якщо до кожного значення вибірки додати або відібрати одну і ту ж величину, або помножити і розділити на ту саму величину, то середня арифметична збільшиться або зменшиться на цю величину;

✓ алгебраїчна сума відхилень окремих варіантів сукупності від середньої арифметичної цієї сукупності дорівнює нулю;

✓ сума квадратів відхилень варіант сукупності від середньої арифметичної x менше суми квадратів відхилень від будь-якої іншої величини.

Середня арифметична (x) є узагальненою характеристикою всієї сукупності взагалі. Середня арифметична – дорівнює відношенню суми індивідуальних значень ознаки до кількості ознак у сукупності і розраховується за формулою:

$$x = \Sigma (X_1 + X_2 + X_3 + \dots + X_n) / n,$$

де:

$\Sigma (X_1 + X_2 + X_3 + \dots + X_n)$ – сума кількісних величин ознаки рослин;

$X_1 + X_2 + X_3 + \dots + X_n$ – кількісна величина вираження кількісної ознаки рослини;

n – кількість рослин на яких вивчається кількісна ознака.

Наприклад, описуються і порівнюються два гібрида кукурудзи за ознакою маса зерен однієї рослини. Визначення ознаки виконується на 10 рослинах кожного гібрида (табл. 2.1).

Таблиця 2.1

Маса зерен однієї рослини гібридів кукурудзи

Гібрид	Маса зерен кукурудзи однієї рослини, г										Середня арифметична, г
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	
гібрид 1	121	129	127	122	128	121	130	124	125	127	125,4
гібрид 2	128	142	145	120	127	119	144	140	133	129	132,7

Виконав розрахунки, встановлюємо, що середня арифметична ознаки маса зерен однієї рослини у гібрида 1 становить 125,4 г, у гібрида 2 становить 132,7 г.

Розрахунок показав різні середні арифметичні у гібридів кукурудзи і дозволив виділити гібрид, що формує більшу середню масу зерен на одній

рослині.

За даними маси зерен однієї рослини можливо встановити у гібридів розмах варіації (R) за цією ознакою.

Розмах варіації (R) у гібрида 1 при масі зерна 125,4 дорівнюється: $R = X_{\max} - X_{\min} = X_7 - X_1 = 130 \text{ г} - 121 \text{ г} = 8 \text{ г}$.

Розмах варіації (R) у гібрида 2 при масі зерна 132,7 дорівнюється: $R = X_{\max} - X_{\min} = X_3 - X_6 = 145 \text{ г} - 119 \text{ г} = 26 \text{ г}$.

— Розрахування середньої арифметичної та розмаху варіювання дозволяє встановити більше варіювання ознаки у гібрида з більш високою масою число зерен однієї рослини.

Таким чином, середня арифметична є основною статистичною характеристикою варіаційного ряду, решта лише пояснюють її.

Дисперсія (s^2) і стандартне відхилення (s) є основними параметрами варіації, розсіяння величини кількісної ознаки, яка вивчається в досліді.

Дисперсія (s^2) – уявлює собою середній квадрат відхилень кожного члена варіаційного ряду ($X_1, X_2, X_3 \dots X_n$) від середньої арифметичної ($\bar{x}$). При проведенні повторних досліджень інших вибірок однієї й тієї ж сукупності розмах варіації кількісної ознаки може бути іншим. Це може свідчити, що розмах варіації кількісної ознаки не є характерним показником варіювання. Більш повно варіаційний ряд кількісних ознак характеризує

дисперсія. Дисперсію розраховують за формулою:

$$s^2 = \Sigma (X_1 - \bar{x})^2 + (X_2 - \bar{x})^2 + (X_3 - \bar{x})^2 + \dots + (X_n - \bar{x})^2 / n - 1,$$

де:

s^2 – дисперсія кількісної ознаки;

$\bar{x}$ – середньоарифметична величина кількісної ознаки;

$X_1, X_2, X_3 \dots X_n$ – показники кількісної ознаки рослин, які досліджуються в досліді.

n – кількість рослин на яких вивчається кількісна ознака, кількість усіх варіант.

Для розрахунку дисперсії гібридів кукурудзи за ознакою маса зерен рослини складаємо таблиці 4 і 5.

Вносимо отримані дані в формулу і розраховуємо дисперсію ознаки для гібрида 1:

$$s^2 = 98,4/9 = 10,9 \text{ г}.$$

Вносимо отримані дані в формулу і розраховуємо дисперсію ознаки для гібрида 2:

$$s^2 = 836,1/9 = 92,9 \text{ г}.$$

Дисперсії характеризують не тільки величину зміни варіаційних рядів, а й специфіку варіювання. Дисперсія ознаки гібрида кукурудзи 2 в 8,5 разів перевищує дисперсію тієї ж самої ознаки гібрида кукурудзи 1.

Стандартне відхилення (s). Чим більша дисперсія або стандартне відхилення, тим більше розсіяні біля середньої арифметичної індивідуальні значення ознаки, тобто спостерігається більша мінливість. Дисперсія постійної

величини дорівнює нулю. Недолік дисперсії у тому, що її розмірність відповідає квадратним одиницям зміни. Наприклад, якщо дані виміряні т/га, то дисперсія має розмірність (т²/га²). Щоб повернутися до вихідної шкали вимірювань, необхідно зробити операцію, яка зворотня возведенню у квадрат, тобто витягти квадратний корінь. Результатом є стандартне відхилення s .

Таблиця 2.2

**Розрахунок квадратів відхилень
за ознакою маса зерен рослини кукурудзи по гібриду 1**

Рослина	X	$\bar{x}$	$X - \bar{x}$	$(X - \bar{x})^2$
1	121	125,4	- 4,4	19,36
2	129	125,4	3,6	12,96
3	127	125,4	1,6	2,56
4	122	125,4	-3,4	11,56
5	128	125,4	2,6	6,76
6	121	125,4	-4,4	19,36
7	130	125,4	4,6	21,16
8	124	125,4	-1,4	1,96
9	125	125,4	-0,4	0,16
10	127	125,4	1,6	2,56
Σ				98,4

Таблиця 2.3

**Розрахунок квадратів відхилень
за ознакою маса зерен рослини кукурудзи по гібриду 2**

Рослина	X	$\bar{x}$	$X - \bar{x}$	$(X - \bar{x})^2$
1	128	132,7	-4,7	22,09
2	142	132,7	9,3	86,49
3	145	132,7	12,3	151,29
4	120	132,7	-12,7	161,29
5	127	132,7	-5,7	32,49
6	119	132,7	-13,7	187,69
7	144	132,7	11,3	127,69
8	140	132,7	7,3	53,29
9	133	132,7	0,3	0,09
10	129	132,7	-3,7	13,69
Σ				836,1

Показник стандартного відхилення являє собою квадратний корінь з дисперсії і обчислюється за формулою:

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

Для гібрида 1 стандартне відхилення за ознакою маса зерна рослини рівняється:

$$S = \sqrt{S^2} = \sqrt{10,9} = 3,3 \text{ г}$$

Для гібрида 2 стандартне відхилення за ознакою маса зерна рослини рівняється:

$$S = \sqrt{S^2} = \sqrt{92,9} = 9,63 \text{ г}$$

Як і дисперсія, стандартне відхилення виражається в тих же одиницях виміру, що і кількісна ознака. Чим більше варіює показник стандартного відхилення, тим більше числове значення стандартного відхилення. У розрахунках стандартне відхилення є більш зручною характеристикою, ніж дисперсія.

Стандартне чи нормоване відхилення використовують для розрахунків помилки вибіркової середнього та ступеня варіації.

Дисперсія і стандартне відхилення непридатні, коли у дослідях порівнюють мінливість ознак, що виражені різними одиницями виміру (маса врожаю, ц; кількість насіння, висота рослин, см; площа листків, см² та інш.). У цих випадках користуються коефіцієнтом варіації.

Коефіцієнт варіації (V) – стандартне відхилення, яке виражається у відсотках до середньої арифметичної цієї сукупності. Коефіцієнт варіації можна розраховувати лише тоді, коли всі вимірювані величини невід’ємні.

Коефіцієнт варіації є відносним показником мінливості. Мінливість прийнято вважати незначною, якщо коефіцієнт варіації вбирається у 10,0%. Якщо коефіцієнт варіації вище 10,0%, але менше 20,0% прийнято вважати, що це середня мінливість. Якщо коефіцієнт варіації більше 20,0% – це вже значна мінливість.

Коефіцієнт варіації розраховують за формулою:

$$V = (s / \bar{x}) \times 100,$$

де:

V – коефіцієнт варіації ознаки, %;

S – стандартне відхилення ознаки;

$\bar{x}$ – середньоарифметичне ознаки.

В нашому прикладі коефіцієнт варіації маси зерен рослини кукурудзи гібрида 1 становить:

$$V = (s / \bar{x}) \times 100 = (3,3 / 125,4) \times 100 = 2,63 \text{ \%}.$$

Коефіцієнт варіації ознаки маса зерен рослини 2,63 % у гібрида 1 – незначний, становить менше 10,0%.

Коефіцієнт варіації маси зерен рослини кукурудзи гібрида 2 дорівнюється:

$$V = (s / \bar{x}) \times 100 = (9,63 / 132,7) \times 100 = 7,25 \text{ \%}.$$

Коефіцієнт варіації ознаки маса зерен рослини 7,25 % у гібрида 2 більше, ніж у гібрида 1, але також незначний, так як становить менше 10,0%.

Відносна помилка вибіркової середньої ($S_x\%$) – це помилка вибірки, яка виражена у відсотках відповідної середньої.

Середні арифметичні помилки виникають в результаті неповного представництва вибірки. Ці помилки властиві лише вибіркового методу досліджень. Числове значення залежить як від ступеня мінливості ознаки, яка вивчається в досліді, так і від обсягу вибірки.

Середні арифметичні помилки пов'язані з перенесенням результатів, які отримані при вивченні зразків вибірки всієї генеральної сукупності.

При розрахунках відносної помилки вибіркової середньої на початку встановлюють помилку середньої арифметичної за формулою:

$$S_x = S\sqrt{n}$$

де:

$S_{\bar{x}}$ – помилка середньої арифметичної;

s – стандартне відхилення ознаки;

n – кількість рослин на яких вивчається кількісна ознака у вибірки, кількість усіх варіант.

Помилка середньої арифметичної у нашому прикладі за ознакою маса зерен однієї рослини кукурудзи для гібрида 1 буде рівнятися:

$$S_x = S\sqrt{n} = 3,3\sqrt{10} = 1,04 \text{ г}$$

Помилка середньої арифметичної за ознакою маса зерен однієї рослини кукурудзи для гібрида 2 буде рівнятися:

$$S_x = S\sqrt{n} = 9,63\sqrt{10} = 3,05 \text{ г}$$

Помилка середньої арифметичної вибірки виражається в тих же одиницях виміру як і кількісна ознака, що вивчається в досліді.

Після встановлення помилки середньої арифметичної розраховують відносну помилку вибіркової середньої:

$$S_{x\%} = \left(\frac{S_x}{x}\right) \times 100$$

де:

$S_{x\%}$ – відносна помилка вибіркової середньої, %;

S_x – помилка середньої арифметичної; x – середньоарифметичне ознаки.

Розраховуємо відносну помилку вибіркової середньої за ознакою маса зерна рослини кукурудзи гібрида 1:

$$S_{x\%} = (1,04/125,4) \times 100 = 0,83\%.$$

Розраховуємо відносну помилку вибіркової середньої за ознакою маса зерна рослини кукурудзи гібрида 2:

$$s_x\% = (3,05/132,7) \times 100 = 2,30\%.$$

Відносну помилку середньої іноді позначають літерою **P** і називають "точністю досліду" або "точністю аналізу". Чим менше відносна помилка, тим вища точність середньої арифметичної. Точність прийнято вважати високою при $S_x\% < 3,0\%$, середньою при значенні $S_x\%$ в межах $3,0...6,0\%$ і низькою при значенні $S_x\% > 6,0...7,0\%$.

Ступінь вирівняності матеріалу (B). Для характеристики ступеня вирівняності матеріалу іноді доцільно використовувати коефіцієнт вирівняності (B), який визначають за формулою:

$$B = 100 - V,$$

де:

B – коефіцієнт вирівняності, %; V – коефіцієнт варіації ознаки, %.

За ознакою маса зерен рослини кукурудзи гібрида 1 коефіцієнт вирівняності буде рівнятися:

$$B = 100 - V = 100 - 2,63 = 97,37\%.$$

За ознакою маса зерен рослини кукурудзи гібрида 2 коефіцієнт вирівняності буде рівнятися:

$$B = 100 - V = 100 - 7,25 = 92,75 \ %.$$

2.2. Статистичні характеристики якісної мінливості

Для якісної мінливості характерна проява обмеженої кількості градацій ознаки.

У селекційно-насінницьких експериментах часто доводиться мати справу з якісною мінливістю ознак: різна форма та забарвлення насіння та плодів, кількість пошкоджених та здорових рослин, кількість підмерзлих рослин, наявність чи відсутність ознаки (ості у пшениці), кількість зіпсованих і здорових бульб картоплі у сховищі, розщеплення гібридів тощо.

За якісної мінливості у вибірці є одна з двох можливостей (альтернатив) – це ознака яка є, або відсутня.

Таку мінливість називають ще альтернативною мінливістю.

У разі альтернативної мінливості фіксують лише два результати: відсутність чи наявність ознаки.

Угрупування результатів спостережень при якісному варіюванні зводиться до розподілу сукупності об'єктів на групи з різними якісними ознаками.

Підготовка таких даних до статистичної обробки заключається у їхній систематизації на «природні» класи.

У дослідях з якісною мінливістю замість вимірювання будь-якого показника підраховують кількість об'єктів з тією чи іншою ознакою. Для аналізу якісної мінливості обчислюють наступні статистичні характеристики:

- ✓ частку наявності ознаки p ;
- ✓ частку відсутності ознаки q ;
- ✓ показник мінливості якісної ознаки s ;
- ✓ коефіцієнт варіації V_p ;
- ✓ помилку вибіркової частки S_p .

Загальний обсяг вибірки позначають літерою N , а кількість варіант, об'єктів, яким власна певна ознака, позначають літерою n .

Частка ознаки (p) – це відносна чисельність (частота) варіанти, якісної ознаки даної сукупності рослин, яка вивчається в досліді.

Частка ознаки розраховується за формулою:

$$p = n/N,$$

де, p – частка ознаки, варіанти;

n – кількість рослин, яким наявна ознака, варіанта;

N – загальна чисельність сукупності рослин, що вивчається в досліді. Частка ознаки виражається або у частках одиниці, або у відсотках.

Розглянемо приклад розрахунків частки ознак.

У вибірках двох гібридів соняшнику гібрид 1 та гібрид 2 загальна чисельність сукупності складає по 100 рослин. Із 100 рослин гібрида 1 расою вочку F уражено 14 рослин. Із 100 рослин гібрида 2 расою вочку F уражено 5 рослин. Позначимо характеристики гібрида 1 p_1 , $n_1 = 14$, $N_1 = 100$. Характеристики гібрида 2 будуть відповідно позначені p_2 , $n_2 = 5$, $N_2 = 100$.

Встановлюємо частку ознаки ураження рослин гібрида 1 расою вочку F: $p_1 = n_1/N_1 = 14/100 = 0,14$ (14%).

Встановлюємо частку ознаки ураження рослин гібрида 2 расою вочку F: $p_2 = n_2/N_2 = 5/100 = 0,05$ (5%).

Частка відсутності ознаки (q) – це різниця між цілим, тобто одиницею, і часткою наявності ознаки. При альтернативній мінливості частка наявної ознаки позначається через p , а частка відсутності ознаки через q . Таким чином, частка відсутності ознаки встановлюють за формулою:

$$q = 1 - p.$$

Для гібрида соняшнику 1 частка відсутності ознаки (відсутність неуражених рослин расою вочку F) буде складати:

$$q = 1 - p = 1 - 0,14 = 0,86 \text{ (86\%)}$$

Для гібрида соняшнику 2 частка відсутності ознаки (відсутність неуражених рослин расою вочку F) буде складати:

$$q = 1 - p = 1 - 0,05 = 0,95 \text{ (95\%)}$$

Показник мінливості якісної ознаки (s). Показник мінливості якісної ознаки для альтернативної мінливості розраховується коли об'єкти, рослини, які

вивчаються в досліді, мають дві або більш градацій (як у прикладі з ураженням рослин гібридів соняшнику расою вовчку F – уражені і неурражені рослини).

Показник мінливості якісної ознаки розраховують за формулою:

$$s = \sqrt{p \times q}$$

де:

s – показник мінливості якісної ознаки;

p – частка ознаки, варіанти;

q – частка відсутності ознаки.

Показник мінливості якісної ознаки за ураженням рослин расою вовчку F для гібрида соняшнику 1 буде становити:

$$s = \sqrt{p \times q} = \sqrt{0.14 \times 0.86} = 0.35$$

Показник мінливості якісної ознаки расою вовчку F для гібрида соняшнику 2 буде становити:

$$s = \sqrt{p \times q} = \sqrt{0.05 \times 0.95} = 0.22$$

Залежно від співвідношення p та q значення s змінюється від 0 до 0,5.

Максимальна мінливість якісної ознаки ($s_{\max}$) спостерігається коли $p = q = 0,5$ та $s_{\max} = \sqrt{0,5 \times 0,5} = 0,5$ або 50%.

Якщо кількість градацій однієї якісної ознаки перевищує 2 ($k > 2$), то тоді відповідні частки градацій якісної ознаки позначають $p_1, p_2, p_3 \dots p_k$. Кожна з градацій частки якісної ознаки ($p_1, p_2, p_3 \dots p_k$) враховує частку ознак, об'єктів, варіант (ознак $n_1, n_2, n_3 \dots n_k$ стосовно загального кількісного обсягу (N).

В даному випадку, частка ознаки (p) – це відносна чисельність (частота) окремої кожної варіанти, якісної ознаки загальної сукупності однієї вибірки сукупності рослин, яка вивчається в досліді.

Частку кожної окремої якісної ознаки рослин, варіантів серед загальної чисельності рослин у вибірці розраховують за наступними формулами.

Частка рослин першої ознаки p_1 рівняється:

$$p_1 = n_1 / N.$$

Частка рослин другої ознаки p_2 рівняється:

$$p_2 = n_2 / N.$$

Частка рослин третьої ознаки p_3 рівняється:

$$p_3 = n_3 / N.$$

Частка k-ознаки p_k рівняється:

$$p_k = n_k / N.$$

Наприклад, в період створення вихідного матеріалу при розщепленні рослин соняшнику за ознакою забарвлення язичкових квітів суцвіття в загальній сукупності чисельністю 100 рослин встановлено: 35 рослин з жовтим забарвленням язичкових квітів (n_1), 42 рослини за лимонним забарвленням

язичкових квітів (n_2), 14 рослин з абрикосовим забарвленням язичкових квітів (n_3), 9 рослин з білим забарвленням язичкових квітів (n_4). В досліді спостерігаємо прояву 4-х ознак забарвлення язичкових квітів – 4 градації ознаки (k).

Розраховуємо частку окремої кожної ознаки забарвлення язичкових квітів суцвіття соняшнику від загальної кількості сукупності рослин у вибірці.

Частка рослин за ознакою жовте забарвлення язичкових квітів (p_1) рівняється:

$$p_1 = n_1 / N = 35/100 = 0,35.$$

Частка рослин за ознакою лимонне забарвлення язичкових квітів (p_2) рівняється:

$$p_2 = n_2 / N = 42/100 = 0,42.$$

Частка рослин за ознакою абрикосове забарвлення язичкових квітів (p_3) рівняється:

$$p_3 = n_3 / N = 14/100 = 0,14.$$

Частка рослин за ознакою абрикосове забарвлення язичкових квітів (p_4) рівняється:

$$p_4 = n_4 / N = 9/100 = 0,09.$$

Показник мінливості якісної ознаки забарвлення язичкових квітів суцвіття рослин соняшнику (s) характеризує варіювання величин ряду щодо один одного і визначається за формулою:

$$s = \sqrt[k]{p_1 \times p_2 \times p_3 \times p_4} = \sqrt[4]{0,35 \times 0,42 \times 0,14 \times 0,09} = 0,207$$

Значення максимальної мінливості для розподілів з різним числом градацій якісних ознак наступне:

Число градацій ознаки (k)	$S_{\max}$
2	0,5 (50%)
3	0,333 (33,3%)
4	0,25 (25%)
5	0,2 (20%)
6	0,167 (16,7%)
7	0,143 (14,3%)

Коефіцієнт варіації якісних ознак (V_p) – фактичний показник мінливості, який виражається у відсотках максимально можливої мінливості.

Коефіцієнт варіації якісних ознак показує співвідношення варіації ознаки, яка вивчається і спостерігається в досліді, і її максимально можливої варіації. Коефіцієнт може змінюватися від нуля до 100%, і чим він ближче до 100%, тим більша варіація (неоднорідність) загальної сукупності рослин за ознакою, яка вивчається.

Коефіцієнт варіації якісної ознаки розраховується за формулою:

$$V_p = (s/s_{\max}) \times 100,$$

де:

V_p – коефіцієнт варіації якісної ознаки; s – показник мінливості якісної ознаки;

$s_{\max}$ – показник максимальної можливої мінливості якісної ознаки, константа для встановленого числа градацій ознаки.

Наприклад, коефіцієнт варіації якісної ознаки ураження рослин соняшнику расою вовчку F для гібрида 1 становить:

$$V_p = (s/s_{\max}) \times 100 = (0,35/0,5) \times 100 = 70,0\%.$$

Так як в досліді вивчається дві градації ознаки, то тоді при двох градацій ознаки показник максимальної можливої мінливості якісної ознаки рівняється 0,5 (50%).

Коефіцієнт варіації якісної ознаки ураження рослин соняшнику расою вовчку F для гібрида 2 становить:

$$V_p = (s/s_{\max}) \times 100 = (0,22/0,5) \times 100 = 44,0\%.$$

Оскільки величина коефіцієнта варіації якісної ознаки ураження рослин соняшнику расою вовчку F для гібрида 1 більше і становить 70,0%, то мінливість (варіація) рівня ураження рослин расою вовчку F вище саме у цього гібрида. Рослини гібрида соняшнику 2 характеризуються меншим рівнем ураження расою вовчку F і мають значно менший величина коефіцієнта варіації якісної ознаки 44,0%.

Розрахуємо коефіцієнт варіації якісної ознаки за забарвленням язичкових квітів суцвіття соняшнику в процесі створення вихідного матеріалу. В нашому прикладі при розщепленні спостерігалось 4 градації цієї ознаки при показник мінливості якісної ознаки (s) 0,207. Числу градацій ознаки ($k=4$) відповідає показник максимальної можливої мінливості якісної ознаки $s_{\max} = 0,25$. На основі отриманих даних встановлюємо за формулою коефіцієнт варіації якісної ознаки за забарвленням язичкових квітів:

$$V_p = (s/s_{\max}) \times 100 = (0,207/0,25) \times 100 = 82,8\%.$$

Помилка вибіркової частки (s_p). Для оцінки генеральної частки, частки ознаки у всій сукупності, необхідно встановити помилку вибірки помилку частки ознаки). Помилка вибіркової частки – це міра відхилення від частки наявності ознаки. Помилка вибіркової частки обчислюється за формулою:

$$s_p = \sqrt{(p \times q)/100},$$

де:

s_p – помилка вибіркової частки;

p – частка ознаки, варіанти;
 q – частка відсутності ознаки;
 N – загальна (генеральна) чисельність сукупності рослин, що вивчається в досліді, обсяг вибірки.

За якісною ознакою ураження рослин соняшнику расою вовчку F помилка вибіркової частки для гібрида 1 становить:

$$s_p = \sqrt{(p \times q)/100} = \sqrt{(0,14 \times 0,86)/100} = 0,034.$$

За якісною ознакою ураження рослин соняшнику расою вовчку F помилка вибіркової частки для гібрида 2 становить:

$$s_p = \sqrt{(p \times q)/100} = \sqrt{(0,05 \times 0,95)/100} = 0,022.$$

На рівні довірчої ймовірності $P_{0,95}$ інтервальну оцінку частки дають за формулою $p \pm 2s_p$.

На рівні довірчої ймовірності $P_{0,99}$ інтервальну оцінку частки дають за формулою $p \pm 3s_p$.

На рівні довірчої ймовірності $P_{0,95}$ інтервальна оцінка частки для гібрида соняшнику 1 за якісною ознакою ураження рослин расою вовчку F буде складати:

$$p \pm 2s_p = 0,14 - 2 \times 0,034 = 0,14 - 0,068 = 0,072$$

$$p \pm 2s_p = 0,14 + 2 \times 0,034 = 0,14 + 0,068 = 0,280.$$

Таким чином, для гібрида 1 нижня межа інтервалу становить 0,072 частки, верхня межа інтервалу становить 0,280 частки.

Частка уражених рослин соняшнику гібрида 1 расою вовчку F всієї дослідної ділянки (генеральна частка) перебуватиме в інтервалі 0,072...0,280.

Ураженість рослин соняшнику гібрида 1 расою вовчку F всієї дослідної ділянки буде знаходитися в межах 7,2 % (0,027)... 28,0% (0,280).

Статистична достовірність цього твердження дорівнює 95% ймовірності.

У межах подвійної помилки вибіркової частки всі значення частки укладаються з ймовірністю $P_{0,95}$ У межах потрійної помилки вибіркової частки всі значення частки укладаються з ймовірністю $P_{0,99}$.

Якщо градацій однієї кількісної ознаки більше двох, в такому разі помилка вибіркової частки – це міра відхилення від частки наявності ознаки від частки її по всій генеральній сукупності внаслідок неповної репрезентативності вибірки і помилку вибіркової частки обчислюють за формулою:

$$S_p = \frac{s}{\sqrt{N}}$$

де:

s_p – помилка вибіркової частки;

s – показник мінливості якісної ознаки;

N – загальна (генеральна) чисельність сукупності рослин, що вивчається в досліді, обсяг вибірки.

Помилка вибіркової частки якісної ознаки забарвлення язичкових квітів суцвіття рослин соняшнику в нашому прикладі буде дорівнюватися відповідно формули:

$$S_p = \frac{s}{\sqrt{N}} = \frac{0,207}{\sqrt{100}} = 0,021$$

Контрольні питання

1. Що таке кількісна мінливість?
2. Які види кількісної мінливості розрізняють?
3. Що характеризує середня арифметична?
4. Як визначають розмах варіації?
5. Яке біологічне значення має дисперсія?
6. Чим стандартне відхилення відрізняється від дисперсії?
7. Для чого використовують коефіцієнт варіації?
8. Як оцінюють ступінь мінливості за величиною коефіцієнта варіації?
9. Що характеризує помилка середньої арифметичної?
10. Що таке відносна помилка середньої та як оцінюють точність дослідів?
11. Що таке коефіцієнт вирівняності та як його визначають?
12. Що називають альтернативною мінливістю?
13. Як визначають частку ознаки (p) та частку її відсутності (q)?
14. Які статистичні показники використовують для аналізу якісної мінливості?

Розрахункові задачі

Задача 1. Розрахунок статистичних характеристик кількісної ознаки.

Під час дослідження маси зерна однієї рослини кукурудзи отримано такі результати (г): 125; 129; 131; 124; 128; 126; 127; 130; 123; 129

Необхідно:

1. Обчислити середню арифметичну ($\bar{X}$).
2. Визначити розмах варіації (R).
3. Обчислити дисперсію (s^2).
4. Обчислити стандартне відхилення (s).
5. Зробити висновок щодо мінливості ознаки.

Задача 2. Оцінка мінливості та точності вибірки.

У досліді визначено висоту рослин соняшнику (см): 165; 168; 170; 169; 167; 171; 166; 168; 170; 169

Необхідно:

1. Визначити середню арифметичну.
2. Обчислити стандартне відхилення.
3. Розрахувати коефіцієнт варіації ($V, \%$).
4. Визначити помилку середньої арифметичної (s_x).
5. Розрахувати відносну помилку середньої ($s_{x\%}$).
6. Оцінити точність дослід.

Задача 3. Визначення коефіцієнта вирівняності.

У результаті статистичної обробки встановлено: середня врожайність пшениці – 6,42 т/га; стандартне відхилення — 0,51 т/га.

Необхідно:

1. Обчислити коефіцієнт варіації.
2. Визначити коефіцієнт вирівняності.
3. Охарактеризувати ступінь вирівняності експериментального матеріалу.

Задача 4. Статистичні характеристики альтернативної мінливості.

Під час оцінки стійкості гібрида соняшнику до вовчка обстежено 150 рослин. Уражено 24 рослини.

Необхідно визначити:

1. Частку уражених рослин (p).
2. Частку неуражених рослин (q).
3. Показник мінливості якісної ознаки (s).
4. Коефіцієнт варіації якісної ознаки (V_p).
5. Помилку вибіркової частки (sp).
6. Побудувати довірчий інтервал частки за рівня довіри 95%.

Задача 5. Аналіз багатоградаційної якісної ознаки.

У популяції соняшнику ($N = 200$ рослин) встановлено таке забарвлення язичкових квіток: жовте – 82 рослини; лимонне – 61 рослина; абрикосове – 37 рослин; біле – 20 рослин.

Необхідно:

1. Визначити частку кожної градації.
2. Обчислити показник мінливості якісної ознаки (s).
3. Визначити максимально можливу мінливість ($s_{\max}$).
4. Розрахувати коефіцієнт варіації якісної ознаки (V_p).
5. Зробити висновок про ступінь різноманітності популяції.

Тестові завдання

1. Який показник характеризує центральне значення варіаційного ряду?

А. Дисперсія

Б. Середня арифметична

- В. Коефіцієнт варіації
- Г. Стандартне відхилення

2. Розмах варіації визначають як:

- А. Добуток максимального і мінімального значень
- Б. Суму всіх значень
- В. Різницю між максимальним і мінімальним значеннями
- Г. Відношення середнього до максимального

3. Який показник характеризує розсіювання даних відносно середньої?

- А. Медіана
- Б. Стандартне відхилення
- В. Мода
- Г. Квартиль

4. Коефіцієнт варіації виражають:

- А. У грамах
- Б. У сантиметрах
- В. У відсотках
- Г. У квадратних одиницях

5. Мінливість ознаки вважають незначною, якщо коефіцієнт варіації становить:

- А. Менше 10 %
- Б. 10–20 %
- В. 20–30 %
- Г. Понад 30 %

6. Помилка середньої арифметичної залежить від:

- А. Лише кількості спостережень
- Б. Лише середнього значення
- В. Стандартного відхилення та обсягу вибірки
- Г. Максимального значення ознаки

7. Коефіцієнт вирівняності визначають за формулою:

- А. $V = V \times 100$
- Б. $V = 100 - V$
- В. $V = s/x$
- Г. $V = x/s$

8. Альтернативна мінливість характеризується:

- А. Необмеженою кількістю градацій
- Б. Двома можливими станами ознаки
- В. Лише кількісними показниками
- Г. Нормальним розподілом ознаки

9. Частку наявності ознаки визначають за формулою:

А. $p = N/n$

Б. $p = n/N$

В. $p = n \times N$

Г. $p = N - n$

10. Для двох градацій якісної ознаки максимально можливий показник мінливості ($s_{\max}$) дорівнює:

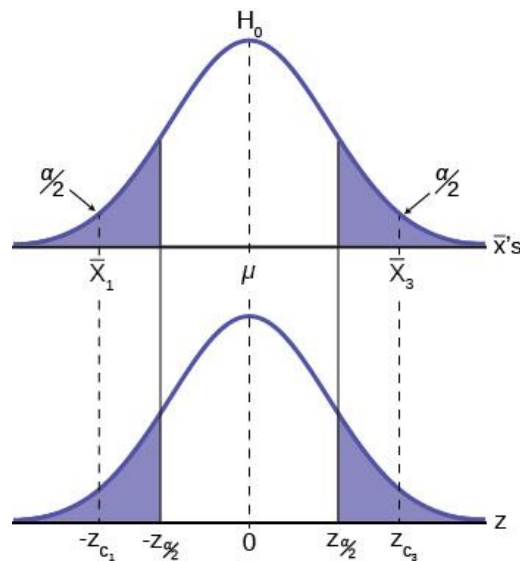
А. 0,25

Б. 0,33

В. 0,50

Г. 1,00

РОЗДІЛ 3. СТАТИСТИЧНІ ГІПОТЕЗИ І ТЕСТИ



- 3.1. Нульова та альтернативна гіпотези
- 3.2. Алгоритм статистичного теста
- 3.3. Вибір способу статистичної обробки даних досліджу
- 3.4. Оцінка суттєвих відмінностей між значеннями вибірок досліджу

3.1. Нульова та альтернативна гіпотези

Після проведення експерименту дослідник повинен встановити, чи є виявлені відмінності між дослідними варіантами випадковими, чи вони зумовлені дією досліджуваного фактора. Для цього застосовують методи перевірки статистичних гіпотез.

Статистична гіпотеза – це науково обґрунтоване припущення щодо параметрів генеральної сукупності або характеру взаємозв'язку між досліджуваними ознаками, яке перевіряється за допомогою статистичних методів. Інакше кажучи статистичні гіпотези це певні припущення про закономірності та параметри розподілу випадкової змінної в сукупності.

У біологічних і агрономічних дослідженнях статистичні гіпотези використовують для оцінювання ефективності добрив, біопрепаратів, сортів, технологій вирощування культур, засобів захисту рослин та інших факторів.

Перевірка статистичних гіпотез є наріжним каменем статистичного висновку, забезпечуючи формальні механізми для оцінки доказів і прийняття рішень на основі даних. Цей процес дозволяє дослідникам визначити, чи є спостережувані дані узгодженими із запропонованою моделлю, чи виявлені відхилення є статистично значущими. Мета статистичних гіпотез – перевести у професійну (агрономічну) постановку проблеми у статистичне русло.

Відмінності середніх за вибірками досліджу можуть бути як істинними, так і випадковими. Звідси дослідник повинен виконати перевірки статистичної

достовірності, чи суттєвості, відмінності середніх за вибірками. Порівнянню будь-яких двох вибірок перед усім є припущення, що вони взяті з однієї або однакових сукупностей, а відмінності, які спостерігаються у досліді, становлять випадкові відхилення, тобто різниця між середніми повинна бути близька до нуля. Звідси походить назва цього припущення – "нульова гіпотеза" (H_0), яка зазвичай представляє базове припущення про відсутність ефекту, різниці або зв'язку. Перевірка H_0 , або статистичний тест, може підтвердити нульову гіпотезу або відкинути на користь альтернативної гіпотези (H_A), яка відображає твердження про наявність ефекту або зв'язку, яке дослідник прагне підтвердити.

Нульову гіпотезу символізує наступна запис:

$$H_0: d = 0 (\mu_1 = \mu_2) \text{ або } y_1 \approx y_2$$

Альтернативну гіпотезу символізує наступний запис:

$$H_A: d \neq 0$$

Альтернативною, або конкуруючою гіпотезою, може бути двостороння робоча гіпотеза: $H_1: y_1 \neq y_2$ і дві односторонні гіпотези: $H_1: y_1 > y_2$ або $y_1 < y_2$.

Наприклад, при порівнянні врожайності сільськогосподарської культури у контрольному варіанті без внесення добрив та у варіанті з внесенням дози мінеральних добрив практично немає сумнівів, що робоча гіпотеза буде односторонньою: при внесенні дози мінеральних добрив врожайність підвищиться.

При порівнянні двох нових сортів або гібридів гіпотеза буде двосторонньою, оскільки ефективність вирощування обох сортів (гібридів) при їх сортовивченні невідома.

В якості нульової гіпотези (H) можна взяти також відомий факт чи конкретно сформульоване припущення. Наприклад, встановлену польову схожість насіння люцерни махрової в 70% передбачається збільшити на 2,0% за допомогою обробки насіння новим препаратом. В цьому випадку за "нульова гіпотеза" (H_0) приймають $p = 0,7$ а за альтернативну гіпотезу (робочу) H_A приймають гіпотезу: $p > 0,7$.

Інший приклад: гібрид соняшнику (стандарт при сортовивченні) формує врожайність умовах Степу на рівні 28,1 ц/га, а новий гібрид, що передається на державне сортовипробування формує врожайність на рівні 32,7 ц/га. Нульова гіпотеза H_0 записується як $y_0 = 28,1$; альтернативна робоча гіпотеза H_A записується як $y_1 = 32,7$ або $y_1 > 28,1$.

3.2. Алгоритм статистичного теста

Під статистичним тестом розуміють порівняння отриманих даних двох вибірок (варіантів) досліді з урахуванням статистичних показників. Алгоритм статистичного тесту включає чотири послідовні етапи.

Перший етап.

Трансформація агрономічної проблеми на статистичне русло шляхом висування нульової та робочих гіпотез. Наприклад, два сорти озимої пшениці розрізняються за врожайністю: $H_0: (\mu_1 = \mu_2)$; $H_1: (\mu_1 \neq \mu_2)$.

Другий етап.

Вибір методу перевірки H_0 і розрахунок фактичного критерію, наприклад: $t = y_1 - y_2 / S_d$.

Третій етап.

Зіставлення розрахункового критерію з контрольним (табличним), наприклад: $t < t_{0,01}$ та $t > t_{0,05}$.

Четвертий етап. Висновок.

У своєму висновку дослідник може зробити помилку I і II роду, коли відповідно відкидає вірну H_0 і приймає невірну H_0 : знаходить випадкові відмінності суттєвими і, навпаки, суттєві – випадковими.

Помилка I роду (α): Відхилення істинної нульової гіпотези («хибнопозитивний результат»). Імовірність такої помилки позначають символом α . Найчастіше використовують: $\alpha = 0,05$; $\alpha = 0,01$; $\alpha = 0,001$.

Помилка II роду (β): Невідхилення хибної нульової гіпотези («хибнонегативний результат»).

Потужність тесту ($1-\beta$): Ймовірність правильно відхилити нульову гіпотезу, коли вона дійсно хибна. Збільшення розміру вибірки часто підвищує потужність тесту. Чим більша потужність тесту, тим вища ймовірність виявити реальний ефект (рис. 3.1).

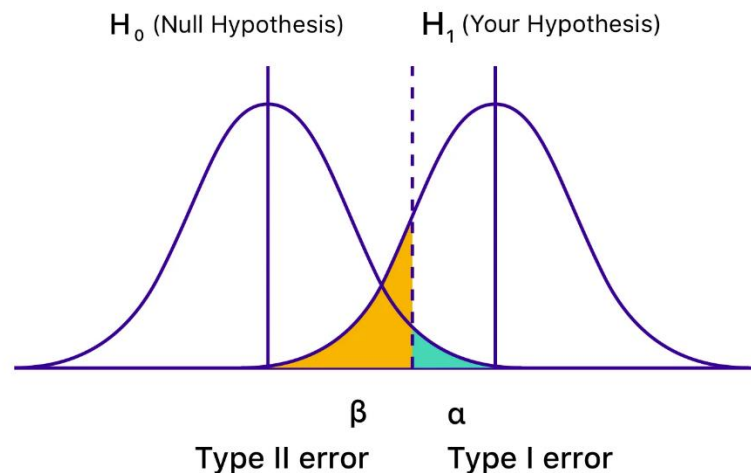


Рис. 3.1. Визначення помилки I та II типу.

Крім цих двох помилкових висновків існує і два правильні висновки, тобто при перевірці нульової гіпотези можливі чотири варіанти результату.

A. Вибірки взяті з однієї (однакової) сукупності. Відмінності між середніми випадкові:

- 1) прийнята правильна $H_0: y_1 \approx y_2$, тобто відмінності між середніми

значеннями не суттєві, або близькі до нуля;

- 2) відкинута H_0 і прийнята неправильна H_A : $y_1 \neq y_2$, тобто відмінності між вибірковими середніми значеннями є суттєві. Такий висновок класифікується помилкою *I* роду (α). Ймовірність помилки дорівнює прийнятого рівня значущості $\alpha = 100 - P$ (P – рівень довіри, чи надійності). В агрономічних дослідженнях найчастіше використовують рівень значущості $\alpha = 0,05$. Якщо $p < 0.05$, нульову гіпотезу відхиляють. Якщо $p > 0.05$, підстав для відхилення H_0 немає.

Б. Вибірки взяті з різних сукупностей та середні значення повинні розрізнятися:

- 1) відкинута H_0 і прийнята правильна H_A : $y_1 \neq y_2$, тобто відмінності між вибірками суттєві;
- 2) не встановлено суттєвих відмінностей та прийнята неправильна H_0 , тобто скоєно помилку II рода (β).

Ймовірність прийняття правильної H_A : $(1 - \beta)$ називають **робастністю** – роздільною здатністю тесту (експерименту).

У зв'язку з тим, що помилку II роду важко встановити, її практично не враховують у науковому дослідженні. Тому ймовірність помилки I роду (α) називають просто ймовірністю помилки. Прийняття H_0 не виключає наявності суттєвих відмінностей разом, а визначає лише відхилення H_A у цьому дослідженні.

У науковій практиці зустрічається «помилка III роду», коли для неправильно сформульованої проблеми робиться правильний статистичний висновок (відкидається або приймається нульова гіпотеза). Ця помилка обумовлена неправильною біологічною моделлю експерименту та професійною некомпетентністю дослідника. Спочатку проблема (тема) була неправильною або недостатньо повно сформульована. Наприклад, за темою дослідження «Вплив густоти стояння рослин на врожайність гібридів соняшнику» у перший рік досліджень була сформульована робоча гіпотеза (H_1): «при загущенні агроценозу гібридів соняшнику знижується врожайність гібридів соняшнику та вміст жиру в насінні». Дослідником зроблено правильний статистичний висновок (прийнято суттєвій відмінності не врожайності гібридів соняшнику, а й вмісту жиру в насінні за всіма варіантами досліду).

Але з погляду проведення дослідів в рослинництві цей висновок дослідника є помилкою III роду, оскільки не були враховано погодні умови періоду вегетації. Період вегетації характеризувався зниженням кількості випадіння опадів у порівнянні зі багаторічними показниками, сильною повітряною та ґрунтовою посухами, які негативно вплинули на ріст та розвиток рослин соняшнику, призвели до зниження у загущених агроценозах рівня врожайності та вмісту жиру в насінні.

3.3. Вибір способу статистичної обробки даних дослідів

Перед проведення статистичного аналізу отриманих експериментальних

даних досліджу науковий співробітник повинен здійснити *вибір способу статистичної обробки даних досліджу*.

Якщо дані щодо кожної експериментальної одиниці досліджу по дослідній ділянці не здаються сумнівними, дослідники обчислюють середні арифметичні варіанти і проводять агрономічний аналіз усього досвіду. Якщо дані досліджу не викликають сумнівів, результати досліджу піддають відповідного аналізу, наприклад, дисперсійного аналізу.

Метод аналізу результатів польових дослідів залежить від способу розміщення варіантів. При розміщенні методом рандомізованих повторень використовують дисперсійний аналіз рандомізованих повторень, при повній рандомізації – дисперсійний аналіз повної рандомізації, при розміщенні латинським квадратом – дисперсійний аналіз латинського квадрата і т.д.

Багатофакторні досліді можна розміщувати методами рандомізованих повторень, розщеплених ділянок, змішування. Результати піддають відповідним дисперсійним аналізам – рандомізованих повторень, розщеплених ділянок, змішування.

Дані вегетаційних дослідів, а також дослідів за технологією зберігання продукції обробляють тими самими методами, що й дані польових дослідів, розміщених методом повної рандомізації. Результати дослідів, у яких варіанти розміщені стандартними методами, обробляють різницевою методом, призначеним для пов'язаних вибірок.

Якщо варіанти в досліді розміщені систематично (що взагалі не рекомендується при проведенні дослідів), результати обробляють дробовим методом. При обробці показників якісної мінливості достовірність різниці визначають за критерієм Стюдента.

Відповідність між спостережуваними і теоретичними розподілами в генетичних дослідженнях оцінюють за критерієм χ^2 . Залежності між різними показниками рослин, між рослинами та їх середовищем визначають за допомогою кореляційних та регресійних аналізів. Для вивчення кореляції якісних ознак використовують формулу Юлла.

Кінцева мета більшості наукових досліджень полягає в знаходженні зв'язків (залежностей) між факторами (змінними) досліджу.

Будь-який закон природи може бути виражений, в кінцевому рахунку, у вигляді описання характеру або структури взаємозв'язків (залежностей) між змінними величинами, які вивчаються дослідником. Залежно від типу залежних та незалежних змінних розрізняють методи статистичного аналізу залежностей (табл. 3.1).

Таблиця 3.1

**Застосування основних методів статистичного аналізу
в залежності від типу вихідних змінних**

Мета статистичного аналізу		Типи змінних		Методи статистичного аналізу	
		залежна	незалежна	параметричні (розподіл ознак наближається до нормального)	непараметричні (розподіл ознак суттєво відрізняється від нормального або невідомо)
оцінка відмінностей між двома вибірками за значенням ознаки	вибірки незалежні	кількісна	номінальна	t -критерій для незалежних вбірок	критерій серій Вальда-Вольфовиця, U-критерій Манна- Уїтні, двовибірковий критерій Колмогорова- Смирнова
		порядкова	Номінальна	-	
		номінальна	Номінальна	-	критерій χ^2 (хі- квадрат) Макне- мара
	вибірки залежні	кількісна	номінальна	t -критерій для залежних вибірок	критерій знаків, критерій Уїлкоксона
		порядкова	номінальна	-	
оцінка відмінностей між декількома вибірками за значенням ознаки	вибірки незалежні	кількісна	номінальна	дисперсійний аналіз	аналіз Краскела- Уоллеса, медіанний тест
		порядкова	номінальна	-	
	вибірки залежні	кількісна	номінальна	дисперсійний аналіз з повторними вимірами	ранговий дисперсійний аналіз Фрідмана
		порядкова	номінальна	-	
аналіз щільності зв'язку між змінними		кількісна	кількісна	кореляційний аналіз (коефіцієнт кореляції Пірсона)	-
		порядкова	порядкова	-	аналіз рангових кореляцій (коефекти кореляції Спірмена та Кендалла)

Статистичний критерій (Statistical Test) – це статистичний інструмент (набір формул та числових показників), за яким вираховується рівень значущості (p-level). Статистичний критерій – метод розрахунку певного числа, яке позначають як емпіричне значення критерію. Статистичні критерії поділяються на параметричні і непараметричні.

Параметричні критерії включають у формулу розрахунку середнє арифметичне і дисперсії і застосовуються при аналізі метричних даних, що вписуються в криву нормального розподілу.

При роботі з непараметричними критеріями оперують частотами і рангами. Непараметричний критерій рекомендується використати також для аналізу метричних даних, розподіл яких значно відрізняється від нормального.

Вибір статистичного критерію залежить від типу даних, їх розподілу та кількості груп, що порівнюються. Важливо також враховувати, чи є вибірки незалежними чи залежними. Якщо розподіл невідомий або вибірка менше 30 спостережень – краще обирати непараметричні критерії.

Нормальний розподіл (розподіл Гауса) має дзвоноподібну форму і характеризується тим, що крайні значення ознаки в нім зустрічаються досить рідко, а значення, близькі до середньої величини, - часто. Отож, найчастіше зустрічаються середні показники, тоді як чим більше показник відхиляється від середнього значення, тим рідше він зустрічається (при чому відхилення від середнього у більшу чи меншу сторони зустрічаються однаково рідше, порівняно з серединою). Доведено, що нормальному закону підпорядковуються більшість агробіологічних показників.

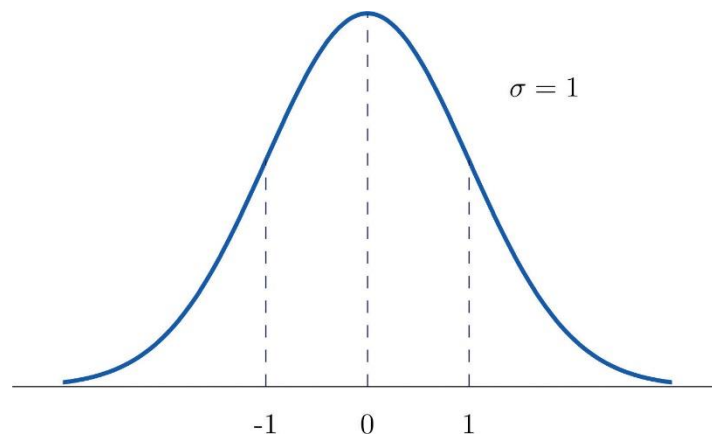


Рис. 3.2. Графічне зображення стандартного нормального розподілу.

Прикладами нормального розподілу значень в рослинництві є висота рослин у посіві на однорідній ділянці; маса 1000 зерен для одного сорту за однакових умов вирощування; довжина колоса пшениці у межах одного сорту та однорідного агрофону; діаметр кошика соняшнику за достатньої кількості вимірювань (30-50 рослин); вміст білка або олії для зерна одного сорту, вирощеного в однорідних умовах; урожайність ділянок у повторностях, якщо дослід закладений правильно (рандомізація, однорідний ґрунт, однакова агротехніка).

Виміряна висота рослин гороху (см) становила: 58, 60, 61, 61, 62, 62, 63, 63, 64, 64, 65, 66, 67, 68. Середнє $\approx 63,4$ см. Більшість значень зосереджена біля 62-64 см, а крайні значення (58 і 68 см) трапляються рідко - це типовий вигляд нормального розподілу (рис. 3.3).

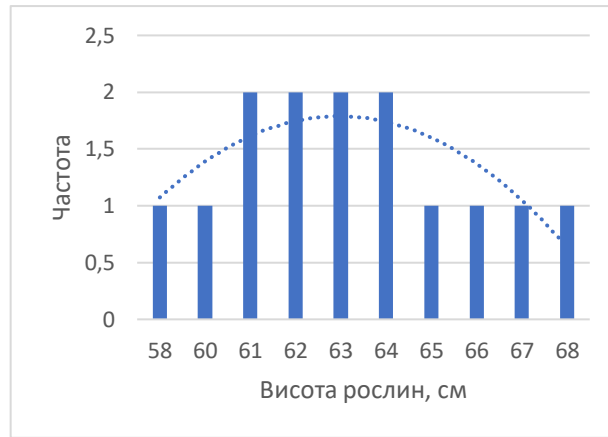


Рис. 3.3. Частотна гістограма розподілу висоти рослин гороху у виборці.

3.4. Оцінка суттєвих відмінностей між значеннями вибірок дослідів

Дані проведеного експерименту, який відповідає конкретній статистичній моделі, або принаймні має повторність дослідів підлягають математичній обробці.

Найпростішими та найбільш ефективними методами математичної обробки слід вважати парні порівняння з методу точечної оцінки шляхом зіставлення фактично розрахованих критеріїв із контрольними: критерій Ст'юдента (t), критерій Фішера (F), критерій хі-квадрат (χ^2) (табл. 3.2).

Таблиця 3.2

Найпоширеніші статистичні тести в агрономії

Статистичний тест	Призначення
t-критерій Ст'юдента	Порівняння двох середніх
Парний t-тест	Порівняння двох залежних вибірок
Однофакторний ANOVA	Порівняння трьох і більше середніх
Двофакторний ANOVA	Аналіз впливу двох факторів
Кореляція Пірсона	Оцінка лінійного зв'язку між кількісними ознаками
Кореляція Спірмена	Оцінка монотонного зв'язку між ранговими даними
χ^2 -критерій Пірсона	Аналіз якісних ознак
Критерій Манна-Уїтні	Порівняння двох незалежних вибірок без нормального розподілу
Критерій Вілкоксона	Аналіз двох залежних вибірок без нормального розподілу

Точечна оцінка включає критерій НСР – *найменшу суттєву різницю* між середніми значеннями (або межа випадкових відхилень). Фактичну різницю між середніми значеннями дослідів зіставляють з НСР. Формула НСР представляє добуток табличного значення t на помилку різниці, наприклад для 95% рівня

ймовірності:

$$HCP_{0,05} = t_{0,05} \times S_d, \text{ де } S_d = \sqrt{\frac{2S^2}{n}}$$

Формула показує, що $HCP_{0,05}$ рівноцінна потрійній помилці експерименту: $HCP_{0,05} = 2 \times 1,42 S_y = 3S_y$.

Приблизною оцінкою істотності відмінностей двох вибірок служить спосіб зіставлення довірчих інтервалів для генеральної середньої (частки). Якщо інтервали не перекриваються, то відмінності суттєві і, навпаки, відмінності оцінюються випадковими, коли інтервали перекриваються.

Нюанси статистичного тесту обумовлені такими причинами:

- ✓ однакові чи неоднакові обсяги вибірок та їх дисперсії;
- ✓ вибірки незалежні або залежні, тобто пов'язані об'єкти вибірок загальними умовами – місцем (установка, фітотрон, ділянка) або об'єктом (рослини або листя верхнього чи нижнього ярусу) випробування варіанту.

Прикладом незалежної вибірки може бути виконане порівняння засміченості двох ділянок: обробленої гербіцидом та контрольної ділянки, де гербіцид не застосовували.

Прикладом залежної вибірки може бути виконання у досліді порівняння засміченості на одній ділянці – до обробки гербіцидом і через 2...3 тижні після обробки на тій же самій ділянці.

Залежними вважають вибірки з однієї сукупності, коли об'єкти вибірки (елементи) піддаються різним випробуванням (двом або великим) у часовому чи просторовому інтервалі.

Облік засміченості посівів, уражених рослин хворобами і шкідниками до обробки пестицидами і після, повторні укоси трав є ідеальними прикладами залежних вибірок. Спостереження за 1-ою та 2-ою вибірками проводять парами.

Для порівняння середніх значень використовують критерій Ст'юдента t .

Критерій Ст'юдента (Student's t-test) - один із найпоширеніших статистичних методів, який застосовують для перевірки статистичної значущості відмінностей між середніми значеннями двох вибірок. Метод був розроблений англійським статистиком Вільямом Сілі Госсетом (William Sealy Gosset), який публікував свої роботи під псевдонімом Student, звідки і походить назва критерію. У сільськогосподарських дослідженнях t -критерій використовують для оцінки впливу різних факторів на кількісні показники рослин, урожайність, біометричні характеристики, вміст поживних речовин, ефективність добрив, біопрепаратів тощо. Критерій Ст'юдента дозволяє встановити, чи є різниця між середніми арифметичними двох вибірок наслідком дії досліджуваного фактора або вона виникла випадково внаслідок природної мінливості.

Для коректного використання критерію Ст'юдента повинні виконуватися такі умови: досліджувана ознака є кількісною; вибірки сформовані випадковим способом; спостереження незалежні; розподіл даних наближений до нормального; дисперсії вибірок приблизно однакові (для класичного t -критерію).

Якщо ці умови не виконуються, рекомендується застосовувати непараметричні критерії (Манна-Уїтні, Вілкоксона тощо).

Залежно від типу експерименту використовують три різновиди критерію.

1. *Одновибірковий t-критерій*: застосовується для порівняння середнього значення вибірки з відомим або теоретичним значенням генеральної сукупності. Наприклад, для порівняння фактичної урожайності з нормативною, для перевірки відповідності середньої маси насіння стандарту.

2. *t-критерій для незалежних вибірок*: використовується найчастіше у польових дослідках і дозволяє порівняти середні двох незалежних груп. Наприклад, дозволяє порівняти контроль і дослід; два сорти; два гібриди; дві технології вирощування.

3. *Парний t-критерій*: використовують тоді, коли однакові об'єкти вимірюються двічі. Наприклад: до та після внесення препарату; урожайність однієї ділянки до і після меліорації; вміст хлорофілу до та після стресового впливу.

Якщо дисперсії двох вибірок вважаються однаковими і відомі середні помилки середніх (m_1 і m_2), що часто трапляється в агрономічних публікаціях, використовують формулу:

$$t = \frac{(y_1 - y_2)}{\sqrt{m_1^2 + m_2^2}}$$

де:

$y_1 - y_2$ – різниця між середніми значеннями вибірок;

$S_d (m_1^2 + m_2^2)$ – помилка середніх значень.

Розглянемо практичний приклад. Вивчали вплив інокуляції насіння гороху на урожайність. Встановили, що в контрольній групі урожайність дорівнювала $3,28 \pm 0,11$ т/га, а в дослідній – $3,74 \pm 0,12$ т/га. Зробимо перевірку гіпотези за критерієм Ст'юдента.

$$t = \frac{(3,74 - 3,28)}{\sqrt{0,11^2 + 0,12^2}} = 2,82$$

При числі ступенів свободи $df = 18$ табличне значення $t_{0,05} = 2,101$. Оскільки $2,82 > 2,101$, нульову гіпотезу відхиляють. Тобто, передпосівна інокуляція насіння статистично достовірно підвищила урожайність гороху ($p < 0,05$).

Контрольні питання

1. Що називають статистичною гіпотезою?
2. У чому полягає відмінність між нульовою та альтернативною гіпотезами?
3. Що таке рівень статистичної значущості?
4. Що характеризує p-value?

5. Які помилки називають помилками I та II роду?
6. Що розуміють під статистичною потужністю (робастністю) тесту?
7. За яких умов застосовують параметричні статистичні критерії?
8. Коли доцільно використовувати непараметричні критерії?
9. Які умови необхідні для застосування t-критерію Ст'юдента?
10. У чому різниця між незалежними та залежними вибірками?
11. Які статистичні методи застосовують для порівняння двох, трьох і більше вибірок?
12. Як визначити, чи підпорядковуються дані нормальному розподілу?
13. Чому перед вибором статистичного критерію необхідно оцінити характер розподілу даних?
14. Яке практичне значення статистичних тестів у рослинництві та агрономії?

Розрахункові задачі

Задача 1. Перевірка статистичної гіпотези за t-критерієм Ст'юдента (незалежні вибірки)

У польовому досліді оцінювали вплив біостимулятора на урожайність гороху.

Повторення	Контроль	Біостимулятор
1	3,42	3,71
2	3,36	3,80
3	3,51	3,76
4	3,44	3,82
5	3,39	3,75

Необхідно:

1. Обчислити середню урожайність кожної групи.
2. Визначити стандартне відхилення та стандартну помилку середньої.
3. Розрахувати t-критерій Ст'юдента для незалежних вибірок.
4. Визначити число ступенів свободи.
5. Порівняти отримане значення з критичним при $\alpha = 0,05$.
6. Зробити статистичний і біологічний висновок.

Задача 2. Парний t-критерій

Визначали площу листової поверхні рослин сої до та після обробки біостимулятором.

Рослина	До	Після
1	31,2	34,6
2	30,8	33,5
3	32,5	35,8
4	31,7	34,1
5	30,9	33,9
6	32,0	35,0

Необхідно:

1. Обчислити різницю для кожної пари.
2. Знайти середню різницю.
3. Обчислити стандартне відхилення різниць.
4. Визначити t-критерій для залежних вибірок.
5. Перевірити гіпотезу при $\alpha = 0,05$.
6. Інтерпретувати результат.

Задача 3. Вибір статистичного критерію

Для кожної ситуації визначте:

- нульову гіпотезу;
- альтернативну гіпотезу;
- тип вибірок;
- параметричний або непараметричний метод;
- конкретний статистичний критерій.

Ситуації:

1. Порівняння урожайності трьох сортів ячменю.
2. Порівняння висоти рослин до і після внесення добрив.
3. Аналіз зв'язку між площею листкової поверхні та урожайністю.
4. Порівняння ураженості рослин двох сортів хворобою.
5. Порівняння вмісту білка у двох незалежних групах, якщо розподіл не є нормальним.

Задача 4. Аналіз нормальності розподілу

Отримано результати вимірювання висоти рослин кукурудзи (см): 164; 166; 167; 168; 169; 170; 170; 171; 171; 172; 173; 174; 175; 176; 178.

Необхідно:

1. Побудувати гістограму частот.
2. Обчислити середню арифметичну.
3. Визначити медіану.
4. Обчислити коефіцієнт асиметрії.
5. Зробити висновок щодо наближеності розподілу до нормального.
6. Визначити, чи можна застосовувати параметричні статистичні критерії.

Задача 5. Інтерпретація результатів статистичного аналізу

У досліді отримано такі результати.

Варіант	Урожайність, т/га
Контроль	$5,84 \pm 0,17$
Добриво А	$6,08 \pm 0,16$
Добриво Б	$6,32 \pm 0,18$

Після аналізу отримано: $F = 7,83$; $p = 0,003$; $НСР_{0,05} = 0,28$ т/га.

Необхідно:

1. Сформулювати H_0 та H_1 .

2. Встановити, чи існують статистично значущі відмінності між варіантами.
3. Визначити, між якими варіантами різниця перевищує НСР.
4. Зробити біологічний висновок.
5. Запропонувати, який статистичний тест слід застосувати після ANOVA для попарного порівняння середніх.

Тестові завдання

1. Нульова гіпотеза (H_0) стверджує, що:

- А. Між вибірками існує достовірна різниця.
- Б. Між вибірками відсутня статистично значуща різниця.
- В. Дані не мають нормального розподілу.
- Г. Вибірки є залежними.

2. Якщо $p < 0,05$, то:

- А. Приймають нульову гіпотезу.
- Б. Відхиляють нульову гіпотезу.
- В. Дані не потребують подальшого аналізу.
- Г. Обов'язково використовують непараметричний тест.

3. Помилка I роду полягає у:

- А. Прийнятті хибної альтернативної гіпотези.
- Б. Відхиленні правильної нульової гіпотези.
- В. Прийнятті правильної альтернативної гіпотези.
- Г. Неправильному визначенні середнього значення.

4. Який критерій застосовують для порівняння двох незалежних вибірок із нормальним розподілом?

- А. χ^2 -критерій Пірсона.
- Б. t-критерій Стюдента для незалежних вибірок.
- В. Критерій Фрідмана.
- Г. Критерій Вілкоксона.

5. Який статистичний метод використовують для порівняння трьох і більше незалежних середніх?

- А. Кореляційний аналіз.
- Б. Парний t-критерій.
- В. Однофакторний дисперсійний аналіз (ANOVA).
- Г. Критерій Манна-Уїтні.

6. Яка умова є обов'язковою для застосування параметричних критеріїв?

- А. Дані повинні бути ранговими.
- Б. Обсяг вибірки повинен бути більше 100.

- В. Розподіл ознаки має бути наближеним до нормального.
- Г. Кількість варіантів повинна бути не менше п'яти.

7. Який критерій рекомендують застосовувати для двох незалежних вибірок, якщо розподіл суттєво відрізняється від нормального?

- А. Парний t-критерій.
- Б. Критерій Манна-Уїтні.
- В. ANOVA.
- Г. Кореляцію Пірсона.

8. Для чого використовують критерій χ^2 Пірсона?

- А. Для аналізу середніх значень.
- Б. Для визначення кореляції.
- В. Для аналізу якісних (категоріальних) ознак.
- Г. Для побудови регресійної моделі.

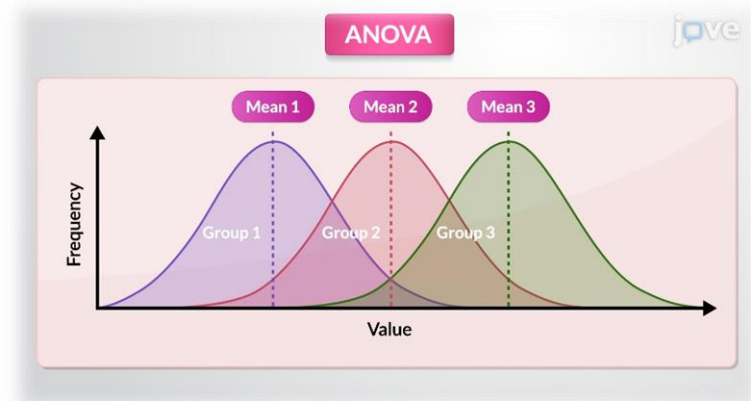
9. Парний t-критерій застосовують, коли:

- А. Порівнюють три незалежні групи.
- Б. Вибірki мають різний обсяг.
- В. Одні й ті самі об'єкти вимірюють двічі.
- Г. Дані мають ненормальний розподіл.

10. Яке твердження щодо вибору статистичного тесту є правильним?

- А. Завжди застосовують t-критерій Ст'юдента.
- Б. Вибір тесту залежить лише від обсягу вибірки.
- В. Усі непараметричні критерії менш точні за параметричні.
- Г. Вибір тесту залежить від типу даних, характеру їх розподілу, кількості груп та незалежності вибірок.

РОЗДІЛ 4. ДИСПЕРСІЙНИЙ АНАЛІЗ



- 4.1. Сутність і алгоритм дисперсійного аналізу
- 4.2. Дисперсійний аналіз даних однофакторного досліджу
- 4.3. Дисперсійний аналіз даних багатфакторного досліджу

4.1. Сутність і алгоритм дисперсійного аналізу

Дисперсійний аналіз – це розчленування загальної суми квадратів відхилень та загальної кількості ступенів свободи на частини – компоненти, що відповідають структурі експерименту, і оцінка значущості дії та взаємодії факторів, що вивчаються по F-критерію. Міжнародним позначенням дисперсійного аналізу (ДА) служить англійська аббревіатура **ANOVA** (Analysis of Variance, аналіз варіації).

Інакше кажучи, дисперсійний аналіз призначений для оцінки впливу одного або кількох факторів на варіювання досліджуваної (результативної) ознаки. Метод дозволяє визначити, чи є статистично значущими відмінності між середніми значеннями декількох груп.

Сутність дисперсійного аналізу полягає у розкладанні загальної варіації (загальної суми квадратів відхилень) та загальної кількості ступенів свободи по структурним елементам експерименту з метою оцінки суттєвості дії (взаємодії) факторів, що вивчаються (варіантів) шляхом зіставлення відповідних дисперсій та розрахункових F-критеріїв з контрольними (табличними).

Основна ідея ANOVA полягає у розкладанні загальної мінливості досліджуваної ознаки на окремі складові:

- факторну (міжгрупову) дисперсію – обумовлену дією досліджуваного фактора;
- випадкову (внутрішньогрупову, залишкову) дисперсію – спричинену неконтрольованими факторами та випадковими помилками.

Якщо факторна дисперсія значно перевищує випадкову, роблять висновок, що досліджуваний фактор істотно впливає на результативну ознаку.

Сьогодні дисперсійний аналіз є одним із найпоширеніших методів

статистичної обробки експериментальних даних у біології, медицині, агрономії, екології, економіці та соціальних науках.

Дисперсійний аналіз дає можливість отримати уявлення про ступінь, або частку впливу того чи іншого фактора у спільній дисперсії ознаки, яку приймають за одиницю або 100%.

Дисперсійний аналіз був розроблений англійським статистиком і генетиком Р. Фішером (Ronald Aylmer Fisher), який відкрив закон розподілу відносин дисперсій у 1920-х роках:

$$\frac{S_v^2}{S_z^2} = F$$

Фактичне (розрахункове) значення F розраховують по відношенню дисперсії фактора s_v , який досліджується в досліді, до дисперсії помилки s_z .

Слово «дисперсія» означає розсіювання даних досліді та розчленування загального варіювання досліджуваних показників на складові. Звідси і назва методу дисперсійний аналіз.

Співвідношення дисперсій – критерій, який дозволяє досліднику надати загальну оцінку достовірності відмінностей між середніми арифметичними виконаного досліді.

На честь автора дисперсійного аналізу критерій позначили першою літерою його прізвища F (*критерій Фішера*).

Для дисперсійного аналізу становлять інтерес дисперсія варіантів, – s_v , дисперсія повторень – s_p і дисперсія помилки – s_z . Підсумовування дисперсій дає суму квадратів загального розсіювання: $s_y^2 = s_v^2 + s_p^2 + s_z^2$.

Для кожного розсіювання обчислюють число ступенів свободи v за формулами: $v_y = N - 1$ (для суми квадратів загального розсіювання); $v_v = l - 1$ (розсіювання варіантів); $v_p = n - 1$ (розсіювання повторень); $v_z = (l - 1) \times (n - 1)$ (розсіювання помилки).

Шляхом розподілу суми квадратів на відповідне число ступенів свободи одержують дисперсію s^2 .

Дисперсійний аналіз широко використовують для планування експерименту та статистичної обробки отриманих результатів виконаного експерименту. При дисперсійному аналізі одночасно обробляють дані кількох вибірок (варіантів), які становлять єдиний статистичний комплекс, оформлений у вигляді спеціальної робочої таблиці. Структура статистичного комплексу та його наступний аналіз визначаються схемою та методикою експерименту.

Дисперсійний аналіз швидко увійшов у вживання при обробці експериментальних даних досліджень завдяки наступним основним перевагам його перед методом попарних порівнянь за критерієм Стьюдента:

- ✓ замість індивідуальних помилок, середніх за кожним варіантом, у дисперсійному аналізі використовується узагальнена помилка середніх, яка спирається на більше спостережень і, отже, є більш надійною базою для оцінок;
- ✓ методом дисперсійного аналізу можна обробляти дані простих та

складних, однорічних та багаторічних, однофакторних та багатфакторних дослідів.

✓ аналіз дозволяє уникнути громіздких обчислень при великій кількості варіантів у досвіді і дозволяє компактно у вигляді суттєвих різниць уявити підсумки статистичної обробки.

Оскільки дисперсійний аналіз є параметричним методом, його застосування обумовлено наступними передумовами:

- ✓ нормальним розподілом значень ознаки у сукупності;
- ✓ випадковим розміщенням варіантів за експериментальними одиницями (ділянками у польовому досліді);
- ✓ незалежністю, чи неспряженістю, вибірок (варіантів);

У статистиці є критерії контролю «нормальності» даних, однак підстав для їх застосування немає доти, доки обсяги вибірок не будуть достатньо більшими. Незалежність передбачає відсутність зв'язку між величинами помилок за варіантами.

Оскільки сусідні ділянки у польовому досвіді схильні до більшої залежності між собою, ніж з іншими ділянками, то важливо уникнути розміщення однойменних варіантів на сусідніх ділянках.

Залежно від природи фактору, що вивчається в досліді, розрізняють дві важливі моделі дисперсійного аналізу.

Модель 1: градаціями фактору, або варіантами, служать чітко фіксовані способи обробітку ґрунту, дози добрив, сорти, гібриди, строки сівби і т. д.

Модель 2: градаціями фактору є випадкові вибірки. Наприклад, варіабельність сорту, гібрида (відмінності за врожайністю) за випадково вибраними регіонами вирощування.

Для багатфакторних дослідів можливі змішані моделі, наприклад сорти, гібриди та зони їх вирощування. Аналогічна ситуація складається в однофакторному польовому досвіді за наявності повторень, контрольованих джерел варіації.

Розрахунковий фактичний критерій $F_{\text{fact.}}$ порівнюють з теоретичним $F_{\text{theor.}}$, який знаходять за таблицями. Якщо $F_{\text{fact.}} \geq F_{\text{theor.}}$, то достовірність відмінностей у дослідженні доведена, тобто є одна або кілька пар варіантів, середні арифметичні яких достовірно різняться. Якщо $F_{\text{fact.}} < F_{\text{theor.}}$, то достовірних відмінностей між варіантами немає.

Зустрічаються випадки, коли $F_{\text{fact.}}$ лише трохи менше $F_{\text{theor.}}$. Строго дотримуючись правила, викладеного вище, можна зробити висновок про відсутність достовірної різниці в досліді. Однак подальший аналіз дозволяє іноді знайти цю різницю хоча б в одному-двох варіантах. У таких випадках, не обмежуючись розрахунком F , слід продовжити обчислення до найменшої суттєвої різниці (НСР). З цим показником порівнюють різницю між парами варіантів ($d = \bar{x}_1 - \bar{x}_2$). Якщо $d \geq \text{НСР}$, то різницю між варіантами, що аналізуються в досліді, вважають доведеною. Докази найчастіше ведуть на рівнях довірчої ймовірності $P_{0,95}$ і $P_{0,99}$.

Дисперсійний аналіз – найбільш досконалий метод статистичної обробки даних.

Переваги дисперсійного аналізу полягають у відокремленні із загального варіювання його компонентів, у обчисленні узагальненої помилки всього досліджуваного Z на основі більшої кількості спостережень, ніж індивідуальні помилки окремих варіантів у недисперсійних методах.

Дисперсійний аналіз особливо цінний для багатofакторних дослідів, оскільки дозволяє визначити достовірність як дії, але і взаємодію факторів, які вивчаються в досліді.

Наприкінці дисперсійного аналізу обов'язково обчислюють відносну помилку всього досліджуваного.

Сучасна теорія планування експерименту та статистичний аналіз базується на принципах рандомізації. Теорія вимагає, щоб усі спостереження були незалежними. В цьому разі дисперсійний аналіз дає правильну, незміщену оцінку проведеному досліді.

Дисперсійний аналіз неможливий для простих вегетаційних чи польових дослідів, які проводились дослідником без повторностей. В досліді мінімум повторностей повинен бути не менше двох повторностей.

Однак, особливо при проведенні польових дослідів, бувають великі помилки і статистично неможливо із-за цього встановити ефекти варіантів, які вивчаються в досліді. Тому однофакторні досліді в рослинництві проводять зазвичай у 2-6 кратній повторності залежно від площі ділянок.

Існує закономірність, яка обґрунтована теоретично та експериментально: ефекти взаємодії факторів в досліді у міру зростання їх порядків зменшуються (ефект $AB > ABC > ABCD$ тощо). У цьому парні взаємодії, які називають взаємодії першого порядку, дають зазвичай значні, а взаємодії вищих порядків між трьома і більше чинниками, зазвичай, незначні і статистично несуттєві ефекти (прирости врожайності). Отже, у двофакторних дослідіх без повторностей втрачається цінна інформація про парні взаємодії і статистично не доводяться навіть дуже значні різниці за варіантами. Інша небезпека постановки багатofакторних дослідів у тому, що випадкове випадання з обліку хоча б однієї ділянки позбавляє експериментатора можливості статистично обробити отримані дані.

Простий одноблоковий (однофакторний) дисперсійний аналіз (дані досліді підсумовують лише за рядками) виконують відповідно до наступного алгоритму.

1. Підготовка таблиці для вихідних даних, сум та середніх значень проведеного досліді.

2. Визначення загальної кількості даних досліді (спостережень): $N = l \times n$.

Суми квадратів відхилень за даними польового експерименту (статистичного комплексу з l -варіантами та n -повтореннями) знаходять у наступній послідовності. У вихідній таблиці визначають суми по повторенням P , по варіантам V та загальну суму всіх спостережень ΣX .

3. Знаходження сум:

- за кожним варіантом;
- середніх значень по варіантах;
- загальної середньої за досліді.

4. Розрахунок коригувального члена або поправки:

5. Знаходження квадратів вихідних дат та сум квадратів відхилень (сума

квадратів мінус поправка), або загальної варіації.

6. Зведення сум за варіантами у квадрат та їх підсумовування.
7. Усереднення суми квадратів для варіантів шляхом розподілу на повторність – n (число складових) і віднімання поправки (знаходження варіації варіантів).
8. Визначення залишкової (випадкової) варіації.
9. Оцінка часткової (долевої) варіації у % від загальної варіації.
10. Розкладання загальної кількості ступенів свободи (ссО).
11. Визначення середніх квадратів (дисперсії) варіантів та залишку шляхом розподілу сум квадратів на число ступенів свободи.
12. Розрахунок F-критерію: $F = S_V^2 / S_E^2$ та визначення його контрольного (табличного) значення відповідно до числа ступенів свободи чисельника та знаменника та рівнем довіри: F_T (ссV та ссЕ; $1 - \alpha$). У загальному випадку кількість ступенів свободи для більшої дисперсії (чисельник) символізує сс1, а для меншої дисперсії (знаменника) символізує сс2. ссV приймають за сс1, а ссЕ приймають за сс2.
13. Зведення всіх сум квадратів, дисперсій, коефіцієнту F у підсумкову таблицю дисперсійного аналізу (ДА).
14. Зіставлення F з F_α (табличне значення) та відхилення або прийняття H_0 .

Для оцінки точності дослідження додатково розраховують рівень варіації V та точність дослідження (відносну помилку середніх) Z. Якщо $V < 7\%$, а $Z < 3\%$, точність дослідження вважається високою.

Оцінка дисперсії проводиться шляхом порівняння дисперсії варіантів (s^2) з дисперсією помилки (s^2) за критерієм Фішера:

$$\frac{S_V^2}{S_Z^2} = F$$

За одиницю порівняння приймають середній квадрат випадкової дисперсії, що визначає випадкову помилку експерименту. При цьому нульовою гіпотезою, яка перевіряється, служить припущення: всі вибіркові середні значення дослідження є оцінками однієї генеральної середньої, отже відмінності між ними не суттєві. Якщо $F_{\text{факт.}} = s_V^2 / s_Z^2 < F_{\text{теор.}}$, то нульова гіпотеза не відкидається (між усіма вибірковими середніми немає істотних відмінностей). На цьому перевірка закінчується.

Нульова гіпотеза відкидається, якщо $F_{\text{факт.}} = s_V^2 / s_Z^2 \geq F_{\text{теор.}}$. У цьому випадку додатково оцінюють суттєвість приватних відмінностей по НСР (найменшої суттєвої різниці) і визначають, якими середніми є значні різниці.

Теоретичні значення критерію F для прийнятого у дослідженні рівня значущості знаходять за спеціальними таблицями з урахуванням числа ступенів свободи для дисперсії варіантів (чисельник) та випадкової дисперсії (знаменник).

Дисперсійний аналіз надає досліднику можливість провести аналіз та отримати уявлення про ступінь, або частку впливу, того чи іншого фактора у спільній дисперсії ознаки, яку приймають за одиницю або 100%:

$$\mu_v^2 = C_v / C_y - \text{вплив варіантів дослідження};$$

$\mu_p^2 = C_p / C_y$ – вплив повторень дослідів;

$\mu_z^2 = C_z / C_y$ – вплив випадкових факторів при проведенні досліджень.

$\mu_y^2 = \mu_v^2 + \mu_p^2 + \mu_z^2 = 1$ (або 100%) – вплив усіх факторів.

При цьому, кореляційне відношення, що характеризує тісноту зв'язку результативної ознаки дослідів з факторіальною, визначається за формулою:

$$\mu_v = \sqrt{\mu_v^2}$$

де:

μ_v – кореляційне відношення, що характеризує тісноту зв'язку результативної ознаки дослідів з факторіальною.

μ_v^2 – індекс детермінації, що показує частку його варіювання під дією досліджуваних факторів.

Двофакторну або двоблочну модель дисперсійного аналізу даних дослідів (дані підсумовують рядками та колонками) використовують за наявності у плані дослідів двох досліджуваних факторів або досліджуваного та контрольованого факторів.

Кожен варіант (доза фактору) зустрічається один раз у кожному блоці (повторенні), а дози двох факторів поєднуються за рядком та стовпцем у взаємоперпендикулярному напрямку.

Таблицю з подібним розміщенням градацій двох факторів називають таблицею-решіткою. За допомогою повторень «блокують» вплив родючості ґрунту (польовий дослід) на досліджуваний в досліді фактор. Це сприяє зниженню впливу випадкової варіації та усуває систематичну помилку виконаного експерименту.

Визначення точності є обов'язковим правилом при статистичному аналізі результатів будь-яких досліджень.

Основна відмінність дисперсійних аналізів полягає в переліку обчислюваних сум квадратів.

Для рандомізованих повторень обчислюють:

$$C_y = C_v + C_p + C_z.$$

Для повної рандомізації обчислюють:

$$C_y = C_v + C_z.$$

Для латинського квадрата та латинського прямокутника обчислюють:

$$C_y = C_v + C_p + C_c + C_z.$$

Для двофакторного дослідів обчислюють:

$$C_y = C_p + C_A + C_B + C_{AB} + C_z.$$

Для трифакторного дослідів обчислюють:

$$C_y = C_p + C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC} + C_z.$$

Для методу змішування обчислюють:

$$C_y = C_{p1} + C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_z.$$

Для подвійного розщеплення ділянок обчислюють:

$$C_y = C_p + C_A + C_B + C_{AB} + C_{z1} + C_{z2}.$$

Для потрійного розщеплення ділянок обчислюють:

$$C_y = C_p + C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC} + C_{z1} + C_{z2} + C_{z3}.$$

В дослідженнях з рослинництва та біології планування експерименту та статистичний аналіз базуються на засадах рандомізації. Отримання вірогідних даних вимагає, щоб усі спостереження були незалежні. У цьому випадку дисперсійний аналіз дає правильну, незміщену оцінку помилки експерименту. Якщо експеримент не рандомізований, то експериментатор може отримати зміщену оцінку помилки досвіду, і зазвичай критерії значущості, які використовуються в дисперсійному аналізі, втрачають законну силу і не можуть використовуватися як аргументи строго статистичного доказу ефектів варіантів.

4.2. Дисперсійний аналіз даних однофакторного досліду

Однофакторний дисперсійний аналіз дозволяє виявити вплив лише одного фактору. При цьому загальне варіювання результативної ознаки при умовах обов'язкового дотримання вимог до проведення польових дослідів і вирівняності дослідного поля за родючістю ґрунту за бажанням дослідника розкладається на два компоненти – варіювання варіантів і випадкове варіювання: $C_y = C_v + C_z$. Варіювання за повтореннями (C_p) в цьому випадку до дисперсійного аналізу не залучається.

В той же час, дослідник повинен урахувати, що в польових дослідях, які виконується навіть при повній рандомізації повторностей, загальна мінливість результативної ознаки може створюватися не тільки під впливом фактора, що вивчається в досліді, та випадкових помилок, але і під впливом строкатості родючості ґрунту дослідного поля. Строкатість родючості дослідного поля може спостерігатися як окремо на ділянці одного повторення, так і може спостерігатися між ділянками повторностей одного варіанту. Мінливість родючості ґрунту в межах однієї ділянки менш значна, ніж між ділянками повторностей одного варіанту.

Мінливість результативної ознаки, яка відбувається під впливом дії строкатості родючості ґрунту в межах однієї ділянки варіанту дисперсійним аналізом окремо не виділяється і відноситься до випадкових помилок.

Мінливість результативної ознаки, яка відбувається під впливом строкатості родючості ґрунту дослідного поля серед різних повторностей варіанту дисперсійним аналізом виділяється і приймається до уваги при проведенні дисперсійного аналізу по встановленню помилки досліду та вірогідності отриманих даних виконаних досліджень.

Якщо дослідник рахує доцільним урахувати вплив строкатості родючості ґрунту серед повторень на мінливість результативної ознаки, то тоді мінливість результативної ознаки розкладається на три компоненти – варіювання повторностей, варіювання варіантів, випадкове варіювання:

$$C_y = C_p + C_v + C_z.$$

Однофакторний дисперсійний аналіз неможливий для польових

експериментів, які були проведені без повторностей. Мінімум повторностей в одно факторному досліді – дві, максимальна кількість повторностей шість.

Розглянемо приклад дисперсійного аналізу одно факторного досліді.

Дослідником розробляється сортова технологія для нового створено гібрида соняшнику. Вивчається реакція гібриду на зміну густоти стояння рослин, при якій густоті стояння рослин гібрид формує максимальний рівень врожайності і на якій густоті стояння можливо отримувати вірогідний приріст врожайності. Гібрид соняшнику в досліді випробується за наступною густотою стояння рослин 40 тис./га (контроль), 50 тис./га, 60 тис./га, 70 тис./га. Вивчаємий фактор – густота стояння рослин. Кожна густота стояння рослин є кожен окремий варіант. Варіантів в досліді (l) – чотири, $l = 4$. Повторностей в досліді (n) – чотири, $n = 4$. Дослід виконано за моделлю повною рандомізації.

Отримані дані за врожайністю гібрида соняшнику заносимо в таблицю за варіантами та повторностями (табл. 4.1).

Таблиця 4.1

**Робоча таблиця підготовки до дисперсійного аналізу
вихідних даних досліді по встановленню ефективності впливу
густоти стояння рослин на формування врожайності гібрида соняшнику**

Варіант	Врожайність за повторностями, ц/га				Сума даних врожайності по варіанту, V	Середня врожайність по варіанту, $\bar{x}$
	1	2	3	4		
1	29,7	31,0	30,9	31,4	$123,0 - V_1$	$30,8 - \bar{x}_1$
2	34,0	34,8	36,5	35,6	$140,9 - V_2$	$35,2 - \bar{x}_2$
3	30,8	31,9	29,0	30,0	$121,7 - V_3$	$30,4 - \bar{x}_3$
4	25,7	26,8	25,0	24,6	$102,1 - V_4$	$25,5 - \bar{x}_4$
	$P_1 = 120,2$	$P_2 = 124,5$	$P_3 = 121,4$	$P_4 = 121,6$	$\sum V = 487,7$	$\bar{\bar{x}} = 30,5$

В таблиці фактор густоти стояння рослин відповідає наступним варіантам:

40 тис./га – 1 варіант;

50 тис./га – 2 варіант;

60 тис./га – 3 варіант;

70 тис./га – 4 варіант.

Після занесення в таблицю даних врожайності по повторностям варіантів дослідник визначає загальну кількість даних досліді (спостережень) (N):

$$N = l \times n = 4 \times 4 = 16.$$

Встановлюється сума даних врожайності окремо по кожному варіанту по повторностях досліді.

Сума даних врожайності кожному варіанту становить:

$$1 \text{ варіант} - V_1 = 123,0;$$

$$2 \text{ варіант} - V_2 = 140,9;$$

3 варіант – $V_3 = 121,7$;

4 варіант – $V_4 = 102,1$.

Сума даних врожайності за повтореннями варіантів становить:

1 повторення – $P_1 = 120,2$;

2 повторення – $P_2 = 124,5$;

3 повторення – $P_3 = 121,4$;

4 повторення – $P_4 = 121,6$.

Визначаємо суму врожайності за всіма варіантами дослідів $\sum V$:

$$\sum V = V_1 + V_2 + V_3 + V_4 = 123,0 + 140,9 + 121,7 + 102,1 = 487,7.$$

Встановлюється середня врожайність окремо по кожному варіанту по повторностях дослідів.

Середня арифметична врожайності по кожному варіанту буде становити:

1 варіант – $x_1 = 30,8$;

2 варіант – $x_2 = 35,2$;

3 варіант – $x_3 = 30,4$;

4 варіант – $x_4 = 25,5$.

Визначаємо середню арифметичну врожайність по всіх варіантах дослідів ($\bar{x}$):

$$\bar{x} = (x_1 + x_2 + x_3 + x_4) / l = (30,8 + 35,2 + 30,4 + 25,5) / 4 = 30,5.$$

Для подальшого проведення дисперсійного аналізу необхідно встановити суму квадратів всіх значень врожайності по дослідів:

$$\begin{aligned} \sum X^2 = & (29,7)^2 + (31,0)^2 + (30,9)^2 + (31,4)^2 + (34,0)^2 + (34,8)^2 + (36,5)^2 + (35,6)^2 \\ & + (30,8)^2 + (31,9)^2 + (29,0)^2 + (30,0)^2 + (25,7)^2 + (26,8)^2 + (25,0)^2 + (24,6)^2 = 882,1 + \\ & 961,0 + 954,8 + 986,0 + 1156,0 + 1211,0 + 1332,3 + 1267,4 + 948,6 + 1017,6 + \\ & 841,0 + 900,0 + 660,5 + 718,2 + 625,0 + 605,2 = 15066,7 \end{aligned}$$

Розрахунок коригувального члена або поправки:

$$\Delta = (\sum V)^2 / N = (487,7)^2 / 16 = 237851,3 / 16 = 14865,7.$$

Отримання цього показника дозволяє далі розрахувати суму квадратів розсіювання загального (C_y), варіантів (C_v), помилки (C_z).

Розраховують суму квадратів відхилень за наступними формулами.

Суму квадратів розсіювання загального (C_y) розраховують наступним чином:

$$C_y = \sum X^2 - \Delta = 15066,7 - 14865,7 = 201.$$

Суму квадратів повторень варіантів (C_p) розраховують наступним чином:

$$C_p = \sum P^2 / l - \Delta = (P_1^2 + P_2^2 + P_3^2 + P_4^2) / l - \Delta = (120,2^2 + 124,5^2 +$$

$$121,4^2 + 121,6^2) / 4 - 14865,7 = (14448,0 + 15500,3 + 14738,0 + 14786,6) / 4 - 14865,7 = 2,5.$$

Суму квадратів варіантів (C_v) розраховують наступним чином:

$$C_v = \sum V^2 / n - \Delta = (V_1^2 + V_2^2 + V_3^2 + V_4^2) / n - \Delta = (123,0^2 + 140,9^2 + 121,7^2 + 102,1^2) / 4 - 14865,7 = (15129,0 + 19852,8 + 14810,9 + 10424,4) / 4 - 14865,7 = 189.$$

Суму квадратів помилки (C_z) розраховують наступним чином:

$$C_z = C_y - C_p - C_v = 201 - 2,5 - 189 = 9,5.$$

Обчислюють число ступенів свободи загального розсіювання (v_y), повторень (v_p), варіантів (v_v), помилку (залишку) (v_z).

Кількість ступенів свободи загального розсіювання буде рівнятися:

$$v_y = N - 1 = 16 - 1 = 15.$$

Кількість ступенів свободи повторень буде рівнятися: $v_p = n - 1 = 4 - 1 = 3$.

Кількість ступенів свободи варіантів буде рівнятися: $v_v = l - 1 = 4 - 1 = 3$.

Кількість ступенів свободи помилки (залишку) буде рівнятися:

$$v_z = (l - 1) \times (n - 1) = (4 - 1) \times (4 - 1) = 9.$$

Розклад загальної кількості ступенів свободи рівняється:

$$v_y (N - 1) = v_p + v_v + v_z = 3 + 3 + 9 = 15.$$

Якщо при дисперсійному аналізі не визначається варіювання повторення, кількість ступенів свободи помилки (залишку) розраховують наступним чином:

$$v_z = l \times (n - 1).$$

Після встановлення суми квадратів та кількості ступенів свободи розраховують дисперсію варіантів і дисперсію помилки (залишків) шляхом розподілу суми квадратів на кількість ступенів свободи.

Дисперсія варіантів (s_v^2) розраховується за формулою:

$$s_v^2 = C_v / v_v = 189 / 3 = 63,0.$$

Дисперсія помилки (s_z^2) розраховується за формулою:

$$s_z^2 = C_z / v_z = 9,5 / 9,0 = 1,1.$$

Після виконання таких обчислювань встановлюють фактичний F-критерій:

$$F_{\text{факт.}} = s_v^2 / s_z^2 = 63,0 / 1,1 = 57,3.$$

Після проведення розрахунків результати дисперсійного аналізу заносять в таблицю (табл. 4.2).

Теоретичне значення критерію Фішера знаходять за даними додатку 2 та додатку 3 по кількості ступенів свободи помилки (залишку) – 9 (дев'ять) рядок ($v_z = 9$) та за кількістю ступенів свободи варіантів – колонка з числом 3 (три) ($v_v = 3$). На їхньому перетині знаходять значення на 5%-му рівні значущості для $F_{0,95} = 3,86$ (додаток 2) і на 1%-му рівні значущості для $F_{0,99} = 6,99$ (додаток 3).

Таблиця 4.2

Результати дисперсійного аналізу однофакторного дослід

Джерела варіації	Сума квадратів	Частка від загальної варіації, %	Ступінь свободи	Дисперсія, s^2	Критерій		
					F _{fact.}	F _{theor.}	
						F _{0,95}	F _{0,99}
Загальні C _y	201	100	vy = 15				
Повторення C _p	2,5	1,24	vp = 3				
Варіанти C _v	189	94,04	vv = 3	S _v ² = 63,0	57,3	3,86	6,99
Помилки C _z	9,5	4,72	vz = 9	s _z ² = 1,1			

Проводять оцінку нульової гіпотези

Порівнюють фактичний (розрахунковий) критерій Фішера $F_{\text{fact.}} = 57,3$ з теоретичним критерієм $F_{0,95} = 3,86$ і $F_{0,99} = 6,99$. Оскільки $F_{\text{fact.}} > F_{\text{theor.}}$, то нульова гіпотеза H_0 відкидається, так як у досліді є суттєві відмінності за варіантами на 5%-му на 1%-му рівні значущості.

На основі порівняння критеріїв Фішера роблять висновок про достовірність проведеного дослід.

Так як фактичне значення критерію Фішера становить 57,3, що більше теоретичних значень $F_{0,95}$ (3,86) і $F_{0,99}$ (6,99), проведений дослід достовірний на обох рівнях довірчої ймовірності – $F_{0,95}$ і $F_{0,99}$.

Це означає, що між середніми арифметичними парами варіантів різниці в урожайності залежно від густоти стояння рослин будуть достовірними і дисперсійний аналіз треба продовжувати. Якщо критерій Фішера розрахунковий набагато менше теоретичного значення, тоді всі розрахунки припиняють і роблять висновок про відсутність достовірних різниць між будь-якими парами варіантів дослід.

Визначають помилку дослід. Точність дослід встановлюють саме за розрахунками узагальненої помилки дослід (E):

$$E = \sqrt{\frac{S_z^2}{n}} = \sqrt{\frac{1,1}{4}} = 0,52 \text{ ц/га}$$

В процентному виразі стандартна помилка дослід буде становити:

$$E = \sqrt{\frac{S_z^2}{n \times (\frac{100}{x})}} = \sqrt{\frac{1,1}{4 \times (\frac{100}{30,5})}} = 1,71\%$$

Дослід по встановленню ефективності впливу густоти стояння рослин на

формування врожайності гібрида соняшнику виконано с досить високою точністю, узагальнена (стандартна) помилка досліду E рівняється 1,71% ($E < 3\%$).

Далі встановлюють помилку різниці середніх значень результативної ознаки (врожайність соняшнику) в досліді (s_d):

$$s_d = E \times 1,41 = 0,52 \times 1,41 = 0,73 \text{ ц/га.}$$

$$1,41 - \text{постійне число} = \sqrt{2}$$

Найменшу суттєву різницю (НСР) розраховують, як правило, на двох рівнях довірчої ймовірності за формулами:

$$НСР_{0,95} = s_d \times t_{0,95};$$

$$НСР_{0,99} = s_d \times t_{0,99}.$$

Теоретичне значення t -критерію Стьюдента знаходять у додатку 4 за кількістю, ступенів свободи залишкового розсіювання v_z , яке у проведеному досліді за дисперсійним аналізом становить – 9 (дев'ять).

У графі додатку 4 для кількості ступенів свободи помилки, в нашому досліді – 9 (дев'ять), вибирають значення t -критерію Стьюдента на 5% -му рівні довірчої ймовірності (ймовірності помилки) $P_{0,95}$ ($\alpha = 0,05$) та на 1%-му рівні ймовірності помилки $P_{0,99}$ ($\alpha = 0,01$). На перетині цих показників значення t -критерію Стьюдента становить $t_{0,95} = 2,26$, $t_{0,99} = 3,25$.

Розраховують найменшу суттєву різницю НСР для 5%-го рівня довірчої ймовірності ($НСР_{0,95}$) і 1%-го рівня ймовірності ($НСР_{0,99}$):

$$НСР_{0,95} = s_d \times t_{0,95} = 0,73 \times 2,26 = 1,65 \text{ ц/га. } НСР_{0,99} = s_d \times t_{0,99} = 0,73 \times 3,25 = 2,37 \text{ ц/га.}$$

Кінцевим етапом дисперсійного аналізу є встановлення відносної помилки досліду:

$$S_x = 100 \times E / \bar{x} = 100 \times 0,73 / 30,5 = 2,39\%.$$

Після закінчення проведення дисперсійного аналізу складають підсумкову таблицю в яку заносять середні арифметичні значення результативної ознаки (врожайність), дані НСР, відносну помилку досліду (табл. 4.3).

Таблиця 4.3

Підсумкова таблиця дисперсійного аналізу досліду

Варіант, густота стояння рослин	Середня врожайність по варіанту, ц/га	± приріст врожайності до контролю, ц/га	НСР		Відносна помилка досліду, %
			НСР _{0,95}	НСР _{0,99}	
40 тис./га (контроль)	30,8	-	1,65 ц/га	2,37 ц/га	2,39
50 тис./га	35,2	+ 5,6			
60 тис./га	30,4	+ 0,4			
70 тис./га	25,5	- 5,3			

Після складання підсумкової таблиці дисперсійного аналізу порівнюють різниці між парами середніх арифметичних за варіантами досліду зі значеннями НСР та роблять висновки про суттєвість цих різниць.

При цьому необхідно дотримуватися правила: якщо різниця між будь-якими парами середніх арифметичних показників результативної ознаки більша, ніж значення НСР, або дорівнює їм, тоді ці різниці суттєві.

Висновки дисперсійного аналізу однофакторного досліду по встановленню ефективності впливу густоти стояння рослин на формування врожайності гібрида соняшнику.

В другому варіанті при густоті стояння рослин 50 тис./га приріст врожайності соняшнику у порівнянні з контрольною густотою стояння рослин 40 тис./га склав + 5,6 ц/га. Приріст врожайності більше значень показників найменшої суттєвої різниці $НСР_{0,95} = 1,65$ ц/га і $НСР_{0,99} = 2,37$ ц/га. Приріст врожайності при цій густоті стояння достовірний на обох рівнях довірчої ймовірності.

В третьому варіанті при густоті стояння рослин 60 тис./га приріст врожайності соняшнику у порівнянні з контрольною густотою стояння рослин 40 тис./га склав + 0,4 ц/га. Приріст врожайності менше значень показників найменшої суттєвої різниці $НСР_{0,95} = 1,65$ ц/га і $НСР_{0,99} = 2,37$ ц/га. Приріст врожайності при цій густоті стояння недостовірний на обох рівнях довірчої ймовірності.

В четвертому варіанті при густоті стояння рослин 70 тис./га зниження врожайності соняшнику у порівнянні з контрольною густотою стояння рослин 40 тис./га складає 5,3 ц/га. Показник зниження врожайності менше значень показників найменшої суттєвої різниці $НСР_{0,95} = 1,65$ ц/га і $НСР_{0,99} = 2,37$ ц/га. Зниження врожайності при цій густоті стояння достовірне на обох рівнях довірчої ймовірності.

Таким чином, найбільш ефективною густотою стояння рослин в технології вирощування соняшнику є густота стояння рослин 50 тис. /га, яка забезпечує формування максимального рівня врожайності 35,2 ц/га. При густоті стояння рослин 60 тис. /га формується один теж самий рівень врожайності 30,8 ц/га як на контрольній густоті. На цій густоті стояння рослин не відбувається ні зниження ні приросту врожайності. Густота стояння рослин 70 тис/га призводить до надмірного загущення посіву соняшнику і формування мінімального рівня врожайності 25,5 ц/га при рівні зниження врожайності 5,3 ц/га.

Значення відносної помилки досліду – 2,39% свідчить про дуже високу точність досліду.

Цими висновками закінчується дисперсійний аналіз однофакторного досліду.

4.3. Дисперсійний аналіз даних багатфакторного досліду

Дисперсійний аналіз даних багатфакторного експерименту проводять у два етапи.

Перший етап. Виконується розкладання загальної варіації результативної

ознаки на два компонента – варіювання варіантів та випадкове варіювання (варіювання помилки): $C_y = C_v + C_z$.

Другий етап. Виконується розкладання суми квадратів відхилень варіантів на компоненти, відповідні джерелам варіювання – головним ефектам факторів, які вивчаються в досліді, та їх взаємодій.

У двофакторному експерименті при вивченні факторів А та В розрахунки відбуваються за формулою:

$$C_v = C_A + C_B + C_{AB}.$$

У трифакторному експерименті при вивченні факторів А, В та С розрахунки відбуваються за формулою:

$$C_v = C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}.$$

Приклад. «Двофакторний польовий експеримент із вивчення впливу попередників на врожайність озимої пшениці на фоні основного обробітку ґрунту».

Фактор А – основний обробіток ґрунту (два способи обробітку ґрунту: оранка 16-18 см, поверхневий обробіток – 10-12 см).

Фактор В – попередники озимої пшениці у сівозміні (три попередника: сидеральний пар, горох, озимий ріпак).

На основі даних за врожайністю (результативна ознака) заповнюють робочу таблицю (табл. 4.4).

Таблиця 4.4

Робоча таблиця підготовки до дисперсійного аналізу вихідних даних досліді по встановленню ефективності впливу попередників на врожайність озимої пшениці на фоні основного обробітку ґрунту

Обробіток ґрунту (фактор А)	Попередник (фактор В)	Врожайність за повтореннями варіантів, ц/га				Сума даних врожайності по варіанту, V	Середня врожайність по варіанту, x
		1 п.	2 п.	3 п.	4 п.		
оранка 16-18 см	сидеральний пар	60	62	66	64	252,0 – V ₁	63,0 – x ₁
	горох	58	54	56	57	225,0 – V ₂	56,3 – x ₂
	озимий ріпак	55	56	54	52	217,0 – V ₃	54,3 – x ₃
Поверхневий обробіток 10-12 см	сидеральний пар	62	57	55	61	235,0 – V ₄	58,8 – x ₄
	горох	51	57	50	54	212,0 – V ₅	53,0 – x ₅
	озимий ріпак	52	54	46	48	200,0 – V ₆	50,0 – x ₆
		P ₁ = 338	P ₂ = 340	P ₃ = 327	P ₄ = 336	ΣV = 1341,0	$\bar{x}$ = 55,9

Після занесення в таблицю даних врожайності за повторностями варіантів дослідник визначає загальну кількість даних досліді (спостережень) (N).

В досліді: 2 (два) способи обробітку ґрунту – $l_A = 2$; 3 (три) попередника озимої пшениці – $l_B = 3$; 4 (чотири) повторності у кожному варіанті – $n = 4$.

Кількість варіантів по фактору А: $l_A = 2$; по фактору В: $l_B = 3$.

Загальна кількість варіантів: $l = l_A \times l_B = 2 \times 3 = 6$.

Загальна кількість дослідних ділянок в досліді (спостережень): $N = l \times n = 6 \times 4 = 24$.

Виконують розшифровку варіантів:

Варіант	Спосіб основного обробітку ґрунту (фактор А)	Попередник озимої пшениці (фактор В)
1	оранка 16-18 см	сидеральний пар
2	оранка 16-18 см	горох
3	оранка 16-18 см	озимий ріпак
4	поверхневий обробіток 10-12 см	сидеральний пар
5	поверхневий обробіток 10-12 см	горох
6	поверхневий обробіток 10-12 см	озимий ріпак

Встановлюється сума даних врожайності окремо по кожному варіанту за повторностями досліді.

Сума даних врожайності кожному варіанту становить:

- 1 варіант – $V_1 = 252,0$;
- 2 варіант – $V_2 = 225,0$;
- 3 варіант – $V_3 = 217,0$;
- 4 варіант – $V_4 = 235,0$;
- 5 варіант – $V_5 = 212,0$;
- 6 варіант – $V_6 = 200,0$;

Сума даних врожайності за повтореннями варіантів становить:

- 1 повторення – $P_1 = 338,0$;
- 2 повторення – $P_2 = 340,0$;
- 3 повторення – $P_3 = 327,0$;
- 4 повторення – $P_4 = 336,0$.

Визначаємо суму врожайності за всіма варіантами досліді $\sum V$:

$$\sum V = V_1 + V_2 + V_3 + V_4 + V_5 + V_6 = 252,0 + 225,0 + 217,0 + 235,0 + 212,0 + 200,0 = 1341,0.$$

Встановлюється середня врожайність окремо по кожному варіанту по повторностях досліді.

Середня арифметична врожайності по кожному варіанту буде становити:

- 1 варіант – $x_1 = 63,0$;
- 2 варіант – $x_2 = 56,3$;
- 3 варіант – $x_3 = 54,3$;
- 4 варіант – $x_4 = 58,8$;
- 5 варіант – $x_5 = 53,0$;
- 6 варіант – $x_6 = 50,0$.

Визначають середню арифметичну врожайність по всіх варіантах досліді

($\bar{x}$):

$$\bar{x} = (x_1 + x_2 + x_3 + x_4 + x_5 + x_6) / (l_A \times l_B) = \\ = (63,0 + 56,3 + 54,3 + 58,8 + 53,0 + 50,0) / (2 \times 3) = 55,9 \text{ ц/га.}$$

Встановлюють суму квадратів всіх значень врожайності по досліді:

$$\sum X^2 = (60)^2 + (62)^2 + (66)^2 + (64)^2 + (58)^2 + (54)^2 + (56)^2 + (57)^2 + (55)^2 + \\ (56)^2 + (54)^2 + (52)^2 + (62)^2 + (57)^2 + (55)^2 + (61)^2 + (51)^2 + (57)^2 + (50)^2 + (54)^2 + \\ (52)^2 + (54)^2 + (46)^2 + (48)^2 = (3600) + (3844) + (4356) + (4096) + (3364) + \\ (2916) + (3136) + (3249) + (3025) + (3136) + (2916) + (2704) + (3844) + (3249) + \\ (3025) + (3721) + (2601) + (3249) + (2500) + (2916) + (2704) + (2916) + (2116) + \\ (2304)^2 = 75487.$$

Дисперсійний аналіз виконують в такій послідовності.

Виконують розрахунок коригувального члена (фактору) або поправки:

$$\Delta = (\sum V)^2 / N = (1341,0)^2 / 24 = 1798281 / 24 = 74928,4.$$

Наступний етап розрахувати суму квадратів розсіювання: загального (C_y), повторень (C_p), варіантів (C_v), помилки (C_z).

Сума квадратів розсіювання загального (C_y):

$$C_y = \sum X^2 - \Delta = 75487,0 - 74928,4 = 558,6.$$

Суму квадратів повторень варіантів (C_p) розраховують наступним чином:

$$C_p = \sum P^2 / l - \Delta = (P_1^2 + P_2^2 + P_3^2 + P_4^2) / l - \Delta = \\ (338^2 + 340^2 + 327^2 + 336^2) / 6 - 74928,4 = \\ (114244 + 115600 + 106929 + 112896) / 6 - 74928,4 = \\ 449669 / 6 - 74928,4 = 74944,8 - 74928,4 = 16,4.$$

Суму квадратів варіантів (C_v) розраховують за формулою:

$$C_v = \sum V^2 / n - \Delta = (V_1^2 + V_2^2 + V_3^2 + V_4^2 + V_5^2 + V_6^2) / n - \Delta = \\ (252,0^2 + 225^2 + 217^2 + 235^2 + 212^2 + 200^2) / 4 - 74928,4 = \\ (63504 + 50625 + 47089 + 55225 + 44944 + 40000) / 4 - 74928,4 = \\ 301387 / 4 - 74928,4 = 75346,7 - 74928,4 = 418,3.$$

Після встановлення сум квадратів загального розсіювання (C_y), розсіювання повторень (C_p), розсіювання варіантів (C_v) розраховують суму квадратів помилки (C_z):

$$C_z = C_y - C_p - C_v = 558,6 - 16,4 - 418,3 = 123,9.$$

Наступним етапом дисперсійного аналізу двофакторного досліді є розрахунок сум квадратів окремо для фактору А (обробіток ґрунту), фактору В (попередник озимої пшениці) та їх взаємодії. З цією метою складають таблицю для обчислювання дії цих факторів та їх взаємодії. В таблицю заносять із таблиці 9 суму даних врожайності по кожному варіанту (табл. 4.5).

Таблиця 4.5

**Дані для обчислювання дії факторів та їх взаємодії
у двофакторному досліді**

Спосіб обробітку грунту (фактор А)	Попередник озимої пшениці (фактор В)			Сума даних врожайності за фактором А, ΣV_A
	сидеральний пар	горох	озимий ріпак	
оранка 16-18 см	$V_1 = 252$	$V_3 = 225$	$V_5 = 217$	$\Sigma V_{A1} = 694$
поверхневий обробіток 10-12 см	$V_2 = 235$	$V_4 = 212$	$V_6 = 200$	$\Sigma V_{A2} = 647$
сума даних врожайності за фактором В, ΣV_B	$\Sigma V_{B1} = 487$	$\Sigma V_{B2} = 437$	$\Sigma V_{B3} = 417$	$\Sigma = 1341$

ΣV_{A1} – сума даних врожайності озимої пшениці по оранки (фактор А) за всіма попередниками = 694.

ΣV_{A2} – сума даних врожайності озимої пшениці по поверхневому обробітку ґрунту (фактор А) за всіма попередниками = 647.

ΣV_{B1} – сума даних врожайності озимої пшениці по сидеральному пару (фактор В) за всіма обробітками ґрунту = 487.

ΣV_{B2} – сума даних врожайності озимої пшениці по гороху (фактор В) за всіма обробітками ґрунту = 437.

ΣV_{B3} – сума даних врожайності озимої пшениці по озимому ріпаку (фактор В) за всіма обробітками ґрунту = 417.

На основі даних врожайності озимої пшениці за варіантами, які закладено по двом обробіткам ґрунту за різними попередниками за різними строками сівби, розраховують суму врожайності за факторами А та В:

$$\Sigma = 1341.$$

За наведеною формулою $C_A = \Sigma V_{A1}^2 + \Sigma V_{A2}^2 / (l_B \times n) - \Delta$ обчислюють суму квадратів дослідів для результативного фактору А (сорт):

$$C_A = \Sigma V_{A1}^2 + \Sigma V_{A2}^2 / (l_B \times n) - \Delta = (694^2 + 647^2) / (3 \times 4) - 74928,4 = (481636 + 418609) / 12 - 74928,4 = 900245 / 12 - 74928,4 = 75020,4 - 74928,4 = 92,0.$$

Сума квадратів для фактору В (строк сівби) буде наступною:

$$C_B = (\Sigma V_{B1}^2 + \Sigma V_{B2}^2 + \Sigma V_{B3}^2) / (l_A \times n) - \Delta = (487^2 + 437^2 + 417^2) / (2 \times 4) - 74928,4 = (237169 + 190969 + 173889) / (2 \times 4) - 74928,4 =$$

$$602027 / 8 - 74928,4 = 75253,4 - 74928,4 = 325.$$

Сума квадратів взаємодій факторів А і В (сорт рідини і строк сівби) визначають за формулою:

$$C_{AB} = C_v - C_A - C_B = 418,3 - 92,0 - 325,0 = 1,3.$$

Наступний етап дисперсійного аналізу двофакторного дослідження спрямовано на встановлення кількості ступенів свободи всіх розсіювань:

Кількість ступенів свободи загального розсіювання: $v_y = N - 1 = 24 - 1 = 23$.

Кількість ступенів свободи повторень: $v_p = n - 1 = 4 - 1 = 3$.

Кількість ступенів свободи фактору А (сорт): $v_A = l_A - 1 = 2 - 1 = 1$.

Кількість ступенів свободи фактору В (строк сівби): $v_B = l_B - 1 = 3 - 1 = 2$.

Кількість ступенів свободи взаємодії факторів А і В (сорт і термін сівби):

$$v_{AB} = (l_A - 1) \times (l_B - 1) = (2 - 1) \times (3 - 1) = 1 \times 2 = 2.$$

Кількість ступенів свободи помилки (залишку) дослідження рівняється:

$$v_z = v_y - v_p - v_A - v_B - v_{AB} = 23 - 3 - 1 - 2 - 2 = 15.$$

Після цих розрахунків переходять до етапу встановлення дисперсії факторів А (сорт), В (строк сівби) і їх взаємодії (сорт рідини – термін сівби).

Дисперсія фактору А: $s_A^2 = C_A / v_A = 92,0 / 1 = 92,0$.

Дисперсія фактору В: $s_B^2 = C_B / v_B = 325,0 / 2 = 162,5$.

Дисперсія взаємодії факторів фактору А і В: $s_{AB}^2 = C_{AB} / v_{AB} = 1,3 / 2 = 0,65$.

Дисперсія помилки (залишку) дослідження: $s_z^2 = C_z / v_z = 123,9 / 15 = 8,26$.

Критерії Фішера розраховують наступним чином.

Критерій Фішера для фактору А: $F_A = s_A^2 / s_z^2 = 92,0 / 8,26 = 11,1$.

Критерій Фішера для фактору В: $F_B = s_B^2 / s_z^2 = 162,5 / 8,26 = 19,7$.

Критерій Фішера для взаємодії факторів А і В: $F_{AB} = s_{AB}^2 / s_z^2 = 0,65 / 8,26 = 0,08$.

Складають таблицю дисперсійного аналізу дослідження (табл. 4.6).

Таблиця 4.6

Результати дисперсійного аналізу

Розсіювання	Сума квадратів	Кількість ступенів свободи, v	Дисперсія, s^2	Ffact.	Ftheor.	
					0,95	0,99
загальне	$C_y = 558,6$	$v_y = 23$				
повторень	$C_p = 16,4$	$v_p = 3$				
фактору А	$C_A = 92,0$	$v_A = 1$	$S_A^2 = 92,0$	$F_A = 11,1$	4,54	8,68
фактору В	$C_B = 325$	$v_B = 2$	$S_B^2 = 162,5$	$F_B = 19,7$	3,67	6,36
взаємодії факторів АВ	$C_{AB} = 1,3$	$v_{AB} = 2$	$S_{AB}^2 = 0,65$	$F_{AB} = 0,08$	3,67	6,36
Залишку (помилки)	$C_z = 123,9$	$v_z = 15$	$S_z^2 = 8,26$			

Теоретичне значення критерію Фішера знаходять за даними додатку 2 та додатку 3 і заносять в таблицю 4.6 для порівняння з фактичним критерієм Фішера.

По фактору А (спосіб обробітку ґрунту) теоретичне значення критерію Фішера визначають по кількості ступенів свободи помилки – ($v_z = 15$), в графі кількість ступенів свободи помилку рядок – 15, та за кількістю ступенів свободи фактору – колонка з числом – 1 ($v_A = 1$), в графі кількість ступенів свободи фактору – колонка з числом 1. На перетині цих цифр 15 і 1 знаходять значення на 5%-му рівні значущості для $F_{0,95} = 4,54$ (додаток 2) і на 1%-му рівні значущості для $F_{0,99} = 8,68$ (додаток 3).

По фактору В (попередник озимої пшениці) теоретичне значення критерію Фішера визначають по кількості ступенів свободи помилки – ($v_z = 15$), в графі кількість ступенів свободи помилку рядок – 15, та за кількістю ступенів свободи фактору – колонка з числом – 2 ($v_B = 2$), в графі кількість ступенів свободи фактору – колонка з числом 2. На перетині цих цифр 15 і 2 знаходять значення на 5%-му рівні значущості для $F_{0,95} = 3,67$ (додаток 2) і на 1%-му рівні значущості для $F_{0,99} = 6,36$ (додаток 3).

По взаємодії факторів АВ (спосіб обробітку ґрунту та попередник озимої пшениці) теоретичне значення критерію Фішера визначають по кількості ступенів свободи помилки – ($v_z = 15$), в графі кількість ступенів свободи помилку рядок – 15, та за кількістю ступенів свободи взаємодії факторів – колонка з числом – 2 ($v_A = 2$), в графі кількість ступенів свободи фактору – колонка з числом 2. На перетині цих цифр 15 і 2 знаходять значення на 5%-му рівні значущості для $F_{0,95} = 3,67$ (додаток 2) і на 1%-му рівні значущості для $F_{0,99} = 6,36$ (додаток 3).

По кожному фактору досліду та по взаємодії факторів порівнюють показники фактичного критерію Фішера з теоретичними значеннями критерію Фішера $F_{0,95}$ і $F_{0,99}$ на 5%-му рівні значущості та 1%-му рівні значущості відповідно.

За фактором А порівнюють фактичний (розрахунковий) критерій Фішера $F_{\text{факт.}} = 11,1$ з теоретичними критеріями $F_{0,95} = 4,54$ і $F_{0,99} = 8,68$. Фактичне значення критерію Фішера $F_{\text{факт.}} = 11,1$ більше теоретичного критерію $F_{0,95} = 4,54$ при 5%-ому рівні значущості і теоретичного критерію $F_{0,99} = 8,68$ при 1%-ому рівні значущості. Так як, фактичне значення критерію Фішера більше теоретичних значень критерію при 5%-му рівні значущості $F_{0,95}$ (4,54) і 1%-му рівні значущості $F_{0,99}$ (8,68) роблять висновок, що в досліді при цих рівнях значущості отримано достовірні дані за врожайністю озимої пшениці і в досліді відмічається за дисперсійним аналізом суттєва різниця у даних за врожайністю по обох способах обробітку ґрунту. Дія обробітку ґрунту на врожайність озимої пшениці достовірна. Нульова гіпотеза по фактору А відкидається.

За фактором В порівнюють фактичний (розрахунковий) критерій Фішера $F_{\text{факт.}} = 19,7$ з теоретичними критеріями $F_{0,95} = 3,67$ і $F_{0,99} = 6,36$.

Фактичне значення критерію Фішера $F_{\text{факт.}} = 19,7$ значно більше теоретичного критерію $F_{0,95} = 3,67$ при 5%-ому рівні значущості і теоретичного

критерію $F_{0,99} = 6,36$ при 1%-ому рівні значущості. Так як, фактичне значення критерію Фішера більше теоретичних значень критерію при 5%-му рівні значущості $F_{0,95} (3,67)$ і 1%-му рівні значущості $F_{0,99} (6,36)$ роблять висновок, що в досліді при цих рівнях значущості отримано достовірні дані за врожайністю озимої пшениці і в досліді відмічається за дисперсійним аналізом суттєва різниця у даних за врожайністю по всіх попередниках озимої пшениці. Дія попередників на врожайність озимої пшениці достовірна. Нульова гіпотеза по фактору В відкидається.

За взаємодією факторів АВ порівнюють фактичний (розрахунковий) критерій Фішера $F_{\text{факт.}} = 0,08$ з теоретичними критеріями $F_{0,95} = 3,67$ і $F_{0,99} = 6,36$. Фактичне значення критерію Фішера $F_{\text{факт.}} = 19,7$ значно менше теоретичного критерію $F_{0,95} = 3,67$ при 5%-ому рівні значущості і теоретичного критерію $F_{0,99} = 6,36$ при 1%-ому рівні значущості. Так як, фактичне значення критерію Фішера менше теоретичних значень критерію при 5%-му рівні значущості $F_{0,95} (3,67)$ і 1%-му рівні значущості $F_{0,99} (6,36)$ роблять висновок, що в досліді при цих рівнях значущості взаємодія способу обробітку ґрунту та попередника не достовірна. Для розгляду результатів по взаємодії факторів АВ може бути прийнята нульова гіпотеза.

Наступним етапом статистичної обробки результатів двофакторного досліді є встановлення узагальненої помилки досліді та помилку різниці в досліді.

Розрахунок узагальненої помилки досліді (E).

$$E = \sqrt{\frac{S_z^2}{n}} = \sqrt{\frac{8,26}{4}} = 1,43 \text{ ц/га}$$

Узагальнена помилка по фактору А.

$$E_A = \sqrt{\frac{S_z^2}{l_B \times n}} = \sqrt{\frac{8,26}{3 \times 4}} = 0,83 \text{ ц/га}$$

Узагальнена помилка по фактору В.

$$E_B = \sqrt{\frac{S_z^2}{l_B \times n}} = \sqrt{\frac{8,26}{2 \times 4}} = 1,02 \text{ ц/га}$$

Далі встановлюють помилку різниці середніх значень результативної ознаки (врожайність озимої пшениці) в досліді (s_d):

$$s_d = E \times 1,41 = 1,43 \times 1,41 = 2,00 \text{ ц/га.}$$

де: 1,41 – постійне число квадратний корінь з 2 .

Помилка різниці по фактору А:

$$s_{dA} = E_A \times 1,4 = 0,83 \times 1,4 = 1,16 \text{ ц/га.}$$

Помилка різниці по фактору В:

$$s_{dB} = E_B \times 1,4 = 1,02 \times 1,4 = 1,43 \text{ ц/га.}$$

Найменшу суттєву різницю (НСР) розраховують на двох рівнях довірчої ймовірності на 5%-му рівні значущості та 1% рівні значущості за формулами:

$$НСР_{0,95} = s_d \times t_{0,95}; \quad НСР_{0,99} = s_d \times t_{0,99}.$$

Теоретичне значення t -критерію Стюдента знаходять у додатку 3 за кількістю, ступенів свободи залишкового розсіювання v_z , яке у проведеному досліді за дисперсійним аналізом становить – $v_z = 15$.

У графі додатку 4 для кількості ступенів свободи помилки, в нашому досліді – 15, вибирають значення t -критерію Стюдента на 5%-му рівні довірчої ймовірності (ймовірності помилки) $P_{0,95}$ ($\alpha = 0,05$) та на 1%-му рівні ймовірності помилки $P_{0,99}$ ($\alpha = 0,01$). На перетині цих показників значення t -критерію Стюдента становить $t_{0,95} = 2,12$ для 5%-ти рівня значущості, $t_{0,99} = 2,92$ для 1%-ого рівня значущості.

Розраховують найменшу суттєву різницю НСР для 5%-го рівня довірчої ймовірності ($НСР_{0,95}$) і 1%-го рівня ймовірності ($НСР_{0,99}$).

Всього досліді:

$$НСР_{0,95} = s_d \times t_{0,95} = 2,00 \times 2,12 = 4,24 \text{ ц/га.}$$

$$НСР_{0,99} = s_d \times t_{0,99} = 2,00 \times 2,92 = 5,84 \text{ ц/га.}$$

По фактору А (обробіток ґрунту):

$$НСР_{0,95} = s_{dA} \times t_{0,95} = 1,16 \times 2,12 = 2,46 \text{ ц/га.}$$

$$НСР_{0,99} = s_{dA} \times t_{0,99} = 1,16 \times 2,92 = 3,39 \text{ ц/га.}$$

По фактору В (попередник озимої пшениці):

$$НСР_{0,95} = s_{dB} \times t_{0,95} = 1,43 \times 2,12 = 3,03 \text{ ц/га.}$$

$$НСР_{0,99} = s_{dB} \times t_{0,99} = 1,43 \times 2,92 = 4,17 \text{ ц/га.}$$

Кінцевим етапом дисперсійного аналізу є встановлення відносної помилки досліді:

$$S_{\bar{x}} = 100 \times E / \bar{x} = 100 \times 1,43 / 55,9 = 2,56\%.$$

Відносна помилка досвіду 2,56% свідчить про високу точність проведення досліді.

На основі даних дисперсійного аналізу складають підсумкову таблицю (табл. 4.7).

При застосуванні оранки різниця по врожайності озимої пшениці за фактором А (спосіб обробітку ґрунту) 3,3-4,3 ц/га на рівні довірчої ймовірності $НСР_{0,95}$ перевищує значення НСР 2,46 ц/га при всіх попередниках вирощування. Отже різниця за фактором А суттєва і достовірна.

Різниця по врожайності за фактором А (спосіб обробітку ґрунту) 4,2-4,3 ц/га на рівні довірчої ймовірності $НСР_{0,99}$ перевищує значення 3,39 ц/га на попередниках сидеральний пар та озимий ріпак. Отже різниця за фактором А при вирощуванні озимої пшениці за цими попередниками суттєва і достовірна.

Таблиця 4.7

Підсумкова таблиця дисперсійного аналізу

Фактор А (спосіб обробітку грунту)	Фактор В (попередник озимої пшениці)	Врожайність, ц/га	Різниця по врожайності (ц/га) за факторами		Найменша суттєва різниця (НСР), ц/га	
			фактор А	фактор В	НСР _{0,95}	НСР _{0,99}
оранка 16-18 см	Сидеральний пар	63,0 – x ₁	–	–		
	горох	56,3 – x ₂	–	-6,7 (x ₁ - x ₂)		
	озимий ріпак	54,3 – x ₃	–	-8,7 (x ₁ - x ₃)	4,24	5,84
поверхневий обробіток 10- 12 см	сидеральний пар	58,8 – x ₄	-4,2 (x ₁ - x ₄)	–		
	горох	53,0 – x ₅	-3,3 (x ₂ - x ₅)	-5,8 (x ₄ - x ₅)		
	озимий ріпак	50,0 – x ₆	-4,3 (x ₃ - x ₆)	-8,8 (x ₄ - x ₆)		
НСР _{0,95} НСР _{0,99}			2,46 3,39	3,03 4,17		

При вирощуванні озимої пшениці за різними попередниками (фактор В) різниця по врожайності на оранці 6,7-8,7 ц/га і на поверхневому обробітку ґрунту 5,8-8,8 ц/га на обох рівнях довірчої ймовірності НСР_{0,95} і НСР_{0,99} перевищує значення НСР 2,46 і 4,17 ц/га. Отже різниця за фактором В при двох способах обробітку ґрунту за двома рівнями довірчої ймовірності суттєва і достовірна.

Якщо за розрахунками в досліді фактичне значення критерію Фішера $F_{\text{fact.}}$ більше або дорівнюється ($\geq$) значенню теоретичного критерію Фішера $F_{\text{theor.}}$, то достовірність відмінностей у дослідженні доведена, тобто є одна або кілька пар варіантів, середні арифметичні яких достовірно різняться.

Якщо фактичне значення критерію Фішера $F_{\text{fact.}}$ менше ($<$) теоретичного значення критерію Фішера $F_{\text{theor.}}$, то достовірних відмінностей між варіантами немає.

Зустрічаються випадки, коли фактичне значення критерію Фішера $F_{\text{fact.}}$ лише трохи менше теоретичного значення критерію Фішера $F_{\text{theor.}}$. Строго дотримуючись правила, викладеного вище, можна зробити висновок про відсутність достовірної різниці в досліді. Однак подальший аналіз дозволяє іноді знайти цю різницю хоча б в одному-двох варіантах.

У таких випадках, не обмежуючись розрахунком F , слід продовжити обчислення до найменшої суттєвої різниці (НСР). З цим показником порівнюють різницю між парами варіантів ($d = x_1 - x_2$). Якщо $d \geq \text{НСР}$, то різницю між варіантами, що аналізується в досліді, вважають доведеною. Докази найчастіше ведуть на рівнях довірчої ймовірності $P_{0,95}$ і $P_{0,99}$.

Контрольні питання

1. У чому полягає сутність дисперсійного аналізу?
2. Яка основна мета застосування ANOVA в агрономічних дослідженнях?
3. Які припущення необхідно виконати перед проведенням дисперсійного аналізу?
4. Що характеризує загальна сума квадратів (S_s або C_y)?
5. Як визначають коригувальний фактор (Δ) і яке його призначення?
6. Який фізичний зміст мають факторна та залишкова дисперсії?
7. Як визначають число ступенів свободи для факторів і залишку?
8. За якою формулою обчислюють критерій Фішера?
9. Як інтерпретують результати порівняння $F_{\text{факт}}$ і $F_{\text{теор}}$?
10. Що характеризує взаємодія факторів у двофакторному дисперсійному аналізі?
11. Для чого використовують найменшу істотну різницю (НСР, LSD)?
12. Що таке кореляційне відношення (η) і яке його практичне значення?
13. Які переваги має двофакторний дисперсійний аналіз порівняно з однофакторним?
14. У яких випадках після ANOVA необхідно застосовувати множинні порівняння середніх?

Розрахункові задачі

Задача 1. Однофакторний дисперсійний аналіз (ANOVA)

У польовому досліді вивчали вплив чотирьох норм азотних добрив на урожайність ярого ячменю (ц/га).

Повторення	N_0	N_{30}	N_{60}	N_{90}
1	45	50	56	58
2	47	52	55	60
3	46	51	57	59
4	44	49	56	61

Необхідно:

1. Обчислити середню урожайність кожного варіанта.
2. Визначити загальну середню.
3. Розрахувати коригувальний фактор (Δ).
4. Обчислити:
 - загальну суму квадратів (C_y);
 - суму квадратів фактора (C_v);
 - суму квадратів помилки (C_z).
5. Визначити ступені свободи.

6. Розрахувати середні квадрати.
7. Обчислити критерій Фішера.
8. Зробити висновок про вплив норм азотних добрив на урожайність.

Задача 2. Двофакторний дисперсійний аналіз

Досліджували вплив способу обробітку ґрунту (фактор А) і попередника (фактор В) на урожайність озимої пшениці.

Фактор А: А₁ – оранка; А₂ – поверхневий обробіток.

Фактор В: В₁ – чорний пар; В₂ – горох; В₃ – ріпак.

Кожна комбінація мала чотири повторення.

Необхідно:

1. Обчислити суму врожайності для кожного рівня факторів.
 - Розрахувати: С_А; С_В; С_{АВ}; С_z.
2. Визначити ступені свободи.
3. Обчислити дисперсії факторів.
 - Розрахувати критерії Фішера: F_А; F_В; F_{АВ}.
4. Зробити висновок щодо значущості факторів та їх взаємодії.

Задача 3. Визначення НСР (LSD)

Після однофакторного дисперсійного аналізу отримано:

- дисперсія помилки $s_z^2=4,25$; кількість повторень $n=4n$; ступені свободи помилки $v=12v$; $t_{0,05}=2,18$.

Середні врожайності (ц/га):

Варіант	Урожайність
Контроль	48,2
Добриво А	50,1
Добриво Б	52,4
Добриво В	55,0

Необхідно:

1. Обчислити НСР_{0,05}.
2. Порівняти всі середні між собою.
3. Визначити, які варіанти достовірно відрізняються.
4. Побудувати матрицю статистичних відмінностей.

Задача 4. Кореляційне відношення (η)

У результаті дисперсійного аналізу встановлено: загальна сума квадратів $C_y=680$; сума квадратів, обумовлена фактором, $C_v=476$.

Необхідно:

1. Визначити кореляційне відношення $\eta = \sqrt{\frac{C_v}{C_y}}$
2. Обчислити коефіцієнт детермінації η^2 .
3. Пояснити біологічний зміст отриманих результатів.
4. Оцінити силу впливу фактора.

Задача 5. Повний аналіз результатів польового експерименту

У трифакторному досліді досліджували вплив: фактор А – сорт; фактор В – густота стояння рослин; фактор С – норма азотних добрив. Після статистичної обробки отримано:

Джерело варіації	SS	df
Фактор А	125,4	2
Фактор В	314,6	3
Фактор С	248,8	2
А×В	26,4	6
А×С	12,5	4
В×С	31,7	6
А×В×С	15,2	12
Помилка	168,3	36

Необхідно:

1. Обчислити середні квадрати.
2. Визначити F-критерій.
3. Встановити статистично значущі фактори.
4. Визначити фактор із найбільшим впливом.
5. Запропонувати подальший статистичний аналіз (LSD або тест Тьюкі).
6. Сформулювати науковий висновок.

Тестові завдання

1. Основною метою дисперсійного аналізу є:

- А. Побудова рівняння регресії.
- Б. Оцінювання впливу одного або кількох факторів на результативну ознаку.
- В. Визначення коефіцієнта варіації.
- Г. Побудова гістограми.

2. Який критерій використовують для оцінювання значущості факторів у дисперсійному аналізі?

- А. Критерій Стюдента.
- Б. Критерій Манна-Уїтні.
- В. Критерій Фішера.
- Г. χ^2 -критерій Пірсона.

3. Загальна сума квадратів (SS або S_y) характеризує:

- А. Варіацію лише між варіантами дослідів.
- Б. Варіацію лише між повтореннями.
- В. Загальну мінливість результативної ознаки.
- Г. Лише випадкові похибки.

4. Якщо $F_{\text{факт}} > F_{\text{теор}}$, то:

- А. Фактор не впливає на результативну ознаку.
- Б. Вплив фактора є статистично значущим.
- В. Необхідно повторити експеримент.
- Г. Дані не підлягають аналізу.

5. Яка складова дисперсії характеризує невраховані або випадкові причини мінливості?

- А. Факторна дисперсія.
- Б. Загальна дисперсія.
- В. Залишкова (помилкова) дисперсія.
- Г. Міжгрупова дисперсія.

6. Двофакторний дисперсійний аналіз дозволяє оцінити:

- А. Лише вплив першого фактора.
- Б. Лише вплив другого фактора.
- В. Тільки кореляцію між факторами.
- Г. Вплив кожного фактора окремо та їх взаємодії.

7. Найменшу істотну різницю (HCP, LSD) використовують для:

- А. Перевірки нормальності розподілу.
- Б. Попарного порівняння середніх після дисперсійного аналізу.
- В. Обчислення коефіцієнта варіації.
- Г. Визначення коефіцієнта кореляції.

8. Кореляційне відношення (η) показує:

- А. Ступінь асиметрії розподілу.
- Б. Однорідність дисперсій.
- В. Силу впливу факторіальної ознаки на результативну.
- Г. Точність середньої арифметичної.

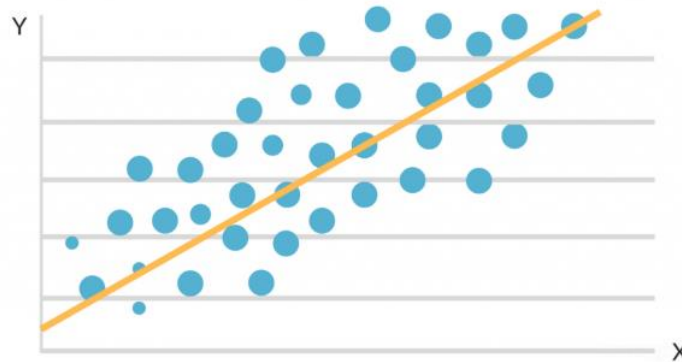
9. Яке значення має кореляційне відношення η ?

- А. Від -1 до $+1$.
- Б. Від $-\infty$ до $+\infty$.
- В. Від 0 до 1 .
- Г. Від 0 до 100 .

10. Яка умова є обов'язковою перед застосуванням параметричного дисперсійного аналізу?

- А. Дані повинні бути лише якісними.
- Б. Вибірка повинна містити не менше 100 спостережень.
- В. Дані мають бути незалежними, а розподіл залишків – наближеним до нормального з однорідними дисперсіями.
- Г. Усі середні значення повинні бути однаковими.

РОЗДІЛ 5. КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ



- 5.1. Застосування та використання кореляційного аналізу
- 5.2. Аналіз парної лінійної кореляції
- 5.3. Парна лінійна кореляція між якісними (альтернативними) ознаками
- 5.4. Аналіз множинної кореляції
- 5.5. Аналіз криволінійної кореляції
- 5.6. Застосування та використання регресійного аналізу
- 5.7. Парна лінійна залежність
- 5.8. Множинна лінійна залежність
- 5.9. Криволінійна залежність

5.1. Застосування та використання кореляційного аналізу

Кореляційно-регресійний аналіз – це сукупність методів, спрямованих на встановлення статичними критеріями суттєвості сили і форми залежності між множинними явищами, ознаками, факторами і оцінка їх взаємозв'язку.

Кореляційно-регресійний аналіз застосовується для вивчення зв'язків і залежності між явищами, що спостерігаються.

Причинно-наслідкові зв'язки при математичному аналізі можуть виражатися у двох формах: функціональній та кореляційній.

Під функціональним зв'язком розуміють такий зв'язок між якими-небудь показниками ознаки рослин, коли за зміни одного показника на певну величину інший показник змінюється також на певну величину.

Необхідно врахувати, що у чистому вигляді функціональні зв'язки зустрічаються лише у ідеальних умовах, коли передбачається, що ніяких сторонніх впливів немає.

Але на практиці це недосяжно. Завжди діють сторонні цієї функціональної залежності чинники, які порушують точність цієї залежності у різних випадках по-різному. Поки такі порушення залишаються настільки незначними, що їх практично не можна враховувати, зв'язок вважається функціональним. При вивченні рослинних об'єктів доводиться мати справу зі зв'язками іншого.

Рослинний організм розвивається у зв'язку з умовами його росту та

розвитку під дією нескінченно великої кількості факторів, які по-різному визначають розвиток різних ознак. У рослин зв'язок між будь-якими двома ознаками настільки часто і сильно порушується і модифікується, що не завжди навіть може бути виявлено.

Рослини у процесі свого розвитку постійно взаємодіють із чинниками довкілля, змінюються під впливом різноманітних умов існування. Тому зв'язок між ознаками рослин проявляється у вигляді кореляційної залежності, чи кореляції.

Ця форма зв'язку характеризується тим, що зміна одного показника на певну величину супроводжується зміною не одного значення показника, а цілий розподіл цих значень при певних визначених основних показниках цього розподілу – середньої величини та ступеня різноманітності.

Аналіз кореляції у науковому дослідженні складається з двох елементів: власне кореляції та регресії.

Розглянемо питання застосування кореляції в статистичній обробці отриманих наукових результатів.

Термін кореляція, який було введено у науку Кюв'є, у перекладі з французької означає співвідношення, зв'язок (принцип співвідношення у кожному рослинному організмі між різними органами).

В досліджах з рослинництва для визначення зв'язку між параметрами ознак, явищ, факторів, що впливають на ріст, розвиток рослин, формування продуктивності сортами та гібридами використовують саме кореляційний аналіз.

У практиці досліджень часто виникає необхідність вивчити характер зв'язку між двома (або більше) ознаками рослин, що варіюють, або вивчити характер зв'язку між властивостями ґрунтів з різним рівнем їх родючості.

Багато ознак і властивостей рослин перебувають між собою у певному взаємозв'язку. Деякі їх є взаємозалежними, інші – змінюються у напрямі під впливом загальних умов.

Між варіюючими ознаками, явищами, об'єктами, умовами середовища, зростанням, продуктивністю рослин та іншими показниками існують певні взаємозв'язки: значення середньої величини однієї ознаки змінюється при зміні іншої ознаки. Коли певному значенню незалежної змінної X відповідає кілька значень іншої залежної змінної ознаки Y , залежність набуває стохастичного характеру.

Взаємозв'язки між ознаками, що варіюють, називають кореляцією.

При аналізі отриманих даних дослідів у рослинництві науковий співробітник зустрічається із трьома ситуаціями:

- між величинами досліджуваних ознак та властивостей X та Y зв'язок відсутній;
- між величинами ознак та властивостей X та Y існує функціональний зв'язок (коли кожному значенню однієї величини відповідає строго певне значення іншої величини);
- між величинами ознак та властивостей X та Y існує стохастичний (імовірнісний, кореляційний) зв'язок (коли кожному значенню однієї величини

ознаки або властивості відповідає безліч можливих значень іншої величини).

Кореляційний аналіз застосовують саме для вивчення стохастичного зв'язку між показниками ознак або властивостей, які вивчаються в досліді. Стохастичний зв'язок – зв'язок особливого роду, при якому зі зміною однієї величини ознаки змінюється розподіл величини іншої ознаки. Такий зв'язок між величинами ознак або властивостей називається стохастичним.

Таким чином, якщо при визначенні значення X змінна Y може набувати різних значень з певною ймовірністю, то залежність між X і Y носить характер кореляційного або стохастичного зв'язку.

Найпростішим візуальним способом виявити наявність взаємозв'язку між кількісними змінними є побудова діаграми розсіювання (scatterplot). Це графік, на якому по горизонтальній осі (X) відкладається одна змінна, по вертикальній осі (Y) інша змінна.

Змінні X та Y – це ті показники величин ознак, властивостей, що можна вимірювати, контролювати або що можна змінювати в дослідженнях. Змінні X та Y відрізняються багатьма аспектами, особливо тією роллю, яку вони грають у дослідженнях, шкалою виміру тощо.

У період проведення дослідження та аналізу отриманих результатів дослідів науковий співробітник (дослідник) встановлює, що відбувається з явищем, яке його цікавить (з y –результативною ознакою або функцією відгуку), якщо в досліді змінюється незалежна змінна величина ознаки або властивості (x –факторна ознака). Зазвичай неявно передбачається, що зміна величини ознаки x є причина зміни показника ознаки y .

Вважається, що показники ознак або властивостей повинні мати достатній рівень варіабельності, який може характеризуватися коефіцієнтом варіації (зазвичай не менше 0,1).

При розрахунках кореляційних зв'язків передбачається, що між результативною та факторною ознаками існує залежність типу $y=f(x)$, де кожному припустимому значенню однієї змінної відповідає одне і лише одне значення іншої змінної. Така залежність називається функціональною, вона виявляється за допомогою строгих логічних доказів і не потребує досвідченої перевірки.

Якщо $y = const$ при зміні x , то кажуть, що величина ознаки y не залежить від величини ознаки x .

В той же час, реальні спостереження в силу неврахованих факторів під час виконання дослідів відрізняються від теоретичного значення y на величину ε – випадкову помилку: $y=f(x)+\varepsilon$.

У дослідженні кореляцій (встановлення залежностей, зв'язків) дослідник не впливає (або принаймні намагається не впливати) на змінні X та Y , а тільки вимірює параметри змінних та намагається знайти залежності (кореляції) між деякими визначеними змінними.

Наприклад, дослідник намагається встановити залежність та зв'язок між запасами продуктивної вологи ґрунту (змінна X) та рівнем врожайності агроценозу кукурудзи (змінна Y).

В експериментальних дослідженнях, навпаки, дослідник варіює деякі змінні

та вимірює вплив цих змін на інші змінні. Наприклад, дослідник може штучно змінювати на зрошенні вміст продуктивної вологи (змінна X) за рахунок проведення поливів, а далі на різних запасах продуктивної вологи у ґрунті визначати рівень врожайності кукурудзи (змінна Y).

Аналіз даних в експериментальному дослідженні також приходить до обчислення «кореляцій» (залежностей) між змінними, а саме між змінними, на які впливають, і змінними, на які впливає цей вплив. Проте експериментальні дані потенційно забезпечують нас більш якісною інформацією. Тільки експериментально можна переконливо довести причинний зв'язок між змінними.

Наприклад, якщо виявлено, що кожного разу, коли змінюється змінна X , змінюється і змінна Y , можна зробити висновок – «змінна X впливає на змінну Y », тобто між змінними X і Y є причинна залежність.

Кореляції поділяють за формою, напрямом, та кількістю, силою зв'язків між ознаками рослин (рис. 5.1).

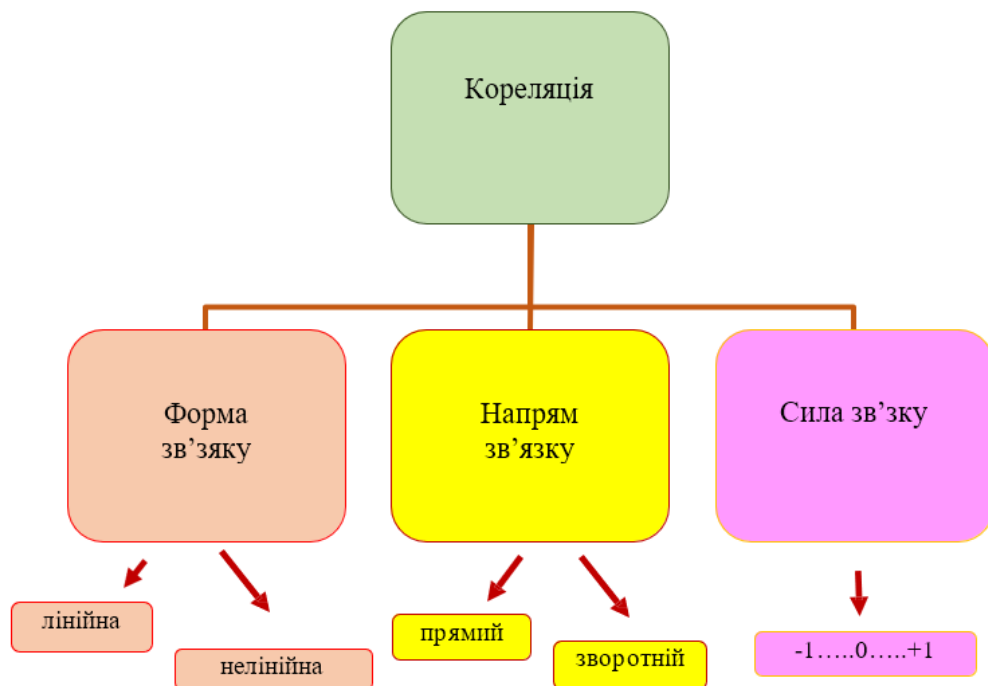


Рис. 5.1. Характер зв'язку між змінними.

Лінійна кореляція за формою зв'язку між ознаками – зміна кількісного параметра незалежної ознаки рослини (*незалежна змінна X*) призводить до зміни величини, кількісного параметра залежної результативної ознаки рослини (*залежна змінна Y*). Наприклад, площа листків гібридів або сортів кукурудзи зростає за одночасним збільшенням їх довжини; врожайність гібриду соняшнику зростає при збільшенні кількості виконаного насіння у кошику; інтенсивність ростових процесів сорту або гібриду зростає при збільшенні площі живлення рослин. Лінійний зв'язок виражається коефіцієнтом кореляції r .

Нелінійна (криволінійна) кореляція за формою зв'язку між ознаками – однакові кількісні зміни незалежної ознаки рослини (*незалежна змінна X*)

призводять до неоднакових кількісних змін величини залежної результативної ознаки рослини (*залежна змінна Y*).

При криволінійній кореляції значення ознак *X* та *Y* змінюється спочатку в одному напрямку, а потім у протилежному напрямку. Наприклад, при постійно зростаючих нормах внесення мінерального або органічного добрива (*незалежна змінна X*) врожайність кормового буряку (*залежна змінна Y*) спочатку має тенденцію к зростанню, далі стабілізується, а після подальшого збільшення внесення добрив (зростанні *незалежної змінної X*) знижується.

Нелінійний зв'язок між ознаками виражається кореляційним відношенням η (буква «ета»).

Лінійна кореляція за напрямом зв'язків між ознаками є прямою та зворотною.

Пряма (позитивна) лінійна кореляція спостерігається, коли зі збільшенням значення незалежної ознаки (*незалежна змінна X*) збільшується значення залежної, результативної ознаки (*залежна змінна Y*).

При прямій (позитивній) лінійній кореляції:

- вищим значенням однієї ознаки відповідають вищі значення іншої;
- більше низьким значенням однієї ознаки відповідають низькі значення іншої ознаки.

При позитивному відхиленні величини незалежної змінної *X* від своєї середньої, величина залежна змінної *Y* також здебільшого відхиляється в позитивну сторону від своєї середньої.

Приклади прямої кореляції: збільшення запасів продуктивної вологи ґрунту в період цвітіння рослин кукурудзи в посушливих умовах обумовлює зростання врожайності зерна у кукурудзи; збільшення маси зерна однієї рослини озимої пшениці обумовлює зростання врожайності всього агроценозу.

Зворотня (негативна) лінійна кореляція – вищим значенням однієї ознаки відповідають нижчі значення іншої, а більше низьким значенням однієї ознаки – високі значення іншої ознаки рослин.

При зворотній кореляції зі збільшенням або зменшенням кількісного значення незалежної ознаки *X* значення результативної ознаки *Y* відповідно зменшується або збільшується. Наприклад, зростання лущинності лузги у соняшнику або у сафлору обумовлює зниження вмісту олії в насінні; зниження щільності орного шару ґрунту обумовлює зростання врожайності гібрида кукурудзи; зниження вмісту білка в насінні соняшнику, кунжуту призводить до збільшення вмісту олії в насінні; при постійному збільшенні маси коренеплодів цукрового буряку зменшується вміст цукру в коренеплодах.

Показником напрямку зв'язку між ознаками є знак коефіцієнта кореляції:

- пряма (позитивна) лінійна кореляція має знак + (плюс);
- зворотня (негативна) лінійна кореляція має знак – (мінус).

Позитивне значення свідчить у тому, що зміни однієї величини ознаки відповідають змінам величини іншої ознаки, тобто із збільшенням значення однієї ознаки відбувається збільшення іншої ознаки, а зв'язок між ознаками – прямий. При негативній кореляції (зворотна залежність), зі збільшенням

значення однієї величини ознаки інша зменшується, і з зменшенням – збільшується величина іншої ознаки.

Кореляція за кількісною мірою сили та напряму імовірного взаємозв'язку двох змінних приймає значення в діапазоні від -1 до $+1$.

Сила зв'язку не залежить від її спрямованості та визначається по абсолютному значенню коефіцієнта кореляції.

Основний показник ступеня сполученості між значеннями однієї та іншої ознаки та форми зв'язку – коефіцієнт кореляції (r).

Коефіцієнт кореляції (r) – безрозмірна величина, що змінюється в межах від -1 до $+1$. При $r = 0$ лінійний зв'язок відсутній, при $r = \pm 1$ кореляційний зв'язок перетворюється на функціональний.

Якщо коефіцієнт кореляції дорівнює 0 (нуль), обидві змінні характеризуються лінійною незалежністю одна від одної.

Коефіцієнт кореляції 1 (один) показує повний (функціональний) лінійний позитивний взаємозв'язок між ознаками.

Коефіцієнт кореляції -1 (мінус один) означає повний (функціональний) лінійний зворотний взаємозв'язок між ознаками.

Коефіцієнт кореляції 0 (нуль) означає відсутність лінійної кореляції (але не обов'язково взаємозв'язку).

Насправді завжди виходять проміжні значення.

При встановленні зв'язків за розрахунками коефіцієнт кореляції характеризується наступниками властивостями:

- коефіцієнт кореляції незалежних величин ознак чи некорельованих величин дорівнює нулю;
- коефіцієнт кореляції не змінюється від додатку до X або Y будь-яких постійних (невипадкових) доданків, від множення їх на позитивні числа;
- якщо одну з випадкових величин, не змінюючи іншої, помножити на -1 , то на -1 множиться і коефіцієнт кореляції;
- чисельно коефіцієнт кореляції знаходиться у межах -1 до $+1$. Якщо коефіцієнт кореляції відмінний від нуля, він своєю величиною характеризує як наявність, так й силу стохастичного зв'язку між величинами ознак X і Y . Чим більша абсолютна величина коефіцієнта кореляції, тим сильніша кореляція між величинами ознак X і Y . Максимальна кореляція відповідає показникам коефіцієнта кореляції -1 та $+1$. Це можливо, коли між випадковими величинами існує чіткий функціональний зв'язок;
- якщо коефіцієнт кореляції > 0 , то величини X та Y з точністю до випадкових похибок одночасно зростають або зменшуються, якщо коефіцієнт кореляції < 0 , то зі зростанням однієї величини інша зменшується. Але це справедливо лише для лінійної залежності Y від X . Тобто, залежність між величинами ознак X і Y може бути строго функціональною (наприклад, квадратичною) без сліду випадковості, а коефіцієнт кореляції все ще буде меншим за 1 ;
- коефіцієнт кореляції є показник того, наскільки зв'язок між випадковими величинами близька до строгої лінійної залежності. Він однаково відзначає і надто велику частку випадковості, і надто велику криволінійність

цього зв'язку;

○ якщо заздалегідь, із загальних міркувань, дослідник може можна передбачити лінійну залежність, то коефіцієнт кореляції є достатнім показником тісноти зв'язку між X та Y ;

Фактичний коефіцієнт кореляції використовується тільки для опису лінійної залежності за умови двовимірного нормального розподілу досліджуваних ознак та характеризує взаємозв'язок ознак лише у цій досліджуваній сукупності.

За кількістю зв'язків між ознаками кореляція може бути:

- простою, коли існує зв'язок між двома ознаками.
- множинною, коли існує зв'язок між трьома і більше ознаками.

Кореляція, яка встановлюється між двома змінними параметрами ознак при постійному значенні параметра третьої ознаки, називається приватною.

Приклад множинної кореляції: врожайність сільськогосподарської культури залежить від норма внесення азоту, фосфору, калію, норм поливної води на зрошенні та інших факторів.

Для встановлення наряду та тісноти залежності між змінними (ознаками) обчислюють парний чи множинний коефіцієнти кореляції.

За силою зв'язку кореляція буває повною, сильною, середньою, слабкою (табл. 5.1).

Таблиця 5.1

Сила зв'язку між ознаками

Значення коефіцієнту кореляції (модуль)	Інтерпретація
0	кореляція відсутня
$r < 0,3$	слабка кореляція
від 0,31 до 0,70	середня кореляція
від 0,71 до 0,90	сильна кореляція
від 0,91 до 0,99	дуже висока кореляція
1,00	повна кореляція

Чим вище за абсолютною величиною коефіцієнт кореляції (модуль), тим сильніший зв'язок між ознаками.

Прийнято вважати, що коефіцієнти кореляції, які за модулем більше 0,70, обумовлюють сильний зв'язок (при цьому коефіцієнти детермінації $> 50\%$, тобто одна ознака визначає іншу більш ніж наполовину).

Коефіцієнти кореляції, які за модулем менше 0,70, але більше 0,50, показують посилення між ознаками середньої сили (при цьому коефіцієнти детермінації менше 50%, але більше 20%).

Нарешті, коефіцієнти кореляції, які за модулем менше 0,30, показують слабкий зв'язок між ознаками (коефіцієнти детермінації менше 20%).

Завдання при проведенні дослідження по встановленню кореляційного зв'язку спрямовано на визначення характеру і вимірювання тісноти взаємозв'язку між ознаками, одна з яких є факторіальною ознакою, інша є

результативною ознакою.

Наприклад, існує певна кореляція між кількістю листя на рослині у кукурудзи (факторіальна ознака) та тривалістю вегетаційного періоду (результативна ознака): чим більше листя утворює рослина, тим пізніше вона дозріває, тим більше має тривалість вегетаційного періоду.

Перед початком проведення кореляційного аналізу, дослідник повинен виділити умови його застосування та обмеження на проведення аналізу:

- ✓ кореляційний зв'язок не слід розраховувати при значенні коефіцієнта кореляції понад 30% у кожному з рядів спостережень;

- ✓ кількість спостережень має бути не менше ніж у 5-6 разів більше від кількості факторів;

- ✓ необхідно, щоб сукупності за результативними та факторіальними ознаками підпорядковувалися нормальному закону розподілу ймовірностей;

Дослідження кореляції зводиться до такого.

По-перше. Встановлюють факт залежності змін однієї ознаки від зміни іншої та визначають форму зв'язку між ними (тип кореляції). Кореляцію називають простою, якщо досліджується зв'язок між двома ознаками, або множинною, коли на величину однієї результативної ознаки впливають кілька факторіальних.

По-друге. Знаходять тісноту зв'язку ознак, тобто. ступінь сполученості між значеннями однієї та іншої ознаки.

По-третє. Залежно від характеру змін результативної ознаки під факторіальним впливом розрізняють такі форми кореляції:

а) лінійна кореляція – зі збільшенням середнього значення однієї ознаки також збільшується середнє значення іншого, або зі збільшенням середнього значення однієї ознаки зменшується середнє значення іншого. У першому випадку кореляцію називають прямою, у другому – зворотною;

б) криволінійна кореляція – при зростанні значення однієї ознаки іншої приймає значення, що зростають до певної величини, а потім спадають, або навпаки.

Таким чином:

- ✓ вивчення зв'язків між змінними, цікавить дослідника з погляду відображення відповідних причинно-наслідкових відносин між ознаками;

- ✓ кореляційна залежність – це узгоджені зміни двох (парний кореляційний зв'язок) або більшої кількості ознак (множинний кореляційний зв'язок);

- ✓ суть кореляційної залежності між ознаками полягає в тому, що при зміні значення однієї змінною відбувається закономірна зміна (зменшення або збільшення) іншої (-их) змінної (-их);

- ✓ кореляційний аналіз – статистичний метод, що дозволяє з використанням коефіцієнтів кореляції визначити, чи існує залежність між змінними та наскільки сильна залежність між ознаками;

- ✓ коефіцієнт кореляції – двовимірна описова статистика, кількісна міра взаємозв'язку (спільної мінливості) двох змінних.

5.2. Аналіз парної лінійної кореляції

Кореляція називається парною, якщо на величину результативної ознаки впливає одна факторіальна ознака. Чисельним заходом парної кореляції при лінійному зв'язку є коефіцієнт кореляції. Коефіцієнт кореляції розраховують за формулою:

$$r = \frac{\Sigma(X - \bar{x}) \times (Y - \bar{y})}{\sqrt{\Sigma(X - \bar{x})^2 \times (Y - \bar{y})^2}}$$

де:

r – коефіцієнт кореляції;

X – показник факторної незалежної ознаки;

Y – показник результативної залежної ознаки;

$\bar{x}$ – середня для показників факторної незалежної ознаки X ;

$\bar{y}$ – середня для показників результативної залежної ознаки Y .

Необхідно підкреслити, що величина коефіцієнта кореляції ще не гарантує його значущості: навіть великий коефіцієнт може стати статистично незначимим (наприклад, при малому обсязі вибірки), а невеликий коефіцієнт (якщо вибірка велика) – значущим.

Слід також застерегти дослідника від зайвого зловживання математичним описом аналізованих залежностей, якщо природа зв'язку невідома і теоретично виправданих припущень про вид рівняння немає. Звичайно, процес відбору може завершитися знаходженням рівняння, яке непогано описує явище, що вивчається. Але не слід забувати, що рівняння, що не має фізичного сенсу, може загасати дуже важливі особливості явища або процесу.

Слід пам'ятати, що математичний вираз лише тоді має користь, коли дозволяє додатково розкрити якісь особливості процесу або допомагає прогнозувати процес, що протікає в аналогічних умовах.

На основі положень про кореляційний зв'язок між показниками ознак в наведеному прикладі виконаємо кореляційний аналіз на основі аналізу залежності між кількістю насіння в кошику соняшнику та масою насіння в кошику лінії відновлювача фертильності пилку GE-3-38Rf.

Після збирання в лабораторних умовах було виконано роботу по встановленню продуктивності рослин за такими показниками – кількість насіння в одному кошику, маса насіння в одному кошику.

Отримані дані заносяться в таблицю для подальшого виконання кореляційного аналізу (табл. 5.2).

Таблиця 5.2

**Дані для розрахунку кореляційної залежності між ознаками
кількість насіння у кошику (шт.) та маса насіння кошику (г)
лінії відновлювача фертильності пилку соняшнику GE-3-38Rf**

№ рослини, кореляючих пар	Кількість насіння в кошику (X), шт.	Маса насіння кошику (Y), г	$X-\bar{x}$	$Y-\bar{y}$	$(X-\bar{x}) \times (Y-\bar{y})$	$(X-\bar{x})^2$	$(Y-\bar{y})^2$
1	706	26,0	-140,1	-9,4	1316,94	19628,01	88,36
2	676	26,9	-170,4	-8,5	1448,4	29036,16	72,25
3	1007	46,3	160,6	10,9	1750,54	25792,36	118,81
4	1019	42,7	172,6	7,3	1259,98	29790,76	53,29
5	874	40,2	27,6	4,8	132,48	761,76	23,04
6	857	33,8	10,6	-1,6	-16,96	112,36	2,56
7	896	46,2	49,6	10,8	535,68	2460,16	116,64
8	565	21,4	-281,4	-14,0	3939,6	79185,96	196
9	794	34,5	-52,4	-0,9	47,16	2745,76	0,81
10	494	20,1	-352,4	-15,3	5391,72	124185,76	234,09
11	872	34,5	25,6	-0,9	-23,04	655,36	0,81
12	891	37,6	44,6	2,2	98,12	1989,16	4,84
13	877	36,2	30,6	0,8	24,48	936,36	0,64
14	546	20,4	-300,4	-15,0	4506	90240,16	225,0
15	1179	47,9	332,6	12,5	4032,5	110622,76	156,25
16	755	32,6	-91,4	-2,8	255,92	8353,96	7,84
17	1363	53,0	516,6	17,6	9092,16	266875,56	309,76
18	1074	49,3	227,6	13,9	3858,64	51801,76	193,21
19	1046	43,0	199,6	7,6	1516,96	39840,16	57,76
20	437	16,0	-409,4	19,4	-9513,76	167608,36	376,36
	$\bar{x} = 846,4$	$\bar{y} = 35,4$			$\sum(X-\bar{x}) \times (Y-\bar{y}) = 29653,52$	$\sum(X-\bar{x})^2 = 1052623$	$\sum(Y-\bar{y})^2 = 2238,32$

п – кількість рослин в аналізі або кількість кореляючих пар = 20. Використовується для розрахунків середньої.

Після розрахунків та занесення даних в таблицю розраховують коефіцієнт кореляції зв'язку між кількістю насіння в кошику та масою насіння в кошику для даної самозапиленої лінії її соняшнику:

$$r = \frac{29653,52}{\sqrt{2356107113,36}} = +0,61$$

Розрахунками встановлено коефіцієнт кореляції між кількістю насіння у кошику та масою насіння кошику лінії відновлювача фертильності пилку GE-3-38Rf – $r = +0,61$.

Наступним етапом розрахунків є встановлення помилки коефіцієнта кореляції за формулою:

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

де:

s_r – помилка коефіцієнта кореляції; r – коефіцієнт кореляції;

n – кількість корелюючих пар або рослин.

$$s_r = \sqrt{(1 - 0,61^2) / (20 - 2)} = \sqrt{1 - 0,37 / 18} = 0,035.$$

В даному випадку помилка коефіцієнта кореляції дорівнюється 0,035.

Критерій достовірності коефіцієнта кореляції при розрахунках в досліді буде рівнятися:

$$t_r = r / s_r = 0,61 / 0,035 = 17,4.$$

Теоретичне значення критерію Стюдента знаходять за кількістю ступенів свободи $v_r = n - 2 = 20 - 2 = 18$. Відповідно даних додатку 4 при ступені свободи 18 критерій Стюдента за рівнем значущості αF_{095} становить 2,26, за рівнем значущості αF_{099} становить 3,25.

Про достовірність зв'язків між ознаками досліді роблять висновок: *фактичний критерій достовірності коефіцієнта кореляції t_r більший теоретичного критерію t . Встановлений зв'язок достовірний.*

Після визначення коефіцієнту кореляції встановлюють коефіцієнт (індекс) детермінації D . Коефіцієнт детермінації D показує частку впливу величини факторної ознаки x на величину результативної ознаки y (змінна y). Індекс детермінації дорівнюється квадрату коефіцієнта кореляції:

$$D = r^2 \times 100\% = +0,61^2 \times 100\% = 0,37 \times 100\% = 37,0\%.$$

Аналізуємо величину коефіцієнта кореляції.

Чим ближче значення коефіцієнта кореляції до нуля, тим ступінь зв'язку між ознаками слабше.

Чим ближче до одиниці, тим зв'язок сильніше. Якщо $r=1$, то зв'язок функціональний.

Якщо $r=0$, то зв'язок між ознаками відсутній.

Про силу зв'язків роблять висновок згідно з показниками таблиці 13:

- якщо коефіцієнт кореляції дорівнює одиниці, то зв'язок повний;
- якщо коефіцієнт кореляції становить від 0,91 до 0,99 між ознаками дуже висока кореляція;
- якщо коефіцієнт кореляції від 0,71 до 0,90 між ознаками сильна кореляція;
- якщо коефіцієнт кореляції від 0,31 до 0,70 між ознаками середня кореляція;
- якщо коефіцієнт кореляції менше 0,30 між ознаками слабка кореляція
- якщо коефіцієнт кореляції має показники до 0,20 відмічається дуже слабка кореляція між ознаками.

Коефіцієнт кореляції також показує напрямок зв'язку між ознаками. Якщо коефіцієнт кореляції $r > 0$ і число має знак «+», то кореляційний зв'язок між ознаками прямий. Якщо коефіцієнт кореляції $r < 0$ і число має знак «-», то кореляційний зв'язок зворотний.

Висновки.

1. Коефіцієнт кореляції $r = +0,61$.
 2. Фактичний критерій Стюдента (t_r) рівняється 17,4, що більше теоретичного критерію Стюдента за рівнем значущості $\alpha F_{095} t = 2,09$ та за рівнем значущості $\alpha F_{099} t = 2,86$. Таким чином кореляційний зв'язок між ознакою кількість насіння в кошику та ознакою маса насіння кошику істотний (достовірний) на найвищих рівнях довірчої ймовірності.
 3. В досліді вивчали кореляційну залежність між двома ознаками. Форма зв'язку (тип кореляції) між двома ознаками, в залежності змін результативної ознаки від факторіальної ознаки, проста.
 4. Сила кореляційного зв'язку між ознаками лінії відновлювача фертильності пилку GE-3-38Rf кількість насіння у кошику (шт.) та маса насіння кошику (г) має середню кореляцію.
 5. Коефіцієнт кореляції між ознаками має позитивний знак «плюс». Кореляція пряма.
 6. Між ознаками лінії відновлювача фертильності пилку GE-3-38Rf – кількість насіння кошику та маса насіння кошика існує середній, лінійний, позитивний взаємозв'язок. Зростання кількості насіння в кошику сприяє зростанню маси насіння кошику.
 7. Кореляція між факторіальною ознакою кількість насіння кошику та результативною ознакою маса насіння кошику лінії відновлювача фертильності пилку GE-3-38Rf – проста, пряма, середня, лінійна.
 8. Так як Індекс детермінації дорівнюється 37,0%, частка впливу (зміни) на формування результативної ознаки (маса насіння кошику) на 37,0% визначається факторною ознакою (кількістю насіння в кошику лінії соняшнику). Частка впливу (зміни) на результативну ознаку на 67,0% визначається іншими ознаками, факторами: виконаність насіння, маса 1000 насінин і так далі.
 9. При доборі рослин лінії відновлювача фертильності пилку GE-3-38Rf за ознакою маса насіння кошику одночасно за ознакою кількість насіння в кошику повинно враховуватися інші ознаки та фактори середовища.
- Розглянемо інший приклад, заснований на визначенні кореляційного взаємозв'язку між показниками щільності орного шару ґрунту та вологості ґрунту, які є важливими факторами при застосуванні способів основного обробітку та визначенні їх впливу на продуктивність сортів і гібридів сільськогосподарських культур.
- Після визначення у досліді показників щільності та вологості ґрунту, дані показників водно-фізичних властивостей ґрунту заносять в таблицю для проведення розрахунків (табл. 5.3).

Таблиця 5.3

Дані для розрахунку кореляційної залежності між показниками щільності орного шару ґрунту (X) та вологістю ґрунту перед сівбою кукурудзи (Y)

№ корелюючих пар	Щільність орного шару ґрунту (X), г/см ³ .	Вологість ґрунту (Y), %	$X - \bar{x}$	$Y - \bar{y}$	$(X - \bar{x}) \times (Y - \bar{y})$	$(X - \bar{x})^2$	$(Y - \bar{y})^2$
1	1,07	17,4	-0,02	2,0	-0,04	0,0004	4,0
2	1,06	14,1	-0,03	-0,7	0,021	0,0009	0,49
3	1,03	15,7	-0,06	0,3	-0,018	0,0036	0,09
4	1,13	15,5	0,04	0,1	0,004	0,0016	0,01
5	1,10	16,2	0,01	0,8	0,008	0,0001	0,64
6	1,12	15,2	0,03	-0,2	-0,006	0,0009	0,04
7	1,04	14,5	-0,05	-0,9	0,045	0,0025	0,81
8	1,03	23,0	-0,06	7,6	-0,456	0,0036	57,76
9	1,05	14,0	-0,04	-1,4	0,056	0,0016	1,96
10	1,12	12,1	0,03	-3,3	-0,099	0,0009	10,89
11	1,25	11,1	0,16	-4,3	-0,688	0,0256	18,49
	$\bar{x} = 1,09$	$\bar{y} = 15,4$			$\sum(X - \bar{x}) \times (Y - \bar{y}) = -1,173$	$\sum(X - \bar{x})^2 = 0,0417$	$\sum(Y - \bar{y})^2 = 95,18$
n – кількість корелюючих пар = 11. Використовується для розрахунків середньої.							

Далі розраховується коефіцієнт кореляції за формулою:

$$r = \frac{-1,173}{\sqrt{0,0417 \times 95,18}} = -0,59$$

Коефіцієнт кореляції між щільністю орного шару ґрунту та наявністю води у ґрунті перед сівбою кукурудзи за проведеними розрахунками рівняється -0,59.

Наступним етапом розрахунків є встановлення помилки коефіцієнта кореляції за формулою:

$$s_r = \sqrt{\frac{(1 - (-0,59)^2)}{(11 - 2)}} = 0,072$$

В даному випадку помилка коефіцієнта кореляції дорівнюється 0,072.

Критерій достовірності коефіцієнта кореляції при розрахунках в досліді буде рівнятися:

$$t_r = r / s_r = -0,59 / 0,072 = 8,19.$$

Теоретичне значення критерію Стюдента знаходять за кількістю ступенів свободи $v_r = n - 2 = 11 - 2 = 9$. Відповідно даних додатку 4 при ступені свободи 9 критерій Стюдента за рівнем значущості α $F_{0,95}$ становить 2,09, за рівнем

значущості α $F_{0,99}$ становить 2,86.

Достовірність зв'язків між ознаками в досліді роблять на основі правила: якщо фактичний критерій достовірності коефіцієнта кореляції t_r більший теоретичного критерію t , то встановлений зв'язок достовірний.

Коефіцієнт (індекс) детермінації D дорівнюється:

$$D = r^2 \times 100\% = -0,59^2 \times 100\% = 0,35 \times 100\% = 35,0\%.$$

Аналізуємо величину коефіцієнта кореляції.

Силу зв'язку між ознаками встановлюють відповідно до даних таблиці 13: якщо коефіцієнт кореляції від 0,51 до 0,70 між ознаками існує середня кореляція.

В досліді коефіцієнт кореляції між ознаками $r < 0$ і має знак «-», тому кореляційний зв'язок між щільністю орного шару ґрунту і вологістю ґрунту перед сівбою кукурудзи зворотний.

Висновки.

1. Коефіцієнт кореляції $r = -0,59$.
2. Фактичний критерій Стюдента (t_r) рівняється 8,19. Він більше теоретичного критерію Стюдента, який за рівнем значущості α $F_{0,95}$ становить $t=2,09$, за рівнем значущості α $F_{0,99}$ становить $t=2,86$. Таким чином кореляційний зв'язок між факторіальною ознакою щільність орного шару ґрунту та результативною ознакою вологість ґрунту перед сівбою кукурудзи істотний (достовірний) на найвищих рівнях довірчої ймовірності.
3. В досліді вивчали кореляційну залежність між двома ознаками. Форма зв'язку (тип кореляції) між двома ознаками, в залежності змін результативної ознаки від факторіальної ознаки, проста.
4. Сила кореляційного зв'язку між ознаками щільність орного шару ґрунту та вологістю ґрунту перед сівбою кукурудзи має середню кореляцію.
5. Коефіцієнт кореляції між ознаками має негативний знак «мінус». Кореляція зворотна.
6. Між ознаками щільність орного шару ґрунту та вологістю ґрунту перед сівбою кукурудзи існує середній лінійний негативний взаємозв'язок. Зростання щільність орного шару ґрунту сприяє зниженню вологості ґрунту.
7. Кореляція між факторіальною ознакою щільність орного шару ґрунту та результативною ознакою вологістю ґрунту перед сівбою кукурудзи за такими показниками щільності ґрунту – проста, зворотна, середня, лінійна.
8. Так як Індекс детермінації дорівнюється 35,0%, частка впливу (зміни) на формування результативної ознаку (вологість ґрунту перед сівбою кукурудзи) на 37,0% визначається факторною ознакою (щільність орного шару ґрунту). Частка впливу (зміни) на результативну ознаку на 65,0% визначається іншими ознаками, факторами: фізичні властивості орного шару ґрунту, кількістю випадіння опадів, погодними умовами і так далі.

5.3. Парна лінійна кореляція між якісними (альтернативними) ознаками

При проведенні досліджень набуває актуальності визначення кореляційних зв'язків для цінних агрономічних ознак, встановлення кореляційного зв'язку між альтернативними ознаками рослин або встановлення залежності між якісними та кількісними показниками.

Для вирішення подібних завдань та розрахунку коефіцієнта кореляції досліді застосовують **формулу Юла (Yule)**.

Розглянемо приклад встановлення сили зв'язку між альтернативними ознаками рослин сояшнику. Так, ураження рослин сояшнику хворобою як іржа (*Russinia helianthi*) призводить до зниження врожайності. Для зниження чисельності ураження рослин іржею у технології вирощування сояшника застосовують передпосівну обробку насіння фунгіцидами. З метою зниження ураження рослин іржею насіння перед сівбою обробили фунгіцидом **Амістар Екстра**. Для цього науковому співробітнику необхідно встановити ефективність застосування даного фунгіциду та зв'язок між дією вибраного фунгіциду **Амістар Екстра** та ураженням рослин сояшнику іржею. В період вегетації науковий співробітник виконує підрахунок уражених та неуражених іржею рослин сояшнику. Кількість неуражених рослин (+) і уражених рослин (-) рослин із обробленого фунгіцидом варіанту насіння (+) та необробленого фунгіцидом варіанту насіння (-) заноситься у спеціальну відведену для цих розрахунків таблицю (табл. 5.4).

Таблиця 5.4

Дія фунгіциду Амістар Екстра при передпосівному обробітку насіння сояшнику на ураження рослин в період вегетації

Рослини	Насіння оброблено перед посівом фунгіцидом Амістар Екстра (+)	Насіння перед сівбою не обробляли фунгіцидом (-)	Сума по строках
Рослини не уражені іржею (+)	132 (n ₁)	59 (n ₂)	191 (N ₁)
Рослини уражені іржею (-)	27 (n ₃)	110 (n ₄)	137 (N ₂)
Сума за стовбцями	159 (N ₃)	169 (N ₄)	328 (N)

При альтернативної мінливості коефіцієнт кореляції за господарсько цінними ознаками обчислюють за формулою Юла (Yule):

$$R = \sqrt{\frac{n_1 \times n_4 - n_2 \times n_3}{N_1 \times N_2 \times N_3 \times N_4}} = \sqrt{\frac{132 \times 110 - 59 \times 27}{191 \times 137 \times 159 \times 169}} = \sqrt{\frac{12927}{26517}} = 0,49$$

Коефіцієнт кореляції альтернативних показників 0,49. Розраховується помилка коефіцієнта кореляції:

$$s_r = \sqrt{\frac{(1 - (0,49)^2)}{(328 - 2)}} = 0,048$$

Помилка коефіцієнта кореляції дорівнюється 0,048. Критерій достовірності коефіцієнта кореляції рівняється:

$$t_r = r / s_r = 0,49 / 0,048 = 10,20.$$

Теоретичне значення критерію Стюдента знаходять за кількістю ступенів свободи $v_r = N - 2 = 328 - 2 = 326$. Відповідно даних додатку 4 при ступені свободи 326 критерій Стюдента за рівнем значущості $\alpha F_{0,95}$ становить $t = 1,96$, за рівнем значущості $\alpha F_{0,99}$ становить $t = 2,58$.

Висновки

- ✓ фактичний критерій достовірності коефіцієнта кореляції $t_r = 10,20$ більший теоретичного критерію за рівнем значущості $\alpha F_{0,95}$ $t = 1,96$, за рівнем значущості $\alpha F_{0,99}$ $t = 2,58$. Встановлений зв'язок достовірний;
- ✓ сила кореляційного зв'язку між дією фунгіциду Амістар Екстра та кількістю неуражених рослин сояшнику іржею має середню кореляцію;
- ✓ коефіцієнт кореляції між ознаками має позитивний знак «плюс».

Кореляція пряма;

- ✓ між дією фунгіциду Амістар Екстра та кількістю неуражених рослин сояшнику іржею існує слабкий, лінійний, позитивний взаємозв'язок;
- ✓ обробка насіння сояшнику перед сівбою фунгіцидом сприяє зростанню неуражених іржею рослин сояшнику;
- ✓ кореляція – проста, пряма, середня, лінійна;
- ✓ незважаючи на те, що в дослідженні встановлено середній зв'язок між ознаками, але у зв'язку з достовірністю отриманого коефіцієнту кореляції можна припустити, що рослини, де насіння перед сівбою оброблено фунгіцидом Амістар Екстра, будуть в меншій мірі уражені іржею, ніж без обробки насіння. На основі цього можна рекомендувати в технології вирощування сояшнику виконувати передпосівну обробку насіння фунгіцидом Амістар Екстра.

5.4. Аналіз множинної кореляції

Коли в досліді для кожного значення залежної (результативної) змінної Y є відповідні значення двох інших факторних змінних X і Z , дослідник повинен розуміти, що він вивчає множинну кореляцію і повинен розрахувати коефіцієнт множинної кореляції, який вимірює об'єднаний зв'язок факторних ознак (змінних) X і Z з результативною ознакою (змінної) Y .

Численний коефіцієнт кореляції трьох змінних – це показник тісноти лінійного зв'язку між однією з ознак та сукупністю двох інших ознак, які вивчаються в досліді. При проведенні розрахунків встановлюють коефіцієнти кореляції між показниками змінних ознак досліді і далі розраховують сам множинний коефіцієнт кореляції за формулою:

Розраховується коефіцієнт кореляції між запасами продуктивної вологи в шарі ґрунту 0-30 см на початку сівби та врожайністю ячменю ярого:

$$r_{yz} = \frac{\sum(Z - z) \times (Y - y)}{\sqrt{\sum(Z - z)^2 \times \sum(Y - y)^2}} = \frac{0,796}{0,84} = 0,95$$

Коефіцієнт кореляції між запасами продуктивної вологи в шарі ґрунту 0-30 см на початку сівби та врожайністю ячменю ярого складає $r = 0,95$.

Помилка та критерій значущості приватної кореляції визначають за тими ж формулам, як і парної кореляції.

Приватні коефіцієнти кореляції можуть набувати значення між -1 і +1. Приватні коефіцієнти детермінації знаходять шляхом зведення у квадрат приватних коефіцієнтів кореляції.

Другим етапом розрахунків є встановлення сили зв'язку між кількістю продуктивних стебел та урожайністю ячменю ярого.

Отримані дані дослідів заносяться в таблицю і виконуються розрахунки (табл. 5.6).

Таблиця 5.6

**Дані для розрахунку кореляційної залежності між показниками
врожайності ячменю ярого (Y) та
кількістю продуктивних стебел на рослині (X)**

№ корелюючих пар	Кількість продуктивних стебел на рослині, (X) шт.	Врожайність ячменю ярого (Y), т/га	$X - \bar{x}$	$Y - \bar{y}$	$(X - \bar{x}) \times (Y - \bar{y})$	$(X - \bar{x})^2$	$(Y - \bar{y})^2$
1	3,31	3,25	0,06	0,02	0,0012	0,0036	0,0004
2	3,14	3,16	-0,11	-0,07	0,0077	0,0121	0,0049
3	3,06	3,03	-0,19	-0,20	0,0380	0,0361	0,0400
4	3,02	3,21	-0,23	-0,02	0,0046	0,0529	0,0004
5	3,12	3,11	-0,13	-0,12	0,0156	0,0169	0,0144
6	3,10	3,15	-0,15	-0,08	0,0120	0,0225	0,0064
7	3,20	3,22	-0,05	-0,01	0,0005	0,0025	0,0001
8	3,84	3,44	0,59	0,21	0,1239	0,3481	0,0441
9	3,40	3,40	0,15	0,17	0,0255	0,0225	0,0289
10	3,35	3,30	0,10	0,07	0,0070	0,0100	0,0049
	$\bar{x} = 3,25$	$\bar{y} = 3,23$			$\sum(X - \bar{x}) \times (Y - \bar{y}) = 0,236$	$\sum(X - \bar{x})^2 = 0,5272$	$\sum(Y - \bar{y})^2 = 0,1445$
n – кількість корелюючих пар = 10. Використовується для розрахунків середньої.							

Розраховується коефіцієнт кореляції між кількістю продуктивних стебел на рослині та врожайністю ячменю ярого:

$$r_{yx} = \frac{\sum (X - \bar{x}) \times (Y - \bar{y})}{\sqrt{\sum (X - \bar{x})^2 \times \sum (Y - \bar{y})^2}} = \frac{0,236}{0,28} = 0,84$$

Коефіцієнт кореляції між кількістю продуктивних стебел на рослині та врожайністю ячменю ярого складає $r = 0,84$.

Третім етапом розрахунків є встановлення сили зв'язку між кількістю продуктивних стебел та запасами продуктивної вологи в шарі ґрунту 0-30 см на початку сівби.

Запасами продуктивної вологи в шарі ґрунту 0-30 см на початку сівби (Z) є факторною ознакою по відношенню до результативної ознаки – кількість продуктивних стебел на одній рослині.

Отримані дані дослідів заносяться в таблицю і виконуються розрахунки (табл. 5.7).

Таблиця 5.7

Дані для розрахунку кореляційної залежності між кількістю продуктивних стебел на рослині (X) та вологістю ґрунту перед сівбою (Z)

№ корелюючих пар	Запаси продуктивної вологи в шарі ґрунту 0-30 см на початку сівби, (Z) %	Кількість продуктивних стебел на рослині, (X) шт.	$Z - \bar{z}$	$X - \bar{x}$	$(Z - \bar{z}) \times (X - \bar{x})$	$(Z - \bar{z})^2$	$(X - \bar{x})^2$
1	9,5	3,31	-0,06	0,06	-0,0036	0,0036	0,0036
2	9,1	3,14	-0,46	-0,11	0,0506	0,2116	0,0121
3	8,9	3,06	-0,66	-0,19	0,1254	0,4356	0,0361
4	9,3	3,02	-0,26	-0,23	0,0598	0,0676	0,0529
5	8,9	3,12	-0,66	-0,13	0,0858	0,4356	0,0169
6	9,1	3,10	-0,46	-0,15	0,069	0,2116	0,0225
7	9,4	3,20	-0,16	-0,05	0,008	0,0256	0,0025
8	11,1	3,84	1,54	0,59	0,9086	2,3716	0,3481
9	10,6	3,40	1,04	0,15	0,1560	1,0816	0,0225
10	9,7	3,35	0,14	0,10	0,014	0,0196	0,0100
	$\bar{z} = 9,56$	$\bar{x} = 3,25$			$\sum (Z - \bar{z}) \times (X - \bar{x}) = 1,4736$	$\sum (Z - \bar{z})^2 = 4,864$	$\sum (X - \bar{x})^2 = 0,5272$
n – кількість корелюючих пар = 10. Використовується для розрахунків середньої.							

Розрахується коефіцієнт кореляції між кількістю продуктивних стебел на рослині ячменю ярого та вологістю ґрунту перед сівбою:

$$r_{xz} = \frac{\sum (Z - \bar{z}) \times (X - \bar{x})}{\sqrt{\sum (Z - \bar{z})^2 \times \sum (X - \bar{x})^2}} = \frac{1,4736}{1,60} = 0,92$$

Коефіцієнт кореляції між кількістю продуктивних стебел на рослині ячменю ярого та вологістю ґрунту перед сівбою $r_{xz} = 0,92$.

Четвертим етапом розрахунків є встановлення коефіцієнту множинної кореляції, який покаже сили зв'язку між врожайністю ячменю ярого (результативна ознака) та кількістю продуктивних стебел, запасами продуктивної вологи в шарі ґрунту 0-30 см на початку сівби (факторні ознаки).

Встановлення множинного коефіцієнта кореляції доцільно тільки в тому випадку, коли приватні коефіцієнти кореляції виявилися значущими, а зв'язок між результативною ознакою та факторними ознаками, які вивчалися в досліді, дійсно існує.

Отримані коефіцієнти кореляції за попередніми етапами розрахунків вставляють в формулу:

$$R = \sqrt{\frac{(r_{yz}^2 + r_{yx}^2) - 2(r_{yz} \times r_{yx} \times r_{xz})}{(1 - r_{xz}^2)}} = \sqrt{\frac{(0,95^2 + 0,84^2) - 2(0,95 \times 0,84 \times 0,92)}{(1 - 0,92^2)}} = \sqrt{0,91} = 0,95$$

Коефіцієнт множинної кореляції $R = 0,95$ показує, що зв'язок між результативною ознакою Y та факторними ознаками факторами X та Z дуже високий, позитивний, прямий.

Значущість отриманого коефіцієнта множинної кореляції оцінюють за F -критерієм (критерій Фішера), ураховуючи, що загальна кількість корелюючих пар в досліді $n=30$, кількість вивчаємих ознак $k=3$:

$$F_R = (R / 1 - R^2) \times \{(n - k) / (k - 1)\} = (0,95 / 1 - 0,95^2) \times \{(30 - 3) / (3 - 1)\} = (0,95 / 1 - 0,9025) \times (27 / 2) = (0,95 / 0,0975) \times (27 / 2) = 9,74 \times 13,5 = 131,49.$$

Фактичне значення критерію Фішера F_R становить 131,49. Далі встановлюють теоретичне значення критерію Фішера. Для цього розраховують:

- число ступенів свободи результативної ознаки Y (врожайність ячменю ярого): $v_y = k - 1 = 3 - 1 = 2$;
- число ступенів свободи помилки результативної ознаки та факторних ознак X, Z (всього три ознаки): $v_z = n - k = 30 - 3 = 27$.

Теоретичні значення критерію Фішера за числом ступенів свободи результативної ознаки $v_y=2$, числом критерію ступенів свободи помилки результативної ознаки та факторних ознак X, Z $v_z = 27$:

- на 5%-ному рівні значущості (вірогідність 95%) $F_{0,95}$ дорівнюється 3,37 (додаток 2);
- на 1%-ному рівні значущості (вірогідність 99%) $F_{0,99}$ дорівнюється 5,53 (додаток 3).

Оскільки фактичне значення критерію Фішера $F_R = 131,49$ значно перевищує теоретичні значення критерію Фішера на 5%-ному та 1%-ному рівні значущості – 3,37 і 5,53, то розрахований коефіцієнт множинної кореляції є високо достовірним.

В досліді коефіцієнт детермінації дорівнюється: $R^2 = 0,95 = 0,90$, тобто. у 90% випадків зміни факторних ознак X (кількість продуктивних стебел на

рослині), Z (вологість ґрунту перед сівбою) призводять до зміни результативної ознаки Y (врожайність ячменю ярого).

Квадрат множинного коефіцієнта кореляції R^2 , що називається множинним коефіцієнтом детермінації; показав, яка частка дисперсії результативної ознаки Y в досліді пояснюється впливом факторних ознак X, Z .

Слід зауважити, що формула для обчислення коефіцієнта детермінації через співвідношення залишкової та загальної дисперсії результативної ознаки дасть той же самий результат.

Коефіцієнт детермінації змінюються в межах від 0 до 1. Якщо значення множинного коефіцієнта кореляції невелике (менше 0,3), це означає, що досліджувані факторні ознаки недостатньо впливають на варіацію результативної ознаки або зв'язок між факторними і результативною змінними є нелінійним.

5.5. Аналіз криволінійної кореляції

Якщо в досліді однакові збільшення факторної ознаки (X) супроводжуються в середньому неоднаковими змінами результативної ознаки (Y), то науковий співробітник про проведенні статистичного обробітку результатів досліджень має діло з криволінійною залежністю між ознаками, які вивчаються в досліді. У цьому випадку коефіцієнт r зменшує тісноту зв'язку і, більше, може заперечувати її ($r = 0$).

Наприклад, при вивченні результатів досліду та аналізу впливу норм висіву насіння на величину врожайності, науковий співробітник спостерігає закономірність: зі збільшенням норми висіву врожайність сільськогосподарської культури зростає. Зростання врожайності при збільшенні норми висіву відмічається в досліді до якоїсь певної норми висіву насіння і відбувається стабілізація врожайності. Подальше збільшення норми висіву насіння вже призводить до зниження врожайності через більш високий рівень загущення агроценозу сорту, гібриду.

Такий зв'язок між ознаками, властивостями називається криволінійним. Якщо такі зв'язки аналізувати за застосуванням методу парної кореляції та встановлення коефіцієнту парної кореляції r , дослідник може виявити помилкову відсутність зв'язку або наявність дуже слабкої кореляції між ознаками та властивостями.

Для аналізу криволінійної залежності в статистичному обробітку обробці результатів дослідів обчислюють кореляційні відносини η_{xy} або η_{yx} .

Кореляційне співвідношення іноді визначає залежність між випадковими змінними X (факторна ознака) і Y (результативна ознака) двосторонньо, тобто, Y стосовно X (η_{yx}) і навпаки X по відношенню до Y (η_{xy}).

Кореляційне відношення змінюється не більше $0 < \eta_{yx} < 1$. При $\eta_{yx} = 0$ між ознаками, що вивчаються в досліді, немає ніякого зв'язку.

На відміну від коефіцієнта лінійної кореляції (r):

- коефіцієнт криволінійної кореляції має величину завжди позитивну – ($0 < \eta_{yx} < 1$);

○ коефіцієнт криволінійної кореляції – величина рівнозначна для обох ознак, а $\eta_{yx} \neq \eta_{xy}$. Якщо в досліді отримано результат, який показує $\eta_{yx} = \eta_{xy}$ – то це вже буде лінійний зв'язок між ознаками.

Попереднє судження про форму зв'язку між двома ознаками можна отримати на основі таблиці парних спостережень (для малих вибірок: $n < 20$).

При криволінійній кореляції ступінь залежності між двома величинами ознак або властивостей визначають за допомогою формули:

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

де:

η_{yx} – коефіцієнт криволінійної кореляції;

$\sum(Y-y)^2$ – сума квадратів відхилень індивідуальних значень Y від середньої арифметичної y ;

$\sum(Y-y_x)^2$ – сума квадратів відхилень варіант від приватних середніх $\bar{y}_x$ відповідних певним, фіксованим значенням незалежної змінної X .

Розглянемо приклад встановлення коефіцієнту криволінійної кореляції. Науковий співробітник провів дослід і виконує аналіз по встановленню зв'язку між врожайністю силосної маси кукурудзи (результативна ознака Y) та середньодобовим водоспоживанням росли кукурудзи (факторіальна ознака X).

В досліді отримані наступні пари спостережень за вказаними ознаками:

Факторна ознака: середньодобо- ве сумарне водоспоживан- ня рослин кукурудзи мм, (X)	0,5	0,8	1,2	1,3	1,7	1,9	2,3	2,5	2,8	3,0	3,2	3,4
Результативна ознака: врожайність силосної маси кукурудзи, ц/га (Y)	120	160	350	330	520	550	520	550	430	410	370	350

На основі наведених даних необхідно розрахувати кореляційне відношення η_{yx} , що характеризує залежність врожайності силосної маси кукурудзи (Y , т/га) від середньодобового сумарного водоспоживання рослин кукурудзи (X , мм).

Дані досліді заносять в таблицю для виконання розрахунків по встановленню коефіцієнта криволінійної кореляції між ознаками (табл. 5.8):

Таблиця 5.8

**Дані для розрахунку кореляційної залежності
між врожайністю силосної маси кукурудзи (Y, ц/га) та
середньодобовим сумарним водоспоживанням рослин кукурудзи (X, мм)**

Середньодобове сумарне водоспоживання рослин кукурудзи мм, (X)	Врожайність силосної маси кукурудзи, ц/га (Y)	y_x	$Y-y_x$	$(Y-y_x)^2$	$Y-y$	$(Y-y)^2$
0,5	120	140	-20	400	-269	72361
0,8	160		20	400	-229	52441
1,2	350	340	10	100	-39	1521
1,3	330		-10	100	-59	3481
1,7	520	535	-15	225	131	17161
1,9	550		+15	225	161	25921
2,3	520	535	-15	225	131	17161
2,5	550		15	225	161	25921
2,8	430	420	10	100	41	1681
3,0	410		-10	100	21	1681
3,2	370	360	10	100	-19	361
3,4	350		-10	100	-39	1521
x	y=389			$\sum(Y-y_x)^2$ = 2300		$\sum(Y-y)^2$ = 221212

Середньодобове водоспоживання рослин кукурудзи як незалежної змінної X розташовують у таблиці у порядку, що зростає. Варіаційний ряд розбивають на 4-7 груп так, щоб у кожній із них було не менше двох спостережень. Число спостережень у групах може бути різним.

Після проведення розрахунків даних в таблиці за формулою встановлюють коефіцієнт криволінійної кореляції, що дозволить вирішити характер зв'язку між ознаками:

$$\eta_{yx} = \sqrt{\frac{\sum(Y-y)^2 - \sum(Y-y_x)^2}{\sum(Y-y)^2}} = \sqrt{\frac{221212 - 2300}{221212}} = \sqrt{\frac{218912}{221212}} = 0,99.$$

Кореляційне відношення між середньодобовим водоспоживанням рослин кукурудзи і врожайністю силосної маси за розрахунками дорівнюється $\eta_{yx} = 0,99$.

Для вірогідності отриманого кореляційного відношення розраховують помилку кореляційного відношення за формулою з урахуванням вивчення двох ознак при вибраних корелюючих пар $n=12$:

$$s_{\eta} = \sqrt{(1-\eta_{yx}^2) / n-2} = \sqrt{(1-0,99^2) / 12-2} = \sqrt{0,02 / 10} = 0,044.$$

Фактичний критерій достовірності кореляційного відношення дослідів буде становити:

$$t_{\eta} = \eta_{yx} / s_{\eta} = 0,99 / 0,044 = 22,5.$$

Теоретичне значення критерію t встановлюють за наведеними даними додатку 4 (Критерій Стьюдента t).

Контрольне значення теоретичного критерію t при ступеню свободи $n-2=12-2=10$ за 1% рівнем значущості складає $F_{0,99}=2,23$, за 5% рівнем значущості складає $F_{0,95}=3,17$.

Висновки.

Фактичний критерій кореляційного відношення $t_{\eta}=22,5$ більше теоретичних значень критерію $t=2,23$ за 1% рівнем значущості $F_{0,99}$ та $t=3,17$ за 5% рівнем значущості складає $F_{0,95}$. Оскільки $t_{\eta} > t$, то кореляційне відношення між ознаками дослідів достовірно і криволінійний зв'язок між ознаками, що вивчаються, підтверджується.

Значення коефіцієнту криволінійної кореляції $\eta_{yx} = 0,99$ свідчить про позитивний дуже високий кореляційний зв'язок між факторною ознакою (середньодобове водоспоживання рослин кукурудзи) та результативною ознакою (врожайність силосної маси кукурудзи).

Характер кореляційного зв'язку між ознаками або властивостями, які вивчаються в досліді також можливо відображати графічно (рис. 5.2).

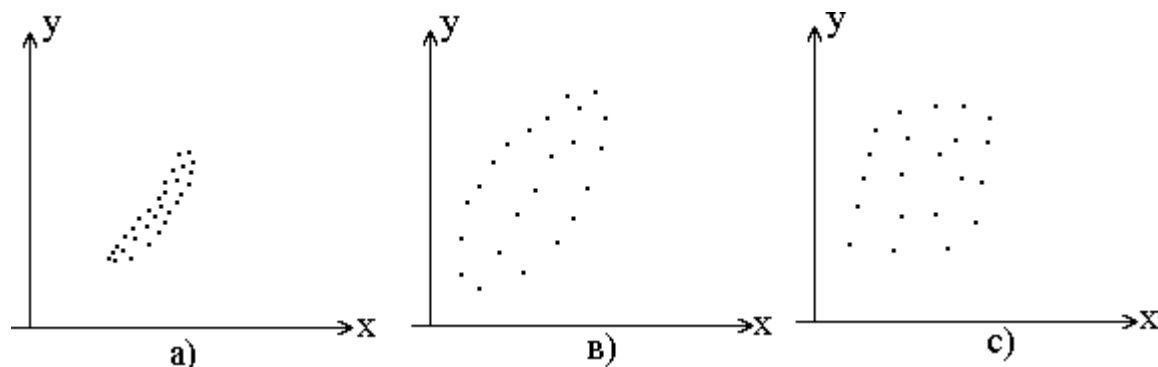


Рис. 5.2. Графічне відображення характеру кореляційних зв'язків.
а – сильна кореляція; в – слабка кореляція; с – кореляція відсутня.

5.6. Застосування та використання регресійного аналізу

В статистичному обробітку результатів досліджень аналіз кореляції у науковому дослідженні складається з двох елементів: власне кореляції та регресії.

Коефіцієнт кореляції вказує на напрям і ступінь сполученості мінливості ознак, але не дозволяє судити про те, як кількісно змінюється результативна ознака при зміні факторіального на одиницю виміру. При постановці та

вирішенні подібних завдань дослідник повинен використовувати регресійний аналіз, основне завдання якого полягає у визначенні формули кореляційної залежності.

В цьому випадку, мета аналізу регресії визначити та дати кількісну оцінку залежності між ознаками, які вивчаються в досліді. Передбачається, що залежна змінна Y визначається відомою функцією (моделлю) і залежить від незалежної ознаки X та деякого параметра.

На підставі регресійного аналізу дослідники повинні знайти такі значення цього параметра, щоб отримана модель найкраще описала б дані експерименту.

Моделі можуть бути як лінійної регресії, так і складніших – нелінійних (криволінійних) зв'язків між ознаками.

Регресія дозволяє:

- ✓ встановити кількісну зміну однієї ознаки, обумовлену зміною іншого;
- ✓ визначити коефіцієнт регресії, форму зв'язку між ознаками, рівняння регресії.

Якщо кореляційний аналіз показав наявність сильного і достовірного зв'язку, тобто такого, який встановлений на рівнях ймовірності $F_{0,95}$ і $F_{0,99}$, виконують регресійні аналізи, обчислюючи коефіцієнт 'регресії b_{xy} або b_{yx} .

Регресія – це характер і ступінь зміни однієї з незалежних ознак X на одиницю виміру іншої залежної, результативної ознаки Y .

Коефіцієнт регресії – це число, яке показує напрямок і величину зміни результативної ознаки (функція) при зміні факторіальної ознаки (аргумент) на одиницю зміни. Наприклад, збільшення довжини листа рослини на 0,9 см приводить до зростання площі листової пластини рослини на 3,8 см².

Проста лінійна регресія базується на рівнянні прямої лінії $y = a + bx$.

Після кореляційних і регресійних аналізів розраховують рівняння регресії, які використовують як опис: виробничої функції (кривий відгук), або оптимізаційної моделі:

- ✓ для обчислення невідомого показника за відомим, наприклад площі листів за їх довжиною;
- ✓ встановлення залежності врожайності сільськогосподарської культури від строку сівби та збирання, норм та доз внесення добрив, щільності агроценозу сортів та гібридів та інших агротехнічних прийомів, які застосовуються в технології вирощування;
- ✓ для прогнозування майбутнього врожаю за кількістю квіток чи зав'язей, запасів продуктивної вологи у ґрунті та кількістю випадіння опадів за період вегетації;
- ✓ для прогнозування якості врожаю щодо норм поливної води на зрошенні, норм внесення мінеральних добрив;
- ✓ для прогнозування поширення шкідників та хвороб за зовнішніми умовами середовища вирощування сільськогосподарської культури;
- ✓ встановлення характеру та особливостей розвитку та прояви ознаки рослин сортів, гібридів, факторів явищ у часі вегетації. На осі абсцис відкладають відрізки часу вегетації: дні, декади, місяці. На осі ординат відкладають значення ознаки, параметри факторів, явищ;

✓ для визначення причинно-наслідкової залежності, або прогностичної моделі. Наприклад, залежність урожайності сорту або гібрида від кількості опадів та температурного режиму у періоди вегетації рослин;

✓ для встановлення залежності між ознаками одного об'єкта (сорту, гібрида, сортозразку). Наприклад: зв'язок між масою насіння та технічною довжиною соломки льону, кількістю зерен у колосі пшениці та масою 1000 зерен, щільністю орного шару ґрунту та вмістом у ньому гумусу;

У статистичній моделі лінійної регресії $y = a + bx + \varepsilon$ незалежною змінною, або факторною ознакою є X , залежною змінною, результативною ознакою є Y .

Коефіцієнти регресії a та b оцінюють параметри лінійної залежності у всій сукупності:

a – місце перетину лінії з віссю Y ($y = a$ при $x = 0$);

b – кут підйому (крутизни) лінії по відношенню до осі X .

Останній доданок ε – це випадкові відхилення фактичних спостережень від очікуваних y .

Тому варіювання значень результативної ознаки (Y) слід розглядати не тільки як наслідок залежності від факторної ознаки (X), але як наслідок випадкового чинника.

Варіювання ознаки y що виражається загальною сумою квадратів відхилень:

$$CK_y = \sum (y - \bar{y})^2.$$

На відміну від одного коефіцієнта кореляції, при встановленні регресії існують два коефіцієнти регресії: b_{yx} або b_{xy} , що показують регресію X за Y і, навпаки, Y за X .

Коефіцієнт b_{yx} є основним коефіцієнтом регресії, оскільки цей коефіцієнт показує кількісну зміну результативної ознаки (Y) при зміні факторної ознаки (X) на одиницю.

У скороченому вигляді формулу можна представити як $b_{yx} = S_{yx}/S_x$,

Коефіцієнт регресії b_{yx} результативної ознаки Y щодо факторної ознаки X_x дорівнює середньої коваріації двох ознак до дисперсії ознаки X , де:

$$S_{yx} = \frac{(x - \bar{x}) \times (y - \bar{y})}{n - 1}$$

$$S_x = \frac{(x - \bar{x})^2}{n - 1}$$

Коефіцієнт регресії b_{yx} показує, як змінюється ознака Y при зміні X на одиницю виміру і вимірюється в одиницях ознаки Y .

Коефіцієнт регресії b_{xy} вказує регресію між ознаками X на Y і виражається в одиницях ознаки X .

Коефіцієнт b_{xy} має сенс лише в оцінці зв'язку двох випадкових змінних (II модель кореляції).

Коефіцієнти регресії мають знак коефіцієнта кореляції. Для перевірки

розрахунків використовують таку рівність: $r^2 = b_{xy} \times b_{yx}$.

Добуток коефіцієнтів регресії дорівнює квадрату коефіцієнта кореляції $b_{yx} \times b_{xy} = r^2$

Наступним етапом вивчення регресійного аналізу є визначення конкретних видів регресійного аналізу.

5.7. Парна лінійна залежність

Під парною лінійною (прямолінійною) кореляційною залежністю між двома ознаками X і Y в регресійному аналізі розуміють таку залежність, яка має лінійний характер та у загальному вигляді виражається рівнянням прямої лінії:

$$Y = a + bX.$$

Таке рівняння називається рівнянням регресії Y на X , а відповідно рівнянню регресії пряма лінія називається вибірковою лінією регресії Y на X . Величина b – вибірковий коефіцієнт регресії.

Коефіцієнти регресії стосовно відношень між результативною ознакою і факторіальною ознакою обчислюють за наступними формулами:

$$b_{yx} = \sum(X - \bar{x}) \times (Y - \bar{y}) / \sum(X - \bar{x})^2,$$

$$b_{xy} = \sum(X - \bar{x}) \times (Y - \bar{y}) / \sum(Y - \bar{y})^2.$$

Регресійний аналіз та розрахування коефіцієнту регресії виконують при умові встановленні в результаті кореляційного аналізу вірогідного зв'язку між ознаками, які вивчаються в досліді.

Розглянемо приклад обчислення коефіцієнту регресії при парній лінійній залежності.

Проведеними розрахунками між ознаками лінії відновлювача фертильності пилку соняшнику GE-3-38Rf – кількість насіння кошику (факторна ознака X) та маса насіння кошика (результативна ознака Y) встановлено кореляційний середній, лінійний, позитивний взаємозв'язок.

Коефіцієнт кореляції між ознаками кількість насіння кошику (шт.) та маса насіння кошика (г) самозапиленої лінії соняшнику GE-3-38Rf дорівнює $r = +0,61$.

За даними таблиці 5.9 встановлюємо коефіцієнт регресії за результативною ознакою Y (функції).

В таблиці вже наведено розрахункові дані, які використовувалися для розрахунку кореляційних зв'язків цих ознак.

Коефіцієнт регресії ознаки Y за ознакою X за даними таблиці буде рівнятися:

$b_{yx} = \sum(X - \bar{x}) \times (Y - \bar{y}) / \sum(X - \bar{x})^2 = 29653,52 / 1052623 = 0,028$ г насіння на 1 шт. насінини лінії соняшнику.

Оскільки добуток коефіцієнтів регресії повинен дорівнюватися квадрату коефіцієнта кореляції $b_{yx} \times b_{xy} = r^2$, розраховують коефіцієнт регресії ознаки X за ознакою Y :

$$b_{xy} = \sum(X - \bar{x}) \times (Y - \bar{y}) / \sum(Y - \bar{y})^2 = 29653,52 / 2238,32 = 13,25 \text{ шт. насінин на 1 г.}$$

Встановлення коефіцієнтів регресії надає змогу перевірити дійсність рівності добутку коефіцієнтів регресії до квадрату коефіцієнта кореляції:

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

$$r^2 = 0,61^2 = 0,37 \quad b_{yx} \times b_{xy} = 0,028 \times 13,25 = 0,37.$$

$$0,37 = 0,37.$$

Таблиця 5.9

**Розрахункові дані для обчислення регресії між ознаками
кількість насіння у кошику (шт.) та маса насіння кошику (г)
лінії відновлювача фертильності пилку соняшнику GE-3-38Rf**

№ рослини, корелюючих пар	Кількість насіння в кошику (X), шт.	Маса насіння кошику (Y), г	$X-\bar{x}$	$Y-\bar{y}$	$(X-\bar{x}) \times (Y-\bar{y})$	$(X-\bar{x})^2$	$(Y-\bar{y})^2$
1	706	26,0	-140,1	-9,4	1316,94	19628,01	88,36
2	676	26,9	-170,4	-8,5	1448,4	29036,16	72,25
3	1007	46,3	160,6	10,9	1750,54	25792,36	118,81
4	1019	42,7	172,6	7,3	1259,98	29790,76	53,29
5	874	40,2	27,6	4,8	132,48	761,76	23,04
6	857	33,8	10,6	-1,6	-16,96	112,36	2,56
7	896	46,2	49,6	10,8	535,68	2460,16	116,64
8	565	21,4	-281,4	-14,0	3939,6	79185,96	196
9	794	34,5	-52,4	-0,9	47,16	2745,76	0,81
10	494	20,1	-352,4	-15,3	5391,72	124185,76	234,09
11	872	34,5	25,6	-0,9	-23,04	655,36	0,81
12	891	37,6	44,6	2,2	98,12	1989,16	4,84
13	877	36,2	30,6	0,8	24,48	936,36	0,64
14	546	20,4	-300,4	-15,0	4506	90240,16	225,0
15	1179	47,9	332,6	12,5	4032,5	110622,76	156,25
16	755	32,6	-91,4	-2,8	255,92	8353,96	7,84
17	1363	53,0	516,6	17,6	9092,16	266875,56	309,76
18	1074	49,3	227,6	13,9	3858,64	51801,76	193,21
19	1046	43,0	199,6	7,6	1516,96	39840,16	57,76
20	437	16,0	-409,4	19,4	-9513,76	167608,36	376,36
	$\bar{x} = 846,4$	$\bar{y} = 35,4$			$\Sigma(X-\bar{x}) \times (Y-\bar{y}) = 29653,52$	$\Sigma(X-\bar{x})^2 = 1052623$	$\Sigma(Y-\bar{y})^2 = 2238,32$
n – кількість рослин в аналізі або кількість корелюючих пар = 20. Використовується для розрахунків середньої.							

Для встановлення помилки коефіцієнта регресії результативної ознаки (Sb_{yx}) застосовують формулу:

$$Sb_{yx} = s_r \times \sqrt{\Sigma(Y-\bar{y})^2 / \Sigma(X-\bar{x})^2} = 0,035 \times \sqrt{2238,32 / 1052623} = \sqrt{0,0021}$$

$$0,035 \times \quad = 0,035 \times 0,046 = 0,0016.$$

s_r – помилка коефіцієнта кореляції дорівнюють 0,035 і визначена при розрахунках помилки коефіцієнта кореляції між ознаками кількість насіння кошику та маса насіння кошика самозапиленої лінії соняшнику GE-3-38Rf за формулою за формулою:

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

де:

s_r – помилка коефіцієнта кореляції; r – коефіцієнт кореляції;

n – кількість корелюючих пар або рослин.

$$s_r = \sqrt{(1-0,61^2) / (20-2)} = \sqrt{1-0,37 / 18} = 0,035.$$

Фактичний критерій суттєвості коефіцієнта регресії результативної ознаки маса насіння кошику b_{yx} визначають за формулою:

$$tb_{yx} = b_{yx} / Sb_{yx} = 0,028 / 0,0016 = 17,5.$$

Теоретичне значення критерію t встановлюють за наведеними даними додатку 4 (Критерій Стьюдента t). Контрольне значення теоретичного критерію t при ступеню свободи $n-2 = 20-2 = 18$ за 1% рівнем значущості складає $F_{0,99} = 2,86$, за 5% рівнем значущості складає $F_{0,95} = 2,09$.

Висновок.

Фактичний критерій кореляційного відношення $tb_{yx} = 17,5$ більше теоретичних значень критерію $t = 2,86$ за 1% рівнем значущості $F_{0,99}$ та $t = 2,09$ за 5% рівнем значущості $F_{0,95}$. Оскільки $tb_{yx} > t$, то регресія за результативною ознакою маса насіння кошику вірогідна.

Встановлення коефіцієнту регресії дозволяє визначити рівняння регресії для результативної ознаки Y (маса насіння кошику, г):

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

де:

$\bar{y}$ – усереднене значення результативної ознаки;

b_{yx} – коефіцієнт регресії результативної ознаки Y щодо факторіальної ознаки X ;

$\bar{y}$ – середнє значення результативної ознаки;

$\bar{x}$ – середнє значення факторіальної ознаки;

X – середня значення факторіальної ознаки варіанту дослідження (береться при проведенні дослідження із конкретного варіанту, що досліджується).

Введення показника факторіальної ознаки, яка визначає прояву результативної ознаки, вирішую питання про визначення у варіанті дослідження

встановити усереднене теоретичне значення результативної ознаки для цього варіанту. Наприклад у варіанті середнє значення кількості насіння в кошику складає 1025 насінин. Цей показник вводиться в рівняння регресії і проводиться розрахунок значення маси насіння кошику в рослині варіанту:

$$Y = y + b_{yx}(X - \bar{x}) = 35,4 + 0,028(1025 - 846,4) = 35,4 + 28,7 - 23,69 = 40,41 \text{ г.}$$

Застосування цієї форми дозволило встановити в окремому варіанті за значенням факторіальної ознаки кількість насіння кошику теоретичне середнє значення маси одного кошику рослини лінії GE-3-38Rf 40,41 г.

Інший шлях використання рівняння регресії – побудова лінії регресії. Для побудови лінії регресії результативної ознаки Y за факторіальною ознакою

X необхідно розрахувати теоретичне усереднене значення результативної ознаки Y щодо мінімального та максимального показників факторіальної ознаки X . В досліді мінімальне значення факторіальної ознаки кількість насіння кошику $X_{\min}$ складає 437 шт., максимальне значення факторіальної ознаки $X_{\max}$ складає 1363 шт. (див. дані табл. 5.9).

Значення теоретично усередненого значення результативної ознаки Y для мінімальної факторіальної ознаки $X_{\min} = 437$ насінин:

$$\begin{aligned} Y_{X_{\min} = 437} &= \bar{y} + b_{yx}(X_{\min} - \bar{x}) = 35,4 + 0,028(437 - 846,4) = 35,4 + (12,24 - 23,69) \\ &= 35,4 + (-11,45) = 35,4 - 11,45 = 23,95 \text{ г.} \end{aligned}$$

Значення теоретично усередненого значення результативної ознаки Y для максимальної факторіальної ознаки $X_{\max} = 1363$ насінин:

$$Y_{X_{\max} = 1363} = \bar{y} + b_{yx}(X_{\max} - \bar{x}) = 35,4 + 0,028(1363 - 846,4) = 35,4 + (38,16 - 23,69) = 35,4 + 14,47 = 49,87 \text{ г.}$$

Знайдені рівнянням регресії точки ($X_{\min} = 437$ насінин, $Y_{\min} = 23,95$ г) і ($X_{\max} = 1363$ насінин, $Y_{\max} = 49,87$ г) наносять на графік та з'єднують прямою лінією. Це буде теоретична лінія регресії результативної ознаки Y за факторіальною ознакою X (рис. 5.3).

На графіку доцільно вказати рівняння регресії, коефіцієнт регресії, коефіцієнт кореляції, довірчу зону для справжньої лінії регресії в сукупності.

Щоб обмежити довірчу зону, необхідно вгору і вниз від теоретичної лінії регресії, відкласти величину однієї (68% зона) або двох (95% зона) помилок відхилення від регресії, тобто $\pm s_{yx}$ або $\pm 2s_{yx}$, і з'єднати знайдені точки пунктирними лініями. Область, укладена між цими лініями, називається довірчою зоною регресії.

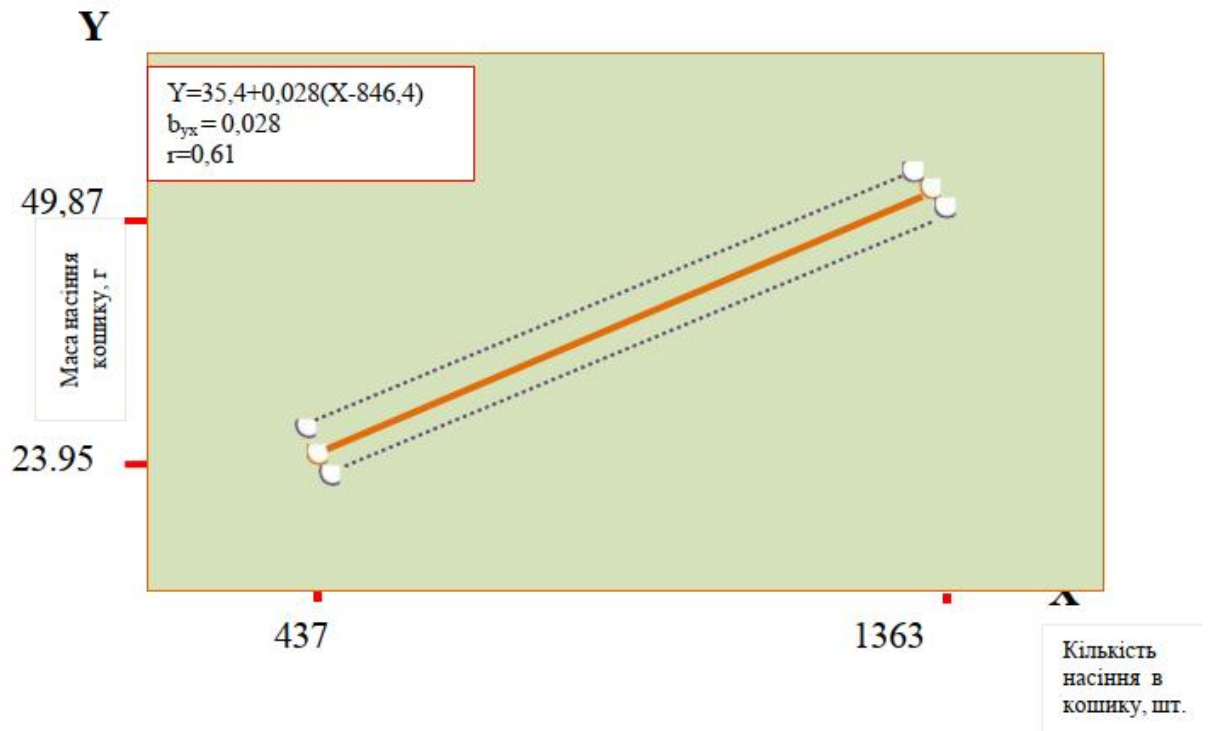


Рис. 5.3. Лінія регресії при позитивній прямолінійній кореляції між масою насіння кошику та кількістю насіння кошику лінії соняшнику GE-3-38Rf.

Для побудови графіку величину помилки (68% зона) відхилення від регресії розраховують за формулою:

$$\pm s_{yx} = s_r \times \sqrt{\sum (Y - y)^2} = 0,035 \times \sqrt{2238,32} = 0,035 \times 47,31 = 1,65.$$

Для побудови графіку величину двох помилок (95% зона) відхилення від регресії розраховують за формулою:

$$\pm 2s_{yx} = 2 \{ s_r \times \sqrt{\sum (Y - y)^2} \} = 2 \{ 0,035 \times \sqrt{2238,32} \} = 2(0,035 \times 47,31) = 2 \times 1,65 = 3,3.$$

Лінія регресії на графіку показує пряму, лінійну середню кореляцію між показниками кількість насіння в кошику та маса насіння в кошику самозапиленої лінії соняшнику GE-3-38Rf.

Збільшення насіння в кошику на одну одиницю обумовлює зростання маси насіння кошику на 0,028 г.

Розкид показників факторіальної ознаки кількість насіння в кошику близько теоретичної лінії регресії результативної ознаки маса насіння в кошику незначний.

При негативній, прямій, лінійній кореляції лінія регресії буде мати на графіку інше відображення.

Розглянемо приклад обчислення коефіцієнту регресії при негативному коефіцієнті кореляції.

Проведеним розрахунками встановлена негативна, пряма, середня кореляція між результативною ознакою вологість ґрунту перед сівбою кукурудзи (Y) та факторіальною ознакою щільність орного шару ґрунту (X) $r = -0,59$.

Для обчислення регресії застосовують дані таблиці 5.10.

Таблиця 5.10

Розрахункові дані для обчислення регресії між показниками щільності орного шару ґрунту (X) та вологістю ґрунту перед сівбою кукурудзи (Y)

№ корелюючих пар	Щільність орного шару ґрунту (X), г/см ³ .	Вологість ґрунту (Y), %	$X - \bar{x}$	$Y - \bar{y}$	$(X - \bar{x}) \times (Y - \bar{y})$	$(X - \bar{x})^2$	$(Y - \bar{y})^2$
1	1,07	17,4	-0,02	2,0	-0,04	0,0004	4,0
2	1,06	14,1	-0,03	-0,7	0,021	0,0009	0,49
3	1,03	15,7	-0,06	0,3	-0,018	0,0036	0,09
4	1,13	15,5	0,04	0,1	0,004	0,0016	0,01
5	1,10	16,2	0,01	0,8	0,008	0,0001	0,64
6	1,12	15,2	0,03	-0,2	-0,006	0,0009	0,04
7	1,04	14,5	-0,05	-0,9	0,045	0,0025	0,81
8	1,03	23,0	-0,06	7,6	-0,456	0,0036	57,76
9	1,05	14,0	-0,04	-1,4	0,056	0,0016	1,96
10	1,12	12,1	0,03	-3,3	-0,099	0,0009	10,89
11	1,25	11,1	0,16	-4,3	-0,688	0,0256	18,49
	$\bar{x} = 1,09$	$\bar{y} = 15,4$			$\sum(X - \bar{x}) \times (Y - \bar{y}) = -1,173$	$\sum(X - \bar{x})^2 = 0,0417$	$\sum(Y - \bar{y})^2 = 95,18$
n – кількість корелюючих пар = 11. Використовується для розрахунків середньої.							

Коефіцієнт регресії результативної ознаки Y (вологість ґрунту перед сівбою кукурудзи, %) за факторіальною ознакою X щільність орного шару ґрунту г/см³) за даними таблиці 22 буде рівнятися:

$$b_{yx} = \sum(X - \bar{x}) \times (Y - \bar{y}) / \sum(X - \bar{x})^2 = -1,173 / 0,0417 = -28,13\%.$$

Встановлюють помилку коефіцієнта регресії результативної ознаки (Sb_{yx}):

$$s_r = \sqrt{\frac{(1 - r^2)}{(n - 2)}} = \sqrt{\frac{(1 - (-0,59)^2)}{(11 - 2)}} = 0,072$$

$$Sb_{xy} = s_r \times \sqrt{\frac{\sum(Y - \bar{y})^2}{\sum(X - \bar{x})^2}} = 0,072 \times \sqrt{95,18/0,0417} = 0,072 \times 47,77 = 3,44$$

Фактичний критерій суттєвості коефіцієнта регресії результативної ознаки b_{yx} визначають за формулою:

$$tb_{yx} = b_{yx} / S_{b_{yx}} = -28,13 / 3,44 = 8,17.$$

Теоретичне значення критерію t встановлюють за наведеними даними додатку 4 (Критерій Стюдента t). Контрольне значення теоретичного критерію t при ступеню свободи $n-2 = 11-2 = 9$ за 1% рівнем значущості складає $F_{0,99}=3,25$, за 5% рівнем значущості складає $F_{0,95}=2,26$.

Висновок.

Фактичний критерій кореляційного відношення $tb_{yx}=8,17$ більше теоретичних значень критерію $t=3,25$ за 1% рівнем значущості $F_{0,99}$ та $t=2,26$ за 5% рівнем значущості $F_{0,95}$. Оскільки $tb_{yx} > t$, то регресія за результативною ознакою (вологість ґрунту перед сівбою кукурудзи, %) вірогідна.

Встановлення коефіцієнту регресії дозволяє визначити рівняння регресії для результативної ознаки Y (вологість ґрунту перед сівбою кукурудзи, %):

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

де:

$\bar{Y}$ – усереднене значення результативної ознаки;

b_{yx} – коефіцієнт регресії результативної ознаки Y щодо факторіальної ознаки X ;

$\bar{y}$ – середнє значення результативної ознаки;

$\bar{x}$ – середнє значення факторіальної ознаки;

$\bar{X}$ – середня значення факторіальної ознаки варіанту досліджується (береться при проведенні досліджу із конкретного варіанту, що досліджується).

Для побудови лінії регресії результативної ознаки Y за факторіальною ознакою X розраховують теоретичне усереднене значення результативної ознаки Y за відношенням до мінімального та максимального показників факторіальної ознаки X .

В досліді мінімальне значення факторіальної ознаки щільність орного шару ґрунту $X_{\min}$ складає $1,03 \text{ г/см}^3$, максимальне значення факторіальної ознаки $X_{\max}$ складає $1,25 \text{ г/см}^3$ (див. дані табл. 5.10).

Значення теоретично усередненого значення результативної ознака Y для мінімальної факторіальної ознаки $X_{\min} = 1,03 \text{ г/см}^3$:

$$\begin{aligned} Y_{X_{\min} = 1,03} &= \bar{y} + b_{yx}(X_{\min} - \bar{x}) = 15,4 + \{-28,13 \times (1,03 - 1,09)\} \\ &= 15,4 + (-28,97 + 30,66) = 15,4 + 1,69 = 17,09 \%. \end{aligned}$$

Значення теоретично усередненого значення результативної ознака Y для максимальної факторіальної ознаки $X_{\max} = 1,25 \text{ г/см}^3$:

$$\begin{aligned} Y_{X_{\max} = 1,25} &= \bar{y} + b_{yx}(X_{\max} - \bar{x}) = 15,4 + \{-28,13 \times (1,25 - 1,09)\} \\ &= 15,4 + (-35,16 + 30,66) = 15,4 - 4,5 = 10,90 \%. \end{aligned}$$

Знайдені за допомогою рівняння регресії точки ($X_{\min} = 1,03 \text{ г/см}^3$, $Y_{\min} = 17,09$

%) і ($X_{\max} = 1,25 \text{ г/см}^3$, $Y_{\max} = 10,9\%$) наносять на графік та з'єднують прямою лінією. Це буде теоретична лінія регресії результативної ознаки Y за факторіальною ознакою X (рис. 5.4).

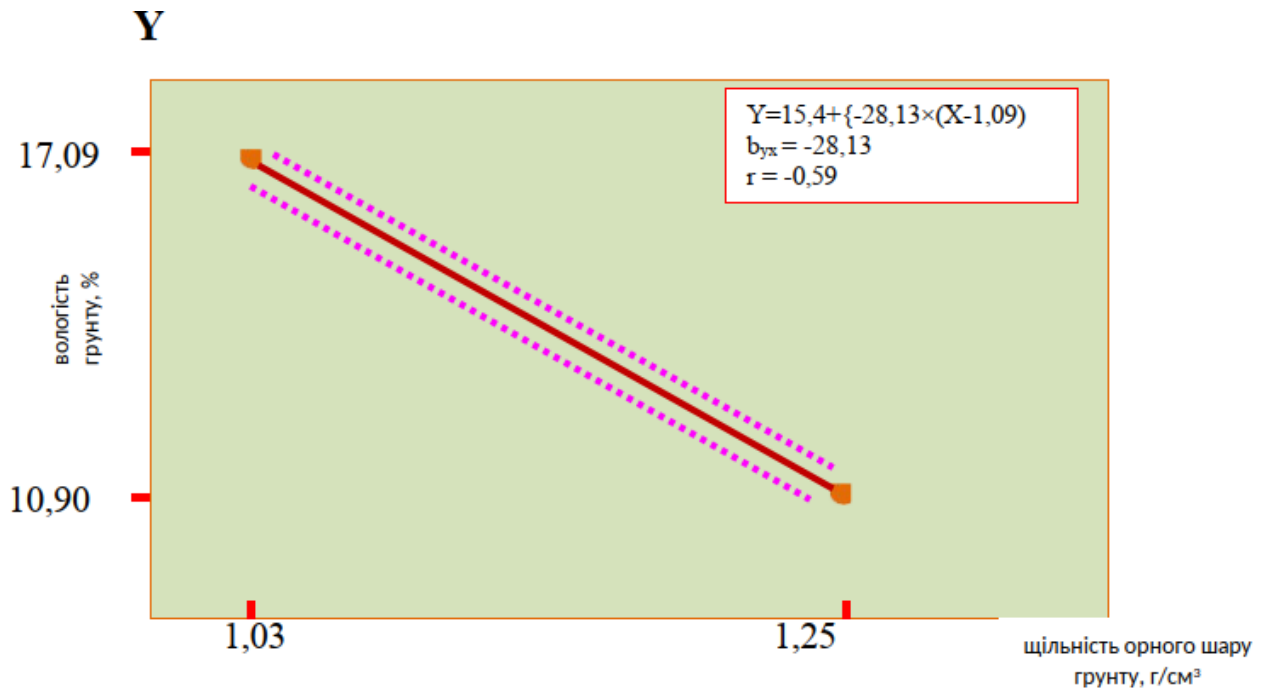


Рис. 5.4. Лінія регресії при негативній, прямолінійній кореляції між вологістю ґрунту перед сівбою кукурудзи та щільністю орного шару ґрунту.

На графіку вказують рівняння регресії, коефіцієнт регресії, коефіцієнт кореляції, довірчу зону для справжньої лінії регресії в сукупності.

Щоб обмежити довірчу зону, вгорі та внизу від теоретичної лінії регресії, відкладають величину однієї (68% зона) або двох (95% зона) помилок відхилення від регресії, тобто $\pm s_{yx}$ або $\pm 2s_{yx}$, і з'єднують знайдені точки пунктирними лініями. Між цими лініями знаходиться довірча зона регресії.

Величина помилки (68% зона) відхилення від регресії становить:

$$\pm s_{yx} = s_r \times \sqrt{\sum(Y - \bar{y})^2} = 0,072 \times \sqrt{95,18} = 0,072 \times 9,76 = 0,70.$$

Для побудови графіку величину двох помилок (95% зона) відхилення від регресії розраховують за формулою:

$$\pm 2s_{yx} = 2 \left\{ s_r \times \sqrt{\sum(Y - \bar{y})^2} \right\} = 2 \{ 0,072 \times \sqrt{95,18} \} = 2(0,072 \times 9,76) = 1,4$$

Лінія регресії на графіку показує негативну пряму, лінійну середню кореляцію між показниками вологості ґрунту перед сівбою кукурудзи та щільністю орного шару ґрунту.

Збільшення щільності орного шару ґрунту обумовлює зменшення вологості ґрунту перед сівбою кукурудзи. Лінія регресії показує зворотню, негативну

залежність між вологістю ґрунту та щільністю орного шару ґрунту.

Розкид показників факторіальної ознаки кількість насіння в кошику близько теоретичної лінії регресії результативної ознаки маса насіння в кошику незначний.

Для порівняння графіків регресії на рис. 5.5 наведено напрям лінії регресії при позитивній кореляції і напрям лінії регресії при негативній, зворотній кореляції між результативною ознакою Y та факторіальною ознакою X .

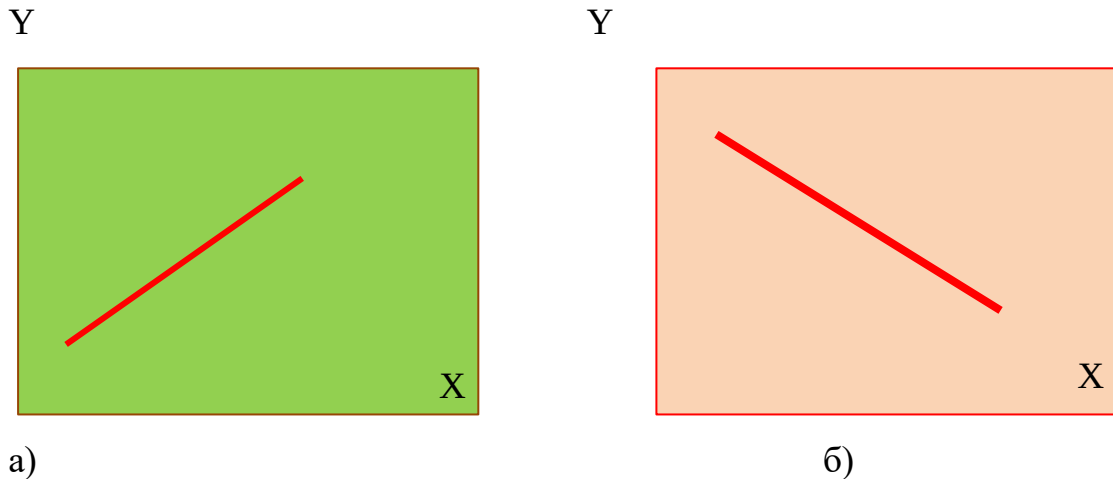


Рис. 5.5. Напрямок лінії регресії : а) при позитивній кореляції; б) при негативній, зворотній кореляції.

5.8. Множинна лінійна залежність

Множинна регресія – особливий метод статистики, який може допомогти досліднику проаналізувати та зрозуміти взаємозв'язок між однією результативною (залежною) ознакою та двома або більше факторіальними (незалежними) ознаками. Оскільки множинна регресія припускає велику дисперсію, вона дає науковим співробітникам можливість робити оптимальні прогнози результатів змінної (Y) відгуку.

Форма множинного регресійного аналізу розширює лінійну регресію, яка є найпростішою формою регресії. Проста лінійна регресія створює лінійну математичну залежність між однією незалежною змінною та однією залежною змінною.

Завдяки же збільшенню кількості змінних, які можуть бути лінійними або нелінійними, модель множинного регресійного аналізу дозволяє дослідникам отримувати велику дисперсію і точність при прогнозуванні результатів досліджень, а також зрозуміти вплив кожної факторіальної ознаки на зміну значень факторіальної ознаки.

На відміну від лінії регресії двох змінних, які відображають показники результативної ознаки та факторіальної ознаки, зв'язок між трьома змінними описується площиною тривимірного простору.

При вивченні кореляції науковому співробітнику важливо визначити не

лише силу зв'язку між змінними досліду, а й саме природу цього зв'язку. Дослідник повинен вивчити та оцінити питання наскільки змінюються значення результативної ознаки Y в разі зміни значень незалежних змінних на одиницю вимірів.

Це питання вирішується за допомогою математичного рівняння.

В статистичному обробітку результатів дослідів математичне рівняння для обчислення прямолінійної залежності між трьома змінними називається множинним лінійним рівнянням площини регресії. Множинне лінійне рівняння площини регресії відображається формулою:

$$Y = a + (b_1 \times X) + (b_2 \times Z),$$

де:

Y – залежна змінна результативної ознаки,

X та Z – незалежні змінні факторіальних ознак, a – загальний початок відліку,

b_1 та b_2 – коефіцієнти приватної регресії.

Коефіцієнт b_1 вказує на яку величину збільшується результативна ознака Y при кожному збільшенні одну одиницю факторіальної ознаки X при постійному значенні іншої факторіальної ознаки Z .

Коефіцієнт b_2 показує на яку величину збільшується результативна ознака Y при кожному збільшенні на одну одиницю факторіальної ознаки Z при постійному значенні факторіальної ознаки X .

Параметри a , b_1 , b_2 обчислюють методом найменших квадратів, який дозволяє знайти таке положення площини регресії у просторі, коли сума квадратів відхилень емпіричних точок від неї є мінімальною:

$$\sum [Y - \{a + (b_1 \times X) + (b_2 \times Z)\}]^2 = \min.$$

Самі параметри рівняння регресії a , b_1 , b_2 визначають за наступними формулами:

$$b_1 = \frac{\{\sum (Z - \bar{z})^2\} \times \{\sum (X - \bar{x}) \times (Y - \bar{y})\} - \{\sum (X - \bar{x}) \times (Z - \bar{z})\} \times \{\sum (Y - \bar{y}) \times (Z - \bar{z})\}}{\{\sum (X - \bar{x})^2 \times \sum (Z - \bar{z})^2\} - \{\sum (X - \bar{x}) \times (Z - \bar{z})\}^2}.$$

$$b_2 = \frac{\{\sum (X - \bar{x})^2\} \times \{\sum (Y - \bar{y}) \times (Z - \bar{z})\} - \{\sum (X - \bar{x}) \times (Z - \bar{z})\} \times \{\sum (X - \bar{x}) \times (Y - \bar{y})\}}{\{\sum (X - \bar{x})^2 \times \sum (Z - \bar{z})^2\} - \{\sum (X - \bar{x}) \times (Z - \bar{z})\}^2}.$$

$$a = \bar{y} - (b_1 \times \bar{x}) - (b_2 \times \bar{z}).$$

Дані для розрахунків коефіцієнтів регресії вибирають із таблиць 5.11, 5.12, 5.13.

$$\sum(Z-\bar{z})^2 = 4,864.$$

$$\sum(X-\bar{x}) \times (Y-\bar{y}) = 0,236.$$

$$\sum(X-\bar{x}) \times (Z-\bar{z}) = 1,4736.$$

$$\sum(Y-\bar{y}) \times (Z-\bar{z}) = 0,796.$$

$$\sum(X-\bar{x})^2 = 0,5272.$$

Таблиця 5.11

Розрахункові дані для обчислення регресії між показниками врожайності ячменю ярого (Y) та вологістю ґрунту перед сівбою (Z)

№ корелюючих пар	Запаси продуктивної вологи в шарі ґрунту 0-30 см на початку сівби, (Z) %	Врожайність ячменю ярого (Y), т/га	$Z-\bar{z}$	$Y-\bar{y}$	$(Z-\bar{z}) \times (Y-\bar{y})$	$(Z-\bar{z})^2$	$(Y-\bar{y})^2$
1	9,5	3,25	-0,06	0,02	-0,0012	0,0036	0,0004
2	9,1	3,16	-0,46	-0,07	0,0322	0,2116	0,0049
3	8,9	3,03	-0,66	-0,20	0,132	0,4356	0,040
4	9,3	3,21	-0,26	-0,02	0,0052	0,0676	0,0004
5	8,9	3,11	-0,66	-0,12	0,0792	0,4356	0,0144
6	9,1	3,15	-0,46	-0,08	0,0368	0,2116	0,0064
7	9,4	3,22	-0,16	-0,01	0,0016	0,0256	0,0001
8	11,1	3,44	1,54	0,21	0,3234	2,3716	0,0441
9	10,6	3,40	1,04	0,17	0,1768	1,0816	0,0289
10	9,7	3,30	0,14	0,07	0,0098	0,0196	0,0049
	$\bar{z} = 9,56$	$\bar{y} = 3,23$			$\sum(Z-\bar{z}) \times (Y-\bar{y}) = 0,796$	$\sum(Z-\bar{z})^2 = 4,864$	$\sum(Y-\bar{y})^2 = 0,1445$

Таблиця 5.12

Розрахункові дані для обчислення регресії між показниками врожайності ячменю ярого (Y) та кількістю продуктивних стебел на рослині (X)

№ корелюючих пар	Кількість продуктивних стебел на рослині, (X) шт.	Врожайність ячменю ярого (Y), т/га	$X-\bar{x}$	$Y-\bar{y}$	$(X-\bar{x}) \times (Y-\bar{y})$	$(X-\bar{x})^2$	$(Y-\bar{y})^2$
1	3,31	3,25	0,06	0,02	0,0012	0,0036	0,0004
2	3,14	3,16	-0,11	-0,07	0,0077	0,0121	0,0049
3	3,06	3,03	-0,19	-0,20	0,0380	0,0361	0,040
4	3,02	3,21	-0,23	-0,02	0,0046	0,0529	0,0004
5	3,12	3,11	-0,13	-0,12	0,0156	0,0169	0,0144

6	3,10	3,15	-0,15	-0,08	0,012	0,0225	0,0064
7	3,20	3,22	-0,05	-0,01	0,0005	0,0025	0,0001
8	3,84	3,44	0,59	0,21	0,1239	0,3481	0,0441
9	3,40	3,40	0,15	0,17	0,0255	0,0225	0,0289
10	3,35	3,30	0,10	0,07	0,007	0,0100	0,0049
	$\bar{x} = 3,25$	$\bar{y} = 3,23$			$\sum(\bar{X}-x) \times (\bar{Y}-y) = 0,236$	$\sum(X-\bar{x})^2 = 0,5272$	$\sum(Y-\bar{y})^2 = 0,1445$

n – кількість корелюючих пар = 10. Використовується для розрахунків середньої.

Таблиця 5.13

Розрахункові дані для обчислення регресії між кількістю продуктивних стебел на рослині (X) та вологістю ґрунту перед сівбою (Z)

№ корелюючих пар	Запаси продуктивної вологи в шарі ґрунту 0-30 см на початку сівби, (Z) %	Кількість продуктивних стебел на рослині, (X) шт.	$Z-\bar{z}$	$X-\bar{x}$	$(Z-\bar{z}) \times (X-\bar{x})$	$(Z-\bar{z})^2$	$(X-\bar{x})^2$
1	9,5	3,31	-0,06	0,06	-0,0036	0,0036	0,0036
2	9,1	3,14	-0,46	-0,11	0,0506	0,2116	0,0121
3	8,9	3,06	-0,66	-0,19	0,1254	0,4356	0,0361
4	9,3	3,02	-0,26	-0,23	0,0598	0,0676	0,0529
5	8,9	3,12	-0,66	-0,13	0,0858	0,4356	0,0169
6	9,1	3,10	-0,46	-0,15	0,069	0,2116	0,0225
7	9,4	3,20	-0,16	-0,05	0,008	0,0256	0,0025
8	11,1	3,84	1,54	0,59	0,9086	2,3716	0,3481
9	10,6	3,40	1,04	0,15	0,1560	1,0816	0,0225
10	9,7	3,35		0,10	0,014	0,0196	0,0100
	$\bar{z} = 9,56$	$\bar{x} = 3,25$			$\sum(\bar{Z}-z) \times (X-\bar{x}) = 1,4736$	$\sum(Z-\bar{z})^2 = 4,864$	$\sum(X-\bar{x})^2 = 0,5272$

n – кількість корелюючих пар = 10. Використовується для розрахунків середньої.

На основі виділених даних із таблиць розраховують коефіцієнти регресії b_1 та b_2 .

$$b_1 = \frac{\{\sum(Z-\bar{z})^2\} \times \{\sum(X-\bar{x}) \times (Y-\bar{y})\} - \{\sum(X-\bar{x}) \times (Z-\bar{z})\} \times \{\sum(Y-\bar{y}) \times (Z-\bar{z})\}}{\{\sum(X-\bar{x})^2 \times \sum(Z-\bar{z})^2\} - \{\sum(X-\bar{x}) \times (Z-\bar{z})\}^2} =$$

$$= \frac{(4,864 \times 0,236) - (1,4736 \times 0,796)}{(0,5272 \times 4,864) - (1,4736)^2} = \frac{1,148 - 1,173}{2,56 - 2,17} = \frac{-0,025}{0,39} = -0,064.$$

$$b_2 = \frac{\{\sum(X-x)^2\} \times \{\sum(Y-y) \times (Z-z)\} - \{\sum(X-x) \times (\bar{Z}-z)\} \times \{\sum(X-x) \times (Y-y)\}}{\{\sum(\bar{X}-x)^2 \times \sum(\bar{Z}-z)^2\} - \{\sum(\bar{X}-x) \times (\bar{Z}-z)\}^2} =$$

$$= \frac{(0,5272 \times 0,796) - (1,4736 \times 0,236)}{(0,5275 \times 4,864) - (1,4736)^2} = \frac{0,41 - 0,35}{2,56 - 2,17} = \frac{0,06}{0,39} = 0,154.$$

Визначають параметр регресії а:

$$a = \bar{y} - (b_1 \times \bar{x}) - (b_2 \times \bar{z}) = 3,23 - \{(-0,064) \times 3,25\} - (0,154 \times 9,56) = 3,23 - (-0,208) - 1,47 = 3,23 + 0,208 - 1,47 = 1,97.$$

Розраховані параметри вставляють в рівнянні регресії. Для даного дослідження рівняння регресії буде таким:

$$Y = a + (b_1 \times X) + (b_2 \times Z) = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z.$$

Встановлене рівняння регресії відношення залежності корелюючих ознак зображують графічно в вигляді ліній та поверхні регресії.

З цієї метою строять регресійну решітку і вносять дані для розрахунків.

Наприклад для даного дослідження, розрахована при встановленні коефіцієнта кореляції, сполученість врожайності ячменю ярого Y (ц з 1 га) щодо запасів вологи перед сівбою Z (%) та кількістю продуктивних стебел на одній рослині ячменю X (шт.) близька до лінійної. Сполученість між урожайністю цими ознаками виражається рівнянням регресії в інтервалі запасів продуктивної вологи в шарі ґрунту від 9,0 до 14% (9,0%; 10,5%; 12,0%; 14,0%) і кількістю продуктивних стебел на 1 рослині від 3,03 до 4,10 (3,03 продуктивних стебла; 3,07 продуктивних стебла; 4,04 продуктивних стебла; 4,10 продуктивних стебла). Дані дослідження факторіальних ознак заносять в регресійну решітку (табл. 5.14).

Таблиця 5.14

Розрахункові дані регресійної решітки факторіальних ознак дослідження по встановленню прогнозуємої врожайності за рівнянням
 $Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z$

Запаси продуктивної вологи в шарі ґрунту 0-30 см на початку сівби, (Z) %	Кількість продуктивних стебел на рослині, (X) шт.			
	3,03	3,07	4,04	4,10
9,0				
10,5				
12,0				
14,0				

Підставляючи в рівняння значення ознак рівняння регресії отримують координати точок врожайності ярого ячменю для побудови ліній та поверхні регресії відповідно до конкретних значень факторіальних ознак Z та X , які наведено в регресійній решітці.

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 9,0% і кількістю продуктивних стебел на 1 рослині – 3,03 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,03\} + 0,154 \times 9,0 = 1,97 - 0,19 + 1,39 = 3,17 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 9,0% і кількістю продуктивних стебел на 1 рослині – 3,07 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,07\} + 0,154 \times 9,0 = 1,97 - 0,20 + 1,39 = 3,16 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 9,0% і кількістю продуктивних стебел на 1 рослині – 4,04 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 4,04\} + 0,154 \times 9,0 = 1,97 - 0,26 + 1,39 = 3,10 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 9,0% і кількістю продуктивних стебел на 1 рослині – 4,10 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 4,10\} + 0,154 \times 9,0 = 1,97 - 0,26 + 1,39 = 3,10 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 10,5% і кількістю продуктивних стебел на 1 рослині – 3,03 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,03\} + 0,154 \times 10,5 = 1,97 - 0,19 + 1,61 = 3,39 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 10,5% і кількістю продуктивних стебел на 1 рослині – 3,07 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,07\} + 0,154 \times 10,5 = 1,97 - 0,20 + 1,61 = 3,38 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 10,5% і кількістю продуктивних стебел на 1 рослині – 4,04 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 4,04\} + 0,154 \times 10,5 \\ = 1,97 - 0,26 + 1,61 = 3,32 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 10,5% і кількістю продуктивних стебел на 1 рослині – 4,10 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 4,10\} + 0,154 \times 10,5 \\ = 1,97 - 0,26 + 1,61 = 3,32 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 12,0% і кількістю продуктивних стебел на 1 рослині – 3,03 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,03\} + 0,154 \times 12,0 \\ = 1,97 - 0,19 + 1,85 = 3,63 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 12,0% і кількістю продуктивних стебел на 1 рослині – 3,07 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,07\} + 0,154 \times 12,0 \\ = 1,97 - 0,20 + 1,85 = 3,62 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 12,0% і кількістю продуктивних стебел на 1 рослині – 4,04 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,03\} + 0,154 \times 12,0 \\ = 1,97 - 0,26 + 1,85 = 3,56 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 12,0% і кількістю продуктивних стебел на 1 рослині – 4,10 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,03\} + 0,154 \times 12,0 \\ = 1,97 - 0,26 + 1,85 = 3,56 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 14,0% і кількістю продуктивних стебел на 1 рослині – 3,03 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,03\} + 0,154 \times 14 \\ = 1,97 - 0,19 + 2,16 = 3,94 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 14,0% і кількістю продуктивних стебел на 1 рослині – 3,07 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 3,07\} + 0,154 \times 14 = 1,97 - 0,20 + 2,16 = 3,90 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 14,0% і кількістю продуктивних стебел на 1 рослині – 4,04 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 4,04\} + 0,154 \times 14 = 1,97 - 0,26 + 2,16 = 3,87 \text{ т/га.}$$

При запасах продуктивної вологи в шарі ґрунту 0-30 см – 14,0% і кількістю продуктивних стебел на 1 рослині – 4,10 шт. прогнозована врожайність (результативна ознака Y за рівнянням регресії при округленні чисел буде становити:

$$Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z = 1,97 + \{(-0,064) \times 4,10\} + 0,154 \times 14 = 1,97 - 0,26 + 2,16 = 3,87 \text{ т/га.}$$

Розраховані дані прогнозованої врожайності ярого ячменю за показниками факторіальних ознак запасах продуктивної вологи в шарі ґрунту 0-30 см і кількістю продуктивних стебел на 1 рослині заносять в регресійну решітку (табл. 5.15).

Таблиця 5.15

Прогнозована врожайність ячменю ярого за факторіальними ознаками відповідно за рівнянням регресії $Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z$

Запаси продуктивної вологи в шарі ґрунту 0-30 см на початку сівби, (Z) %	Кількість продуктивних стебел на рослині, (X) шт.			
	3,03	3,07	4,04	4,10
9,0	3,17	3,16	3,10	3,10
10,5	3,39	3,38	3,32	3,32
12,0	3,63	3,62	3,56	3,56
14,0	3,94	3,90	3,87	3,87

Отримані та занесені в регресійну решітку дані врожайності вирішують питання графічного відображення множинної регресії.

На початку відображаєм графічно лінії регресії результативної ознаки Y (врожайність ячменю ярого) на факторіальній ознаці Z (запаси продуктивної

вологи в шарі ґрунту 0-30 см) для чотирьох фіксованих значень факторіальної ознаки X (кількість продуктивних стебел на одній рослині ячменю ярого) (рис. 5.6).

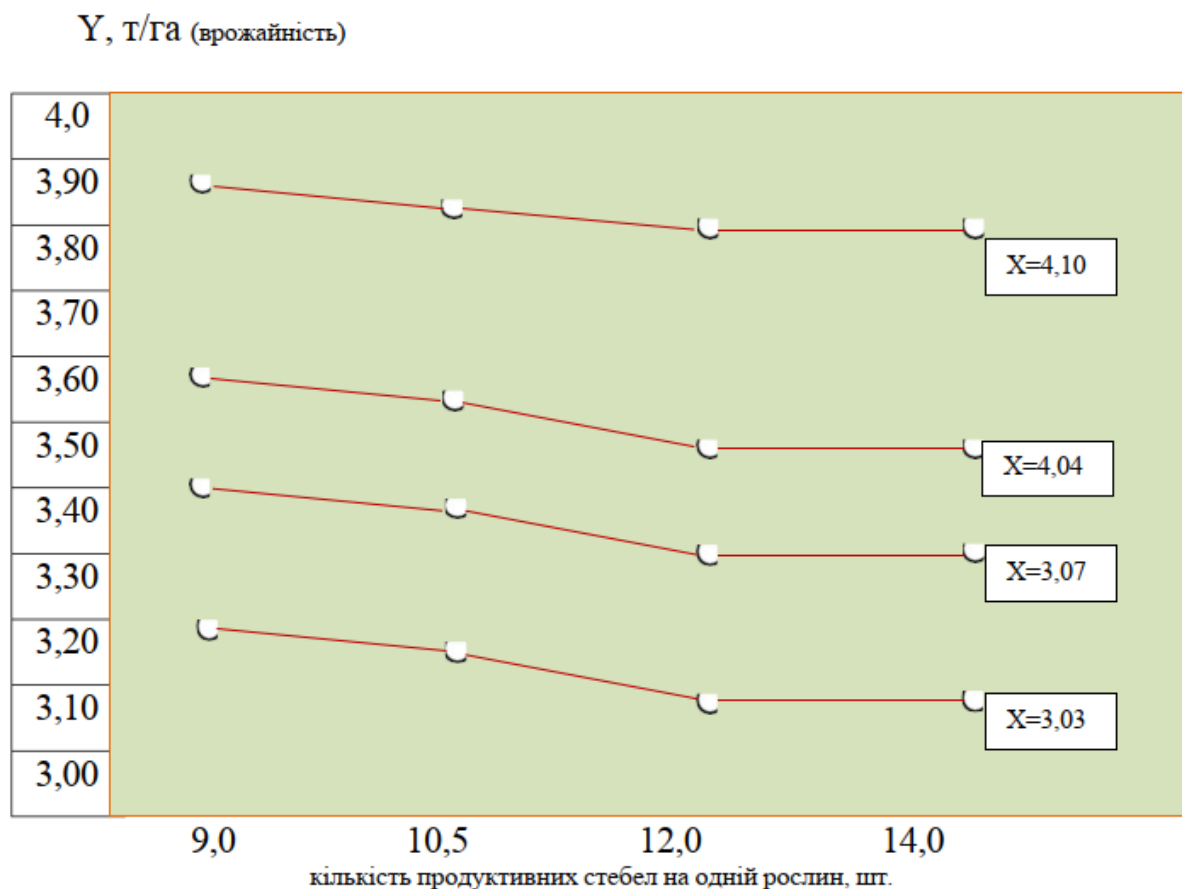


Рисунок 5.6. Лінія регресії рівняння $Y = 1,97 + \{(-0,064) \times X\} + 0,154 \times Z$ для залежності врожайності ячменю ярого від запасів води у ґрунті Z та кількості продуктивних стебел на одній рослині ячменю X.

Далі, за даними регресійної решітки (табл. 5.15) дослідник виконує роботу з графічної побудови поверхні прямолінійної регресії результативної ознаки Y на факторіальні ознаки X і Z у тривимірному просторі – площина регресії (рис. 5.7).

На графіку врожайність ячменю ярого Y, яка за показниками відповідає визначеним параметрам X і Z, відображена по вертикальному ребру куба. Запаси води у ґрунті відображені на правій стороні куба. Кількість продуктивних стебел на одній рослині ячменю відображена на лівій стороні куба. Поверхня регресії дає чітке уявлення про ефект комбінованого впливу досліджуваних факторів на результативну ознаку.

Математичні рівняння для парної та множинної регресії мають сенс тільки в області фактичних значень ознак X, Y і Z тільки в тому разі, коли кореляційний зв'язок між ознаками дослідження значно відрізняється від нуля.

Y, т/га (врожайність)

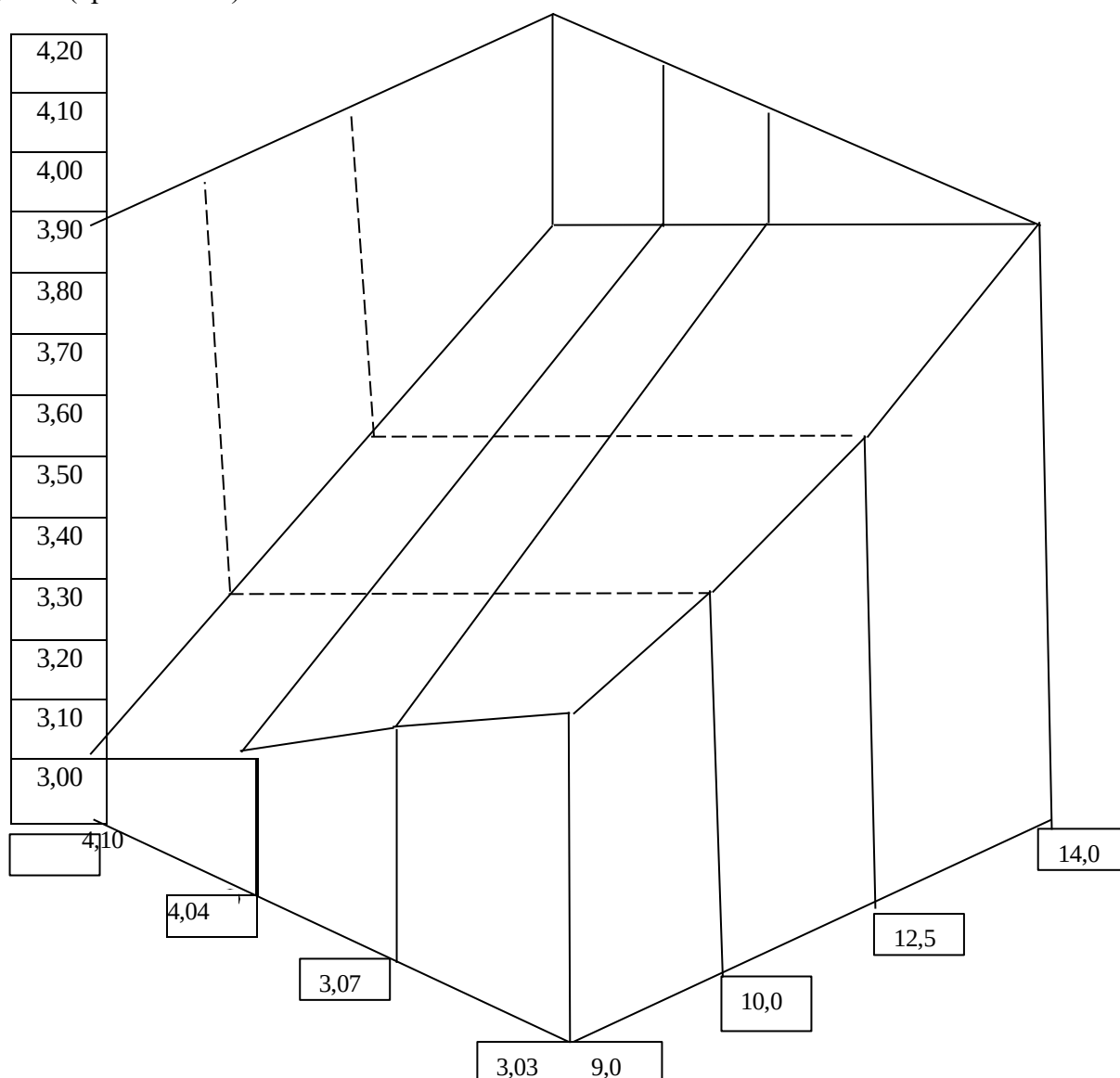


Рис. 5.7. Залежність урожайності ячменю ярого від спільної дії запасів води в шарі ґрунту 0-30 см та кількості продуктивних стебел на одній рослині ячменю (площина регресії).

5.9. Криволінійна залежність

Криволінійна регресія – залежність, коли в досліді при однакових збільшеннях параметрів незалежної змінної X залежна змінна Y характеризується неоднаковими збільшеннями параметрів.

Криволінійні залежності між двома змінними зазвичай виражаються при регресійному аналізі отриманих даних досліді у вигляді математичних рівнянь та відповідно до параметрів рівняння кривих на графіку регресії.

Емпіричні точки кореляційного поля при криволінійній кореляції розташовуються близько кривих різного типу – парабол, гіпербол, логарифмічних кривих тощо. Таким чином, на графіку криві можуть бути у

вигляді парабол, гіпербол, логарифмічних кривих.

На графіках регресії, лінії регресії є кривими і показують різні типи залежності (рис. 5.8).

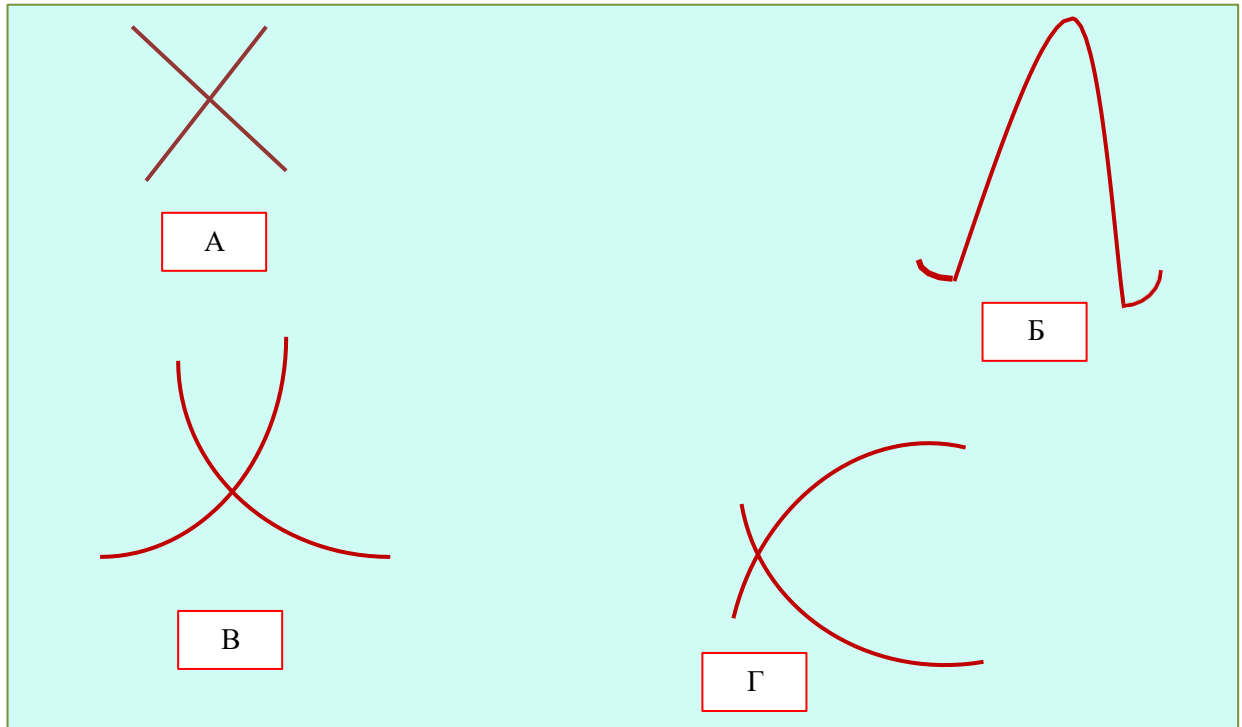


Рисунок 5.8. Лінії регресії, які характеризують типи залежності між ознаками: А) прямолінійна залежність типу $Y = a + (b \times X)$; Б) крива лінія типу параболу $Y = a + (b_1 \times X) + (b_2 \times X^2)$; В) крива лінія типу $\lg Y = a - (b \times X)$; Г) крива лінія типу $Y = a + (b \times \lg X)$.

Основний метод встановлення математичних рівнянь для побудови графіку полягає у підборі типу формули та знаходження коефіцієнта до цієї формули.

Для полегшення обчислення параметрів рівнянь криволінійної регресії способом найменших квадратів при криволінійній залежності, обране рівняння регресії доцільно звести до лінійного вигляду відповідними перетвореннями.

Статистичний обробіток отриманих даних дослідів з застосуванням методу найменших квадратів приводить наукового співробітника до побудови рівняння, яке близьке до типу квадратичної параболу:

$$Y = a + (b_1 \times X) + (b_2 \times X^2).$$

Рівняння криволінійної регресії для квадратичної параболу розраховують за такою формулою:

$$Y = y + \frac{\sum(X - x) \times Y}{\sum(X - x)^2} \times (X - x) + \left\{ \frac{\sum(X - x)^2 \times Y - (n \times C \times y)}{\sum(X - x)^4 - (n \times C^2)} \right\} \times \{(X - x)^2 - C\}$$

X – перемінна змінна, яка в даній формулі не обчислюється.

Розраховуємо С: $C = \sum(X - x)^2 / n$.

Так як в досліді встановлено дуже висока кореляційне відношення при криво ліній залежності між середньодобовим водоспоживанням рослин кукурудзи і врожайністю силосної маси кукурудзи ($\eta_{yx} = 0,99$), отримані дані досліді використовують для складання рівняння регресії криволінійної залежності між цими ознаками кукурудзи.

Дані досліді за сумарним водоспоживанням рослин кукурудзи та врожайністю силосної маси кукурудзи заносять в таблицю і виконують розрахунки для встановлення показників формули рівняння криволінійної регресії для подальшої побудови квадратичної параболи (табл. 5.16).

Таблиця 5.16

Вихідні та розрахункові дані для складання рівняння регресії криволінійної залежності між середньодобовим водоспоживанням рослин кукурудзи і врожайністю силосної маси кукурудзи

Середньодобове сумарне водоспоживання рослин кукурудзи, мм, (X)	Врожайність силосної маси кукурудзи, ц/га (Y)	(X-x)	(X-x) ²	(X-x) ⁴	(X-x)×Y	(X-x) ² ×Y
0,5	120	-1,55	2,40	5,77	-186	288
0,8	160	-1,25	1,56	2,44	-200	249,6
1,2	350	-0,85	0,72	0,52	-297,5	252
1,3	330	-0,75	0,56	0,32	-247,5	184,8
1,7	520	-0,35	0,12	0,014	-182	62,4
1,9	550	-0,15	0,022	0,00048	-82,5	12,1
2,3	520	0,25	0,062	0,0039	130	32,24
2,5	550	0,45	0,202	0,041	247,5	111,10
2,8	430	0,75	0,56	0,32	322,5	240,8
3,0	410	0,95	0,90	0,81	389,5	369
3,2	370	1,15	1,32	1,75	425,5	488,4
3,4	350	1,35	1,82	3,32	472,5	637
x=2,05	y=389		$\sum(X-x)^2 = 10,24$	$\sum(X-x)^4 = 15,31$	$\sum(X-x) \times Y = 792$	$\sum(X-x)^2 \times Y = 2927,4$
n – кількість корелюючих пар = 12.						

З таблиці вибираємо розрахункові дані для обчислення рівняння регресії.

$$\sum(X-x) \times Y = 792.$$

$$\sum(X-x)^2 = 10,24.$$

$$x=2,05.$$

$$y=389.$$

$$\sum(X - x)^2 \times Y = 2927,4.$$

$$\sum(X - x)^4 = 15,31.$$

$$n = 12.$$

$$C = \sum(X - x)^2 / n = 10,24 / 12 = 0,85.$$

За наведеними вихідними та розрахунковими даними обчислюють рівняння регресії:

$$\begin{aligned} Y &= y + \frac{\sum(X - x) \times Y}{\sum(X - x)^2} \times (X - x) + \left\{ \frac{\sum(X - x)^2 \times Y - (n \times C \times y)}{\sum(X - x)^4 - (n \times C^2)} \right\} \times \{(X - x)^2 - C\} \\ &= 389 + \frac{792}{10,24} \times (X - 2,05) + \left\{ \frac{2927,4 - (12 \times 0,85 \times 389)}{15,31 - (12 \times 0,85^2)} \right\} \times \{(X - 2,05)^2 - 0,85\} \\ &= 389 + 77,34 \times (X - 2,05) + \left\{ \frac{2927,4 - 3967,8}{15,31 - 8,67} \right\} \times \{(X^2 - 4,1X + 4,2) - 0,85\} \\ &= 389 + 77,34X - 158,55 + \left\{ \frac{-1040,4}{6,64} \right\} \times \{(X^2 - 4,1X + 4,2) - 0,85\} \\ &= 389 + 77,34X - 158,55 + (-156,69) \times (X^2 - 4,1X + 3,35) \\ &= 389 + 77,34X - 158,55 + (-156,69X^2 + 641,49X - 524,14) \\ &= 389 + 77,34X - 158,55 - 156,69X^2 + 642,43X - 524,14 \\ &= 719,77X - 156,69X^2 - 294,46 \end{aligned}$$

Математичне значення $(X - 2,05)^2$ вирішується за формулою:
 $(x - 2 \times x \times y + y^2).$

Так як в нашому прикладі у дорівнюється 2,05, в формулу вставляємо показник 2,05 і отримуємо наступне значення:

$$(x - 2 \times x \times y + y^2) = (X - 2 \times X \times 2,05 + 2,05^2) = (X - 4,1X + 4,20).$$

Таке значення застосовано було при рішенні рівняння регресії.

Рівняння регресії для криволінійної залежності між середньодобовим сумарним водоспоживанням рослин кукурудзи та врожайністю силосної маси за результатами обчислення рівняється:

$$Y = 719,77X - 156,69X^2 - 294.$$

За даним рівнянням регресії криволінійної залежності науковий співробітник набуває можливості розрахувати при проведенні досліджень теоретично прогнозовані рівні врожайності силосної маси кукурудзи за різними рівнями водоспоживання рослин кукурудзи, а також графічно відобразити залежність між цими ознаками у вигляді теоретичної лінії регресії.

За показниками середньодобового сумарного водоспоживання рослин

кукурудзи: $X=0,5$ мм; $X=0,8$ мм; $X=1,2$ мм; $X=1,3$ мм; $X=1,7$ мм; $X=1,9$ мм; $X=2,3$ мм; $X=2,5$ мм; $X=2,8$ мм; $X=3,0$ мм; $X=3,2$ мм; $X=3,4$ мм обчислюється теоретично прогнозований рівень врожайності силосної маси кукурудзи з застосуванням вже встановленого рівняння регресії криволінійної залежності (табл. 5.17).

Таблиця 5.17

Розрахування теоретично прогнозованої врожайності силосної маси кукурудзи за рівнянням регресії $Y = 719,77X - 156,69X^2 - 294$

Вихідні дані середньодобового сумарного водоспоживання рослин кукурудзи, мм, (X)	Рівняння регресії криволінійної залежності $Y = 719,77X - 156,69X^2 - 294$	Теоретично прогнозований рівень врожайності силосної маси, Y, ц/га
0,5	$Y = 719,77 \times 0,5 - 156,69 \times 0,5^2 - 294 = 359,5 - 39,17 - 294 = 26,3$	26,3
0,8	$Y = 719,77 \times 0,8 - 156,69 \times 0,8^2 - 294 = 575,82 - 100,28 - 294 = 125,1$	125,1
1,2	$Y = 719,77 \times 1,2 - 156,69 \times 1,2^2 - 294 = 863,72 - 225,63 - 294 = 344,1$	344,1
1,3	$Y = 719,77 \times 1,3 - 156,69 \times 1,3^2 - 294 = 935,70 - 264,80 - 294 = 376,9$	376,9
1,7	$Y = 719,77 \times 1,7 - 156,69 \times 1,7^2 - 294 = 1223,61 - 452,83 - 294 = 476,3$	476,3
1,9	$Y = 719,77 \times 1,9 - 156,69 \times 1,9^2 - 294 = 1367,56 - 556,65 - 294 = 516,9$	516,9
2,3	$Y = 719,77 \times 2,3 - 156,69 \times 2,3^2 - 294 = 1655,47 - 828,89 - 294 = 532,6$	532,6
2,5	$Y = 719,77 \times 2,5 - 156,69 \times 2,5^2 - 294 = 1799,42 - 979,31 - 294 = 526,1$	526,1
2,8	$Y = 719,77 \times 2,8 - 156,69 \times 2,8^2 - 294 = 2015,36 - 1228,45 - 294 = 492,9$	492,9
3,0	$Y = 719,77 \times 3,0 - 156,69 \times 3,0^2 - 294 = 2159,31 - 1410,21 - 294 = 455,1$	455,1
3,2	$Y = 719,77 \times 3,2 - 156,69 \times 3,2^2 - 294 = 2303,26 - 1604,51 - 294 = 404,8$	404,8
3,4	$Y = 719,77 \times 3,4 - 156,69 \times 3,4^2 - 294 = 2447,22 - 1811,33 - 294 = 341,9$	341,9

Розрахункові дані теоретично прогнозованого рівня врожайності силосної маси кукурудзи (Y) та значення вихідних даних середньодобового сумарного водоспоживання рослин кукурудзи (X) використовують для графічного відображення криволінійної залежності між цими ознаками (рис. 5.9).

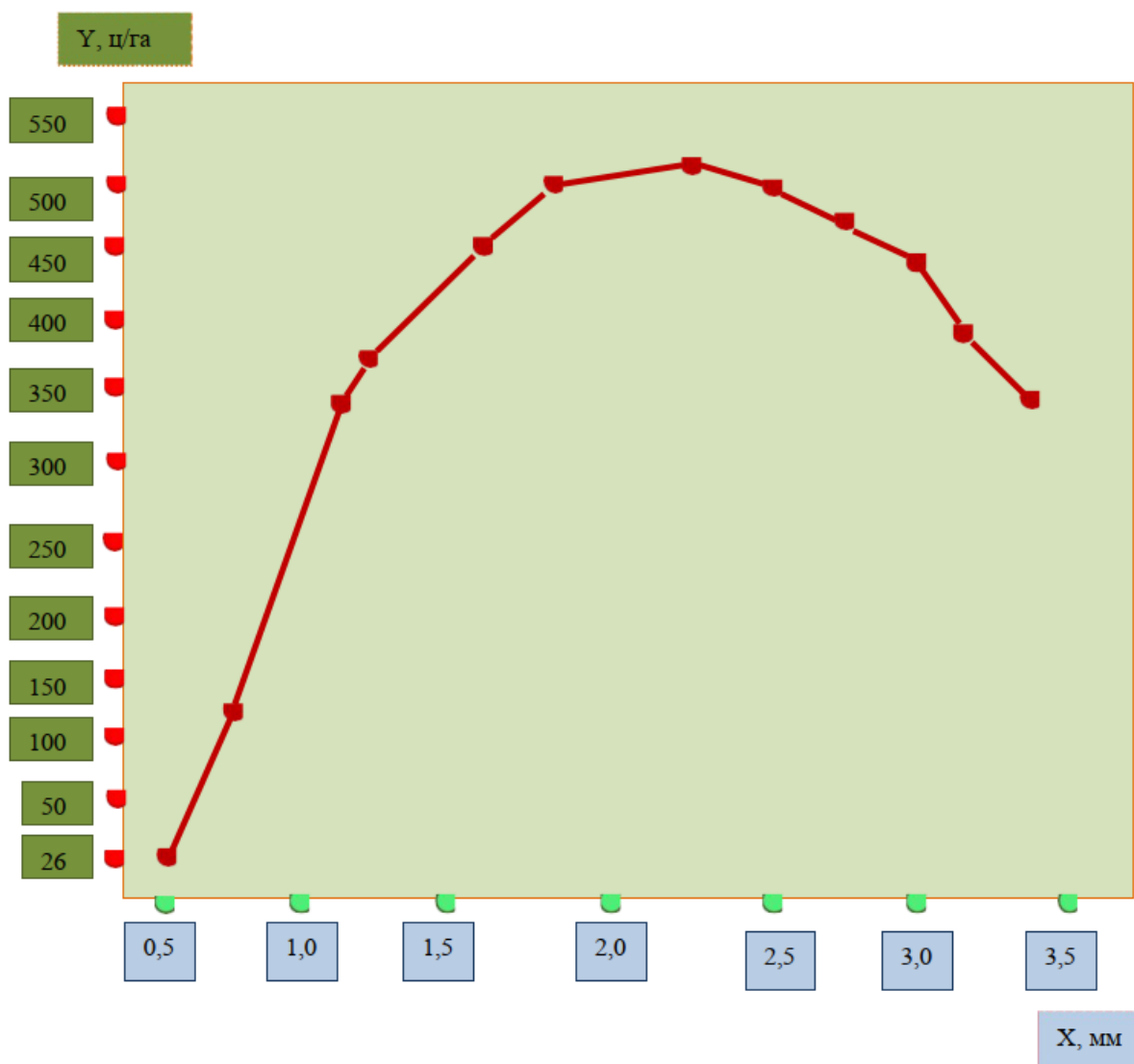


Рис. 5.9. Теоретична прогнозована крива або парабола, отримана на основі рівняння регресії: $Y = 719,77X - 156,69X^2 - 294$. Теоретична лінія регресії на графіку відображає криволінійну залежність між ознаками врожайності силосної маси кукурудзи (Y) та середньодобовим сумарним водоспоживанням рослин кукурудзи (X).

Парабола криволінійної залежності на графіку показує, що теоретично розрахункова врожайність силосної маси кукурудзи в межах від 26,3 ц/га до 532,6 ц/га відображає практичну варіацію цієї ознаки не більше середньодобового сумарного водоспоживання рослин кукурудзи 0,5-2,3 мм.

Контрольні питання

1. Що таке кореляційний аналіз?
2. Чим відрізняється функціональний зв'язок від кореляційного?
3. Які види кореляції використовують у біологічних дослідженнях?

4. Який коефіцієнт застосовують для оцінювання лінійного зв'язку між кількісними ознаками?
5. Який діапазон значень має коефіцієнт кореляції Пірсона?
6. Як інтерпретують знак коефіцієнта кореляції?
7. Що характеризує коефіцієнт детермінації (R^2)?
8. За яких умов застосовують кореляційний аналіз?
9. Яка різниця між парною та множинною кореляцією?
10. Для чого використовують регресійний аналіз?
11. Як перевіряють статистичну значущість коефіцієнта кореляції?
12. У яких випадках доцільно застосовувати криволінійну регресію?
13. Як побудувати діаграму розсіювання та лінію регресії в Excel або GraphPad Prism?
14. Яке практичне значення кореляційно-регресійного аналізу для агрономічних досліджень?
15. Чому високий коефіцієнт кореляції не завжди свідчить про причинно-наслідковий зв'язок?

Розрахункові задачі

Задача 1. Обчислення коефіцієнта кореляції Пірсона

У досліді вивчали залежність урожайності гороху від кількості бульбочок на коренях.

Варіант	Кількість бульбочок (X)	Урожайність, т/га (Y)
1	18	2,45
2	21	2,61
3	24	2,84
4	27	3,02
5	30	3,20
6	33	3,31

Необхідно:

1. Обчислити середні значення X і Y .
2. Розрахувати коефіцієнт кореляції Пірсона.
3. Обчислити коефіцієнт детермінації.
4. Оцінити силу та напрям зв'язку.
5. Зробити біологічний висновок.

Задача 2. Побудова рівняння лінійної регресії

За даними задачі 1:

1. Побудувати рівняння регресії $Y=a+bX$.
2. Визначити коефіцієнти a і b .
3. Побудувати діаграму розсіювання.
4. Нанести лінію регресії.
5. Спрогнозувати урожайність за 35 бульбочок на рослині.

Задача 3. Перевірка статистичної значущості кореляції

За результатами дослідів отримано: $r=0,82$; $n=12$.

Необхідно:

1. Розрахувати критерій Стюдента $t = \frac{r\sqrt{n-2}}{1-r^2}$
2. Визначити ступені свободи.
3. Порівняти розраховане значення з табличним.
4. Зробити висновок про статистичну значущість кореляції.

Задача 4. Порівняння двох моделей регресії

Для прогнозування урожайності соняшнику отримано дві моделі.

Лінійна

$$Y=1,92+0,082X,$$

$$R^2=0.$$

Поліноміальна

$$Y=1,35+0,146X-0,0014X^2,$$

$$R^2=0,93.$$

Необхідно:

1. Порівняти точність моделей.
2. Пояснити, яку модель доцільніше використовувати.
3. Обчислити прогноз урожайності при $X=45$.
4. Пояснити біологічний зміст отриманих результатів.

Задача 5. Аналіз множинної регресії

Отримано рівняння

$$Y=2,84+0,021X_1+0,014X_2-0,006X_3,$$

де

- Y – урожайність, т/га;
- X_1 – площа листової поверхні, тис. м²/га;
- X_2 – кількість бульбочок;
- X_3 – забур'яненість, шт./м².

Для варіанта дослідів:

- $X_1=38$;
- $X_2=27$;
- $X_3=14$.

Необхідно:

1. Визначити прогнозовану урожайність.
 2. Пояснити вплив кожного фактора.
 3. Встановити фактор, що найбільше впливає на результат.
 4. Інтерпретувати знак кожного коефіцієнта регресії.
-

Тестові завдання

1. Основною метою кореляційного аналізу є:

- А. Визначення середньої арифметичної.
- Б. Оцінювання сили та напрямку зв'язку між ознаками.
- В. Обчислення дисперсії.
- Г. Перевірка нормальності розподілу.

2. Коефіцієнт кореляції Пірсона набуває значень у межах:

- А. Від 0 до 1.
- Б. Від 0 до 100.
- В. Від -1 до $+1$.
- Г. Від -100 до $+100$.

3. Значення $r = -0,92$ свідчить про:

- А. Слабкий прямий зв'язок.
- Б. Відсутність зв'язку.
- В. Сильний обернений зв'язок.
- Г. Функціональний зв'язок.

4. Якщо $r = 0$, то:

- А. Існує сильний прямий зв'язок.
- Б. Лінійний зв'язок між ознаками відсутній.
- В. Ознаки повністю взаємозалежні.
- Г. Дані мають нормальний розподіл.

5. Який показник характеризує частку варіації результативної ознаки, пояснену моделлю?

- А. Стандартне відхилення.
- Б. Коефіцієнт варіації.
- В. Коефіцієнт детермінації (R^2).
- Г. Критерій Фішера.

6. Яке рівняння описує просту лінійну регресію?

- А. $Y = aX^2 + b$
- Б. $Y = a + bX$
- В. $Y = aX$
- Г. $Y = a + bX + cX^2 + dX^3$

7. Якщо коефіцієнт регресії $b > 0$, це означає, що:

- А. Залежність відсутня.
- Б. Із збільшенням X величина Y зменшується.
- В. Із збільшенням X величина Y зростає.
- Г. Усі значення Y однакові.

8. Для оцінювання статистичної значущості коефіцієнта кореляції використовують:

- А. Критерій Манна—Уїтні.
- Б. t-критерій Стюдента.
- В. χ^2 -критерій.
- Г. Критерій Вілкоксона.

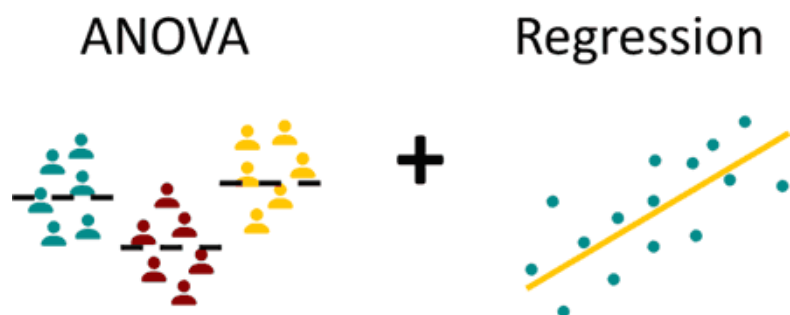
9. Множинна регресія застосовується тоді, коли:

- А. Є одна незалежна змінна.
- Б. Потрібно лише визначити середнє значення.
- В. Результативна ознака залежить від двох і більше факторів.
- Г. Дані мають ненормальний розподіл.

10. Яке твердження є правильним?

- А. Високий коефіцієнт кореляції доводить причинно-наслідковий зв'язок.
- Б. Кореляційний аналіз застосовують лише для якісних ознак.
- В. Регресійний аналіз не використовують для прогнозування.
- Г. Кореляція характеризує силу зв'язку, а регресія дає змогу кількісно описати залежність і прогнозувати значення результативної ознаки.

РОЗДІЛ 6. КОВАРІАЦІЙНИЙ АНАЛІЗ В СТАТИСТИЧНОМУ ОБРОБІТКУ РЕЗУЛЬТАТІВ ДОСЛІДІВ



- 6.1. Загальне поняття про коваріаційний аналіз
- 6.2. Основні формули для розрахунків при коваріаційному аналізі.
- 6.3. Алгоритм виконання коваріаційного аналізу
- 6.4. Приклад коваріаційного аналізу.

6.1. Загальне поняття про коваріаційний аналіз

На достовірність результатів досліджень у дослідах значний вплив може надавати невірність супутніх, неконтрольованих умов. Наприклад, рівень формування продуктивності сортів і гібридів визначається не тільки застосовуваними в досліді агротехнічними прийомами та знанням, генетичних і біологічних особливостей генотипів, а й фізико-механічними параметрами ґрунту, погодними умовами періодів вегетації, реакцією генотипів на зміну умов вегетації та іншими факторами.

Відсутність контролю з боку дослідника за неконтрольованими умовами, що змінюються, в період проведення досліджень призводять до помилок при аналізі отриманих результатів дослідів, до отримання недостовірних даних у досліді, до помилкових висновків досліджень.

При всіх подібних випадках досліднику необхідно брати до уваги, враховувати, оцінювати та вимірювати параметри факторів і умов, що впливають на прояв і формування результативних ознак дослідів і впливають на достовірність отриманих результатів.

Такий підхід до проведення досліджень та отримання достовірних даних дослідів дозволяє застосовувати статистичне вирівнювання умов досвіду і більш точно обчислити та встановити помилку досвіду.

З цією метою в статистичному обробітку результатів експериментів застосовують коваріаційний аналіз.

Коваріаційний аналіз – це поширення методів дисперсійного аналізу на випадок кількох змінних, а також кореляційного та регресійного аналізів на загальні схеми польових, вегетаційних та лабораторних дослідів в рослинництві.

Коваріаційний аналіз відноситься до складних багатовимірних

статистичних аналізів, що поєднує в собі кореляційно-регресійний та дисперсійний методи. Таким чином, коваріаційний аналіз – це значить і одночасний аналіз сум квадратів та сум стандартних відхилень двох або кількох змінних від їх середніх показників.

Головна мета методу — оцінити вплив якісних чинників (категорій) на залежну змінну, одночасно виключивши (контролюючи) вплив сторонніх кількісних змінних, які називаються **коваріатами**. Сутність коваріаційного аналізу – у вирівнюванні умов досвіду на основі обліку супутніх умов його проведення чи обліку первісного стану експериментального матеріалу, так як аналіз використовують для зменшення помилки експерименту за рахунок вирівнювання супутньої експерименту змінної X .

Величину супутньої ознаки, супутнього фактору зазвичай називають незалежною змінною X , а результативна ознака, яка вивчається в досліді, називається залежною змінною – Y .

Залежна змінна (Y) – безперервна кількісна величина, яку ми досліджуємо.

Незалежна змінна (Фактор) – якісна або категоріальна змінна, яка ділить дані на групи.

Коваріата (X) – безперервна кількісна змінна, яка не є предметом дослідження, але впливає на Y і має бути врахована (табл. 6.1).

Таблиця 6.1

Типові приклади залежної, незалежної змінної та коваріати в агрономічних дослідженнях

Приклад дослідження	Незалежна змінна	Залежна змінна	Можливі коваріати
Вплив норми азотних добрив на урожайність пшениці	Норма азоту (0, 60, 120 кг N/га)	Урожайність, т/га	Вміст мінерального азоту в ґрунті до посіву, запаси продуктивної вологи
Вплив біостимуляторів на сою	Варіант обробки (контроль, Стимпо, Регоплант)	Висота рослин, площа листової поверхні, урожайність	Початкова маса насіння, густина стояння рослин
Вплив інокуляції на горох	Наявність інокуляції (так/ні)	Кількість і маса бульбочок	Вміст доступного азоту в ґрунті, температура ґрунту
Вплив строків сівби	Строк сівби	Урожайність, польова схожість	Температура ґрунту під час сівби, вологість ґрунту
Вплив густоти посіву	Густота (тис. рослин/га)	Маса 1000 зерен, урожайність	Родючість ґрунту, кількість опадів

Завдання коваріаційного аналізу, який застосовується в статистичному обробітку, полягає у зрівнянні умов експерименту таким чином, щоб науковий співробітник зміг достовірно порівняти вплив варіантів досвіду на прояву та формування результативної ознаки Y незалежно від факторіальної ознаки X або з урахуванням поправки на цю ознаку.

Наведена характеристика коваріаційного аналізу показує, що цей аналіз використовується при плануванні та статистичній обробці результатів дослідження як спосіб зменшення помилки експерименту, яка, можливо, не піддається безпосередньому контролю (вирівнюванню).

Застосовуваний у цьому випадку, коваріаційний аналіз дозволяє встановити співвідношення між варіацією залежною змінною величиною результативної ознаки, наприклад врожайності Y , та варіацією величини змінної X , яка супроводжує у проведенні експерименту.

Наприклад, супутньою ознакою, фактором, умовою експерименту може бути зрідженість посівів сорту або гібриду, вміст поживних речовин у орному шарі ґрунту, водний режим ґрунту та періоду вегетації, температурний режим періоду вегетації тощо.

За підсумками співвідношення цих умов проводиться статистичне вирівнювання умов експерименту.

Статистичний контроль над супутньою дослідження змінної за умови, що її варіація не пов'язана з фактором, що вивчається, дає можливість досліднику отримати такий кінцевий результат, який був би отриманий за збереження величини X на постійному рівні. Це помітно уточнює результати дослідження, знижує його помилку, підвищує достовірність результатів.

Коваріація в рослинництві може розглядатися як у вузькому, так та широкому аспекті статистичного обробітку результатів дослідів.

При вузькому аспекті значення коваріації, в математичній статистиці під коваріацією (cov) розуміють середній добуток відхилення двох змінних від своїх середніх:

$$cov = \sum(X-x) \times (Y-y) / n-1.$$

Саме в математичній статистиці коваріація (cov) – це середня з здобутку відхилень від їх середніх арифметичних. При цьому аспекті коваріація може бути як позитивною, так і негативною.

У широкому аспекті розуміння коваріації, під коваріацією розуміють сукупність трьох статистичних показників:

- середніх арифметичних x і y ;
- сум квадратів відхилень $\sum(X-x)^2$ і $\sum(Y-y)^2$;
- суми добутоків відхилень $\sum(X-x) \times (Y-y)$.

Паралельне розкладання цих величин за факторами варіювання складає суть коваріаційного аналізу.

Коваріаційний аналіз включає три етапи:

1) дисперсійний аналіз за показниками ознак X , Y та їх взаємодії, яка виражається добутком показників ознак добутоків XY ;

2) розкладання залишкової дисперсії C_z за рядом Y (залишок 1) на суму квадратів відхилень, обумовлену регресією Y за X (позначається C_b) та суму квадратів відхилень від регресії $C_{d \ y \times x}$ (залишок 2), обумовлених іншими причинами, незалежно від регресії;

3) приведення фактичних середніх результативної ознаки Y до повної вирівняності умов експерименту за супутньою ознакою X .

Таким чином, коваріаційний аналіз – це поширення методів дисперсійного аналізу на випадок декількох нечасних, а також кореляційного та регресійного аналізів на загальні схеми польових, вегетаційних та лабораторних експериментів.

Коли між змінною Y , що підлягає вивченню, і супутньої змінної X можна припускати існування лінійного зв'язку, то доцільно запланувати вимірювання величини X . Це дає можливість отримати додаткову інформацію про явище, що вивчається в досліді, і використовувати регресію з метою уточнення експерименту.

У дослідях з рослинництва коваріаційний аналіз доцільно застосовувати для уточнення отриманих результатів експериментів у двох основних випадках:

Перший випадок. Якщо вихідний стан умов експерименту може надати помітний і різний вплив на результативну ознаку, наприклад, вміст поживних речовин у ґрунті, різна кількість бур'янів за варіантами дослідів перед проведенням передпосівної культивуації, тобто такі фактори і умови, які можуть бути оцінені і виміряні на початку проведення дослідів.

Другий випадок. Якщо на досліджувану ознаку в процесі експерименту впливають не залежать від варіантів досвіду і від наукового співробітника причини, умови, наприклад, випадання опадів, пошкодження рослин і зрідженість рослин, що з'являється, і так далі.

Правильне застосування коваріаційного аналізу передбачає незалежний від варіантів досвіду розподіл випадкової величини X .

Якщо супутня ознака, фактор X має відношення до факторів варіантів, які вивчаються, то виключення її ефекту неправомірно, так як це призводить до виключення частини ефекту факторів варіанту та спотворення результатів експерименту.

6.2. Основні формули для розрахунків при коваріаційному аналізі.

Розглянемо формули, які застосовують при коваріаційному аналізі.

Формула визначення суми квадратів відхилень (залишок 1), яка обумовлена регресією Y за X :

$$C_b \text{ (залишок 1)} = \{\sum (X-x) \times (Y-y)\}^2 / \sum (X-x)^2.$$

Суму квадратів відхилень від регресії (залишок 2), тобто суму квадратів випадкового варіювання, знаходять за різницею, як залишок:

$$C_{d \ y \times x} \text{ (залишок 2)} = C_z \text{ (залишок 1)} - C_b.$$

Формула визначення коефіцієнту регресії Y за X : $b_{yx} = \sum (X-x) \times (Y-y) / \sum (X-x)^2$.

Вирівнювання результативної ознаки Y проводиться за співвідношенням:

$$Y_1 = Y + \{b_{yx} \times (x-X)\},$$

де:

Y_1 – кориговане значення дати; Y – фактичне значення дати;

b_{yx} – коефіцієнт регресії Y за X ;

$(x-X)$ – різниця між середнім арифметичним значенням незалежної змінної x в досліді та фактичним її значенням X у варіанті.

Зазвичай вирівнюють лише підсумкові дані, тобто середні, тому у рівнянні регресії показники Y і X будуть відповідати середнім варіантам дослідів.

Помилку відхилення від регресії розраховують наступним чином:

$$s_{yx} = s_r \times \sqrt{(Y-y)^2},$$

де:

s_r – помилка коефіцієнта кореляції.

Внесення поправок до результатів експерименту проводять наступним чином.

Обчислюють величину F -критерію шляхом поділу середнього квадрату для регресії на дисперсію залишку 2.

Якщо фактичне значення F -критерію більше теоретичного значення ($F_{\text{факт.}} > F_{\text{теор.}}$), то зв'язок між змінними Y і X не випадковий – можна проводити коригування результатів експерименту.

Якщо $F_{\text{факт.}} < F_{\text{теор.}}$, то введення поправок марно, це не приведе до уточнення експерименту.

Поправки вводять в такий спосіб (наприклад, для корегування врожайності). До значень врожайності облікових ділянок, які за даними попереднього обліку в досліді виявилися нижчими за середній рівень, додають величину поправки, рівну $b_{yx} (x - X)$.

Якщо показники врожайності ділянок перевищують середній рівень врожайності, то тоді віднімають поправку.

Таким чином, одержують кориговані середні показники врожайності за варіантами дослідів.

6.3. Алгоритм виконання коваріаційного аналізу

Для виконання коваріаційного аналізу дослідник за алгоритмом складає схему виконання аналізу, яка передбачає проведення підготовчу роботу, а далі виконання розрахунків за встановленими для аналізу формулами та корегування даних дослідів.

Підготовча робота для виконання коваріаційного аналізу передбачає занесення вихідних даних дослідів в таблицю (табл. 6.2).

Таблиця 6.2

Вихідні дані дослідів для проведення коваріаційного аналізу

Варіанти дослідів (1)		Повторення дослідів (n)				Сума даних повторень за змінними V_x і V_y	Середня значення даних варіанту: $V_x : n, V_y : n$
		1	2	3	4		
1	X	*	*	*	*	*	*
	Y	*	*	*	*	*	*
2	X	*	*	*	*	*	*
	Y	*	*	*	*	*	*
3	X	*	*	*	*	*	*
	Y	*	*	*	*	*	*
4	X	*	*	*	*	*	*
	Y	*	*	*	*	*	*
Суми	P_x					$\sum X =$	$x =$
	P_y					$\sum Y =$	$y =$

Примітка.

* – показник; l – кількість варіантів; n – кількість повторень;

$l \times n = N$ – загальна кількість ділянок в досліді; $\sum V_x$ – сума даних показників за варіантами дослідів за змінною ознакою X; $\sum V_y$ – сума даних показників за варіантами дослідів за змінною ознакою Y; x – середнє арифметичне значення даних дослідів за ознакою X; y – середнє арифметичне значення даних дослідів за ознакою Y.

Для виконання коваріаційного аналізу на основі занесених даних відповідно до алгоритму аналізу застосовують наступні формули (табл. 6.3).

Таблиця 6.3

Формули для виконання дисперсійного аналізу по встановленню сум квадратів відхилення та їх добутків

Дисперсія	Сума квадратів відхилень та їх добутків		
	$\sum (X - x)^2$	$\sum (X - x) \times (Y - y)$	$\sum (Y - y)^2$
	X^2	$X \times Y$	Y^2
загальна, C_y	$\sum X^2 - C$	$\sum XY - C$	$\sum Y^2 / N$
повторень, C_p	$\sum P_x^2 / l - C$	$\sum P_x P_y / l - C$	$\sum P_y^2 / l - C$
варіантів, C_v	$\sum V_x^2 / n - C$	$\sum V_x V_y / n - C$	$\sum V_y^2 / n - C$
залишку 1, C_z	$C_y - C_p - C_v$	$C_y - C_p - C_v$	$C_y - C_p - C_v$
	$C = \sum (X)^2 / N$	$C = (\sum X) \times (\sum Y) / N$	$C = (\sum Y)^2 / N$

При проведенні розрахунків також використовують формули: для обчислення коефіцієнта регресії: $b_{yx} = \sum X \times Y / X^2$; суми квадратів відхилень, яка обумовлюється регресійною залежністю Y/X між ознаками, що вивчаються в досліді: $C_b = \sum X \times Y^2 / X^2$; залишкової суми квадратів відхилень, обумовлену

випадковими факторами: $C_{dy/x} = C_z - C_b$.

На останньому етапі коваріаційного аналізу обчислення виконують із застосуванням рівняння регресії.

Дані результатів коваріаційного аналізу заносять в стогову таблицю (табл. 6.4).

Таблиця 6.4

Результати коваріаційного аналізу

Дисперсія	Сума квадратів та добутків			Ступінь свободи, ν	Коефіцієнт регресії, b_{yx}	Середній квадрат, S^2	Критерії	
	X^2	$X \times Y$	Y^2				$F_{\text{факт.}}$	F_{05}
загальна, C_y	*	*	*	*	-	-	-	-
повторень, C_p	*	*	*	*	-	-	-	-
варіантів, C_v	*	*	*	*	-	*	*	*
залишок 1, C_z	*	*	*	*	-	*	-	-
регресія, C_b	-	-	*	*	*	*	*	*
залишок 2, $C_{dy/x}$	-	-	*	*	-	*	-	-

Прим.

$$X^2 = \sum (X - x)^2;$$

$$X \times Y = \sum (X - x) \times (Y - y) \quad Y^2 = \sum (Y - y)^2.$$

Критерії дисперсії варіантів та регресії обчислюють таким чином:

$$F_{\text{факт. вар.}} = \frac{\text{середній квадрат для варіантів}}{\text{квадрат для залишку 2}}$$

$$F_{\text{факт. регр.}} = \frac{\text{середній квадрат для регресії}}{\text{квадрат для залишку 2}}$$

$$\text{Залишок 2} = 1 - C_b.$$

Отримані результати коваріаційного аналізу дозволяють виконати корегування врожайності. Дані корегування врожайності заносять до таблиці (табл. 6.5).

Таблиця 6.5

Корегування врожайності

Варіанти	X	x-X	$b_{yx} \times (x-X)$	Врожайність	
				фактична	скорегована $Y_1 = Y + b_{yx}(x-X)$
1	*	*	*	*	*
2	*	*	*	*	*
3	*	*	*	*	*
4	*	*	*	*	*
середнє значення	$x=*$	*	*	$y=*$	$y_1=*$

Заключним етапом розрахунків у коваріаційному аналізі є у відповідності з

алгоритмом аналізу обчислення вирахування помилки середньої, помилки різниці, визначення найменшої суттєвої різниці.

Помилка середньої:

$$S_y = \sqrt{S^2_{\text{залишку 2}} / n}$$

$S^2_{\text{залишку 2}}$ – середній квадрат залишку 2.

Помилка різниці:

$$S_d = \sqrt{2 \times S^2_{\text{залишку 2}} / n}$$

Найменша суттєва різниця: $НСР_{05} = t_{05} \times S_d$.

Усі різниці між середніми за варіантами, які перевищують НСР, суттєві на 5%-му рівні значущості.

Результатом коваріаційного аналізу є висновок, який робить дослідник після завершення заключного етапу розрахунків. На основі наведено алгоритму розрахунків розглянемо та виконаємо приклад коваріаційного аналізу.

6.4. Приклад коваріаційного аналізу

При на ділянках конкурсного сортовивчення нових створених гібридів кукурудзи середньоранньої групи стиглості науковому співробітнику необхідно скорегувати рівень врожайності зерна на фоні зрідження рослин на облікових ділянках при чотирикратному повторенні у досліді. Перед збирання науковий співробітник виконав підрахунок рослин і відповідно визначив густоту стояння рослин (змінна X) на кожній обліковій ділянці і встановив врожайність зерна (змінна Y) кожної облікової ділянки.

Досліднику для уточнення врожайності необхідно виконати коваріаційний аналіз за отриманими даними і зробити висновок про залежність урожайності нових гібридів кукурудзи від густоти стояння рослин і суттєвої відмінності між гібридами.

Алгоритм виконання коваріаційного аналізу.

Встановлені дані за кількістю рослин та врожайності по кожній обліковій ділянці заносять в таблицю (табл. 6.6).

Таблиця 6.6

Вихідні дані дослідів для проведення коваріаційного аналізу за даними густота стояння рослин (X, тис. рослин / га) та врожайності зерна (Y, ц/га) по кожній обліковій ділянці

Варіанти дослідів (1)		Повторення дослідів (n)				Сума даних повторень за змінними V_x і V_y	Середня значення даних варіанту: $V_x : n, V_y : n$
		1	2	3	4		
Гібрид 1 (стандарт)	X	58,1	45,4	55,4	50,2	209,1	55,0
	Y	37,8	35,2	40,0	41,2	154,2	38,6
Гібрид 2	X	50,0	40,5	52,9	55,3	198,7	56,0
	Y	38,1	36,2	39,9	42,0	156,2	39,1
Гібрид 3	X	55,1	40,6	55,0	56,5	207,2	54,5
	Y	42,2	39,8	45,0	46,2	173,2	43,3
Гібрид 4	X	49,9	58,0	56,2	55,3	219,4	54,2
	Y	34,2	38,4	37,9	37,8	148,3	37,1
Суми	P_x	213,1	184,5	219,5	217,3	$\sum X = 834,4$	$\bar{x} = 54,9$
	P_y	152,3	149,6	162,8	167,2	$\sum Y = 631,9$	$\bar{y} = 39,5$

Кількість варіантів $l = 4$.

Кількість повторень у варіанті $n = 4$.

Загальна кількість облікових ділянок $N = l \times n = 4 \times 4 = 16$.

Після занесення даних за густотою стояння та врожайністю виконуємо розрахунок по встановленню суми змінної X (густина стояння рослин) – P_x повторень варіантів.

P_x для першого повторення = $58,1 + 50,0 + 55,1 + 49,9 = 213,1$.

P_x для другого повторення = $45,4 + 40,5 + 40,6 + 58,0 = 184,5$.

P_x для третього повторення = $55,4 + 52,9 + 55,0 + 56,2 = 219,5$.

P_x для четвертого повторення = $50,2 + 55,3 + 56,5 + 55,3 = 217,3$.

Розрахунок з встановлення суми змінної Y (врожайність зерна гібридів кукурудзи) – P_y повторень варіантів.

P_y для першого повторення = $37,8 + 38,1 + 42,2 + 34,2 = 152,3$.

P_y для другого повторення = $35,2 + 36,2 + 39,8 + 38,4 = 149,6$.

P_y для третього повторення = $40,0 + 39,9 + 45,0 + 37,9 = 162,8$.

P_y для четвертого повторення = $41,2 + 42,0 + 46,2 + 37,8 = 167,2$.

Розрахунок з встановлення суми даних повторень змінної X (густина стояння рослин) – V_x по кожному варіанту (гібриду).

V_x по першому варіанту = $58,1 + 45,4 + 55,4 + 50,2 = 209,1$.

V_x по другому варіанту = $50,0 + 40,5 + 52,9 + 55,3 = 198,7$.

V_x по третьому варіанту = $55,1 + 40,6 + 55,0 + 56,5 = 207,2$.

V_x по четвертому варіанту = $49,9 + 58,0 + 56,2 + 55,3 = 219,4$.

Сума даних повторень змінної X (густота стояння рослин) $\sum X$ за всіма варіантами = $209,1+199,3+207,2+219,4 = 834,4$.

Розрахунок з встановлення суми даних повторень змінної Y (врожайність зерна гібридів кукурудзи) – V_y по кожному варіанту (гібриду).

$$V_y \text{ по першому варіанту} = 37,8+35,2+40,0+41,2 = 154,2.$$

$$V_y \text{ по другому варіанту} = 38,1+36,2+39,9+42,0 = 156,2.$$

$$V_y \text{ по третьому варіанту} = 42,2+39,8+45,0+46,2 = 173,2.$$

$$V_y \text{ по четвертому варіанту} = 34,2+38,4+37,9+37,8 = 148,3.$$

Сума даних повторень змінної Y (врожайність зерна гібридів кукурудзи) $\sum Y$ за всіма варіантами = $154,2+156,2+173,2+148,3 = 631,9$.

Розрахунок з встановлення середнього значення даних варіанту змінної X (густота стояння рослин) – V_x : n по кожному варіанту (гібриду).

$$V_x: n \text{ по першому варіанту} = 209,1 : 4 = 55,0.$$

$$V_x: n \text{ по другому варіанту} = 198,7 : 4 = 56,0.$$

$$V_x: n \text{ по третьому варіанту} = 207,2 : 4 = 54,5.$$

$$V_x: n \text{ по четвертому варіанту} = 219,4 : 4 = 54,2.$$

Середнє значення даних варіантів змінної X (густота стояння рослин) в досліді $x = 55,0+56,0+54,5+54,2 = 54,9$

Розрахунок з встановлення середнього значення даних варіанту змінної Y (врожайність зерна гібридів кукурудзи) – V_y : n в кожному варіанту (гібриду).

$$V_y: n \text{ за першим варіантом} = 154,2 : 4 = 38,6.$$

$$V_y: n \text{ за другим варіантом} = 156,2 : 4 = 39,1.$$

$$V_y: n \text{ за третім варіантом} = 173,2 : 4 = 43,3.$$

$$V_y: n \text{ за четвертим варіантом} = 148,3 : 4 = 37,1.$$

Середнє значення даних варіантів змінної Y (врожайність зерна гібридів кукурудзи) $y = 38,6+39,1+43,3+37,1 = 39,5$.

Відповідно до алгоритму проведення коваріаційного аналізу наступним етапом є виконання розрахунків з визначення суми квадратів відхилень в ряду змінних X (густота стояння рослин кукурудзи), Y (врожайність зерна гібридів кукурудзи) та суми добутків змінних XY .

Розрахунок суми квадратів відхилень в ряду змінної X (густота стояння рослин кукурудзи).

Коригуючий фактор:

$$C = (\sum 834,4)^2 : N = 834,4^2 : 16 = 696223,36 : 16 = \mathbf{43513,96}.$$

Сума квадратів відхилень для загальної дисперсії:

$$\begin{aligned} C_y &= \sum X^2 - C = \\ &= (58,1^2+45,4^2+55,4^2+50,2^2+50,0^2+40,5^2+52,9^2+55,3^2+55,1^2+40,6^2+55,0^2+56,5^2+49,9^2+ \\ &+ 58,0^2+ 56,2^2+55,3^2) - 43513,96 = \\ &= (3375,61+2061,16+3069,16+2520,04+2500,00+1640,25+2798,41+3058,09+3036,01+ \\ &+ 1648,36+ 3025,00+3192,25+2490,01+3364,00+3158,44+3058,09) - 43513,96 = \end{aligned}$$

$$43994,88 - 43513,96 = \mathbf{480,92}.$$

Сума квадратів відхилень для дисперсії за повтореннями:

$$C_p = \sum P_x^2 : l - C = \{(213,1^2 + 184,5^2 + 219,5^2 + 217,3^2) : 4\} - 43513,96 = \\ \{(45411,61 + 34040,25 + 48180,25 + 47219,29) : 4\} - 43513,96 = \\ (174851,4 : 4) - 43513,96 = 43712,85 - 43513,96 = \mathbf{198,89}.$$

Сума квадратів відхилень для дисперсії за варіантами:

$$C_v = \sum V_x^2 : n - C = \\ \{(209,1^2 + 198,7^2 + 207,2^2 + 219,4^2) : 4\} - 43513,96 = \\ \{(43722,81 + 39481,69 + 42931,84 + 48136,36) : 4\} - 43513,96 = \\ (174272,7 : 4) - 43513,96 = \\ 43568,175 - 43513,96 = \mathbf{54,22}.$$

Сума квадратів відхилень для залишкової дисперсії (дисперсія – залишок 1, C_z):

$$C_z = C_y - C_p - C_v = 480,92 - 198,89 - 54,22 = \mathbf{227,81}.$$

Розрахунок суми квадратів відхилень в ряду змінної Y (врожайність зерна гібридів кукурудзи).

Коригуючий фактор:

$$C = (\sum Y)^2 : N = 631,9^2 : 16 = 399297,61 : 16 = \mathbf{24956,10}.$$

Сума квадратів відхилень для загальної дисперсії:

$$C_y = \sum Y^2 - C = \\ (37,8^2 + 35,2^2 + 40,0^2 + 41,2^2 + 38,1^2 + 36,2^2 + 39,9^2 + 42,0^2 + 42,2^2 + 39,8^2 + 45,0^2 + 46,2^2 + 34,2^2 + 38,4^2 + 37,9^2 + 37,8^2) - 24956,10 = \\ (1428,84 + 1239,04 + 1600,00 + 1697,44 + 1451,61 + 1310,44 + 1592,01 + 1764,00 + 1780,84 + 1584,04 + 2025,00 + 2134,44 + 1169,64 + 1474,56 + 1436,41 + 1428,84) - \\ 24956,10 = 25117,15 - 24956,10 = \mathbf{161,05}.$$

Сума квадратів відхилень для дисперсії за повтореннями:

$$C_p = \sum P_y^2 : l - C = \{(152,3^2 + 149,6^2 + 162,8^2 + 167,2^2) : 4\} - 24956,10 = \\ \{(23195,29 + 22380,16 + 26503,84 + 27955,84) : 4\} - 24956,10 = \\ (100\,035,13 : 4) - 24956,10 = \\ 25008,78 - 24956,10 = \mathbf{52,68}.$$

Сума квадратів відхилень для дисперсії за варіантами:

$$C_v = \sum V_y^2 : n - C = \\ \{(154,2^2 + 156,2^2 + 173,2^2 + 148,3^2) : 4\} - 24956,10 = \\ \{(23777,64 + 24398,44 + 29998,24 + 21992,89) : 4\} - 24956,10 = \\ (100167,21 : 4) - 24956,10 = \\ 25041,80 - 24956,10 = \mathbf{85,70}.$$

Сума квадратів відхилень для залишкової дисперсії (дисперсія – залишок 1, C_z):

$$C_z = C_y - C_p - C_v = 161,05 - 52,68 - 85,70 = \mathbf{22,67}.$$

Розрахунок суми добутків змінних XY (X – густина стояння рослин кукурудзи, Y – врожайність зерна гібридів кукурудзи).

Коригуючий фактор:

$$C = (\sum X \times \sum Y) : N = (834,4 \times 631,9) : 16 = 527257,36 : 16 = \mathbf{32953,59}.$$

Сума квадратів відхилень для загальної дисперсії:

$$\begin{aligned} C_y &= \sum XY - C = \\ &= (58,1 \times 37,8 + 45,4 \times 35,2 + 55,4 \times 40,0 + 50,2 \times 41,2 + 50,0 \times 38,1 + 40,5 \times 36,2 + 52,9 \times 39,9 + \\ &+ 55,3 \times 42,0 + \\ &+ 55,1 \times 42,2 + 40,6 \times 39,8 + 55,0 \times 45,0 + 56,5 \times 46,2 + 49,9 \times 34,2 + 58,0 \times 38,4 + 56,2 \times 37,9 + \\ &+ 55,3 \times 37,8) - 32953,59 = \\ &= (2196,18 + 1598,08 + 2216,00 + 2068,24 + 1905,00 + 1466,1 + 2110,71 + 2322,60 + 2325, \\ &+ 22 + 1615,88 + 2475,00 + 2610,30 + 1706,58 + 2227,20 + 2129,98 + 2090,34) - \\ &= 32953,59 = 33063,41 - 32953,59 = \mathbf{109,82}. \end{aligned}$$

Сума квадратів відхилень для дисперсії за повтореннями:

$$\begin{aligned} C_p &= \{(\sum P_x \times \sum P_y) : l\} - C = \\ &= \{(213,1 \times 152,3 + 184,5 \times 149,6 + 219,5 \times 162,8 + 217,3 \times 167,2) : 4\} - 32953,59 = \\ &= \{(32455,13 + 27601,20 + 35734,6 + 36332,56) : 4\} - 32953,59 = \\ &= (132123,49 : 4) - 32953,59 = \\ &= 33030,87 - 32953,59 = \mathbf{77,28}. \end{aligned}$$

Сума квадратів відхилень для дисперсії за варіантами:

$$\begin{aligned} C_v &= \{(\sum V_x \times \sum V_y) : n\} - C = \\ &= \{(209,1 \times 154,2 + 198,7 \times 156,2 + 207,2 \times 173,2 + 219,4 \times 148,3) : 4\} - 32953,59 = \\ &= \{(32243,22 + 31036,94 + 35887,04 + 32537,02) : 4\} - 32953,59 = \\ &= (131704,22 : 4) - 32953,59 = \\ &= 32926,06 - 32953,59 = \mathbf{-27,53}. \end{aligned}$$

Сума квадратів відхилень для залишкової дисперсії (дисперсія – залишок 1, C_z):

$$C_z = C_y - C_p - C_v = 109,82 - 77,28 - (-27,53) = 109,82 - 77,28 + 27,53 = \mathbf{60,07}.$$

Усі обчислені суми квадратів відхилень та суми творів заносять до таблиці (табл. 6.7).

Таблиця 6.7

Результати коваріаційного аналізу

Дисперсія	Сума квадратів та добутків			Степень свободи, ν	Коефіцієнт регресії, b_{yx}	Середній квадрат, S^2	Критерії	
	X^2	Y^2	$X \times Y$				$F_{\text{факт.}}$	F_{05}
загальна, C_y	480,92	161,05	109,82	15	-	-	-	-
повторень, C_p	198,89	52,68	77,28	3	-	-	-	-
варіантів, C_v	54,22	85,70	-27,53	3	-	28,56	33,6	4,07
залишок 1, C_z	227,81	22,67	60,07	9	-	2,52	-	-
регресія, C_b	-	15,84	-	1	0,26	15,84	18,6	5,32
залишок 2, $C_{d_{yx}}$	-	6,83	-	8	-	0,85	-	-

Розрахунок ступенів свободи.

Ступені свободи для подальшого проведення розрахунків будуть становити.

$$\text{Ступінь свободи для загальної дисперсії } C_y: \nu_y = N - 1 = 16 - 1 = 15.$$

Ступінь свободи для дисперсії повторень C_p : $\nu_p = n - 1 = 4 - 1 = 3$.

Ступінь свободи для дисперсії варіантів C_v : $\nu_v = n - 1 = 4 - 1 = 3$.

Ступінь свободи для дисперсії залишку 1, C_z : $\nu_z = 15 - 3 - 3 = 9$.

Ступінь свободи для дисперсії регресії C_b : $\nu_b = 2 - 1 = 1$ (так як в досліді вивчаються дві ознаки).

Ступінь свободи для дисперсії залишку 2, C_{dyx} :

$\nu_{dyx} = \text{Ступінь свободи для дисперсії залишку 1, } C_z - 1 = 9 - 1 = 8$.

Розрахунок коефіцієнта регресії b_{yx} за ознаками Y/X .

На підставі отриманих результатів визначають коефіцієнт регресії b_{yx} за ознаками Y/X :

$$b_{yx} = \sum XY : X^2,$$

де:

b_{yx} – коефіцієнт регресії ознаки Y (врожайність зерна гібридів кукурудзи) за ознакою X (густота стояння рослин кукурудзи);

$\sum XY$ – сума квадратів та добутків змінних XY за дисперсією залишок 1, C_z ;

X^2 – сума квадратів та добутків змінної X за дисперсією залишок 1, C_z .

Коефіцієнт регресії b_{yx} за ознаками Y/X становить:

$b_{yx} = \sum XY : X^2 = 60,07 : 227,81 = 0,26$ при ступені свободи $\nu = 1$ (оскільки у досліді досліджуються дві ознаки - врожайність зерна кукурудзи та густота стояння рослин кукурудзи).

Таким чином, за зміни густоти стояння рослин гібридів кукурудзи на 1 тис. рослин/га, врожайність гібридів у середньому змінюється (зростає або знижується) на 0,26 ц/га.

Розрахунок суми квадратів відхилень для регресії C_b за ознакою Y (врожайність зерна кукурудзи).

Суми квадратів відхилень для регресії C_b за ознакою Y (врожайність зерна кукурудзи) визначається за формулою:

$$C_b = (\sum XY)^2 : X^2 = (60,07)^2 : 227,81 = 3608,40 : 227,81 = 15,84.$$

Розрахунок залишкової суми квадратів відхилень за ознакою Y (врожайність зерна кукурудзи).

Залишкова сума квадратів відхилень за ознакою Y (врожайність зерна кукурудзи) зумовлена випадковими факторами.

Залишкову суму квадратів знаходять за різницею:

залишок 2, $C_{dyx} = \text{дисперсія залишку 1, } C_z - C_b = 22,67 - 15,84 = 6,83$ при 8-ми ступенях свободи ($9-1=8$).

Розрахунок середніх квадратів відхилень (дисперсії) за рядом ознаки Y

(врожайність зерна кукурудзи) за варіантами, залишку I, регресії та залишку II.

Суму квадратів відхилень за рядом ознаки Y відповідної дисперсії необхідно розділити на відповідний їй ступінь свободи: Y^2 / v .

Середній квадрат S^2 для варіантів розраховують за формулою:

сума квадратів відхилень за ознакою Y для дисперсії варіантів C_v : ступінь свободи дисперсії варіантів = $85,70 : 3 = \mathbf{28,56}$.

Середній квадрат S_z^2 для залишку 1, C_z розраховують за формулою:

сума квадратів відхилень за ознакою Y для дисперсії залишку 1, C_z : ступінь свободи дисперсії залишку 1 = $22,67 : 9 = \mathbf{2,52}$.

Середній квадрат S_b^2 для регресії, C_b розраховують за формулою:

сума квадратів відхилень за ознакою Y для дисперсії регресії, C_b : ступінь свободи дисперсії регресії, $C_b = 15,8461 = \mathbf{15,84}$.

Середній квадрат S_{dyx}^2 для залишку 2, C_{dyx} розраховують за формулою:

сума квадратів відхилень за ознакою Y для дисперсії для залишку 2, C_{dyx} : ступінь свободи дисперсії для залишку 2 = $6,83 : 8 = \mathbf{0,85}$.

Середній квадрат залишку 1, C_z дорівнюється **2,52**.

Середній квадрат залишку 2, C_{dyx} дорівнюється **0,85**.

Порівнюються показники середній квадратів залишку 1 та залишку 2.

Середній квадрат залишку, C_{dyx} (0,85), який характеризує помилку проведено дослід, після внесення поправок, знизився у порівнянні з середнім квадратом залишку 1, C_z (2,52) знизився у 2,96 разів:

$$2,52 : 0,85 = 2,96.$$

Таким чином, середній квадрат залишку 2, C_{dyx} , що характеризує помилку дослід після внесення поправок, знизився в 2,96 разів (майже в три рази).

Оцінка суттєвості відмінностей за варіантами та обумовлених регресією, тобто обчислити $F_{\text{факт.}}$ і порівняти з $F_{\text{теор.}}$ (F_{05}).

Критерій Фішера для оцінки достовірності відмінностей результатів дослід за варіантами дорівнює:

$$F_{\text{факт.}} = s_v^2 : s_{\text{залишок 2, } C_{dyx}} = 28,56 : 0,85 = \mathbf{33,6}.$$

Теоретичне значення критерію $F_{\text{теор.}}$ знаходимо за таблицею додатку 2 на 5% рівні значущості при ступенях свободи 3 та 8.

Кількість ступенів свободи результативного фактору (варіанту) = 3. Кількість ступенів свободи помилки, $v_z = 8$.

Теоретичне значення критерію $F_{05} = \mathbf{4,07}$.

Так як, при встановленні достовірності відмінностей, $F_{\text{факт.}} (33,6) > F_{05} (4,07)$, то відмінності за врожайністю гібридів кукурудзи на 5% рівні значущості суттєві (вірогідні).

Критерій Фішера для оцінки достовірності коефіцієнта регресії дорівнює:

$$F_{\text{факт.}} = s_b^2 : s_{\text{залишок 2, } C_{dyx}} = 15,84 : 0,85 = \mathbf{18,6}.$$

Теоретичне значення критерію $F_{\text{теор.}}$ знаходимо за таблицею додатку 2 на 5% рівні значущості при ступенях свободи 1 та 8.

Кількість ступенів свободи результативного фактору (варіанту) = 1.
Кількість ступенів свободи помилки, $v_z = 8$.

Теоретичне значення критерію $F_{05} = 5,32$.

Так як, при встановленні достовірності коефіцієнта регресії, $F_{\text{факт.}} (18,6) > F_{05} (5,32)$, то коефіцієнт регресії на 5% рівні значущості достовірний.

Таким чином, при отриманні достовірних даних стає доцільним привести показники врожайності гібридів кукурудзи до однакової густоти стояння рослин.

Дані з приведення показників врожайності гібридів кукурудзи до однакової густоти стояння рослин наведено в таблиці 6.8.

Таблиця 6.8

Внесення поправок для приведення середньої врожайності гібридів кукурудзи Y (т/га), до вирівняних умов густоти стояння X (тис. шт./га)

Гібрид	Середня густота стояння рослин гібрида, тис./га, X	Різниця за густотою стояння рослин, тис./га, $x-X$	$b_{yx} \times (x-X) = 0,26 \times (x-X)^*$	Врожайність гібридів кукурудзи в досліді, ц/га	
				фактична, Y	скоректована (Y_1) за формулою: $Y_1 = Y + b_{yx} \times (x-X)$
Гібрид 1 (стандарт)	55,0	-0,1	-0,026	38,6	38,6
Гібрид 2	56,0	-1,1	-0,286	39,1	38,8
Гібрид 3	54,5	+0,5	0,13	43,3	43,4
Гібрид 4	54,2	+0,7	0,182	37,1	37,3
Середня врожайність дослідів за варіантами	54,9	0	0	39,5	39,5

*Примітка. b_{yx} – розрахований коефіцієнт регресії, який дорівнюється 0,26 (табл. 6.7).

Розрахунок скоректованої врожайності за формулою $Y_1 = Y + b_{yx} \times (x-X)$ за кожним гібридом, який вивчався в досліді.

Для Гібрида 1:

$$Y_1 = Y + b_{yx} \times (x-X) = 38,6 + (-0,026) = 38,6 - 0,026 = 38,57 \approx 38,6.$$

Для Гібрида 2:

$$Y_1 = Y + b_{yx} \times (x-X) = 39,1 + (-0,286) = 39,1 - 0,286 = 38,8.$$

Для Гібрида 3:

$$Y_1 = Y + b_{yx} \times (x-X) = 43,3 + 0,13 = 43,3 + 0,13 = 43,4.$$

Для Гібрида 4:

$$Y_1 = Y + b_{yx} \times (x-X) = 37,1 + 0,182 = 37,1 + 0,182 = 37,282 \approx 37,3.$$

Оцінка суттєвості часних відмінностей.

Для визначення найменшої суттєвої різниці обчислюють помилку середньої:

$$S_y = \sqrt{\frac{\text{Середній квадрат } S^2 \text{ залишку 2, } C_{\text{dyx}}}{n}} = \sqrt{\frac{0.85}{4}} = 0,46 \text{ ц/га}$$

Для визначення найменшої суттєвої різниці обчислюють помилку різниці:

$$S_d = \sqrt{\frac{2 \times \text{Середній квадрат } S^2 \text{ залишку 2, } C_{\text{dyx}}}{n}} = \sqrt{\frac{2 \times 0.85}{4}} = 0,65 \text{ ц/га}$$

Після чого встановлюється найменша суттєва різниця.

Для розрахунку найменшої суттєвої різниці (НСР) необхідно знати критерій Стьюдента. При 8-ми ступенях свободи критерій Стьюдента t при рівні значущості $t_{0,05}$ дорівнюється 2,31 (додаток 4.).

В розрахунку застосовують дані критерію Стьюдента та помилки різниці:

$$\text{НСР}_{05} = t_{0,05} \times S_d = 2,31 \times 0,65 = 1,5 \text{ ц/га.}$$

Всі різниці між середніми показниками врожайності гібридів кукурудзи, які перевищують 1,5 ц/га, суттєві на 5% рівні значущості.

Висновок.

При однаковій густоті стояння рослин створені гібриди кукурудзи Гібрид 2 і Гібрид 4 формують однаковий рівень врожайності як і стандарт – Гібрид 1. У порівнянні із стандартом Гібрид 1 і Гібрида 2 різниця у врожайності складає 0,2 ц/га, у Гібрида 4 різниця у врожайності складає 1,3 ц/га і не перевищують найменшу суттєву різницю 1,5 ц/га.

При однаковій густоті стояння різниця за врожайністю стандарту – Гібрид 1 і Гібрида 3 складає 4,8 ц/га, що істотно перевищує найменшу суттєву різницю 1,5 ц/га. Створений гібрид 3 формує більший рівень врожайності, ніж стандарт Гібрид 1 і може бути переданий на Державне сортовипробування.

Таким чином, в дослідях з рослинництва коваріаційний аналіз використовують у двох основних випадках:

1. Якщо на результативну ознаку може мати помітний вплив різний вихідний стан умов експерименту – родючість ґрунту, потужність багаторічних рослин тощо, які можуть бути виміряні на початку проведення дослідів.

2. Якщо на результативну ознаку, яка вивчається, в досліді, впливають фактори, які не залежать від варіантів експерименту – випадання рослин після перезимівлі або від посухи, пошкодження рослин хворобами, шкідниками, птахами та інше.

Правильне застосування коваріаційного аналізу передбачає незалежний від варіантів дослідів розподіл випадкової величини « x_i ».

Але якщо супутня « x_i » має відношення до варіантів, які вивчаються в

досліді, то виключення її ефекту неправомірно, оскільки це веде до виключення частини ефекту варіанта. Наприклад, при сортовипробуванні окремі сорти або гібриди можуть уражатися більшою чи меншою мірою, і виключення цього впливу неправильно по відношенню до більш стійких сортів. Також у дослідях з просапними, овочевими та плодовими культурами, коли різна густота стояння рослин є результатом дії варіантів, що вивчаються, не можна робити ніяких поправок на зрідженість.

Контрольні питання

1. Що називають коваріаційним аналізом?
2. Яка головна мета застосування коваріаційного аналізу?
3. Чим коваріаційний аналіз відрізняється від дисперсійного аналізу?
4. Що таке коваріата?
5. Які вимоги висуваються до вибору коваріати?
6. За яких умов застосування коваріаційного аналізу є доцільним?
7. Коли застосування коваріаційного аналізу є неправомірним?
8. Які основні етапи проведення коваріаційного аналізу?
9. Як визначають коефіцієнт регресії b_{yx} ?
10. Яке біологічне значення коефіцієнта регресії?
11. Як визначають суму квадратів, обумовлену регресією (C_b)?
12. Що характеризує залишкова дисперсія після коригування ($C_{dy/x}$)?
13. Як виконується коригування результативної ознаки?
14. Як оцінюють статистичну значущість регресії за допомогою критерію Фішера?
15. Які переваги використання коваріаційного аналізу в агрономічних дослідженнях?

Розрахункові задачі

Задача 1. Розрахунок коефіцієнта регресії

У польовому досліді отримано такі статистичні характеристики:

- $\Sigma(X-\bar{x})^2 = 248,4$
- $\Sigma(X-\bar{x})(Y-\bar{y}) = 71,8$

Необхідно:

1. Обчислити коефіцієнт регресії b_{yx} .
2. Пояснити його біологічний зміст.
3. Визначити, як зміниться урожайність при збільшенні коваріати X на 5 одиниць.

Задача 2. Коригування середніх

Середнє значення коваріати в досліді $\bar{x} = 54,7$ тис. рослин/га.

Для одного із сортів: $X = 52,3$; $Y = 6,84$ т/га; $b_{yx} = 0,18$

Необхідно:

1. Обчислити поправку.
2. Визначити скориговану урожайність.
3. Пояснити, чому саме така поправка вноситься до результату.

Задача 3. Визначення доцільності застосування коваріаційного аналізу

Отримано результати: $C_b = 22,6$; $C_{dy/x} = 7,9$; $v_1 = 1$; $v_2 = 8$. Табличне значення $F_{0,05} = 5,32$.

Необхідно:

1. Обчислити середні квадрати.
2. Обчислити $F_{\text{факт}}$.
3. Зробити висновок щодо необхідності коригування результатів.

Задача 4. Оцінка ефективності коригування

До коригування: Середній квадрат залишку $S_z^2 = 3,28$.

Після коригування: $S_{dy/x}^2 = 1,09$.

Необхідно:

1. Визначити, у скільки разів зменшилась експериментальна помилка.
2. Пояснити, що означає цей показник.
3. Зробити висновок про ефективність коваріаційного аналізу.

Задача 5. Повний цикл коваріаційного аналізу

У досліді отримано: $X^2 = 265,8$; $XY = 82,4$; $C_z(Y) = 31,2$

Необхідно:

1. Обчислити коефіцієнт регресії.
2. Обчислити C_b .
3. Обчислити залишок 2.
4. Пояснити біологічний зміст отриманих результатів.
5. Вказати, чи доцільно проводити коригування врожайності.

Тестові завдання

1. Основною метою коваріаційного аналізу є:

- а) збільшення кількості повторень;
- б) усунення впливу супутньої змінної та зменшення експериментальної помилки;
- в) визначення коефіцієнта варіації;
- г) обчислення середньої арифметичної.

2. Коваріата – це:

- а) результативна ознака;
- б) фактор досліджу;

- в) кількісна змінна, яка впливає на результативну ознаку, але не є предметом дослідження;
- г) випадкова помилка.

3. Коефіцієнт регресії визначають за формулою:

- а) $\Sigma Y / \Sigma X$
- б) $\Sigma X^2 / \Sigma XY$
- в) $\Sigma XY / \Sigma X^2$
- г) $\Sigma Y^2 / \Sigma XY$

4. Якщо $F_{\text{факт}} > F_{\text{теор}}$, то:

- а) регресія відсутня;
- б) поправки вносити не потрібно;
- в) дослід необхідно повторити;
- г) коригування результатів є доцільним.

5. Якщо коваріата пов'язана з фактором дослід, то:

- а) її необхідно обов'язково вилучити;
- б) необхідно провести логарифмування;
- в) коваріаційний аналіз застосовувати не можна;
- г) потрібно збільшити число повторень.

6. Позитивне значення коефіцієнта регресії означає:

- а) відсутність зв'язку;
- б) обернений зв'язок;
- в) випадковий зв'язок;
- г) прямий зв'язок між X та Y .

7. Після проведення коваріаційного аналізу експериментальна помилка повинна:

- а) збільшитися;
- б) залишитися без змін;
- в) дорівнювати нулю;
- г) зменшитися.

8. Коригування середніх проводять за формулою:

- а) $Y = X + byx$
- б) $Y_1 = Y + byx(x - X)$
- в) $X = Y + byx$
- г) $Y_1 = X + Y$

9. Який критерій використовують для перевірки достовірності регресії?

- а) критерій Пірсона;
- б) критерій Манна-Уїтні;
- в) критерій Стюдента;

г) критерій Фішера.

10. Яка головна перевага коваріаційного аналізу?

- а) зменшення кількості повторень;
- б) скорочення часу дослідження;
- в) підвищення точності оцінки впливу факторів шляхом усунення впливу супутніх змінних;
- г) заміна дисперсійного аналізу.

РОЗДІЛ 7. ВИЗНАЧЕННЯ КОЕФІЦІЄНТА УСПАДКОВУВАНOSTІ.

$$h^2 = \frac{C_x}{C_y}$$

- 7.1. Поняття про успадкування кількісних ознак та коефіцієнт успадковуваності*
- 7.2. Проведення аналізу мінливості кількісних ознак*
- 7.3. Алгоритм обчислення коефіцієнта успадковуваності за допомогою коефіцієнтів кореляції та регресії*
- 7.4. Використання коефіцієнту успадковуваності для генетичного покращення селекційного матеріалу*
- 7.5. Алгоритм обчислення коефіцієнта успадковуваності у широкому сенсі (H^2) за допомогою дисперсійного аналізу.*
- 7.6. Алгоритм обчислення коефіцієнта успадковуваності у широкому сенсі (H^2) у двофакторному досліді за допомогою дисперсійного аналізу*

7.1. Поняття про успадкування кількісних ознак та коефіцієнт успадковуваності

Більшість ознак і властивостей організмів характеризуються кількісним типом індивідуальної мінливості, для якої є типовою безперервна зміна величини ознаки у рослин будь-якого генотипу.

В селекції величина кількісної ознаки генотипів рослин варіює від мінімального рівня в одній частині рослин до середнього і далі до максимального рівня в іншій частині груп рослин. Навіть у межах одного гомозиготного генотипу рослин може спостерігатися індивідуальна мінливість кількісної ознаки, величину якої науковий співробітник може виміряти в процесі проведення дослідів.

До кількісних ознак відносять господарсько цінні ознаки, наприклад – довжина колоса у зернових колосових культур, довжина початка у кукурудзи, кількість рядів зерна на початку кукурудзи, діаметр кошику у соняшнику, маса 1000 насінин, кількість насіння на одній рослині, маса насіння однієї рослини, вміст білку, жиру в зерні, насінні та так далі.

Кількісні ознаки характеризуються типовою безпервною зміною рівня в рослинах одного генотипу. До кількісних ознак відносять також і ряд фізіологічних відмінностей рослин.

Спадкування кількісних ознак обумовлено однаковою або подібною дією багатьох домінантних неалельних генів на ознаку (полімерія) або багатьма однозначними генами (полігінія).

На наявність двох або трьох пар однозначно діючих полімерних генів, що визначають ступінь виразності ознаки, вказує тип розщеплення ознаки у особин другого покоління. Так, за трьох домінантних генів A_1 ; A_2 і A_3 та їх рецесивних алелів a_1 , a_2 , a_3 у II поколінні буде виявлено 64 варіанти генотипів у співвідношенні 1: 6: 15:20: 15:6: 1.

Якщо загальна можливість розвитку ознаки пов'язані з дією одного гена, його прийнято називати головним (менделюючим) геном (олігогеном) і тоді ознака успадковується відповідно до закону Менделя.

Кількісний характер успадковуваності кількісних ознак є відображенням складності біологічних процесів, що лежать в основі їхнього прояву. Відсутність чітко виражених мутантів дозволяє легко аналізувати біологічні процеси, що визначають прояв кількісних ознак.

Вважається, що прояв кількісних ознак рослин перебуває під контролем полігенних систем та впливом довкілля (рис. 7.1).

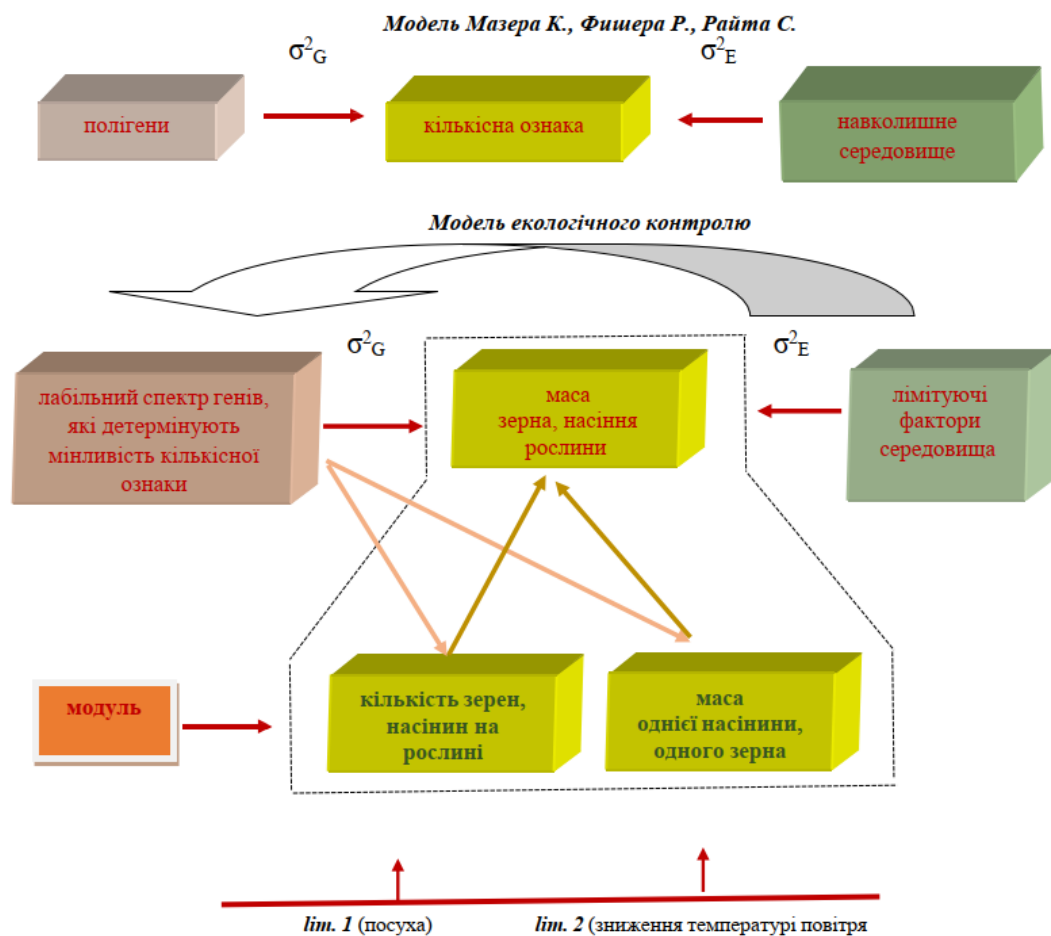


Рисунок 7.1. Моделі організації кількісних ознак. (σ_G^2 – генотипова дисперсія, σ_E^2 – паратипова дисперсія).

Полігени можуть проявляти модифікуючий вплив на кількісні ознаки та становити групу генів-модифікаторів, тобто генів, які, діючи кожен окремо, мають слабкий вплив на зміну у фенотипі, викликане дією головного гена. Гени-модифікатори можуть впливати і за відсутності головного гена.

Полімерні гени сприяють збільшенню мінливості та формуванню різних підгруп (екотипів) усередині виду, оскільки вони забезпечують різноманітні рекомбінації генотипів. Вплив рекомбінації та відбору в різних умовах середовища сприяє утворенню неподібних екотипів та підвищенню пристосованості виду до різноманіття факторів зовнішнього середовища, формуванню спадкової адаптації.

Дослідники при виконанні селекційно-генетичних досліджень мають справу з ознаками рослинного організму. Так як, кількісна ознака формується під впливом генотипу та середовища, науковому співробітнику необхідно та важливо знати, якою мірою прояв ознаки залежить від генотипу рослини. З цією метою проводять розрахунки з встановлення коефіцієнта успадковуваності ознаки, який показує частку генетичної мінливості загальної фенотипічної мінливості.

При проведенні досліджень в селекції та генетики статистичні методи оцінки характеру успадковуваності ознак та комбінаційної спроможності вихідних форм є передумовою для правильного планування та відбору батьківських компонентів для схрещування та прогнозування отримання сортів і гібридів за заданими параметрами, ознаками рослин.

Але як би не створювалася фенотипічна мінливість у популяції рослин при проведенні комбінацій схрещування, з яких компонентів мінливість не складалася, фенотипову мінливість завжди при проведенні досліджень можна встановити практично і виміряти математично за допомогою обчислення дисперсії.

Фенотип кожної рослини залежить від взаємодії спадкових факторів генотипу та умов довкілля в яких вирощується генотип рослин. При цьому, дослідник може визначити яку частку загальної фенотипової мінливості займає дисперсія, яка залежить від генетичних відмінностей у популяції рослин, що вивчаються. На таке питання досліднику відповідає показник, який характеризує успадковуваність ознаки.

Під час проведення досліджень очевидно успадковуваність не можна плутати зі спадковістю чи успадкуванням.

У селекційно-генетичних дослідженнях необхідно розрізняти три близькі, але різні поняття: спадковість, успадкування і спадковість.

Спадковість — це загальнобіологічне явище, що полягає в наявності подібних ознак у рослин батьківських форм та у рослин їх гібридних комбінацій внаслідок передачі через статеві клітини певних матеріальних структур (хромосоми з ДНК, цитоплазматичні елементи). Спадковість можна віднести до двох пов'язаних споріднених генотипів рослин: материнська форма — гібридна комбінація.

В селекції спостерігається також *спадковість цитоплазматична*

(cytoplasmic heredity) – успадкування ознак, генетичні детермінанти (одиниці спадковості) яких локалізовані у цитоплазмі клітин, а не в каріоплазмі. У клітинах рослин присутні два типи внутрішньоклітинних органел (мітохондрії та хлоропласти) з власним генетичним матеріалом – мт-ДНК та хл-ДНК. Обидва типи органел здатні до автономної реплікації та рівномірного розподілу між дочірніми клітинами при каріокінезі. До ознак з цитоплазматичним (материнським) типом успадкування відноситься у рослин – чергування зелених і безбарвних ділянок тканин листя і стебел. Строкатості відповідає наявність у клітинах двох типів пластид: пластиди з наявністю хлорофілу і пластиди без наявності хлорофілу. Іншим прикладом подібного роду служить цитоплазматична чоловіча стерильність у культурних рослин (у цитоплазмі клітин фертильних і стерильних рослин присутні два генетично різні типи мітохондрій).

Таким чином, спадковість – це здатність, властивість рослин передавати з покоління до покоління свої морфологічні, біологічні, біохімічні та фізіологічні ознаки та особливості індивідуального розвитку за певних умов середовища ознаки та виявляється в подібності батьківських форм і їх потомства.

Завдяки цій здатності рослини при схрещуванні материнських та батьківських форм зберігають у своїх нащадках характерні риси виду. Така спадкоємність спадкових якостей забезпечується передачею генетичної інформації. У еукаріотів матеріальними одиницями спадковості є гени, локалізовані в хромосомах ядра і ДНК органел.

Успадкування – це спосіб передачі спадковості батьківських форм потомству рослин за допомогою гамет та їхнього хромосомного апарату.

Успадковуваність – це є частка спадкованої мінливості у загальній фенотипічній мінливості популяції рослин. Успадковуваність – це та частка фенотипічної мінливості фенотипу, яка обумовлена дією генетичних факторів (спадковістю). Успадковуваність є суто статистичне поняттям, яке завжди характеризує прояву кількісної ознаки у рослин популяції та характеризує кількісну ознаку генотипу рослин і є показником для прогнозування ефективності селекції за фенотиповими показниками кількісної ознаки.

Успадковуваність характеризується коефіцієнтом успадковуваності. Мірою вираження успадковуваності (генетичної детермінації мінливості кількісної ознаки) служить величина коефіцієнта успадковуваності (h^2).

Коефіцієнт успадковуваності – частка генетичної мінливості у загальній варіабельності кількісної ознаки або частка генетичних відмінностей в загальній фенотиповій мінливості кількісної ознаки.

Встановлення коефіцієнту успадковуваності кількісних ознак, генетичних механізмів успадковуваності кількісних ознак дозволяють селекціонеру в процесі ведення селекції провести аналітичні дії, вирішити наступні селекційні задачі та питання:

- провести аналіз основних компонентів мінливості кількісних ознак;
- встановити значення ознаки та популяційне середнє;
- встановити фенотипове та генотипне значення ознаки;
- надати генотипна оцінка;

- виявити середній ефект гена, "ефект заміщення гена";
- встановити селекційну цінність домінантного та епістатичного відхилення;
- встановити компоненти фенотипової дисперсії ознак, принципи розкладання дисперсії на "казуальні" компоненти;
- визначити ступінь генетичної детермінації;
- визначити адитивні домінантні та епістатичні компоненти генотипної дисперсії, дисперсії;
- встановити середову дисперсію, механізми взаємодії між генотиповою та середовою компонентами дисперсії;
- надати оцінку ступеня спорідненості між рослинами в популяції та визначити коефіцієнт інбридингу при самозапиленні рослин;
- використання показників успадковуваності у селекційних експериментах та аналізі даних щодо відбору в природних популяціях;
- ефективно проводити відбір рослин за кількісними ознаками;
- встановити співвідношення основних показників відбору, вплив відбору на мінливість і успадковуваність у популяції рослин;
- виконувати відбір з урахуванням інформації щодо родословних рослин вихідного матеріалу;
- встановлення селекційного індексу, його побудова та застосування в селекційному процесі;
- визначити вплив самозапилення та гетерозису на методи та результати відбору, генотипову та фенотипічну дисперсію;
- встановити роль міжлінійних схрещувань при відборі, загальну та специфічну комбінаційну здатність батьківських форм;
- надати оцінку фенотипічних та генотипових кореляцій, їх природі проявлення;
- виконувати побудову профілів пристосованості створених генотипів до умов вирощування;
- встановити співвідношення відбору за кількісними ознаками та за пристосованістю рослин до умов вирощування;
- встановити поведінку кількісних ознак із нелінійною реакцією на відбір;
- визначити основні типи кількісних ознак, які пов'язані з адаптованістю генотипів до умов вирощування, характер їх мінливості та перетворення під дією відбору в процесі селекційної роботи.

Успадковуваність відноситься до прояву ознак у популяціях рослин і обумовлена характером передачі наслідуючої інформації від батьківських форм до їх гібридних комбінацій.

Успадковуваність характеризується двояко: середнім рівнем і певним різноманіттям ознак, які спадкуються.

У зв'язку з цим під успадковуваністю слід розуміти насамперед спадкову обумовленість різноманітності, тобто мінливості досліджуваної ознаки, а під рівнем успадковуваності – ту частку фенотипічного різноманіття кількісної ознаки, яка обумовлена обсягом генетичної інформації, що віддається від

батьківських форм до рослин гібридних комбінацій.

Тому саме, для засвоєння методів встановлення коефіцієнта успадкованості науковому співробітнику необхідно уявити себе, що таке фенотипова мінливість та її види, які зустрічаються в селекційній роботі.

Фенотипічна мінливість – це зміна фенотипу без зміни генотипу. Фенотипічна мінливість є неспадковою формою мінливості. Фенотипова мінливість представляє один із варіантів змін фенотипу рослин, у разі якого не відбуваються зміни генотипу і отримані рослиною зміни не передаються у спадок.

Одним із найпоширеніших підходів для генетичного аналізу кількісних ознак є біометричний.

Хоча властивості полігенів аналогічні властивостям генів, що визначають альтернативні ознаки, біометричний підхід передбачає, що кількісні ознаки контролюються великою кількістю генів, внесок кожного з яких у ознаку, що вивчається, малий.

Таке припущення незначності індивідуальних ефектів генів у порівнянні з розмахом екологічної мінливості кількісних ознак обумовлює необхідність застосування для їх аналізу методів, відмінних від звичайних для аналізу альтернативних ознак (розкладання дисперсій та підступів, визначення коефіцієнта успадкованості тощо).

При використанні методів вивчення мінливості та успадкованості кількісних ознак рослин у статистичній обробці отриманих результатів дослідів загальну фенотипічну мінливість кількісних ознак визначають за допомогою статистичних параметрів (σ , C_v). Фенотиповий зв'язок між ознаками визначають за допомогою статистичних параметрів із застосуванням параметрів r , b , r_s та ін.

Загальна фенотипова мінливість кількісних ознак має наступні види: генотипову (спадкову) мінливість та паратипову (неспадкова, модифікаційна) мінливість.

Загальна фенотипова мінливість кількісних ознак характеризується середнім квадратом відхилень від середньої арифметичної показника кількісної ознаки та виражається через варіансу σ^2_p .

Фенотипова варіанса популяції рослин (σ^2_p) складна. Вона утворюється під впливом різноманітності генотипів у рослин, що обумовлює прояву спадкової мінливості і виражається генотиповою варіансою.

Генотипова (спадкова) мінливість виражається через варіансу σ^2_G . Генотипова варіанса має складну природу, оскільки відбиває результат сумарного впливу генів, які надають різний ефект в розвитку ознаки. У зв'язку з цим вона складається з трьох основних компонентів, кожен із яких відповідає певному типу дії генів: $\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_J$. Адитивна генетична варіанса (σ^2_A) відображає мінливість середніх ефектів генів, домінантна варіанса (σ^2_D) вимірює варіабельність, яка створюється за рахунок взаємодій всередині багатьох алельних пар, епістатична варіанса (σ^2_J) оцінює ефект міжлокусних взаємодій, причому вона являє собою статистичний епістаз.

Паратипова (неспадкова, модифікаційна) мінливість виражається через варіансу σ^2_E .

Отже, для встановлення генетичної структури популяції за кількісними ознаками визначають три типи варіанс: фенотипову, генотипову та середову.

7.2. Проведення аналізу мінливості кількісних ознак

Для проведення селекційно-генетичного аналізу мінливості кількісних ознак досліднику необхідно розкласти загальну фенотипову мінливість ознак (σ^2_P) на генотипову (σ^2_G) і паратипову (σ^2_E).

Тому, загальна фенотипова мінливість дорівнюється таким чином:

$$\sigma^2_P = \sigma^2_G + \sigma^2_E .$$

Якщо у наведеній формулі загальну фенотипову мінливість (σ^2_P) прийняти за одиницю, то відношення дисперсій σ^2_G/σ^2_P та σ^2_E/σ^2_P покаже відсоток (частку) генетичної та паратипової мінливості у загальній фенотиповій мінливості сукупності рослин, за умови, що зв'язок генотипу з середовищем відсутній. Таким чином:

$$(\sigma^2_G / \sigma^2_P) + (\sigma^2_E / \sigma^2_P) = 1 \text{ (або 100\%)}.$$

Такий підхід у статистичному обробітку з встановлення коефіцієнта успадковуваності кількісних ознак дозволяє встановити у групі рослин частку мінливості, обумовлену їх генетичним різноманіттям, і мінливості, пов'язану з впливом чинників середовища.

Популяційна середня величина кількісної ознаки характеризує як фенотиповий її рівень, так й генотиповий її рівень у ряді поколінь за збереження чинників середовища.

Зміна середнього рівня кількісної ознаки при відборі відбувається більш ефективно у бік його збільшення, ніж у бік зменшення, що могло бути результатом дії «сильних» генів та материнського ефекту. Чим більше локусів визначає рівень ознаки і чим триваліший відбір у ряді поколінь, то спостерігається більше виражена реакція рослин на відбір.

Постійний спостерігаємий процес мутування рослин створює нове джерело мінливості та глухого кута відбору не виникає.

Для виявлення успадковуваності ознаки застосовують методи кореляційного та дисперсійного аналізу. Тим самим виявляється частка генетичного впливу на ознаку фенотипової мінливості.

В селекції за фенотиповими показниками кількісної ознаки найбільш важливе значення має вивчення та встановлення характеру успадковуваності кількісних ознак.

Коефіцієнт успадковуваності позначається як H^2 (h^2) та розраховується за формулою:

$$H^2 = \sigma^2_G / \sigma^2_P = \sigma^2_G / (\sigma^2_G + \sigma^2_E) .$$

Величина коефіцієнта успадковуваності може бути більше одиниці і менше нуля (тобто негативної).

Чим більша отримана при розрахунках величина h^2 , тим більша мінливість кількісної ознаки рослин обумовлена генетичними факторами і менша мінливість, яка викликається факторами середовища (σ^2_E).

При величині h^2 менше 0,05 (тобто менше 5%) покращення ознаки за рахунок ведення селекції є малоефективним.

При величині $h^2 > 0,3$ і $< 0,7$ селекція досить ефективна.

Особливості чинників середовища вирощування сортів і гібридів сільськогосподарських культур і ряд показників популяції (напрямок відбору, наявність інбридингу) впливають величину коефіцієнта успадковуваності.

Велика мінливість чинників середовища зменшує h^2 , а однорідність умов збільшує чи стабілізує його величину.

Обчислені коефіцієнти успадковуваності можуть характеризувати лише цей агроценоз рослин й тільки у умовах середовища і відбору.

Оскільки генотипова дисперсія включає аддитивний, домінантний і епістатичний компоненти рекомендовано розрізняти успадковуваність у широкому (H^2) і вузькому (h^2) сенсі.

Відповідно до широкого сенсу коефіцієнт успадковуваності (H^2) є частка генотипової дисперсії. Коефіцієнт успадковуваності (H^2) показує, якою мірою загальна мінливість кількісної ознаки в даному генотипі рослин обумовлена його генетичною різноманітністю.

Відповідно до вузького сенсу коефіцієнт успадковуваності (h^2) – є частка адитивної дисперсії у загальній фенотипової. В цьому сенсі коефіцієнт успадковуваності (h^2) характеризує частку генотипової мінливості, яка обумовлена адитивним ефектом генів у загальній фенотиповій мінливості. У загальній генетичній варіансі (σ^2_G) велику частку займає Адитивна генетична варіанса (σ^2_A). Тому, беручи відношення адитивної генетичної варіанси до загальної фенотипової варіанси, отримують коефіцієнт успадковуваності у вузькому значенні слова (у вузькому сенсі):

$$h^2 = \sigma^2_A / \sigma^2_G + \sigma^2_E = \sigma^2_A / \sigma^2_P.$$

Коефіцієнт успадковуваності у вузькому сенсі (h^2) показує частку адитивних генів генотипової складової, які контролюють кількісні ознаки.

При цьому виходять із таких припущень:

- генотип рослин, який вивчається в досліді, перебуває у стані генотипного рівноваги;
- кількісні ознаки визначаються адитивною дією генів;
- зв'язок між фенотипом батьківських форм та їх отриманих від їх схрещування рослин гібридних комбінацій (потомств) може бути виражений за допомогою коефіцієнта кореляції, тобто цей зв'язок є прямолінійним.

Таким чином, успадковуваність у вузькому сенсі, або як її позначають як дійсна або адитивна успадковуваність, оцінюється коефіцієнтом

успадковуваності у вузькому значенні. Коефіцієнт успадковуваності у вузькому значенні показує ставлення адитивної варіанси до загальної фенотипічної варіанси.

При розрахунках необхідно уявляти різницю між коефіцієнтами успадковуваності в широкому та вузькому сенсі.

Коефіцієнт успадковуваності в широкому сенсі (його іноді називають коефіцієнтом генетичної детермінації) свідчить, наскільки мінливість фенотипової ознаки визначається генетичними відмінностями генотипу рослин. Його величина може змінюватись від 0 до 1. Тобто, теоретично мінливість ознаки може зовсім не залежати від варіативності генотипів і, навпаки, повністю визначатися нею; чим вище значення цього коефіцієнта, тим вище роль успадковуваності у формуванні індивідуальних відмінностей рослин.

Коефіцієнт успадковуваності у вузькому сенсі оцінює лише ту частку мінливості, яка пов'язана з адитивною дією генів; завдяки цьому він дозволяє отримати відомості не про причини мінливості популяційної ознаки, а про властивості гамет і генів, які отримано гібридним потомством від схрещування батьківських форм. Важливо зауважити, що одне з головних припущень при обчисленнях коефіцієнту успадковуваності у вузькому сенсі h^2 – це відсутність коваріації та взаємодії між генетичним значенням та середовищним відхиленням, що також не завжди вірно.

Коефіцієнт успадковуваності вимірюється у відсотках чи частках одиниці. Наприклад, $H^2=0,65$, це означає, що 65% фенотипової мінливості кількісної ознаки, яка вивчається, обумовлено спадковою мінливістю рослин і відбір за цією ознакою в популяції повинен бути ефективним.

У селекційній практиці відомо кілька способів визначення коефіцієнта успадковуваності. Найбільш поширені з них – це способи визначення H^2 за допомогою коефіцієнта кореляції або коефіцієнта регресії між фенотипами споріднених генотипів рослин, а також за допомогою дисперсійного аналізу, який дозволяє розкласти фенотипову мінливість (σ^2_P) на складові її компоненти – дисперсію генотипову (σ^2_G) і дисперсію паратипову (σ^2_E).

Перед проведенням статистичної роботи по встановленню коефіцієнта успадковуваності дослідник повинен скласти алгоритм визначення коефіцієнта успадковуваності та виконувати розрахунки відповідно складено алгоритму.

7.3. Алгоритм обчислення коефіцієнта успадковуваності за допомогою коефіцієнтів кореляції та регресії

Розглянемо алгоритм складання та проведення розрахунків з встановлення коефіцієнту успадковуваності.

Алгоритм обчислення коефіцієнта успадковуваності у широкому сенсі (H^2) за допомогою коефіцієнтів кореляції та регресії за масою насіння кошику соняшнику.

1. Складання таблиці за вихідними даними дослідю.

Складається таблиця і в таблицю заносяться вихідні дані: номер комбінації

схрещування, дані кількісної ознаки маса насіння кошику материнських форм та отриманих гібридних комбінацій відповідно до кожної гібридної комбінації досліду (табл. 7.1).

Таблиця 7.1

Таблиця з розрахунковими допоміжними даними

Номер гібридної комбінації	Материнська форма (X)	Гібридна комбінація (Y)	X ²	Y ²	X×Y
1					
2					
3					
4					
5					
6					
7					
8					
9					
-					
-					
n					
Σn=	Σ X=	Σ Y=	ΣX ² =	ΣY ² =	ΣX×Y

2. Обчислення допоміжних величин на основі вихідних даних, які занесено в таблицю.

За вихідними даними робочої таблиці обчислюють допоміжні величини X², Y², (X×Y) і заносять їх в таблицю (табл. 7.1).

3. Розрахунок коефіцієнта кореляції.

Коефіцієнт кореляції розраховується за формулою:

$$r = \frac{\Sigma X \times Y - \{(\Sigma X \times \Sigma Y) : n\}}{\sqrt{\{\Sigma X^2 - (\Sigma X)^2 : n\} \times \{\Sigma Y^2 - (\Sigma Y)^2 : n\}}}$$

4. Висновок про тісноту та напрям зв'язку за величиною коефіцієнта кореляції.

На основі отриманої величини коефіцієнту кореляції встановлюють зв'язок між кількісними показниками материнських форм та гібридних комбінацій, які отримано за їх участю.

5. Розрахунок помилки коефіцієнту кореляції s_r.

Помилку коефіцієнта кореляції розраховують за формулою:

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

6. Встановлення достовірності коефіцієнту кореляції.

Розрахунки достовірності коефіцієнту кореляції виконують за наступними формулами.

Визначають фактичне значення t-критерію: $t_r = r : s_r$.

Фактичне значення критерію t_r за кількістю ступенів свободи, які встановлено в проведеному досліді ($v = n-2$), порівнюють з теоретичним значенням t-критерію на 5% рівні значущості (додаток 4).

Якщо фактичне значення критерію t_r більше теоретичного значення t-критерію $t_{0,05}$, розрахунковий коефіцієнт кореляції достовірний і можна обчислювати коефіцієнт успадкованості.

7. Розрахунок коефіцієнта успадкованості.

Коефіцієнт успадкованості за кількісною ознакою обчислюють, виходячи з теоретичних припущень про те, що H^2 дорівнює подвоєному коефіцієнту кореляції між фенотипами материнських ліній та гібридних комбінацій, отриманих за їх участю:

$$H^2 = 2 \times r.$$

Наприклад, при визначити ефективності добору рослин кукурудзи за високим вмістом жиру в зерні коефіцієнт успадкованості при розрахунковому коефіцієнті кореляції 0,36 дорівнюється $H^2 = 2 \times 0,36 = 0,72$ або 72,0 %.

Таким чином коефіцієнт успадкованості за величиною $H^2 = 0,72$ (72,0%) показує, що 72,0% фенотипової мінливості за кількісною ознакою «вміст жиру в насінні зерна кукурудзи» визначається спадковою мінливістю рослин кукурудзи і проведення добору зерна за цією ознакою може бути ефективним.

При проведенні статистичного аналізу щодо встановлення коефіцієнта успадкованості прийнято вважати, що показник H^2 , обчислений за коефіцієнтом регресії Y/X , є більш точною величиною, ніж показник, обчислений за коефіцієнтом кореляції.

Розглянемо формули розрахунку коефіцієнта успадкованості за коефіцієнтом регресії. Коефіцієнт регресії розраховується за формулою:

$$b_{yx} = \frac{\Sigma X \times Y - \{(\Sigma X \times \Sigma Y) : n\}}{\{\Sigma X^2 - (\Sigma X)^2 : n\}} = \frac{23018,65 - \{(\Sigma 373,4 \times \Sigma 560,2) : 9\}}{\{15610,74 - (\Sigma 373,4)^2 : 9\}}.$$

Наприклад, при визначити ефективності добору рослин кукурудзи за високим вмістом олії у зерні коефіцієнт успадкованості при розрахунковому коефіцієнті регресії $b_{yx} = 0,30$ дорівнюється $H^2 = 2 \times b_{yx} = 2 \times 0,30 = 0,60$ або 60,0 %. Таким чином, 60% фенотипової мінливості кількісної ознаки визначається спадковою мінливістю (генетичним факторами), 40% – впливом середовища вирощування.

Отже, дослідник бере до уваги величину коефіцієнта успадкованості, який розраховано за коефіцієнтом регресії, та робить висновок: приблизно 60% фенотипової мінливості кількісної ознаки «вміст жиру в зерні кукурудзи» обумовлено спадковою мінливістю рослин, і відбір рослин за цією ознакою буде ефективним.

У разі, коли дослідник визначає успадковуваність за потомством не від материнської форми, а від батьківської форми або від середніх значень ознаки батьківських форм, розрахунок коефіцієнта успадковуваності виконують за наступною формулою:

$$H^2 = 4 \times r \text{ або } H^2 = 4 \times b_{yx}.$$

Отримана величина коефіцієнту успадковуваності виражає міру зв'язку між генотипом і фенотипом, що має велике значення для оцінки та проведення добору рослин в селекції

Якби за фенотипом можна було точно оцінити генотип рослини, проблема селекційної оцінки і самого добору рослин набагато спростилася б. Відпала б необхідність в оцінці за родословною, за потомством.

Для багатьох господарсько цінних ознак коефіцієнт успадковуваності зазвичай невисокий, і тому оцінка за фенотипом рослин виявляється дуже неточною.

Якщо виконувати добір лише на основі індивідуальних морфологічних ознак показників, багато цінних за генотипом рослин виявляться вибракованими, і буде помилково залишено для розмноження багато генотипно посередніх індивідів.

У цих випадках на допомогу селекціонеру приходять родословні записи в польовому журналі, які містять інформацію про прямі і бічні батьківські форми даної рослини. Така поглиблена оцінка рослин за різними групами батьківських форм особливо важлива для ознак з низьким коефіцієнтом успадковуваності.

У практиці результатом статистичного аналізу іноді коефіцієнт успадковуваності виходить дорівнюється нулю або дослідник в результаті розрахунків отримує біологічно безсуттєві коефіцієнти успадковуваності, які більше одиниці або менше нуля.

Нульовий коефіцієнт успадковуваності не обов'язково може бути пов'язаний із повною гомозиготністю рослин. Нульовим коефіцієнтом успадковуваності може мати також популяція рослин, яка має гетерозиготність по всіх локусах.

Наприклад, якщо селекціонер у своєму розпорядженні має дві самозапилені лінії, які досягли повної гомозиготності в період проведення самозапилення рослин і в кожній лінії гомозиготними є різні алелі (1-а самозапилена лінія – AAbbCCDDeeff, 2-а самозапилена лінія – aaBBccddEEFF), то рослини в отриманому поколінні F_1 від схрещування цих ліній будуть гетерозиготними за всіма генами – AaBbCcDdEeFf. Однак, при проведенні опису морфологічних ознак усі рослини будуть характеризуватись ідентичним генотипом, і, отже, коефіцієнт успадковуваності кількісних ознак дорівнюватиме нулю.

Таким чином, успадковуваність пов'язана не з гомозиготністю або гетерозиготністю рослин, а саме з генетичною гетерогенністю, тобто з генотиповим різноманіттям рослин популяції.

Отримання величин коефіцієнтів успадковуваності більше одиниці або менше нуля можливо пояснити за такими причинами:

- ❖ малий, недостатній обсяг вибірки рослин при проведенні дослідів;
- ❖ взаємодія досліджуємого генотипу рослин зі середовищем, при якій ранги батьківських форм та їх потомств (отриманих в результаті комбінацій схрещування) при зміні умов вирощування змінюються по різному, приводячи до негативного коефіцієнту кореляції;
- ❖ міжалельна взаємодія генів, домінування, сверждомінування, епістаз надають більш сильний вплив на прояв ознаки, ніж аддитивна дія генів;
- ❖ порушення рівноважної структури популяції рослин під впливом інбридингу або добору, при яких розрахувати коефіцієнт успадковуваності стає недоцільним;
- ❖ популяція батьківських форм не є випадковою вибіркою рослин;
- ❖ розрахування коефіцієнту успадковуваності виконувалося різними методами.

7.4. Використання коефіцієнту успадковуваності для генетичного покращення селекційного матеріалу

Встановив величину коефіцієнту успадковуваності, визначивши, чи успадковується досліджувана кількісна ознака, які гени контролюють фенотипічний прояв кількісної ознаки, селекціонер відповідно до мета робочої програми досліджень може переходити до наступного етапу науково-дослідної роботи – добору рослин.

Для того щоб визначити генетичне покращення при доборі рослин за кількісною ознакою, яка вивчається в досліді, його розраховують для генетично однорідного матеріалу (чистих ліній, клонів) через коефіцієнт успадковуваності в широкому сенсі H^2 за наступною формулою:

$$R = S \times H^2,$$

де:

R – відповідь на селекцію (генетичний зсув), генетичне вдосконалення;

S – селекційний диференціал; H^2 – коефіцієнт успадковуваності.

Відповідь на селекцію (генетичний зсув) – це спадкова частина селекційного диференціала. У літературі зсув відбору називають респонсом та позначають індексом (R).

Селекційний диференціал (S) є різницею між середньою величиною ознаки в популяції відібраних рослин (x_e) і відповідною середньою його величиною у вихідній популяції (x_p), в якій виконувався добір рослин. Селекційний диференціал розраховують за формулою:

$$S = x_e - x_p,$$

де: x_e – середнє значення ознаки у відібраній групі рослин (від extraction – виділення, екстракція) чи однієї відібраної рослини;

x_p – середнє значення ознаки у популяції, з якої відібрано групу особин (від population – популяція), сприймається як власна вихідна популяція.

Наприклад, якщо у вихідній популяції соняшнику із середньою висотою рослин $x_p = 145$ см відбирають усі низькорослі рослини із середньою висотою $x_e = 110$ см, то отриманий селекційний диференціал становитиме 35 см.

При проведенні контрольованого перезапилення рослин застосовують формулу: $R = S \times H^2$.

При проведенні вільного перезапилення: $R = S$, тобто генетичний зсув дорівнюється селекційному диференціалу.

7.5. Алгоритм обчислення коефіцієнта успадковуваності у широкому сенсі (H^2) за допомогою дисперсійного аналізу.

Розглянемо визначення коефіцієнта успадковуваності кількісної ознаки із застосування методу дисперсійного аналізу на прикладі отриманих даних з урахуванням маси насіння кошику гібридних комбінацій соняшнику.

В досліді вивчається 6 гібридних комбінацій соняшнику. Гібридні комбінації висіяно в 4-кратному повторенні.

1. Складання таблиця та внесення вихідних даних досліді за повтореннями.

Отримані вихідні дані досліді заносять в таблицю і за цими даними розраховують суми та середні значення кількісної ознаки (табл. 7.2).

Таблиця 7.2

Вихідні дані досліді для проведення дисперсійного аналізу

Гібридна комбінація*	Повторення				Суми, V	Середні, x
	1	2	3	4		
GE57A×GE3-38Rf	65,3	65,9	65,9	64,3	261,4	63,35
GE57A×ZL678BRf	63,7	53,3	55,8	58,2	231,0	57,75
GE57A×GE111Rf	63,2	58,2	78,7	66,7	266,8	66,70
GE57A×Z1965Rf	58,9	57,3	59,5	58,6	234,3	58,58
GE57A×KLV80Rf	64,8	54,0	56,0	58,3	233,1	58,28
GE57A×LVO7RF	63,1	61,5	66,4	63,7	254,7	63,68
ΣP	379,0	350,2	382,3	369,8	$\Sigma X = 1481,3$	

*Гібридні комбінації відрізняються батьківською формою. Материнська форма – GE57A.

2. Проведення дисперсійного аналізу даних за типом однофакторного досвіду, розміщеного методом рандомізованих повторень.

Кількість варіантів, гібридних комбінацій – $l = 6$.

Кількість повторень – $n = 4$.

Загальна кількість ділянок в досліді – $N = l \times n = 6 \times 4 = 24$.

$C = (\Sigma X)^2 : N = (1481,3)^2 : 24 = 2194249,69 : 24 = 91427,07$.

$C_y = (\Sigma X^2) - C = (65,3^2 + 65,9^2 + 65,9^2 + 64,3^2 + 63,7^2 + 53,3^2 + 55,8^2 + 58,2^2 + 63,2^2 + 58,2^2 + 78,7^2 + 66,7^2 + 58,9^2 + 57,3^2 + 59,5^2 + 58,6^2 + 64,8^2 + 54,0^2 + 56,0^2 + 63,1^2 + 61,5^2 + 66,4^2 + 63,7^2) - 91427,07$

$$\begin{aligned}
 &+ 58,3^2 + 63,1^2 + 61,5^2 + 66,4^2 + 63,7^2) - 91427,07 = \\
 &= (4264,09 + 4342,81 + 4342,81 + 4134,49 + 4057,69 + 2840,89 + 3113,64 + \\
 &3387,24 + 3994,24 + 3387,24 + 6193,69 + 4448,89 + 3469,21 + 3283,29 + 3504,64 + \\
 &3433,96 + 4199,04 + 2916,00 + 3136,00 + 3398,89 + 3981,61 + 3782,25 + \\
 &4408,96 + 4057,69) - 91427,07 = 92079,26 - 91427,07 = 652,19.
 \end{aligned}$$

$$\begin{aligned}
 CP &= \{(\Sigma P^2) : l\} - C = \{(379,0^2 + 350,2^2 + 382,3^2 + 369,8^2) : 6\} - 91427,07 = \\
 &\{(143641,00 + 122640,04 + 146153,29 + 146153,29) : 6\} - 91427,07 = (549186,37 : \\
 &6) - 91427,07 = 91531,06 - 91427,07 = 103,99.
 \end{aligned}$$

$$\begin{aligned}
 CV &= \{(\Sigma V^2) : n\} - C = \{(261,4^2 + 231,0^2 + 266,8^2 + 234,3^2 + 233,1^2 + 254,7^2) : 4\} - 91427,07 = \\
 &(68329,96 + 53361 + 71182,24 + 54896,49 + 54335,61 + 64872,09) : 4 - 91427,07 = \\
 &(366977,39 : 4) - 91427,07 = 91744,35 - 91427,07 = 317,28.
 \end{aligned}$$

$$Cz = Cy - Cp - Cv = 652,19 - 103,99 - 317,28 = 230,92.$$

3. Занесення отриманих результатів дисперсійного аналізу в таблицю.

Отримані результати проведено дисперсійного аналізу за кількісною ознакою «маса насіння кошику» у гібридних комбінаціях соняшнику вносять в таблицю дисперсійного аналізу та визначають значущість дії генотипів на фенотипічну мінливість кількісної ознаки за F-критерієм (таблиця 7.3).

Таблиця 7.3

Результати дисперсійного аналізу

Дисперсія	Сума квадратів, SS	Ступінь свободи, v	Середній квадрат, σ^2	Fфакт.	Fтеор. (F05)
загальна, Cy	652,19	23			
повторень, CP	103,99	3			
батьківських форм (варіантів), CV	317,28	5	63,456	4,12	
помилки (залишку), Cz	230,92	15	15,394		

Розраховують ступеня свободи для кожного типу дисперсії і результати заносять в таблицю.

Ступінь свободи загальної дисперсії:

$$v_{Cy} = N - 1 = 24 - 1 = 23.$$

Ступінь свободи дисперсії повторень:

$$v_{CP} = n - 1 = 4 - 1 = 3.$$

Ступінь свободи дисперсії батьківських форм (варіантів):

$$v_{CV} = l - 1 = 6 - 1 = 5.$$

Ступінь свободи дисперсії помилки (залишок):

$$v_{Cz} = 23 - 3 - 5 = 15.$$

Середній квадрат батьківських форм (варіантів):

σ^2CV = сума квадратів варіантів, SS : ступінь свободи, $v = 317,28 : 5 = 63,456$.

Середній квадрат помилки:

σ^2Cz = сума квадратів помилки, SS : ступінь свободи, $v = 230,92 : 15 = 15,394$.

Фактичний критерій:

$F_{\text{факт.}}$ = середній квадрат варіантів, σ^2CV : середній квадрат помилки, $\sigma^2Cz = 63,456 : 15,394 = 4,12$.

При кількості ступенів свободи батьківських форм (варіантів) – 5 та кількості ступенів свободи помилки – 15 в додатку 2 знаходять величину теоретичного критерію F_{05} . За даними додатку 2 теоретичний критерій F_{05} дорівнюється 2,90.

При $F_{\text{факт.}} > F_{05}$ вплив батьківських форм на величину кількісного показника «маса насіння кошику» гібридних комбінацій соняшнику виявляється істотним.

4. Розрахунок коефіцієнту успадкованості.

Так як, розрахунками встановлено істотний вплив батьківських форм на формування маси насіння кошику, виконують розрахунок по визначенню коефіцієнта успадкованості кількісної ознаки за наступною формулою:

$$H^2 = \sigma^2_G / \sigma^2_P = \sigma^2_G / \sigma^2_G + \sigma^2_E ,$$

де:

H^2 – коефіцієнт успадкованості;

σ^2_P – дисперсія фенотипової мінливості;

σ^2_G – дисперсія генотипової мінливості;

σ^2_E – дисперсія паратипової мінливості.

Для обчислення коефіцієнта успадкованості кількісної ознаки необхідно розібратися з дисперсіями, які використовуються безпосередньо в аналізі і в формулі.

Дисперсія варіантів (σ^2_V), обумовлена мінливістю групових середніх і має складну природу. Дисперсія варіантів визначається як генотиповою мінливістю кількісної ознаки σ^2_G , так і випадковою, паратиповою мінливістю σ^2_E . Отже, дисперсія варіантів, тобто батьківських форм σ^2_V , складається з двох компонентів:

$$\sigma^2_V = \sigma^2_E + (n \times \sigma^2_G),$$

де:

n – кількість повторень в досліді. В досліді, який наведено в якості прикладу, кількість повторень 4.

Кількість повторень вказується в формулі перед генотиповою мінливістю групових середніх, щоб привести до рівня первинних спостережень.

Таким чином, генотипова дисперсія (σ^2_G) буде дорівнюватися:

$$\sigma^2_G = (\sigma^2_V - \sigma^2_E) : n.$$

За даними таблиці 7.3 дисперсія варіантів $\sigma^2_V = \sigma^2_{C_V} = 63,546$, паратипова дисперсія (дисперсія середовища) $\sigma^2_E = \sigma^2_{C_Z} = 15,394$.

Паратипова мінливість σ^2_E за величиною дорівнюється середньому квадрату дисперсії середовища $\sigma^2_{C_Z}$.

Такі дані застосовуємо у формулі для визначення дисперсії генотипової:

$$\sigma^2_G = (\sigma^2_V - \sigma^2_E) : n = (63,546 - 15,394) : 4 = 48,152 : 4 = 12,038.$$

Генотипова мінливість σ^2_G за величиною дорівнюється дисперсії генотипової σ^2_G .

Фенотипова дисперсія – це сума генотипової та паратипової (середовищної) дисперсії: $\sigma^2_P = \sigma^2_G + \sigma^2_E$.

Визначається фенотипова дисперсія:

$$\sigma^2_P = \sigma^2_G + \sigma^2_E = 12,038 + 15,394 = 27,432.$$

Таким чином, за розрахунками величини мінливості складають: генотипова мінливість (дисперсія генотипова) $\sigma^2_G = 12,038$; фенотипова мінливість (фенотипова дисперсія) $\sigma^2_P = 27,432$.

За величинами генотипової та фенотипової мінливостей розраховують коефіцієнт успадковуваності кількісної ознаки «маса насіння кошику» гібридних комбінацій соняшнику:

$$H^2 = \sigma^2_G : \sigma^2_P = 12,038 : 27,432 = 0,44 (44,0\%).$$

Висновок.

Коефіцієнт успадковуваності, що характеризує ступінь генетичної мінливості кількісної ознаки «маса насіння кошику» при передачі від батьківської форми гібридним комбінаціям соняшнику, дорівнює 0,44 (44,0%). Бракування та добір за цією ознакою новинні бути ефективними.

7.6. Алгоритм обчислення коефіцієнта успадковуваності у широкому сенсі (H^2) у двофакторному досліді за допомогою дисперсійного аналізу

Розглянемо приклад встановлення коефіцієнту успадковуваності кількісної ознаки «маса насіння кошику» соняшнику при проведенні двофакторного досліді.

В досліді вивчаються гібридні комбінації соняшнику, які отримано із

застосуванням трьох материнських форм. Кожна материнська форма схрещувалась з однаковими трьома батьківськими формами.

В досліді вивчається 9 гібридних комбінацій соняшнику. Гібридні комбінації висіяно в 4-кратному повторенні.

1. Складання таблиця та внесення вихідних даних досліді за повтореннями.

Отримані вихідні дані досліді заносять в таблицю і за цими даними розраховують суми та середні значення кількісної ознаки (табл. 7.4).

Таблиця 7.4

**Вихідні дані досліді для проведення дисперсійного аналізу
двофакторного досліді**

Гібридна комбінація		Повторення				Суми, V	Середні, x
материнська форма (A)	батьківська форма (B)	1	2	3	4		
GE57A	GE3-38Rf	65,3	65,9	65,9	64,3	261,4	63,35
	GE111Rf	63,2	58,2	78,7	66,7	266,8	66,70
	ZL678Rf	63,7	53,3	55,8	58,2	231,0	57,75
ZL50A	GE3-38Rf	72,7	55,8	61,0	62,9	252,4	63,1
	GE111Rf	60,4	61,1	48,4	56,2	226,1	56,5
	ZL678Rf	69,8	60,3	59,2	63,6	252,9	63,2
ZL52A	GE3-38Rf	54,1	61,1	56,6	57,5	229,3	57,3
	GE111Rf	67,4	85,8	75,9 ²	76,0	305,1	76,3
	ZL678Rf	58,6	60,0	53,3	57,3	229,2	57,3
ΣP		575,2	561,5	554,8	562,7	$\Sigma X=2254,2$	

Фактор А – материнська форма (лінія).

Фактор В – батьківська форма (лінія).

Варіантів – $l = 9$.

Варіантів за фактором А – $l_A = 3$.

Варіантів за фактором В – $l_B = 3$.

Таким чином, всього варіантів – $l = l_A \times l_B = 3 \times 3 = 9$.

Повторень – $n = 4$

Загальна кількість ділянок $N = l \times n = 9 \times 4 = 36$.

2. Дисперсійний аналіз даних за типом двофакторного досліді, розміщеного методом рандомізованих повторень.

$$C = (\Sigma X)^2 : N = (2254,2)^2 : 36 = 5081417,64 : 36 = 141150,49.$$

$$C_y = \Sigma X^2 - C = (65,3^2 + 65,9^2 + 65,9^2 + 64,3^2 + 63,2^2 + 58,2^2 + 78,7^2 + 66,7^2 + 63,7^2 + 53,3^2 + 55,8^2 + 58,2^2 + 72,7^2 + 55,8^2 + 61,0^2 + 62,9^2 + 60,4^2 + 61,1^2 + 48,4^2 + 56,2^2 + 69,8^2 + 60,3^2 + 59,2^2 + 63,6^2 + 54,1^2 + 61,1^2 + 56,6^2 + 57,5^2 + 67,4^2 + 85,8^2 + 75,9^2 + 76,0^2 + 58,6^2 + 60,0^2 + 53,3^2 + 57,3^2) - 141150,49 = (4264,09 + 4342,81 +$$

$$4342,81 + 4134,49 + 3994,24 + 3387,24 + 6193,69 + 4448,89 + 4057,69 + 2840,89 + 3113,64 + 3387,24 + 5285,29 + 3113,64 + 3721,00 + 3956,41 + 3648,16 + 3733,21 + 2342,56 + 3158,44 + 4872,04 + 3636,09 + 3504,64 + 4044,96 + 2926,81 + 3733,21 + 3203,56 + 3306,25 + 4542,76 + 7361,64 + 5760,81 + 5776,00 + 3433,96 + 3600,00 + 2840,89 + 3283,29) - 141150,49 = 143293,34 - 141150,49 = 2142,85.$$

$$C_p = \{(\Sigma P^2) : (1A \times 1B)\} - C = \\ \{(575,2^2 + 561,5^2 + 554,8^2 + 562,7^2) : (3 \times 3)\} - 141150,49 = . \\ \{(330855,04 + 315282,25 + 307803,04 + 316631,29) : (3 \times 3)\} - 141150,49 = \\ (1270571,62 : 9) - 141150,49 = \\ 141174,62 - 141150,49 = 24,13.$$

$$C_v = \{(\Sigma V^2) : n\} - C = \\ \{(261,4^2 + 266,8^2 + 231,0^2 + 252,4^2 + 226,1^2 + 252,9^2 + 229,3^2 + 305,1^2 + 229,2^2) : 4\} - 141150,49 = \\ \{(68329,96 + 71182,24 + 53361,00 + 63705,76 + 51121,21 + 63958,41 + 52578,49 + 93086,01 + 52532,64) : 4\} - 141150,49 = \\ (569855,72 : 4) - 141150,49 = \\ 142463,93 - 141150,49 = 1313,44.$$

$$C_z = C_y - C_p - C_v = 2142,85 - 24,13 - 1313,44 = 805,28.$$

3. Визначення суми кількісної ознаки "маса насіння кошика" у гібридів за кожною материнською формою.

Для оцінки суттєвості впливу материнських та батьківських форм та їх взаємодії на масу насіння у кошику у гібридів розраховують суму величин кількісної ознаки за всіма варіантами для кожного гібрида окремо за кожною материнською лінією та батьківською лінією за факторами А та В. Дані для розрахунків використовують із таблиці 7.4.

Сума насіння кошика за фактором А – материнська форма.

Розрахунки за материнськими формами та гібридними комбінаціями за їх участю в комбінаціях схрещування.

$$\text{Сума маси насіння кошика варіантів лінії GE57A} = 261,4 + 266,8 + 231,0 = 759,2.$$

$$\text{Сума маси насіння кошика варіантів лінії ZL50A} = 252,4 + 226,1 + 252,9 = 731,4$$

$$\text{Сума маси насіння кошика варіантів лінії ZL52A} = 229,3 + 305,1 + 229,2 = 763,6.$$

Сума насіння кошика за фактором В – батьківська форма.

Розрахунки за батьківськими формами та гібридними комбінаціями за їх участю в комбінаціях схрещування.

$$\text{Сума маси насіння кошика варіантів лінії GE3-38Rf} = 261,4 + 252,4 + 229,3 = 743,1.$$

$$\text{Сума маси насіння кошика варіантів лінії GE111Rf} = 266,8 + 226,1 + 305,1 = 798,0.$$

Сума маси насіння кошика варіантів лінії ZL678Rf = 231,0 + 252,9 + 229,2 = 713,1.

Отримані дані заносять до таблиці 7.5 і розраховують загальну суму $\sum X$.

Таблиця 7.5

Таблиця сум для визначення ефектів факторів А, В та їх взаємодії АВ

Фактор А (материнська форма)	Фактор В (батьківська форма)			Суми за фактором А
	GE3-38Rf	GE111Rf	ZL678Rf	
GE57A	261,4	266,8	231,0	759,2
ZL50A	252,4	226,1	252,9	731,4
ZL52A	229,3	305,1	229,2	763,6
Суми за фактором В	743,1	798,0	713,1	$\sum X = 2254,2$

4. Розрахунок дисперсії факторів А та В, взаємодії факторів АВ.

Проведеними раніше розрахунками поправка досліджується $C = 141150,49$, сума квадратів варіантів дорівнюється $C_v = 1313,44$.

Розраховують суму квадратів для факторів А та В, взаємодії АВ. Сума квадратів фактору А:

$$C_A = \{(\sum A^2) : (l_B \times n)\} - C =$$

$$\{(759,2^2 + 731,4^2 + 763,6^2) : (3 \times 4)\} - 141150,49 =$$

$$\{(576384,64 + 534945,96 + 583084,96) : 12\} - 141150,49 =$$

$$(1694415,56 : 12) - 141150,49 =$$

$$141201,30 - 141150,49 = 50,81.$$

При ступені свободи $v = l_A - 1 = 3 - 1 = 2$.

$$C_B = \{(\sum B^2) : (l_A \times n)\} - C =$$

$$\{(743,1^2 + 798,0^2 + 713,1^2) : (3 \times 4)\} - 141150,49 =$$

$$\{(552197,61 + 636804,00 + 508511,61) : 12\} - 141150,49 =$$

$$(1699513,22 : 12) - 141150,49 =$$

$$141626,10 - 141150,49 = 475,61.$$

При ступені свободи $v = l_B - 1 = 3 - 1 = 2$.

Сума квадратів взаємодії факторів АВ:

$$C_{AB} = C_v - C_A - C_B = 1313,44 - 50,81 - 475,61 = 787,02.$$

При ступені свободи $v = (l_A - 1) \times (l_B - 1) = (3 - 1) \times (3 - 1) = 2 \times 2 = 4$.

5. Складання таблиці дисперсійного аналізу та оцінка суттєвості дії та взаємодії за критерієм F.

Отримані результати проведеного дисперсійного аналізу заносяться до таблиці (табл. 7.6).

Таблиця 7.6

Результати дисперсійного аналізу досліджу

Дисперсія	Сума квадратів, SS	Ступінь свободи, ν	Середній квадрат, σ^2	Fфакт.	Fтеор. (F05)
загальна, S_y	2142,85	35			
повторень, S_p	24,13	3			
материнських форм, фактор А, S_A	50,81	2	25,41	0,75	3,40
батьківських форм, фактор В, S_B	475,61	2	237,81	7,09	3,40
взаємодія АВ, S_{AB}	787,02	4	196,75	5,86	2,78
залишку, S_z	805,28	24	33,55		

Розраховують ступеня свободи для кожного типу дисперсії і результати заносять в таблицю.

Ступінь свободи загальної дисперсії:

$$\nu_{Cy} = N - 1 = 36 - 1 = 35.$$

Ступінь свободи дисперсії повторень:

$$\nu_{CP} = n - 1 = 4 - 1 = 3.$$

Ступінь свободи дисперсії материнських форм:

$$\nu_{CA} = l_A - 1 = 3 - 1 = 2.$$

Ступінь свободи дисперсії батьківських форм:

$$\nu_{CB} = l_B - 1 = 3 - 1 = 2.$$

Ступінь свободи взаємодії факторів АВ:

$$\nu_{CAB} = (l_A - 1) \times (l_B - 1) = (3 - 1) \times (3 - 1) = 2 \times 2 = 4.$$

Ступінь свободи дисперсії помилки (залишок):

$$\nu_{Cz} = 35 - 3 - 2 - 2 - 4 = 24.$$

Розраховують середні квадрати.

Середній квадрат материнських форм:

$$\sigma^2_{CA} = \text{сума квадратів материнських форм, SS} : \text{ступінь свободи, } \nu_{CA} = 50,81 : 2 = 25,41.$$

Середній квадрат батьківських форм:

$$\sigma^2_{CB} = \text{сума квадратів батьківських форм, SS} : \text{ступінь свободи, } \nu_{CB} = 475,61 : 2 = 237,81.$$

Середній квадрат взаємодії батьківських форм АВ:

$$\sigma^2_{CAB} = \text{сума квадратів взаємодії АВ, SS} : \text{ступінь свободи, } \nu_{CAB} = 787,02 : 4 = 196,75.$$

Середній квадрат помилки:

σ^2_{Cz} = сума квадратів помилки, SS : ступінь свободи, $\nu_{Cz} = 805,28 : 24 = 33,55$.

Розраховують фактичні критерії в досліді.

Фактичний критерій для фактору материнські форми:

$F_{\text{факт.}}$ = середній квадрат материнської форми, σ^2_{CA} : середній квадрат помилки, $\sigma^2_{Cz} = 25,41 : 33,55 = 0,75$.

При кількості ступенів свободи материнських форм – 2 та кількості ступенів свободи помилки – 24 в додатку 2 знаходять величину теоретичного критерію F_{05} .

За даними додатку 2 теоретичний критерій F_{05} дорівнюється 3,40.

При $F_{\text{факт.}} < F_{05}$ ($0,75 < 3,40$) вплив материнських форм на величину кількісного показника «маса насіння кошику» гібридних комбінацій соняшнику виявляється неістотним.

Фактичний критерій для фактору батьківські форми:

$F_{\text{факт.}}$ = середній квадрат батьківської форми, σ^2_{CAB} : середній квадрат помилки, $\sigma^2_{Cz} = 237,81 : 33,55 = 7,09$.

При кількості ступенів свободи батьківських форм – 2 та кількості ступенів свободи помилки – 24 в додатку 2 знаходять величину теоретичного критерію F_{05} .

За даними додатку 2 теоретичний критерій F_{05} дорівнюється 3,40.

При $F_{\text{факт.}} > F_{05}$ ($7,09 > 3,40$) вплив батьківських форм на величину кількісного показника «маса насіння кошику» гібридних комбінацій соняшнику виявляється істотним.

Фактичний критерій для фактору взаємодії батьківських форм:

$F_{\text{факт.}}$ = середній квадрат взаємодії факторів АВ, σ^2_{CAB} : середній квадрат помилки, $\sigma^2_{Cz} = 196,75 : 33,55 = 5,86$.

При кількості ступенів свободи взаємодії батьківських форм – 2 та кількості ступенів свободи помилки – 24 в додатку 2 знаходять величину теоретичного критерію F_{05} .

За даними додатку 2 теоретичний критерій F_{05} дорівнюється 2,78.

При $F_{\text{факт.}} > F_{05}$ ($7,09 > 2,78$) вплив взаємодії батьківських форм на величину кількісного показника «маса насіння кошику» гібридних комбінацій соняшнику виявляється істотним.

Дія батьківських форм (фактор В) та взаємодія батьківських форм (АВ) виявилися суттєвими, так як $F_{\text{факт.}} > F_{05}$.

Суттєвий вплив на кількісну ознаку «маса насіння кошику» в гібридних комбінаціях соняшнику в даному досліді чинили батьківські форми та взаємодія материнських та батьківських форм (материнська форма × батьківська форма).

Таким чином, є сенс обчислити коефіцієнти успадкованості, що

характеризують силу генетичного впливу на масу насіння кошику батьківських форм та взаємодії материнських та батьківських форм.

Варіабельність кількісної ознаки гібридних комбінацій соняшнику «маса насіння кошика» суттєво не залежала від дії даних материнських форм.

6. Розрахунок коефіцієнта успадкованості.

У двофакторному досліді дисперсія групових середніх має більш складну природу, ніж у однофакторному досліді. Дисперсія групових середніх у таких досліді визначається як генетичною мінливістю кількісної ознаки, обумовленої генотипами материнських та батьківських форм та їх взаємодією (дисперсія за варіантами), так і випадковою, мінливістю середовища (залишок).

Загальний коефіцієнт успадкованості у такому випадку дорівнює:

$$H^2 = H_A^2 + H_B^2 + H_{AB}^2.$$

Щоб визначити дисперсії, які характеризують вплив генотипів материнських форм (σ_A^2), батьківських форм (σ_B^2) та їх взаємодію (σ_{AB}^2) на фенотипічну мінливість, необхідно оцінити параметри дисперсій (табл. 7.7).

Таблиця 7.7

Компонентний аналіз двофакторного досліді за визначенням успадкованості кількісної ознаки

Дисперсія	Параметри, які оцінюються
материнських форм А	$\sigma^2 + n \times \sigma_{AB}^2 + l_B \times \sigma_A^2$
батьківських форм В	$\sigma^2 + n \times \sigma_{AB}^2 + l_A \times \sigma_B^2$
взаємодії форм АВ	$\sigma^2 + n \times \sigma_{AB}^2$
помилка	σ^2

Керуючись компонентним аналізом досліді стає можливим дисперсії σ_A^2 , σ_B^2 , σ_{AB}^2 розрахувати за наступними формулами.

Дисперсія за фактором материнські форми:

$$\sigma_A^2 = (\sigma^2 + n \times \sigma_{AB}^2 + l_B \times \sigma_A^2) - (\sigma^2 + n \times \sigma_{AB}^2) : (l_B \times n).$$

Значення $(\sigma^2 + n \times \sigma_{AB}^2 + l_B \times \sigma_A^2)$ дорівнюється середньому квадрату материнських форм $\sigma^2 C_A$: $(\sigma^2 + n \times \sigma_{AB}^2 + l_B \times \sigma_A^2) = \sigma^2 C_A = 25,41$.

Значення $(\sigma^2 + n \times \sigma_{AB}^2)$ дорівнюється середньому квадрату взаємодії батьківських форм $\sigma^2 C_{AB}$: $(\sigma^2 + n \times \sigma_{AB}^2) = \sigma^2 C_{AB} = 196,75$.

Формула для розрахунку дисперсії за фактором материнські форми придбає такий вид:

$$\sigma_A^2 = (\sigma^2 C_A - \sigma^2 C_{AB}) : (l_B \times n).$$

В досліді вплив материнських форм на кількісну ознаку «маса насіння кошику» виявився неістотним. Тому дану дисперсію не розраховують.

Дисперсія за фактором батьківські форми:

$$\sigma_B^2 = (\sigma^2 + n \times \sigma_{AB}^2 + l_A \times \sigma_B^2) - (\sigma^2 + n \times \sigma_{AB}^2) : (l_A \times n).$$

Значення $(\sigma^2 + n \times \sigma_{AB}^2 + l_A \times \sigma_B^2)$ дорівнюється середньому квадрату

батьківських форм σ^2_{CB} : $(\sigma^2 + n \times \sigma^2_{AB} + l_B \times \sigma^2_A) = \sigma^2_{CB} = 237,81$.

Значення $(\sigma^2 + n \times \sigma^2_{AB})$ дорівнюється середньому квадрату взаємодії батьківських форм σ^2_{CAB} : $(\sigma^2 + n \times \sigma^2_{AB}) = \sigma^2_{CAB} = 196,75$.

Формула для розрахунку дисперсії за фактором батьківські форми придбає такий вид:

$$\sigma^2_B = (\sigma^2_{CB} - \sigma^2_{CAB}) : (l_B \times n) = (237,81 - 196,75) : (3 \times 4) = 41,06 : 12 = 3,42.$$

Дисперсія за фактором взаємодія батьківських форм: $\sigma^2_{AB} = \{(\sigma^2 + n \times \sigma^2_{AB}) - \sigma^2\} : n$.

Значення $(\sigma^2 + n \times \sigma^2_{AB})$ дорівнюється середньому квадрату взаємодії батьківських форм σ^2_{CAB} : $(\sigma^2 + n \times \sigma^2_{AB}) = \sigma^2_{CAB} = 196,75$.

Значення σ^2 дорівнюється середньому квадрату помилки σ^2_{CZ} : $\sigma^2 = \sigma^2_{CZ} = 33,55$.

Формула для розрахунку дисперсії за фактором взаємодії батьківських форм придбає такий вид:

$$\sigma^2_{AB} = (\sigma^2_{CAB} - \sigma^2_{CZ}) : n = (196,75 - 33,55) : 4 = 163,20 : 4 = 40,8.$$

Визначається фенотипова дисперсія за формулою:

$$\sigma^2_P = \sigma^2_B + \sigma^2_{AB} + \sigma^2_{CZ} = 3,41 + 40,8 + 33,55 = 77,76.$$

Розрахувавши фенотипову мінливість та мінливість, обумовлену впливом генотипу батьківської форми та гібридної комбінації, визначають коефіцієнти успадковуваності кількісної ознаки.

Коефіцієнт успадковуваності маси насіння кошику за фактором батьківські форми:

$$H^2_B = \sigma^2_B : \sigma^2_P = 3,42 : 77,76 = 0,04 (4,0\%).$$

Коефіцієнт успадковуваності маси насіння кошику за фактором взаємодії батьківських форм:

$$H^2_{AB} = \sigma^2_{AB} : \sigma^2_P = 40,8 : 77,76 = 0,52 (52\%).$$

Загальний коефіцієнт успадковуваності кількісної ознаками «маса насіння кошика» в гібридних комбінаціях соняшнику (H^2) при правильному підборі пар для схрещування буде дорівнюватися:

$$H^2 = H^2_B + H^2_{AB} = 0,04 + 0,52 = 0,56 (56,0\%).$$

Висновок.

Таким чином, при підборі батьківських пар для схрещування необхідно мати на увазі, що прояв кількісної ознаки "маса насіння кошика" у гібридних комбінаціях соняшника визначається на 52,0% взаємодією батьківських форм, задіяних у комбінації схрещування. Сумісна дія батьківські форми та їх взаємодія з материнськими формами в гібридних комбінаціях АВ обумовлюють прояву кількісної ознаки «маса насіння кошика» на 56,0%.

Оцінюючи успадковуваність тієї чи іншої кількісної ознаки в селекції рослин досліднику необхідно завжди пам'ятати, що коефіцієнт успадковуваності характеризує не окремі рослини, а взагалі всю популяцію рослин, з якої виконується добір. Це обумовлюється тим, що для розрахунку використовують узагальнені показники популяції.

Статистичні показники ступеня успадковуваності селектованих кількісних

ознак вихідних форм відіграють важливу роль при доборі батьківських пар для схрещування та орієнтовного прогнозування ефективності селекції.

Так, коефіцієнт успадковуваності дозволяє прогнозувати зсув середнього значення кількісної ознаки популяції потомства у відповідь на зміщення середнього значення ознаки батьківської популяції в результаті проведення добору.

При цьому науковий співробітник повинен при статистичному аналізі дотримуватися наступних умов:

- 1) Повинен бути встановлений лінійний зв'язок між кількісними ознаками батьківських форм та їх потомством;
- 2) створено однакові умови вирощування для селекційного матеріалу;
- 3) проведення аналізу обов'язково в однакову фазу вегетації;
- 4) дотримання принципів рандомізації при доборі батьківських пар для схрещування.

Також слід зазначити, що, як і всі статистичні показники вибірки, коефіцієнти успадковуваності в широкому сенсі H^2 і вузькому сенсі h^2 несуть помилки виборочності, що впливають на точність прогнозу. Але, попри це, дані показники широко використовуються, так як є найпростішими та обґрунтованими генетичними параметрами.

Контрольні питання

1. Що таке кількісні ознаки?
2. Чим кількісні ознаки відрізняються від якісних?
3. Які чинники визначають прояв кількісної ознаки?
4. Що таке фенотипова мінливість?
5. Що таке генотипова мінливість?
6. Що таке паратипова (екологічна) мінливість?
7. Дайте визначення коефіцієнта успадковуваності.
8. Чим відрізняються коефіцієнти успадковуваності у широкому (H^2) та вузькому (h^2) сенсі?
9. Як розраховується коефіцієнт успадковуваності за коефіцієнтом кореляції?
10. Як визначають H^2 за коефіцієнтом регресії?
11. Як обчислюють H^2 за допомогою дисперсійного аналізу?
12. Яке біологічне значення має високий коефіцієнт успадковуваності?
13. Чому коефіцієнт успадковуваності залежить від умов середовища?
14. Як використовується коефіцієнт успадковуваності для прогнозування ефективності добору?
15. Що характеризує генетичний прогрес (відгук на добір) і як його визначають?

Розрахункові задачі

Задача 1. Визначення коефіцієнта успадковуваності за коефіцієнтом кореляції
У селекційному досліді між фенотипами материнських рослин і їх потомства отримано коефіцієнт кореляції: $r = 0,41$.

Необхідно:

1. Розрахувати коефіцієнт успадковуваності у широкому сенсі (H^2).
2. Виразити результат у частках одиниці та відсотках.
3. Оцінити ефективність добору за цією ознакою.
4. Визначити частку фенотипової мінливості, обумовлену середовищем.

Задача 2. Визначення H^2 за коефіцієнтом регресії

За результатами аналізу отримано коефіцієнт регресії між батьківськими формами та потомством $b = 0,34$.

Необхідно:

1. Розрахувати коефіцієнт успадковуваності.
2. Визначити частку генетичної та паратипової мінливості.
3. Пояснити біологічний зміст отриманого результату.
4. Зробити висновок щодо доцільності фенотипового добору.

Задача 3.

Під час дисперсійного аналізу отримано: генотипова дисперсія $\sigma^2_G = 15,6$; паратипова дисперсія $\sigma^2_E = 10,4$.

Необхідно:

1. Визначити фенотипову дисперсію.
2. Обчислити H^2 .
3. Інтерпретувати отримане значення.
4. Оцінити можливість ефективного селекційного добору.

Задача 4. Розрахунок генетичного прогресу

У вихідній популяції середня маса 1000 насінин становить 38,4 г. У відібраній групі 42,9 г. Коефіцієнт успадковуваності $H^2 = 0,62$.

Необхідно:

1. Визначити селекційний диференціал.
2. Розрахувати очікуваний генетичний прогрес (R).
3. Обчислити прогнозовану середню величину ознаки у наступному поколінні.
4. Зробити селекційний висновок.

Задача 5. Компонентний аналіз двофакторного досліді

У досліді встановлено: $\sigma^2_A = 18,6$; $\sigma^2_B = 5,1$; $\sigma^2_{AB} = 24,8$; $\sigma^2_P = 64,5$.

Необхідно:

1. Обчислити H^2_A .
2. Обчислити H^2_B .

3. Обчислити H^2_{AB} .
4. Визначити загальний коефіцієнт успадкованості.
5. Встановити, який фактор найбільше впливає на формування ознаки.

Тестові завдання

1. Коефіцієнт успадкованості характеризує:

- а) силу кореляції між ознаками;
- б) частку генетичної мінливості у загальній фенотиповій мінливості;
- в) величину середньої арифметичної;
- г) коефіцієнт варіації.

2. Коефіцієнт успадкованості у широкому сенсі позначають:

- а) r
- б) C_v
- в) H^2
- г) F

3. Фенотипова дисперсія складається з:

- а) лише генотипової дисперсії;
- б) лише екологічної дисперсії;
- в) генотипової та паратипової дисперсій;
- г) залишкової дисперсії.

4. Якщо коефіцієнт кореляції між батьками і потомством дорівнює 0,42, тоді $H^2 =$

- а) 0,21;
- б) 0,84;
- в) 0,42;
- г) 1,42.

5. Високе значення H^2 свідчить про:

- а) переважний вплив середовища;
- б) відсутність спадковості;
- в) значний внесок генотипу у формування ознаки;
- г) випадковий характер мінливості.

6. Генетичний прогрес визначають за формулою:

- а) $R = H^2 / S$;
- б) $R = S \times H^2$;
- в) $R = S + H^2$;
- г) $R = H^2 - S$.

7. Селекційний диференціал – це:

- а) коефіцієнт варіації;
- б) різниця між максимальною і мінімальною величинами;
- в) різниця між середнім значенням ознаки у відібраних рослин і вихідній популяції;
- г) стандартне відхилення.

8. Який метод НЕ застосовують для визначення коефіцієнта успадковуваності?

- а) кореляційний аналіз;
- б) регресійний аналіз;
- в) дисперсійний аналіз;
- г) кластерний аналіз.

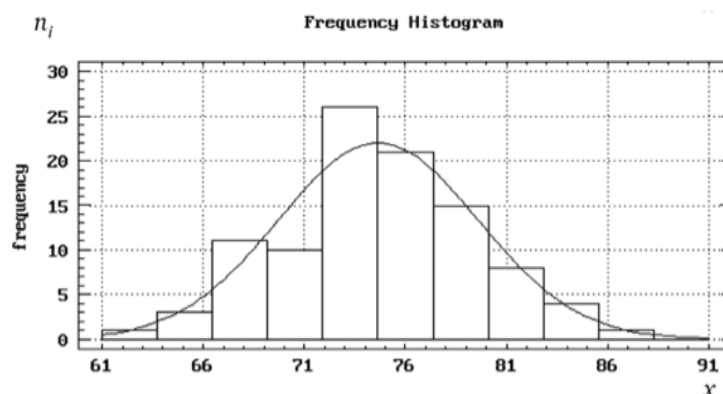
9. Якщо $H^2 = 0,80$, це означає, що:

- а) 20% мінливості обумовлено генотипом;
- б) 80% фенотипової мінливості пояснюється генетичними відмінностями;
- в) ознака повністю визначається середовищем;
- г) успадковуваність відсутня.

10. Основною практичною метою визначення коефіцієнта успадковуваності є:

- а) визначення середньої величини ознаки;
- б) оцінка достовірності дисперсії;
- в) прогнозування ефективності селекційного добору та генетичного покращення популяції;
- г) перевірка нормальності розподілу.

РОЗДІЛ 8. ПЕРЕВІРКА ГІПОТЕЗ ПРО ЗАКОНИ РОЗПОДІЛУ



8.1. *Поняття про критерій відповідності χ^2 (Хі-квадрат)*

8.2. *Використання χ^2 критерію в генетичному та селекційному аналізах*

8.3. *Приклади розрахунків критерію χ^2*

8.1. Поняття про критерій відповідності χ^2 (Хі-квадрат)

Під час проведення досліджень у рослинництві на величині будь-якої ознаки рослин, що варіює, позначається вплив численних, у тому числі і випадкових, факторів, які призводять до спотворення чіткої картини варіювання та отримання недостовірних результатів статистичного аналізу. Тим часом знання закону розподілу дозволяє науковому співробітнику у роботі уникнути можливих помилок в оцінці генеральних параметрів за вибірковими характеристиками.

Гіпотезу про закон розподілу ознак у рослинництві перевіряють за різними способами, а також виробляють за допомогою спеціально вироблених критеріїв.

Один із критеріїв, що знайшов широке застосування в біометрії, – це критерій згоди або відповідності χ^2 . Цей критерій запропонований К. Пірсоном – в 1900 р.

До кінця ХІХ століття нормальний розподіл вважався загальним законом варіації даних. Однак Пірсон зауважив, що емпіричні частоти можуть сильно відрізнятися від нормального розподілу.

Постало питання, як це довести. Потрібно було не тільки графічне зіставлення, яке має суб'єктивний характер, а й суворе кількісне обґрунтування.

Так був винайдений критерій χ^2 (Хі-квадрат), який перевіряє значущість розходження емпіричних (спостережуваних) та теоретичних (очікуваних) частот.

Критерій χ^2 є сумою квадратів відхилень емпіричних частот f від обчислених або очікуваних частот f' , які віднесено до теоретичних частот і символ χ^2 не є квадратом якогось числа, а виражає лише вихідну величину, що визначається формулою.

Критерій Пірсона, або критерій χ^2 (Хі-квадрат) – застосовують для

перевірки гіпотези про відповідність емпіричного розподілу передбачуваного теоретичного розподілу $F(x)$ при великому обсязі вибірки ($n \geq 100$).

Критерій застосовується для будь-яких видів функції $F(x)$, навіть при невідомих значеннях їх параметрів, що зазвичай має місце при аналізі результатів механічних випробувань. У цьому полягає його універсальність.

Використання критерію χ^2 передбачає розбиття розмаху варіювання вибірки на інтервали та визначення числа спостережень (частоти) кожного з інтервалів. Для зручності оцінок параметрів розподілу інтервали вибирають однакової довжини. Кількість інтервалів залежить від обсягу вибірки.

При перевірці гіпотези про відповідність емпіричних розподілів нормальному розподілу бажано мати не менше 50 спостережень (при менш суворому підході мінімум приймають три спостереження).

Якщо крайні групи у ряді розподілу нечисленні, їх необхідно об'єднати. Вважають, що застосування критерію χ^2 вимагає, щоб у крайніх класах варіаційного ряду містилося щонайменше п'ять варіант. Якщо в крайніх класах міститься менше п'яти варіантів, то обчислені та емпіричні частоти об'єднуються до зазначеного мінімуму і відповідно зменшується кількість класів варіаційного ряду.

Число ступенів свободи для $\chi^2_{\text{теор.}}$ щодо відповідності розподілів нормальному закону дорівнює числу груп без трьох ($k-3$), оскільки обчислення теоретичних частот пов'язані трьома умовами які визначають нормальний розподіл:

- обсягом вибірки n ;
- середнім значенням ознаки x ;
- значенням дисперсій, за якими будувався теоретичний нормальний розподіл.

Число ступенів свободи встановлюють за вторинним числом класів з урахуванням обмежень свободи варіації, яка в різних випадках буває різною.

Так, в оцінці емпіричних розподілів, які відповідають нормальному закону, число ступенів свободи $k=N-3$ (з урахуванням трьох обмежень свободи варіації: n , x і s_x).

Якщо ж оцінці підлягає розподіл, який відповідає закону Пуассона, кількість ступенів свободи зменшується на одиницю, тобто $k = N-2$ (з урахуванням двох обмежень свободи варіації n і s^2 або x).

На величині критерію χ^2 позначається ступінь точності, з яким визначено теоретично обчислені чи очікувані частоти. Тому при зіставленні емпіричних частот з обчисленими частотами останні не слід округлювати до цілих чисел.

Застосовується в тих випадках, коли досліднику необхідно визначити відповідність двох порівнюваних рядів розподілу – емпіричного та теоретичного або двох емпіричних розподілів.

8.2. Використання χ^2 критерію в генетичному та селекційному аналізах

Особливо широко критерій відповідності використовується в генетичному аналізі, коли необхідно переконатися, чи є виявлене відхилення від теоретично

очікуваного розщеплення (1: 1; 3: 1; 9: 3: 4; 9: 3: 3: 1 і т. д.) відхиленням закономірним або воно лежить у межах можливих випадкових коливань.

Якщо позначити теоретичні очікувані показники для групи об'єктів рослин через $F_1, F_2, \dots, F_n$, а досліджувані, які емпірично отримані, через $f_1, f_2, \dots, f_n$, то відхилення фактичних даних від теоретичних дорівнюють $f_1-F_1; f_2-F_2, \dots, f_n-F_n$.

Загальною мірою відхилення фактичних даних від теоретичних, тобто критерію відповідності χ^2 , буде сума відносин квадратів різниць між частотами емпіричного і теоретичного розподілів до частот теоретичного розподілу для цієї групи рослин.

Критерій χ^2 використовується щодо якісних ознак рослин з метою оцінки відповідності емпіричних даних певної теоретичної передумові нульової гіпотези (H_0).

Нульова гіпотеза зводиться до припущення, що відмінності, спостерігаються між емпіричними та обчисленими чи очікуваними частотами, носять виключно випадковий характер.

Для перевірки нульової гіпотези потрібно фактично одержану величину $\chi^2_{\text{факт.}}$ порівняти з її критичним значенням $\chi^2_{\text{теор.}}$.

Нульова гіпотеза має бути відкинута, якщо $\chi^2_{\text{факт.}} > \chi^2_{\text{теор.}}$ на прийнятому рівні значності з числом ступенів свободи k . В цьому випадку розглядається альтернативна гіпотеза. Альтернативна гіпотеза – це відхилення в частотах виходять за рамки випадкових коливань, розбіжності статистично значущі.

Нульова гіпотеза не відкидається, якщо $\chi^2_{\text{факт.}} < \chi^2_{\text{теор.}}$.

Коли фактичні та теоретичні очікувані частоти повністю збігаються, то $\chi^2 = 0$. Нульова гіпотеза у тому, що частоти узгоджені, тобто фактичні дані не суперечать очікуваним.

Якщо збіг між фактичними та теоретичними очікуваними частотами неповний, то у цьому випадку χ^2 буде відмінний від нуля. Чим більша відмінність, тим більша розбіжність між теоретичними та емпіричними частотами.

Граничні значення χ^2 , за яких нульова гіпотеза приймається, наведені у додатку 5.

У найбільш типових випадках статистичного аналізу результатів досліджень застосування критерію відповідності числу ступенів свободи визначається за формулою:

$$v = (c-1) \times (k-1),$$

де:

c – кількість рядків в аналітичній таблиці;

k – кількість колонок в аналітичній таблиці.

Критерій χ^2 широко використовується дослідниками при проведенні селекційної роботи в генетичному аналізі відповідності розщеплення гібридів теоретично очікуваному, для оцінки незалежності (або сполученості) у розподілі об'єктів сукупності, визначення ступеня відповідності фактичного розподілу

ознаки, що вивчається нормальному розподілу та оцінки відповідності двох емпіричних розподілів.

В той же час, застосування критерію χ^2 потребує певної обережності, так як критерій має свої недоліки.

Недоліком критерію згоди Пірсона є втрата частини первинної інформації, яка пов'язана з необхідністю угруповання результатів спостережень в інтервали та об'єднання окремих інтервалів з малою кількістю спостережень. У зв'язку з цим рекомендується доповнювати перевірку відповідності розподілів за критерієм χ^2 іншими критеріями. Особливо це необхідно за порівняно малого обсягу вибірки ($n \approx 100$).

Тому, у формулу для розрахунків критерію χ^2 необхідно підставляти лише частоти, а не величини, які отримані виміром, зважуванням тощо.

До того ж, розподіл можливих значень випадкової величини χ^2 є безперервним та асиметричним, розподіл залежить від числа ступенів свободи k і наближається до нормальної кривою зі збільшенням числа випробувань n . Тому застосування критерію χ^2 для оцінки дискретних розподілів пов'язано з деякими помилками, які позначаються на його величині, особливо при малих вибірках.

8.3. Приклади розрахунків критерію χ^2

Розглянемо приклади розрахунків критерію χ^2 .

Приклад.

Селекціонер задіяв в комбінації схрещування сорти соняшнику із жовтим та лимонним забарвленням крайових квітів суцвіття. В F_1 всі рослини характеризувалися жовтим забарвленням крайових квітів.

Після самозапилення в F_2 селекціонер відмічає за домінантною ознакою $f_1 = 767$ рослин соняшнику з жовтими крайовими квітами, за рецесивною ознакою $f_2 = 264$ рослини з лимонним забарвленням крайових квітів. Загальна кількість досліджених рослин з жовтим та лимонним забарвленням крайових квітів 1031.

Чи відповідають результати дослідження теоретично очікуваному відношенню рослин з жовтим забарвленням крайових квіток до рослин з лимонним забарвленням крайових квіток у співвідношенні як 3:1?

Співвідношення 3 : 1 береться як нульова гіпотеза H_0 , яку треба довести.

Керуючись співвідношенням 3 : 1, визначають теоретично очікувані частоти F .

Так як співвідношення домінантної ознаки до рецесивної ознаки складає 3 : 1, то спостерігається в загальній частині $4 = 1 + 3$.

Ураховуючи загальну кількість рослин, яка обстежена в досліді 1031, встановлюють кількість рослин, які приходяться на одну частину співвідношення: $1031 : 4 = 257,75$ рослин.

Розраховують теоретично очікувану частоту рослин в другому поколінні за домінантною ознакою – жовте забарвлення крайових квітів:

$F_1 = 257,75 \times 3 = 773,25$ рослин з жовтим забарвленням крайових квітів.

Розраховують теоретично очікувану частоту рослин в другому поколінні за

рецесивною ознакою – лимонне забарвлення крайових квітів:

$F_2 = 257,75 \times 1 = 257,75$ рослин з лимонним забарвленням крайових квітів.
Загальна кількість, сума = $1031 = 795 + 265$.

Таким чином, емпіричні частоти за ознакою забарвлення крайових квітів дорівнюється: $f_1 = 767$; $f_2 = 264$.

Теоретично передбачаємо частоти за ознакою забарвлення крайових квітів дорівнюється: $F_1 = 773,25$; $F_2 = 257,75$.

Підставляючи емпіричні та теоретичні очікувані частоти у формулу, визначають χ^2 :

$$\begin{aligned}\chi^2 &= \{(f_1 - F_1)^2 : F_1\} + \{(f_2 - F_2)^2 : F_2\} = \\ &= \{(767 - 773,25)^2 : 773,25\} + \{(264 - 257,75)^2 : 257,75\} = \\ &= \{(-6,25)^2 : 773,25\} + \{(6,25)^2 : 257,75\} = \\ &= (39,06 : 773,25) + (39,06 : 257,75) = 0,050 + 0,151 = 0,201.\end{aligned}$$

Фактичне значення $\chi^2_{\text{факт.}} = 0,201$.

Визначають кількість ступенів свободи за формулою: $\nu = (c-1) \times (k-1) = (2-1) \times (2-1) = 1 \times 1 = 1$.

При кількості ступеней свободи $\nu = 1$ теоретичне значення $\chi^2_{0,05} = 3,84$ (додаток 5).

Порівнюють фактичне та теоретичне значення χ^2 :

$$\chi^2_{\text{факт.}} (0,201) < \chi^2_{0,05} (3,841).$$

Висновок. Відмінності між фактичними та теоретичними очікуваними частотами несуттєві $\chi^2_{\text{факт.}} < \chi^2_{0,05}$ не відкидається.

Можливо розрахунки щодо встановлення χ^2 виконувати і за формулою таблиці 2×2 (2 рядки та 2 стовпці) (табл. 8.1).

Таблиця 8.1

Обчислення теоретичних частот F та критерію відповідності χ^2 за таблицею 2×2

Показники	Забарвлення крайових квітів суцвіття соняшнику		Сума
	жовте забарвлення	лимонне забарвлення	
очікуване розщеплення ознак, H_0	3	1	4
частоти, що спостерігаються в досліді, f	767	264	1031
очікуємі частоти, F	773,25	257,75	1031
різниця, f - F	-6,25	6,25	
квадрат різниці $(f - F)^2$	39,06	39,06	
відношення $(f - F)^2 : F$	0,050	0,151	$\chi^2 = 0,201$

За даними таблиці виконують розрахунки:

$$\chi^2 = \{(f_1 - F_1)^2 : F_1\} + \{(f_2 - F_2)^2 : F_2\} =$$

$$\begin{aligned} & \{(767 - 773,25)^2 : 773,25\} + \{(264 - 257,75)^2 : 257,75\} = \\ & \{(-6,25)^2 : 773,25\} + \{(6,25)^2 : 257,75\} = \\ & (39,06 : 773,25) + (39,06 : 257,75) = 0,050 + 0,151 = 0,201. \\ & \text{Фактичне значення } \chi^2_{\text{факт.}} = 0,201. \end{aligned}$$

Приклад визначення χ^2 при проведенні дигібридного схрещування.

В процесі науково-дослідної роботи зі створення вихідного матеріалу в селекції динь у другому поколінні F_2 в результаті розщеплення при дигібридному схрещуванні отримано дині видовженої форми плоду з оранжевим забарвленням м'якоті $f_1 = 321$; видовженої форми плоду з білим забарвленням м'якоті $f_2 = 108$; овальної форми плоду з оранжевим забарвленням м'якоті $f_3 = 110$; овальної форми плоду з білим забарвленням м'якоті $f_4 = 35$. Загальна сума частот = 574. Необхідно встановити, відповідність емпіричного розподілу отриманих у досвіді частот f теоретично очікуваному успадкуванню у співвідношенні 9 : 3 : 3 : 1.

Рішення виконують таким чином.

Співвідношення 9 : 3 : 3 : 1 береться як нульова гіпотеза H_0 , яку треба довести.

Керуючись співвідношенням 9 : 3 : 3 : 1, визначають теоретично очікувані частоти F .

Так як співвідношення ознак складає 9 : 3 : 3 : 1, то спостерігається в загалі 16 частини $16 = 9 + 3 + 3 + 1$.

Ураховуючи загальну кількість рослин, яка обстежена в досліді 574, встановлюють кількість рослин, які приходяться на одну частину співвідношення: $574 : 16 = 35,875$ рослин.

Розраховують теоретично очікувану частоту рослин в другому поколінні за ознакою – видовжена форма плоду з оранжевим забарвленням м'якоті:

$$F_1 = 35,875 \times 9 = 322,875.$$

Розраховують теоретично очікувану частоту рослин в другому поколінні за ознакою – видовжена форма плоду з білим забарвленням м'якоті:

$$F_2 = 35,875 \times 3 = 107,625.$$

Розраховують теоретично очікувану частоту рослин в другому поколінні за ознакою – овальна форма плоду з оранжевим забарвленням м'якоті:

$$F_3 = 35,88 \times 3 = 107,625.$$

Розраховують теоретично очікувану частоту рослин в другому поколінні за ознакою – овальна форма плоду з білим забарвленням м'якоті:

$$F_4 = 35,87 \times 1 = 35,875.$$

$$\text{Загальна кількість, сума} = 574 = 322,875 + 107,625 + 107,625 + 35,875.$$

За бажанням можливо заповнити таблицю для обчислення χ^2 .

Таблиця 8.2

Обчислення теоретичних частот F та критерію відповідності χ^2 за таблицею 2×4 за ознаками форма плоду та забарвлення м'якоті плоду дині

Показники	Видовжена форма плоду		Овальна форма		Сума
	оранжева м'якоть	біла м'якоть	оранжева м'якоть	біла м'якоть	
очікуване розщеплення ознак, H_0	9	3	3	1	
частоти, що спостерігаються в досліді, f	321	108	110	35	574
очікуємі частоти, F	322,875	107,625	107,625	35,875	574
різниця, f - F	-1,875	0,375	2,375	-0,875	
квадрат різниці $(f - F)^2$	3,52	0,14	5,64	0,77	
відношення $(f - F)^2 : F$	0,011	0,001	0,052	0,021	$\chi^2 = 0,085$

За формулою критерій відповідності дорівнюється:

$$\begin{aligned} \chi^2 &= \{(f_1 - F_1)^2 : F_1\} + \{(f_2 - F_2)^2 : F_2\} + \{(f_3 - F_3)^2 : F_3\} + \{(f_4 - F_4)^2 : F_4\} = \\ &= \{(321 - 322,875)^2 : 322,875\} + \{(108 - 107,625)^2 : 107,625\} + \{(110 - 107,625)^2 : 107,625\} + \{(35 - 35,875)^2 : 35,875\} = \\ &= \{(-1,875)^2 : 322,875\} + \{(0,375)^2 : 107,625\} + \{(2,375)^2 : 107,625\} + \{(-0,875)^2 : 35,875\} = \\ &= (3,52 : 322,875) + (0,14 : 107,625) + (5,64 : 107,625) + (0,77 : 35,875) = \\ &= 0,011 + 0,001 + 0,052 + 0,021 = 0,085 \end{aligned}$$

Фактичне значення $\chi^2_{\text{факт.}} = 0,085$.

Визначають кількість ступенів свободи за формулою:

$$v = (c-1) \times (k-1) = (2-1) \times (4-1) = 1 \times 3 = 3.$$

При кількості ступеней свободи $v = 3$ теоретичне значення $\chi^2_{0,05} = 7,815$ (додаток 5).

Порівнюють фактичне та теоретичне значення χ^2 : $\chi^2_{\text{факт.}} (0,085) < \chi^2_{0,05} (7,815)$.

Висновок. Відмінності між фактичними та теоретичними очікуваними частотами несуттєві $\chi^2_{\text{факт.}} < \chi^2_{0,05}$ і нульова гіпотеза H_0 не відкидається.

Контрольні питання

1. Що називають законом розподілу випадкової величини?
2. Чому перевірка закону розподілу є важливою перед проведенням статистичного аналізу?
3. Яка основна мета застосування критерію χ^2 Пірсона?

4. Які гіпотези перевіряють за допомогою критерію χ^2 ?
5. За яких умов застосування критерію χ^2 є коректним?
6. Чому при використанні критерію χ^2 необхідно об'єднувати малочисельні класи?
7. Як обчислюється критерій відповідності χ^2 ?
8. Що таке емпіричні та теоретичні частоти?
9. Як визначають число ступенів свободи при перевірці нормального розподілу?
10. Як визначають число ступенів свободи для таблиці спряженості?
11. Яке правило прийняття або відхилення нульової гіпотези?
12. Які переваги та недоліки критерію χ^2 ?
13. Як застосовується критерій χ^2 у генетичному аналізі?
14. Яке практичне значення перевірки відповідності емпіричного розподілу теоретичному у біометричних дослідженнях?

Розрахункові задачі

Задача 1. Перевірка моногібридного розщеплення (3 : 1)

У другому поколінні (F_2) після самозапилення гібридів пшениці отримано: домінантний фенотип – 615 рослин; рецесивний фенотип – 205 рослин.

Необхідно:

1. Визначити теоретично очікувані частоти за співвідношенням 3 : 1.
2. Обчислити критерій відповідності χ^2 .
3. Визначити число ступенів свободи.
4. За табличного значення $\chi^2_{0.05} = 3,84$ зробити статистичний висновок.
5. Інтерпретувати результат з генетичної точки зору.

Задача 2. Перевірка дигібридного розщеплення (9 : 3 : 3 : 1)

У F_2 отримано такі результати:

Фенотип	Частота
A–B–	362
A–bb	126
aaB–	118
aabb	42

Необхідно:

1. Визначити очікувані частоти.
2. Обчислити χ^2 .
3. Визначити число ступенів свободи.
4. Перевірити нульову гіпотезу.
5. Зробити висновок щодо відповідності закону Менделя.

Задача 3. Перевірка відповідності нормальному розподілу

Під час оцінювання висоти рослин сформовано 8 інтервалів варіювання. Після об'єднання крайніх класів залишилось 6 інтервалів. Обчислений критерій χ^2 факт = 5,91.

Необхідно:

1. Визначити число ступенів свободи.
2. Знайти критичне значення $\chi^2_{0.05}$.
3. Зробити висновок щодо відповідності нормальному розподілу.
4. Пояснити причини об'єднання крайніх класів.

Задача 4. Аналіз таблиці спряженості

У досліді вивчали залежність стійкості рослин до хвороби від способу передпосівної обробки насіння. Отримано таблицю:

	Стійкі	Нестійкі
Варіант 1	58	22
Варіант 2	41	39

Необхідно:

1. Визначити очікувані частоти.
2. Обчислити χ^2 .
3. Визначити число ступенів свободи.
4. Перевірити гіпотезу про незалежність ознак.
5. Інтерпретувати біологічний зміст результатів.

Задача 5. Комплексний аналіз критерію χ^2

У селекційному досліді встановлено: кількість класів після об'єднання – 7; $n = 140$; χ^2 факт = 8,76.

Необхідно:

1. Визначити число ступенів свободи при перевірці нормального розподілу.
2. Встановити критичне значення $\chi^2_{0.05}$.
3. Перевірити нульову гіпотезу.
4. Пояснити практичне значення отриманого результату.
5. Вказати, які подальші статистичні методи можна застосовувати за умови підтвердження нормальності розподілу.

Тестові завдання

1. Основне призначення критерію χ^2 Пірсона полягає у:

- а) визначенні коефіцієнта варіації;
- б) оцінці середньої арифметичної;
- в) перевірці відповідності емпіричного розподілу теоретичному;
- г) визначенні коефіцієнта кореляції.

2. Критерій χ^2 використовують переважно для аналізу:

- а) середніх величин;
- б) дисперсій;
- в) кількісних ознак;
- г) частот розподілу якісних ознак.

3. Якщо $\chi^2_{\text{факт}} < \chi^2_{\text{теор}}$, то:

- а) нульову гіпотезу відхиляють;
- б) приймають альтернативну гіпотезу;
- в) необхідно повторити експеримент;
- г) немає підстав відхиляти нульову гіпотезу.

4. При перевірці відповідності нормальному розподілу число ступенів свободи визначають як:

- а) $k - 1$;
- б) $k - 2$;
- в) $k - 3$;
- г) $n - 1$.

5. Мінімальна рекомендована чисельність очікуваних частот у крайніх класах становить:

- а) 2;
- б) 3;
- в) 4;
- г) 5.

6. Для таблиці спряженості число ступенів свободи визначають за формулою:

- а) $c + k$;
- б) $c \times k$;
- в) $c - k$;
- г) $(c - 1) \times (k - 1)$.

7. Яке співвідношення найчастіше перевіряють при моногібридному схрещуванні?

- а) $1 : 2 : 1$;
- б) $3 : 1$;
- в) $9 : 3 : 3 : 1$;
- г) $1 : 1$.

8. Яке співвідношення характерне для дигібридного схрещування за незалежного успадкування ознак?

- а) $3 : 1$;
- б) $1 : 2 : 1$;
- в) $9 : 3 : 3 : 1$;
- г) $15 : 1$.

9. Основним недоліком критерію χ^2 є:

- а) неможливість використання великих вибірок;
- б) необхідність групування даних, що призводить до втрати частини інформації;
- в) неможливість аналізу якісних ознак;
- г) використання лише для нормального розподілу.

10. Якщо фактичні та очікувані частоти повністю збігаються, то:

- а) $\chi^2 = 1$;
- б) $\chi^2 = k$;
- в) $\chi^2 = n$;
- г) $\chi^2 = 0$.

КОРОТКИЙ ВКАЗІВНИК СИМВОЛІВ

τ – критерій «тау», що використовується для вибракування «сумнівних» даних дослідів;

x_i, y_i - значення (варіанту) варіюючої ознаки; t – нормоване відхилення;

μ – середня генеральна сукупність;

σ^2 – дисперсія, варіанс, середній квадрат;

σ – стандартне (середньоквадратичне) відхилення;

$\bar{x}$ - вибіркова середня, арифметична середня;

X_i – середня групова;

f – частота народження при групуванні даних;

Δ – межа випадкових коливань середньої арифметичної; M_o – мода;

Me – медіана;

p, q – частка ознаки;

R – розмах варіювання;

V – коефіцієнт варіації, мінливості;

V_p – коефіцієнт варіації, мінливості частки ознаки; S_x – середня помилка вибіркової середньої;

$S_{x\%}$ – відносна помилка вибіркової середньої; r – коефіцієнт фенотипічної кореляції;

r^2 – коефіцієнт детермінації;

r_G – коефіцієнт генотипної кореляції; η_{yx}, η_{xy} – кореляційне ставлення;

η^2_{yx} – індекс детермінації;

$r_{xy(z)}$ - приватний коефіцієнт кореляції; $b_{y/x}$ – коефіцієнт регресії;

s - помилка стандартного відхилення; s_r – помилка коефіцієнта кореляції;

s_p – помилка частки ознаки;

$s_{by/x}$ – помилка коефіцієнта регресії;

l – кількість варіантів в досліді;

n – кількість повторень в досліді, обсяг вибірки;

N – загальна кількість дослідних ділянок у досліді; v – кількість ступенів свободи;

$t_{\text{факт.}}$ – фактичне значення критерію Стюдента;

t_{05}, t_{01} – табличні значення критерію для 5%-го та 1%-го рівня значущості;

$F_{\text{факт.}}$ - фактичне значення критерію Фішера;

F_{05}, F_{01} – табличні значення критерію F для 5%-го та 1%-го рівня значущості;

H_0 – нульова гіпотеза;

HSP_{05}, HSP_{01} – найменші суттєві різниці для 5%-го та 1%-го рівня значущості;

H^2, h^2 – коефіцієнт успадкованості;

C – коригуючий фактор (поправка) у дисперсійному аналізі;

S_u, S_p, S_v, S_z і т.д. – суми квадратів відхилень для різних джерел варіювання в дисперсійному аналізі;

Σ - сума, знак підсумовування.

ГЛОСАРІЙ

Біометрія – (лат. bios – життя, metron – міра) – наука про статистичний аналізі масових явищ у біології.

Варіанта (x_i) – окреме числове значення ознаки, що варіює.

Варіаційний ряд – подвійний ряд чисел, який показує, з якого частотою трапляються ті чи інші значення ознаки.

Варіація – коливання величини однієї і тієї ж ознаки, що спостерігається в загальній масі його числових значень.

Імовірність – числова міра об'єктивної можливості появи випадкової події.

Імовірність довірча – рівень ймовірності, який вважається достатнім для судження про достовірність статистичних показників, які одержані на основі вибірових даних.

Взаємодія генотип-середовище – ефект, що відображає відмінність реакції генотипів (за величиною ознаки) на однакові зміни умов вирощування.

Вибіркова сукупність – частина генеральної, відібрана для сумісного вивчення.

Вибірковий метод – метод щодо статистичних сукупностей, заснований на вивченні вибірки.

Виняток – частина облікової ділянки, виключена з обліку, внаслідок випадкових ушкоджень чи помилок, допущених під час проведення науково-дослідної роботи.

Генеральна сукупність – нескінченно велика кількість відносно однорідних одиниць, що складають її зміст.

Гістограма – зображення інтервального варіаційного ряду у вигляді стовпчик діаграми. Показує залежність між класами і частоту зустрічальності в їх відповідних значень ознаки.

Градації – окремі значення фактора, які вивчаються в експерименті. Бувають фіксовані, випадкові та ієрархічні.

Угрупування даних – процес систематизації первинних даних у цілях отримання отриманої інформації та виявлення закономірностей.

Дисперсійний аналіз – метод аналізу результатів експерименту, який заключається у розкладанні загальної мінливості ознаки на складові компоненти, які відповідають мінливості за варіантами, повтореннями та внаслідок дії випадкової помилки. Істотність впливу різних факторів на результативну ознаку оцінюють за допомогою F та НСР.

Дисперсія (від латів. dispersio – розсіювання) чи варіансу – середній квадрат відхилень варіант від своїх середньої величини у цій сукупності.

Довірчий інтервал – це область знаходження окремих значень ознаки у певних межах із заданою ймовірністю.

Індекс детермінації (η^2) – показник, що характеризує частку варіювання ознаки Y, яка обумовлена ступенем коливання ознаки X, тобто. визначає відсоток варіації результативної ознаки Y під впливом факторіального X при

нелінійному зв'язку.

Коваріаційний аналіз – поширення методів дисперсійного аналізу на випадок декількох змінних, а також кореляційного та регресійного аналізу на загальні схеми польових, вегетаційних, лабораторних експериментів.

Коваріація (cov) – середній добуток відхилень значень однієї ознаки на відповідні відхилення значень іншої ознаки від їх середніх арифметичних (коваріація у вузькому значенні).

Кореляційна решітка – допоміжна таблиця, в якій указуються діапазони класів двох ознак з частотою їх появи та проводяться обчислення всіх допоміжних значень для визначення коефіцієнту ефекту кореляції.

Кореляційне відношення (η) – показник, який характеризує тісноту криволінійного зв'язку.

Кореляційні плеяди – складна мережа кореляційних зв'язків, які мають різну тісноту.

Кореляція (лат. correlatio – співвідношення, зв'язок) – взаємозв'язок між ознаками, коли певному значенню однієї ознаки відповідає не одне, а цілий варіаційний ряд числових значень іншої ознаки. Спостерігається лише для групових об'єктів.

Коефіцієнт варіації – відносний показник мінливості ознаки, що характеризує ступінь варіювання ознаки.

Коефіцієнт детермінації (r^2) – показник, що характеризує частку (або відсоток) тих змін однієї ознаки, яка в даному явищі лінійно залежать від іншої ознаки, яка вивчається.

Коефіцієнт кореляції (r) – показник зв'язку, який характеризує тісноту (силу) та напрямок лінійного зв'язку.

Коефіцієнт регресії – показник зв'язку, який характеризує наскільки, в середньому, величина однієї ознаки (Y) змінюється при зміні на одиницю міри іншої ознаки (X), пов'язаної з Y кореляційно.

Криволінійна кореляція – нерівномірні зміни однієї ознаки у відповідь на зміни іншої.

Критерій (від грец. kriterion - мірило, засіб судження) – показник, який дозволяє судити про надійність висновків щодо прийнятої гіпотези, очікуваного результату тощо.

Ліміти – крайні значення ранжованого ряду, що характеризують мінімальне і максимальне значення досліджуваної ознаки у вибірці, вказують на амплітуду варіації.

Успадковуваність – частка (частина) генетичних відмінностей у загальній фенотипічній мінливості у популяції. Визначається за допомогою коефіцієнта успадковуваності.

Нормоване відхилення – показник різноманітності ознаки, характеризує окрему варіанту (або групу варіант). Представляє виражене у частках σ відхилення тієї чи іншої варіанти від середньої арифметичній.

Нульова гіпотеза (H_0) – статистична гіпотеза, яка виходить з припущення про відсутність різниці між параметрами порівнюваних сукупностей.

Обсяг вибірки – кількість одиниць вибіркової сукупності.

Одиниця спостережень – це окремий елемент біологічного об'єкта.

Помилка репрезентативності – це міра відхилення вибіркового показника від показника генеральної сукупності. Вона виникає, коли вибірка недостатньо точно відображає характеристики генеральної сукупності.

Ранг – порядковий номер ранжованих значень ознаки.

Розподіл – сукупність випадкових значень ознаки та відповідних їм ймовірностей.

Регресія (від латів. regressus - зворотний рух) – кількісна зміна функції в залежності від зміни одного або декількох аргументів.

Рівень ймовірності – площа під кривою розподілу, що відмежована від середнього на $\pm t$ стандартних відхилень ($\mu \pm t\sigma$). Виражається в відсотках.

Рівень значущості – величина, якої можна знехтувати у даному дослідженні чи явищі. Вказує можливість відхилення, від встановлених меж варіювання випадкової величини.

Середньоквадратичне (стандартне) відхилення – показник варіювання ознаки для сукупності, що має певну середню величину $\bar{x}$, яке враховує відхилення кожної варіанти від середньої $\bar{x}$. Виражається в абсолютних одиницях виміру будь-якої ознаки.

Статистична сукупність – сукупність одиниць спостережень, об'єднуються для спільного вивчення в однакових умовах.

Частка ознаки – це відношення чисельності кожного з членів низки n_1, n_2, n_3 і т.д. до загального обсягу сукупності n .

ДОДАТКИ

Додаток 1

Контрольні значення критерію τ для перевірки сумнівних дат

n	Контрольні τ	
	При 1-ому рівні значущості 099 (0,01)	При 5-ти рівні значущості 095 (0,05)
4	0,991	0,995
5	0,916	0,807
6	0,805	0,689
7	0,740	0,610
8	0,683	0,554
9	0,635	0,512
10	0,597	0,477
11	0,566	0,460
12	0,541	0,428
13	0,502	0,395
14	0,500	0,400
16	0,472	0,369
18	0,449	0,349
20	0,430	0,334
22	0,414	0,320
24	0,400	0,309
26	0,389	0,299
28	0,378	0,291
30	0,369	0,283

Додаток 2

Квантилі F-розподілу (контрольні значення F-критерію).
 Значення критерію Фішера $F_{0,95}$ на 5%-ному рівні значущості (вірогідність 95%).
 Імовірність помилки $\alpha = 0,05$ (05%)

Кількість ступенів свободи помилки, v_z	Кількість ступенів свободи результативного фактору (варіанту)													
	1	2	3	4	5	6	7	8	9	10	12	24	50	100
1	161	200	216	225	230	234	237	239	241	242	244	249	452	253
2	18,51	19,0	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,41	19,45	19,47	19,49
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,74	8,64	8,58	8,56
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,77	5,70	5,66
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,68	4,53	4,44	4,40
6	5,99	5,14	4,76	4,53	4,39	4,27	4,21	4,15	4,10	4,06	4,00	3,84	3,75	3,71
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,57	3,41	3,32	3,28
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,28	3,12	3,03	2,98
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,07	2,90	2,80	2,76
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,91	2,74	2,64	2,59
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,79	2,61	2,50	2,45
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,69	2,50	2,40	2,35
13	4,64	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,60	2,42	2,32	2,26
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,53	2,35	2,24	2,19
15	4,54	3,67	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,48	2,29	2,18	2,12
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,24	2,13	2,07
17	4,45	3,59	3,20	2,96	2,81	2,69	2,62	2,55	2,50	2,45	2,38	2,19	2,08	2,02
18	4,41	3,55	3,18	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,15	2,04	1,98
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,31	2,11	2,00	1,94
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,28	2,08	1,96	1,90
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,05	1,93	1,87
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,23	2,03	1,91	1,84
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,20	2,00	1,88	1,82
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,18	1,98	1,86	1,80
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,25	2,24	2,06	1,96	1,84	1,77
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,25	2,22	2,15	1,95	1,82	1,76
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,12	1,91	1,78	1,72
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,12	2,09	1,89	1,76	1,69
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,00	1,79	1,66	1,59
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,95	1,74	1,60	1,52
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,85	1,63	1,48	1,39

Додаток 3

Квантилі F-розподілу (контрольні значення F-критерію).
 Значення критерію Фішера $F_{0,99}$ на 1%-ному рівні значущості (вірогідність 99%).
 Імовірність помилки $\alpha = 0,01$ (01%)

Кількість ступенів свободи помилки, V_z	Кількість ступенів свободи результативного фактору (варіанту)													
	1	2	3	4	5	6	7	8	9	10	12	24	50	100
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6106	6234	6302	6334
2	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,42	99,46	99,48	99,49
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,05	26,60	26,35	26,23
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,37	13,93	13,69	13,57
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,89	9,47	9,24	9,13
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,31	7,09	6,99
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,47	6,07	5,85	5,75
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,67	5,28	5,06	4,96
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,11	4,73	4,51	4,41
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,71	4,33	4,12	4,01
11	9,85	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,40	4,02	3,80	3,70
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,16	3,78	3,56	3,46
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,59	3,37	3,27
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,43	3,21	3,11
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,29	3,07	2,97
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,78	3,69	3,61	3,54	3,18	2,96	2,86
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,45	3,08	2,86	2,76
18	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51	3,37	3,00	2,78	2,68
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,68	3,52	3,43	3,30	2,92	2,70	2,63
20	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,23	2,86	2,63	2,53
21	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31	3,17	2,80	2,58	2,47
22	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,75	2,53	2,42
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,70	2,48	2,37
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,03	2,66	2,44	2,33
25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,21	3,13	2,99	2,62	2,40	2,29
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09	2,96	2,58	2,36	2,25
28	7,64	5,45	4,57	4,07	3,76	3,53	3,36	3,23	3,11	3,03	2,90	2,52	2,30	2,18
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,84	2,47	2,24	2,13
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,80	2,66	2,29	2,05	1,94
50	7,17	5,06	4,20	3,72	3,41	3,18	3,02	2,88	2,78	2,70	2,56	2,18	1,94	1,81
100	6,90	4,82	3,98	3,51	3,20	2,99	2,82	2,69	2,59	2,51	2,36	1,98	1,73	1,59

Додаток 4

Критерій Стюдента t

Кількість ступенів свободи	Рівень значущості, α			
	0,10	0,05 F0,95	0,01 F0,99	0,001
1	6,31	12,70	63,70	637,00
2	2,92	4,30	9,92	31,60
3	2,35	3,18	5,84	12,90
4	2,13	2,78	4,60	8,61
5	2,01	2,57	4,03	6,86
6	1,94	2,45	3,71	5,96
7	1,89	2,36	3,50	5,40
8	1,86	2,31	3,36	5,04
9	1,83	2,26	3,25	4,78
10	1,80	2,20	3,11	4,44
11	1,78	2,18	3,05	4,32
12	1,77	2,16	3,01	4,22
13	1,76	2,14	2,98	4,14
14	1,75	2,13	2,95	4,07
15	1,75	2,12	2,92	4,01
16	1,74	2,11	2,90	3,96
17	1,73	2,10	2,88	3,92
18	1,73	2,09	2,86	3,88
19	1,73	2,09	2,85	3,85
20	1,72	2,08	2,83	3,82
21	1,72	2,07	2,82	3,79
22	1,71	2,07	2,81	3,77
23	1,71	2,06	2,80	3,74
24	1,71	2,06	2,79	3,72
25	1,71	2,06	2,78	3,71
26	1,71	2,05	2,77	3,69
28	1,70	2,05	2,76	3,66
29	1,70	2,05	2,76	3,66
30	1,70	2,04	2,75	3,65
40	1,68	2,02	2,70	3,55
60	1,67	2,00	2,66	3,46
120	1,66	1,98	2,62	3,37
∞	1,64	1,96	2,58	3,29

Додаток 5.

Критичні значення розподілу по критерію χ^2

ν^*	0,995	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,005
1	---	---	-0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	18,114	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892	53,672
40	20,707	22,164	24,433	26,509	29,051	51,805	55,758	59,342	63,691	66,766
50	27,991	29,707	32,357	34,764	37,689	63,167	67,505	71,420	76,154	79,490
60	35,534	37,485	40,482	43,188	46,459	74,397	79,082	83,298	88,379	91,952
70	43,275	45,442	48,758	51,739	55,329	85,527	90,531	95,023	100,425	104,215
80	51,172	53,540	57,153	60,391	64,278	96,578	101,879	106,629	112,329	116,321
90	59,196	61,54	65,647	69,126	73,291	107,565	113,145	118,136	124,116	128,299
100	67,328	70,065	74,222	77,929	82,358	118,498	124,342	129,561	135,807	140,169

ν^* - кількість ступенів свободи

Таблиця розподілу хі-квадрату – це таблиця, яка показує критичні значення розподілу хі-квадрат. Щоб використовувати таблицю розподілу хі-квадрат, вам потрібно знати лише два значення:

- Ступені свободи для тесту хі-квадрат
- Альфа-рівень для тесту (зазвичай вибираються 0,01, 0,05 та 0,10).

СПИСОК ЛІТЕРАТУРИ

1. Бабак В. П. Статистична обробка даних : Монографія. / В. П. Бабак, А. Я. Білецький, П.О. Приставка, О.П. Приставка. К.: МІВВЦ, 2001. 388 с.
2. Бабак В. П. Теорія ймовірностей, випадкові процеси та математична статистика / В. П. Бабак, Б. Г. Марченко, М. Є. Фриз. К.: Техніка, 2004. 288 с.
3. Горбатенко І. Ю. Основи наукових досліджень. К.: Вища школа, 2001. 92 с.
4. 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.
5. Дослідна справа в агрономії : навч. посібник у 2 кн. Кн. 1. Теоретичні аспекти дослідної справи / А.О. Рожков, В. К. Пузік, С. М. Каленська та ін. Харків: Майдан, 2016. 314 с. URI <https://repo.btu.kharkov.ua/handle/123456789/33532>
6. 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. Харків: Майдан, 2016. 314 с. URI <https://repo.btu.kharkov.ua/handle/123456789/33533>
7. Єщенко В. О. Основи наукових досліджень в агрономії : підручник / В. О. Єщенко, П. Г. Копитко, В. П. Опришко, П. В. Костогриз; За ред. В.О. Єщенка. – К.: Дія, 2005. – 288 с.
8. Лісовал А. П. Методи агрохімічних досліджень. К.: НАУ, 2001. 247 с.
9. Мойсейченко В. Ф. Методичні рекомендації для проведення польових досліджень у землеробстві/ В. Ф. Мойсейченко, В. О. Єщенко. К.: УСГА, 1985. 84 с.
10. Статистика. Словник термінів і позначень. Частина 1. Загальні статистичні терміни та терміни теорії ймовірностей (ISO 3534-1:2006, IDT): ДСТУ ISO 3534-1:2008. – [чин. від 2010-01-01]. К.: Держспоживстандарт України, 2010. 42 с.
11. Статистика. Словник термінів і позначень. Частина 2. Прикладна статистика (ISO 3534-2: 2006, IDT): ДСТУ ISO 3534-2:2008. – [чин. від 2010-01-01]. К.: Держспоживстандарт України, 2010. 43 с.
12. Статистичне опрацювання даних. Частина 7. Медіана. Оцінювання і довірчі інтервали: ДСТУ ISO 16269-7:2006.– [чин. від 2000-09-25]. К.: Держспоживстандарт України, 2008. 19 с.
13. Статистичний аналіз даних вимірювань: навч. посіб. / Єременко В. С., Куц Ю. В., Мокійчук В. М., Самойліченко О. В. К.: НАУ, 2013. 320 с.
14. Тимошенко І. І. Основи наукових досліджень в агрономії/ І. І. Тимошенко, З. М. Майшук, Г. О. Косилевич. Львів: ЛДАУ, 2004. 111 с.
15. Fisher, R. A. The general theory of natural selection / R. A. Fisher. 2nd ed. N.-Y.: Dover Publ., 1958. 291 p.
16. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution / S. Wright // Proc. 6th Intern. Cong. Genetics. 1932. Vol. 1. P. 356–366.