Бібліотека журналу «Біологія»
Заснована 2003 року

Випуск 12 (96)

Шамрай С. М., Задорожний К. М.

БІОЛОГІЧНІ ДОСЛІДЖЕННЯ. ПЛАНУВАННЯ І ПРОВЕДЕННЯ

Харків
Видавнича група «Основа»
2010

УДК 37.016
ББК 74.262.8
Ш19

Шамрай С. М.

Ш19 Біологічні дослідження. Планування і проведення / С. М. Шамрай, К. М. Задорожний. — Х.: Вид. група «Основа», 2010. — 111, [1] с.: табл. — (Б-ка журн. «Біологія»; Вип. 12 (96)).

ISBN 978-617-00-0755-1.

У посібнику розглядаються особливості проведення біологічних досліджень у шкільних умовах. На конкретних прикладах зроблено аналіз типових помилок під час проведення експериментальних досліджень учнями. Багато уваги приділено питанням організації та планування досліджень і статистичній обробці матеріалів. Посібник дозволить значно підвищити ефективність досліджень біологічних гуртків, робіт МАН та проектів екологічних олімпіад.

Для вчителів біології загальноосвітніх навчальних закладів і студентів біологічних спеціальностей ВНЗ.

УДК 37.016
ББК 74.262.8

Навчальне видання

Бібліотека журналу «Біологія»
Випуск 12 (96)

ШАМРАЙ Сергій Миколайович
ЗАДОРОЖНИЙ Костянтин Миколайович

БІОЛОГІЧНІ ДОСЛІДЖЕННЯ. ПЛАНУВАННЯ І ПРОВЕДЕННЯ

Навчально-методичний посібник

Головний редактор *К. М. Задорожний*
Редактор *Л. В. Мариненко*
Коректор *О. М. Журенко*
Технічний редактор *О. В. Лєбєдєва*
Комп’ютерне верстання *Є. С. Островський*

Підп. до друку 09.12.2010. Формат 60×90/16. Папір газет.
Гарнітура Шкільна. Друк офсет. Ум. друк. арк. 7,0. Зам. № 10—12/13—04.

ТОВ «Видавнича група “Основа”».
Свідоцтво суб’єкта видавничої справи КВ № 11394–267Р від 26.06.2006.
Україна, 61001 Харків, вул. Плеханівська, 66.
Тел. (057) 731-96-32. E-mail: bio@osnova.com.ua

Віддруковано з готових плівок ПП «Тріада+»
Свідоцтво суб’єкта видавничої справи ДК № 1870 від 16.07.2007.
Харків, вул. Киргизька, 19. Тел.: (057) 757-98-16, 757-98-15.

ISBN 978-617-00-0755-1

ВСТУП

XXI ст. ставить нові завдання перед усією системою освіти, й одне з найважливіших — підготовка конкурентоспроможних на ринку праці фахівців. Розв'язання цієї проблеми вимагає вже на рівні школи, принаймні старшої школи, перейти від винятково педагогічної діяльності до розвитку в учнів умінь у сфері «проблема-розв'язок» або в галузі наукового пошуку. Недостатньо «завантажити» пам'ять школяра значним обсягом досліджуваного матеріалу, що в більшості випадків подається механічно, з упором на запам'ятовування, а не на осмислення. У сучасному світі надзвичайно важливо мати навички самостійного отримання нових даних, нових фактів і знань, їх раціонального осмислення і практичного застосування. Це важливо не лише для тих учнів, які прагнуть стати вченими; такі навички не менш необхідні й майбутнім інженерам, політикам, бізнесменам, банкірам, менеджерам і взагалі всім творчим людям, які прагнуть досягти у своїй діяльності якнайбільших успіхів, однаково, чи буде це тюнинг і ремонт автомобілів, організація й обслуговування комп'ютерної мережі, лікування раку чи конструювання мікропроцесора.

Допомогти вчителеві у виконанні сучасних завдань шкільної освіти може науково-дослідна робота (НДР) школярів. Виконуючи наукове дослідження, учень розвиває самостійність, накопичує навички й досвід самореалізації, розвиває творчі здібності, пам'ять, логічне мислення, уміння чітко висловлювати свої думки — як письмово, так і усно, розвиває комунікабельність тощо.

Ініціатором виконання НДР є вчитель. Він є організатором і керівником наукової праці школяра, а також відповідальним за її результати. Дуже важливо, щоб у проведенні роботи був зацікавлений учень. Якщо на першому етапі роботи можна використати елементи обов'язковості, покладатися на почуття обов'язку школяра, то надалі його однаково слід зацікавити цією діяльністю. У протилежному випадку проведення НДР принесе мало користі або може навіть мати негативні наслідки для формування особистості дитини.

У плані зацікавленості біологія дає вчителеві найширші можливості. Справді, мало знайдеться інших наук, настільки тісно пов'язаних з нашим життям і повсякденністю. З цієї причини проведення біологічних досліджень набуло значної популярності в школах.

Важливо відзначити, що всі етапи НДР мають здійснюватися учнем самостійно. Учитель виконує лише функцію контролера й консультанта. Керівництво науковою працею школярів вимагає від учителя певних додаткових знань і досвіду виконання наукової праці. Не всі вчителі мають достатню компетентність у цих питаннях; з цієї причини деякі НДР, які представляють школярі (наприклад, у рамках Малої Академії Наук), мають недоліки й вади, що часом зводять нанівець усі витрачені на виконання роботи зусилля. Водночас найчастіше багато яких з найбільш типових помилок і недоліків легко уникнути, добре розуміючи логіку, зміст і суть наукової діяльності. Завданням цієї книги і є допомогти вчителеві в керуванні НДР школярів.

У ній розглядаються загальні питання організації та проведення наукових досліджень, особливості правильної організації дослідів, і насамперед дослідів з рослинами, а також особливості статистичного опрацювання результатів.

Розділ 1. ЗАГАЛЬНІ ПИТАННЯ НАУКОВОГО ДОСЛІДЖЕННЯ

Наука за своєю суттю поділяється на «шкільну», «студентську», «вузівську», «академічну» тощо. На найдорожчому встаткуванні в найпрестижнішому університеті можна виконати експеримент, що не має жодної цінності, і в той же час школяр може провести дослідження, результати якого будуть ураховувати у своїй роботі імениті професори. Цінність наукового дослідження залежить від дотримання певних загальних вимог і рекомендацій, головні з яких обговорюватимуться в цій книзі. Розуміння їх є важливим для вчителя, який керує науковою працею школярів, і для самих школярів, що виконують НДР.

Перше питання, відповідь на яке важливо чітко розуміти, звучить так: що являють собою результати наукових досліджень в емпіричній науці? Якщо не брати до уваги словесне обговорення й пояснення результатів, то отримані в науковому дослідженні дані є не більш ніж співвідношення між окремими кількостями й/або певними якостями. Плануючи роботу, дослідник формулює питання, відповіді на які має дати дослідження. Наприклад, чи впливає розмір кристалів солі на швидкість їх розчинення? Чи впливає опудрювання насіння гороху бульбочковими бактеріями на врожай рослин? Чи змінюється кількість зоопланктону в певній ділянці річки в різні періоди часу? Чи посилюється експресія певного гена бактерій у відповідь на зміну складу поживного середовища?

Ставлячи подібні прості або значно більш складні запитання, учені врешті-решт вимірюють або підраховують деякі кількісні або якісні *змінні* й знаходять нові або підтверджують / спростовують раніше відомі співвідношення між ними. У наступному розділі ми розглянемо, що являють собою ці змінні, з якими працюють дослідники.

Що ми вивчаємо в процесі дослідження, або ознаки та змінні

Будь-який фізичний об’єкт, живий чи ні, має певні відмітні ознаки. Під словом «ознака» зазвичай розуміють властивість або особливість, завдяки якій певний об’єкт відрізняється від іншого об’єкта. Наприклад, вода відрізняється від граніту, зокрема, тим, що за звичайного тиску за температури 20 °C вода є рідиною, а граніт — ні. Ртуть за цієї температури так само рідка, як і вода, однак ртуть має низку інших ознак, за якими будь-яка людина відрізнить її від води (наприклад, густина ртуті набагато більша, ніж густина води). Етиловий спирт — також рідина за кімнатної температури, однак і його неважко відрізнити як від води (наприклад, за запахом і смаком), так і від ртуті (наприклад, за прозорістю).

У біології ознаки дозволяють класифікувати живі організми, відрізняти один організм від іншого, порівнювати їх між собою. Наприклад, одна людина може відрізнятися від іншої зростом, вагою, кольором волосся, рисами обличчя, групою крові й багатьма іншими ознаками.

Кожна ознака в кожного об'єкта має певну величину, і як саме ця величина визначається, залежить від природи ознаки (у загальному випадку вона визначається або підрахунком, або вимірюванням (*див. нижче*).

Деякі ознаки є дуже стабільними й не змінюються в разі переходу від одного організму до іншого. Наприклад, якщо підрахувати величину такої ознаки, як кількість ніг, то в усіх корів виявиться чотири ноги. Навіть більше, за цією ознакою неможливо відрізнити корову від коня, слона або миші. Однак за наявністю, наприклад, рогів корови легко відрізняються від останніх трьох тварин, хоча й не відрізняються одна від одної. Ознаки, величина яких не змінюється внаслідок переходу від одного об'єкта до іншого, є постійними величинами (*константами*).

Крім стабільних, є безліч інших ознак, величина яких у більш або менш широких межах змінюється між різними організмами навіть одного виду. Напевно, кожен хоча б один раз бачив череду корів. Якщо це корови однієї масті, то спочатку практично неможливо розрізнити їх між собою. Однак якщо ми спостерігатимемо за чередою досить довго, то почнемо помічати ознаки (наприклад, загальний розмір тіла, форма рогів тощо), за якими тварин можна відрізнити одну від одної. Величина таких ознак, як кров'яний тиск або вміст глюкози в крові, різна не лише в різних людей, навіть у тієї самої людини вона може варіювати в широких межах залежно від часу доби, прийому їжі, фізичного навантаження та інших обставин. Отже, є ознаки, величини яких є *змінними*.

Фундаментальною властивістю живих організмів є мінливість. Власне кажучи, мінливість і означає, що дуже багато ознак живих об'єктів, які впадають в око або які ми аналізуємо, змінюють свою величину в разі переходу від одного об'єкта до іншого. Вага зерен пшениці, що росте на певній ділянці, не дорівнює деякій постійній величині, а змінюється від зернини до зернини, не лише для різних рослин, але навіть і для зерен в одному колосі. Активність амілази в слині людини варіюватиме між окремими людьми, навіть якщо вони й проживають в однакових умовах і харчуються подібною їжею. Щодобовий приріст маси тіла свиней буде відрізнятися від однієї свині до іншої, навіть якщо це свині однієї породи, однієї статі й однакового віку та споживають ті самі корми.

Найчастіше ознаки, що варіюють, і є тими самими речами, які є предметом дослідження в біологічних дисциплінах. Конкретна величина будь-якої ознаки, що варіює, у конкретного об'єкта *є змінною випадковою* величиною, або просто змінною. Випадковою таку змінну називають з тієї причини, що в кожного об'єкта (у біології — у кожного живого організму) вона має різні чисельні значення, залежно від *випадкових* обставин, що не піддаються попередньому урахуванню. Через неможливість урахувати вплив *усіх* факторів, що супроводжують ріст і розвиток кожного організму, величина ознаки, що цікавить нас, змінюється залежно від *випадку*.

Дискретні й безперервні змінні

Залежно від властивостей аналізованої ознаки в найбільш загальному плані змінні можуть бути двох типів. До першого типу — *дискретних* змінних — належать змінні, що є результатом підрахунку. Ці змінні можуть діставати винятково цілочислові значення. Наприклад, у родині може бути 0, 1, 2, 3 або інша кількість дітей. Однак кількість дітей може бути лише *цілою*. Подібно до цього в колосі пшениці може бути лише ціле число зернин, у класі може бути тільки ціле число хлопчиків і дівчаток, у виводку свиноматки може бути тільки ціле число кабанчиків і свинок. Точно так само з певного числа черешків може вкоренитися лише ціле їх число, з деякої кількості насінин огірка може прорости також тільки ціле число рослин, у деякій місцевості поділ жителів за групами крові так само може характеризуватися лише цілими числами (хоча *середні* величини можуть бути й дробовими, наприклад, у родинах певної місцевості цілком може бути *в середньому* 1,3 хлопчика).

Інші змінні є результатами вимірювань, і їх називають *безперервними* змінними, оскільки вони можуть діставати будь-які (не обов'язково цілі) значення в певному інтервалі. Безперервними змінними є зріст і вага людини, кров'яний тиск, урожай рослин, активність ферменту, надій молока, приріст ваги свиней і багато інших подібних величин. Прийняті за безперервні змінні конкретні значення залежать від точності вимірювального приладу.

Шкали вимірювання ознак

Для того щоб виміряти конкретну величину конкретної ознаки, ми використаємо певну *шкалу вимірювання*, і тип такої шкали визначатиметься природою вимірюваної ознаки. Ознаки можуть бути *якісними* (*номінальними*), *порядковими* (*ординальними*) і *кількісними*.

Якісні ознаки не можна впорядкувати, виміряти й порівняти, тобто з ними не можна виконати жодних арифметичних дій. Вони попросту поділяються на два або більше класів, що істотно відрізняються один від одного. Прикладами таких ознак є стать, національність, країна проживання, забарвлення, сорт рослин, порода тварин тощо. Ми не можемо говорити про величину такої ознаки, ми можемо лише підрахувати кількість об'єктів у деякій групі, які мають той чи інший клас ознаки. Наприклад, ми можемо порахувати кількість горошин зеленого кольору, або число хлопчиків і дівчаток, які недужали на певне захворювання, або кількість самок і самців у виводку свиноматки. Якісні ознаки завжди пов'язані з дискретними змінними, чисельні значення яких виражаються цілими числами.

Порядкові, або *ординальні*, ознаки дозволяють ранжувати об'єкти залежно від того, більшою чи меншою мірою вони мають якість, виражену цією ознакою. Однак вони не дозволяють сказати, *на скільки* більше чи *на скільки* менше, тобто не дозволяють проводити арифметичних операцій. Хорошим прикладом таких ознак є експертні оцінки, зокрема дегустація. Наприклад, дегустатори можуть визначити, що вино марки А має кращий смак і більш приємний аромат, ніж вино марки Б, однак сказати, що різниця між винами за смаком дорівнює, наприклад, 21 %, не можна. Зазвичай порядковими змінними є різноманітні візуальні оцінки, наприклад, під час визначення Міс Україна на конкурсі краси, ураженості рослин певним захворюванням, вираженої в балах, або стану хворого (тяжкий, середній або легкий) і т. п. У роботі з порядковими ознаками ми також використовуємо дискретні змінні, наприклад, підраховуємо кількість людей, які перехворіли на грип у тяжкій формі, серед вакцинованих і не вакцинованих, або підраховуємо число рослин, що має той чи інший бал ураженості.

Кількісні ознаки можна безпосередньо вимірювати, а їхні величини можуть змінюватися в певних межах. Наприклад, кількість зернин у колосі, вага і зріст людей, прирости ваги свиней, яйценосність курей, активність ферментів тощо. Однак не всі кількісні ознаки піддаються прямому порівнянню. Їх можна поділити на *інтервальні* й *відносні*.

Інтервальні ознаки дозволяють не лише впорядкувати об'єкти, але й виразити їх чисельно й порівняти відмінності між ними. Однак їх не можна безпосередньо зіставити. Гарним прикладом інтервальної шкали є температура, вимірювана в градусах Фаренгейта або Цельсію. Ми з упевненістю можемо стверджувати, що температура 36,6 °С менша, ніж температура 37,3 °С. Навіть більше, якщо температура в одному випадку збільшилася від 15 до 20 °С, а в іншому випадку — від 15 до 30 °С, то ми можемо сказати, що в другому випадку *зміна* температури була втричі більшою, ніж у першому. Однак сказати, що температура 30 °С *удвічі* більша за температуру 15 °С, було б неприпустимою помилкою (насправді вона більша приблизно на 5 %).

Відносні ознаки подібні до інтервальних ознак. Їхньою характерною особливістю є наявність певної нульової точки (початку шкали вимірювання); отже, ці змінні ми можемо безпосередньо порівняти. Типовим прикладом таких змінних є температура за Кельвіном. Ми можемо не лише стверджувати, що температура 200 К вища, ніж 100 К, але й те, що вона вдвічі вища. Інтервальні шкали не дають такої можливості. Відносними ознаками є розмір, вага, активність ферменту, уміст глюкози, плідність, кількість зернин у колосі, кров'яний тиск тощо, тобто більшість кількісних ознак, з якими дослідник має справу.

Величини кількісних ознак можуть вимірюватися дискретними й безперервними змінними.

Залежні й незалежні змінні

Як було відзначено вище, плануючи дослідження, учений формулює питання, відповіді на які він бажає отримати. Часто питання формулюються приблизно так: чи впливає деяка виконувана нами дія на величину ознаки експериментальних об'єктів, що цікавить нас? Наприклад, ми можемо обробити насіння якої-небудь сільськогосподарської культури біостимулятором і перевірити, чи підвищуватиме він схожість або масу рослин. Так само нас може цікавити вплив унесення добрив на врожай, вплив сортових особливостей винограду на вкорінення черешків, вплив інтенсивності фізичного навантаження на частоту пульсу штангіста і т. п.

У такому разі ми чинимо певний вплив на наші об'єкти. Способи, види й дози такого впливу в наукових дослідженнях зазвичай також називають змінними. Однак це не будуть знайомі нам випадкові змінні величини, оскільки під час проведення досліду ми надаємо цим величинам цілком певних значень. Отже, ряд змінних у конкретному дослідженні ми контролюємо та змінюємо на власний розсуд. Наприклад, ми можемо дати хворому ту чи іншу дозу ліків, рослини можна вдобрювати різними дозами або навіть формами нітратних добрив, а свиней можна утримувати на різних раціонах, що відрізняються кількісно або якісно (наприклад, за калорійністю). У дослідах з рослинами ми можемо використовувати різні сорти, дотримуватися певного режиму поливу або освітленості, плодових мушок можна піддавати дії різних контрольованих нами температур, у токсикологічних дослідженнях давати лабораторним тваринам певні дози досліджуваної речовини, і т. д. Змінні, що визначаються й контролюються дослідником, називають *незалежними* змінними. Така назва може здатися дивною, оскільки саме вони й залежать від того, що спаде досліднику на думку. Однак тут важливо інше: вони *не залежать* безпосередньо від об'єкта дослідження, і він не може вплинути на них жодним чином, за будь-якого свого бажання (хто запитує свиню, чим її годувати, або хворого, чим його лікувати?).

Низка змінних, які також зараховують до незалежних, не може контролюватися дослідником безпосередньо, оскільки ними не можна маніпулювати. До подібних незалежних змінних відносять якісні, наприклад стать тварин. У таких випадках ми просто приписуємо об'єкти до певних експериментальних груп, ґрунтуючись на їхніх апріорних властивостях, але впливати на цю змінну в конкретної тварини або людини дослідник не може.

Зовсім інші справи з тими змінними, які ми підраховуємо або вимірюємо, тобто задля визначення величини яких ми й проводимо дослідження. Наприклад, у хворого ми можемо вимірювати кров'яний тиск, у свині — щодобовий приріст, у гороху підрахувати розподіл насінин за забарвленням і формою, а в ячменя — визначити врожай. Ці змінні цілком залежать від об'єкта досліджень, від його реакцій, властивостей і намірів. Подібні змінні називають *залежними*. У науковій літературі залежну змінну часто називають змінною-відгуком. Відзначимо, що величина залежних змінних не залежить від намірів і бажань дослідника.

Експеримент і спостереження

Вище ми говорили про те, що перед дослідженням фахівець формулює питання, відповіді на які воно дає. Зрозуміло, що в реальних наукових дослідженнях ці питання найчастіше не формулюються, наприклад, так: «А як зміниться врожай пшениці, якщо внести в ґрунт нітратні добрива?» або «А що станеться, якщо хворому замість ліків № 1 дати ліки № 2?» Плануючи дослідження, фахівець аналізує наявні в науковій літературі відомості щодо тієї проблеми, у рамках якої проводиться дослідження, і на підставі цих відомостей до здійснення експерименту висловлює певну гіпотезу щодо виду зв'язку між залежними й незалежними змінними. Наприклад, можна припустити, що певна доза нітратних добрив підвищить урожайність пшениці як мінімум на 20 ц/га, або що застосування ліків № 2 скоротить перебування хворого в стаціонарі порівняно з використанням традиційних ліків № 1. Експеримент, або дослід, фактично й проводиться для перевірки подібних гіпотез.

Отже, у багатьох випадках наукове дослідження пов'язане з перевіркою гіпотез. Розглянемо найпростіший приклад подібного типу досліджень.

Можна висловити гіпотезу, що більш дрібні кристали солі розчинятимуться у воді швидше, ніж великі кристали, оскільки маленькі кристали мають відносно більшу площу поверхні. Щоб перевірити цю гіпотезу, потрібно провести експеримент, у якому слід порівняти швидкість розчинення кристалів солі різного розміру. Дослідник повинен чітко ідентифікувати незалежну змінну, величиною якої він маніпулюватиме (розмір кристалів солі), і залежну змінну, величину якої він визначатиме (швидкість, з якою кристали розчиняються, або час, за який вони цілком розчиняться). Крім залежної змінної, слід також контролювати ряд інших змінних, щоб гарантувати однаковий ступінь їхнього впливу на залежну змінну в різних варіантах досліду (наприклад, об'єм води, у яку поміщають кристали, кількість солі з різним розміром кристалів, температура води, швидкість перемішування та ін.). Здійснивши експеримент, дослідник порівнює, наскільки швидко розчиняються кристали різного розміру за цих контрольованих умов, і виконує відповідні обчислення.

Наведений приклад є простим, однак він служить яскравою ілюстрацією того, як працюють учені. Подібне дослідження належить до випадку проведення *експериментального дослідження,* або просто експерименту, що зазвичай виконується в умовах лабораторії або в умовах поля (наприклад, у рослинництві). Важливим є те, що під час експерименту дослідник активно маніпулює незалежними змінними за загального контролю решти умов, які можуть вплинути на залежну змінну або змінні. Досліджувані в експериментах системи є відносно простими, існує лише невелика кількість змінних, на які потрібно звертати увагу, й очікувані взаємодії між залежними й незалежними змінними найчастіше мають характер причини й наслідку. Експерименти підтверджують або спростовують причинно-наслідкові зв'язки.

Однак експерименти, в основу яких покладено перевірку гіпотез і аналіз причинно-наслідкових зв'язків, аж ніяк не є всіма типами досліджень, що проводяться в біології або інших науках.

Багато фахівців формулюють свої гіпотези лише після аналізу даних, а ціла низка спеціалістів, таких, як польові біологи або астрономи, проводять описові дослідження, у яких гіпотези не можна перевірити безпосередньо.

Інакше кажучи, чимало вчених проводять спостереження, або обсерваційні дослідження (від англ. *observation* — «спостереження»). Здійснюючи спостереження, фахівець не впливає на змінні, а лише вимірює та реєструє їх, тобто спостерігає перебіг процесу. Учені в багатьох галузях знань не можуть активно маніпулювати змінними та створювати «контрольні» й «експериментальні» групи. Такими є спостереження, характерні для метеорології (ми *спостерігаємо* за напрямком вітру й температурою повітря, але не можемо *впливати* на них), популяційної біології (ми *спостерігаємо* розподіл алелей різних генів у популяціях живих організмів), в етології тварин (ми *спостерігаємо* за поводженням тварин, намагаючись не налякати їх), флористики й фауністики, астрономії, геології, океанографії тощо. Дослідники в усіх цих галузях створюють моделі явищ, але не контролюють умови, а відбирають проведені в природних умовах спостереження й відшукують у них описові, корелятивні або причинні закономірності.

Наприклад, дослідник може цікавитися взаємозв'язком між якістю повітря й ростом лишайників на деревах. Він не має можливості маніпулювати якістю повітря в посадці дерев. Такий дослідник виділить ділянку з високим забрудненням повітря (можливо, поблизу автостради або в промисловому районі) і ділянку з низьким забрудненням повітря. Далі він вирішуватиме, як урахувати змінні, що потенційно чинять вплив, такі, як вид дерева, вологість зони або кількість сонячного світла. Потім він вибере для вивчення подібні дерева, що ростуть у порівнянних умовах у зоні з високим забрудненням повітря й зоні з низьким забрудненням повітря (ступінь забруднення повітря має бути єдиною відмінністю між обраними деревами), і порівняє кількість лишайників на кожній із двох груп дерев.

Іншою відмінністю між експериментом і спостереженням є те, що під час спостережень часто не передбачається наявності причинного взаємозв'язку між змінними. Зокрема, у біологічних науках досліджувані системи часто є складними і змінні можуть узаємодіяти ймовірнісним способом. Зв'язок може мати вигляд кореляції, але не бути обов'язково причинним. Якщо згадати приклад з лишайниками, то може статися так, що третя змінна, така, як кількість місцевих опадів, впливає як на ступінь забруднення повітря, так і на ріст лишайників на деревах. У такому випадку дослідник, можливо, прагнутиме не стільки вибрати дві окремі ділянки з високим і низьким рівнями забруднення повітря, але, швидше, перевірятиме всі дерева, для яких можна оцінити ступінь росту лишайників і якість повітря в місцях їх зростання.

Спостереження проводяться не лише в причинно-наслідкових і корелятивних дослідженнях. Учені також виконують дослідження, у яких вони намагаються створити винятково описову модель окремих природних явищ. Одним з подібних досліджень є відстеження тварин у їхніх місцях існування з допомогою радіонашийників. Типовим питанням може бути таке: «Де ці тварини проводять більшість свого часу?» або «Як їхній ареал перекривається з ділянками діяльності людини?» Іншим прикладом описового дослідження є складання профілю наявності макроскопічних безхребетних уздовж русла ріки. Такі типи спостережень часто генерують достатню кількість даних, щоб допомогти сформулювати змістовні корелятивні або причинно-наслідкові питання для наступних спостережень. Описові результати можуть бути ефективно представлені на карті.

Отже, якщо під час проведення експерименту ми практично завжди аналізуємо причинно-наслідкові зв'язки й перевіряємо попередньо сформульовані гіпотези, то під час проведення спостережень ми формулюємо три основні типи питань — щодо причинно-наслідкових зв'язків, щодо кореляцій та описові питання. Приклади різних типів питань наведено в таблиці.

Типи питань під час проведення різних типів спостережень

Питання під час проведення описових спостережень	Наскільки багато? Наскільки часто? Що відбувається?
Питання під час проведення спостережень з метою пошуку кореляції	Чи є взаємозв'язок між двома або кількома змінними? Висловлюється гіпотеза про взаємозв'язок, що може виявитися причинно-наслідковим, а може, і ні
Питання під час проведення спостережень з метою пошуку причинно-наслідкових зв'язків	Чи є відмінності між групами, різними періодами часу або певних місць? Висловлюється передбачення або гіпотеза щодо відмінностей

Під час проведення спостережень, які ми здійснюємо з метою встановлення причинно-наслідкових зв'язків, слід дотримуватися обережності. У цьому разі дослідник з'ясовує значення змінних, які він уважає незалежними, і реєструє відповідні їм значення залежних змінних. Виявивши певний узаємозв'язок (кореляцію) між першими та другими, дослідник намагається обґрунтувати причинно-наслідковий зв'язок між змінними. Наприклад, палії частіше за тих, хто не палить, хворіють на рак легенів. Визначивши кількість цигарок, які випалюють палії за день, і стаж паління, а також імовірність занедужати на рак легенів, неважко встановити між ними зв'язок, або кореляцію: чим довше й більше палить людина, тим вище для неї ймовірність захворіти на рак легенів. Це вважається доказом, що паління є причиною раку. Однак наскільки надійним є такий доказ?

В обсерваційному дослідженні ми *ніколи* не можемо бути впевнені, що обстежені нами групи об'єктів розрізняються *лише* тією ознакою, величину якої ми вважаємо незалежною змінною й за якою ці групи були сформовані. Головна причина цього полягає в порушенні фундаментального принципу вибіркового методу — однакової ймовірності будь-якого члена генеральної сукупності потрапити до вибірки (*див. нижче*). Інакше кажучи, під час проведення спостережень вибірки часто бувають невипадковими. Наприклад, відносно паліїв не можна виключити, що підвищена ймовірність захворіти на рак легенів зумовлена генетично й що вона зчеплена з генетичною схильністю до тютюнопаління. В обсерваційному дослідженні спростувати таке припущення неможливо (усе сказане вище є лише ілюстрацією обмеженості обсерваційних досліджень; шкода паління не викликає жодних сумнівів). Отже, результати спостереження можуть інтерпретуватися як причинно-наслідкова залежність лише на ґрунті певної апріорної теорії, але самі по собі не можуть довести причинність.

У деяких випадках причина й наслідок не викликають сумнівів. Наприклад, якщо розглядати вплив опадів на врожай, або доз внесених добрив на врожай, або ступеня ураженості шкідниками плодів на врожай, тут причина зрозуміла. Саме опади, добрива або шкідники є причиною тієї чи іншої зміни врожайності (а вітер — причиною обертання флюгера), але не навпаки. Однак не завжди наявна настільки проста ситуація. Наприклад, між зростом і вагою людини зв'язок існує, але чи можемо ми тут установити причину й наслідок? Зріст — причина ваги, чи вага — причина зросту?

Дуже часто аналізовані нами змінні залежать ще від якогось фактора. І якщо цей фактор не виявлено й не враховано, то під час припису причинно-наслідкових зв'язків ми можемо легко припуститися помилки. Наприклад, якщо зібрати відповідні дані, то в літню пору в Харкові напевно виявився би високий зв'язок між твердістю дорожнього покриття й кількістю проданих прохолодних напоїв, а саме: чим м'якшим є дорожнє покриття, тим більше кожен житель міста в середньому купує прохолодних напоїв. З цього жартівливого прикладу неважко здогадатися, що перша і друга змінні є наслідками третього фактора — температури. Чим спекотніше, тим сильніше розігрівається й розм'якшується асфальт, і тим більше люди купують прохолодних напоїв. У науковій діяльності дуже часто буває не так просто побачити додатковий фактор, що впливає на наші об'єкти, і причину іноді починають шукати там, де насправді вона відсутня. Уявімо собі, що ми нічого не знаємо про літні температури та властивості асфальту, і маємо дані лише про зв'язок твердості дорожнього покриття й кількості напоїв, які випивають жителі міста. Чи зуміємо ми правильно пояснити їх? Розумний ступінь обережності, як і в багатьох інших випадках, дозволить нам не впасти в явну оману, інтерпретуючи результати спостережень.

Розділ 2. БІОЛОГІЧНІ ЕКСПЕРИМЕНТИ ТА ЇХ КЛАСИФІКАЦІЯ

Біологічні експерименти в умовах школи

Участь у біологічних експериментах і спостереженнях є чудовим способом залучення дітей до занять біологією. Воно викликає в дітей зацікавленість обраною для досліджень проблемою, стимулює їхню самостійну роботу з науковою літературою, сприяє розвитку наукового системного мислення, формує в учнів навички аналізу й синтезу інформації. Крім того, біологічні дослідження дозволяють зміцнювати міждисциплінарні зв'язки, особливо з такими предметами, як математика, хімія й фізика.

Усі дослідження в школі проводяться учнями під керівництвом і контролем учителя, який пропонує учнями теми досліджень і забезпечує їх відповідними методичними рекомендаціями щодо пошуку наукової літератури, планування дослідження й аналізу його результатів. Необхідно відзначити, що вибір теми дослідження становить певні труднощі. З одного боку, тема не має вимагати проведення занадто складних експериментів, оскільки всю або майже всю роботу учні мають виконувати самостійно. З іншого боку, робота не має бути й занадто простою й тривіальною, оскільки підтвердження в результаті виконання роботи фактів і закономірностей, які давно й добре всім відомі, не має ніякого сенсу. А от отримання нехай і невеликого за обсягом, але важливого наукового результату позитивно впливає на ставлення учнів до роботи, стимулює її подальше проведення.

База для проведення біологічних експериментів у шкільного вчителя дуже невелика, і навряд чи збільшиться найближчими роками. У типовому випадку в нього є куточок живої природи з рибками, кролями, папугами або іншими дрібними тваринами (якщо він узагалі є в школі), кімнатні рослини в кабінетах і на пришкільній ділянці. Крім того, шкільний учитель може організувати дослідження окремих видів тварин і рослин або навіть екосистем на території свого населеного пункту чи району. Більш складні дослідження можна проводити спільно зі станціями юннатів, близько розташованими науковими установами й вищими навчальними закладами. Однак така можливість є не завжди.

Проте, такого скромного набору можливостей буває цілком достатньо для проведення цікавих і справді наукових досліджень. Необхідно лише чітко сформулювати мету й добре усвідомити, навіщо проводяться дослідження. Нижче буде показано, як слід і як не слід проводити біологічні дослідження.

Головну увагу під час розгляду методики біологічних експериментів буде приділено дослідам з рослинами. Це пов'язано з тим, що такі експерименти та спостереження проводяться в школах найчастіше і їх найлегше виконувати. Водночас багато чого зі сказаного стосується також експериментів із тваринами. Потрібно, однак, ураховувати, що в школі можуть проводитися лише обмежені типи експериментів із тваринами, що не порушують принципів біоетики. Особливо цікавими можуть бути генетичні експерименти з вивчення особливостей спадкування окремих ознак, наприклад, у кролів або риб.

Класифікація експериментів (дослідів)

Під час проведення експериментів з рослинами їх можна вирощувати в полі або в лабораторії, працювати як із цілими рослинами, так і з окремими органами, тканинами й навіть клітинами. У виконанні досліджень можуть використовуватися чотири основні типи експериментів: лабораторний, вегетаційний, лізиметричний і польовий. Слід ураховувати, що відмінності між цими основними типами певною мірою умовні й вони можуть сполучатися й доповнювати один одного в конкретному дослідженні.

Лабораторний експеримент — це дослідження, що виконується за чітко контрольованих умов. Об'єктами досліджень можуть бути не лише цілі рослини, але і їхні окремі частини, органи або навіть окремі клітини, органели й навіть макромолекули.

Звичайно, провести дослідження на рівні макромолекул, органел або клітин в умовах школи практично неможливо, але інші рівні досить легко можна використати.

Вегетаційний експеримент — це дослідження, за якого об'єктом вивчення є рослини, що вирощуються за контрольованих умов вегетаційних будиночків, теплиць, оранжерей, кліматичних камер та інших споруджень.

Сутність вегетаційного методу досліджень полягає в тому, що рослини вирощують у штучній, але обґрунтованій щодо агрономії обстановці, регульованій експериментатором. За умов вегетаційного досліду вплив того чи іншого фактора вивчається в найбільш чистому вигляді, за сталості або відсутності дії інших, не досліджуваних факторів.

Під час вегетаційного досліду рослини вирощують у різних вегетаційних посудинах у вигляді водних, гравійних, піщаних або ґрунтових культур.

Удосконалення техніки вегетаційного методу привело до створення складних інженерних споруджень — фітотронів, у кліматичних камерах яких можна цілорічно працювати з рослинами, моделюючи для них будь-які умови життя.

Очевидно, що використання фітотронів у школах є справою майбутнього (швидше за все, досить віддаленого), але наявність у школах теплиць і оранжерей не є рідкістю, тому проведення подібних досліджень є цілком можливим.

Лізиметричний експеримент — це дослідження динаміки ґрунтових процесів та їх впливу на ті чи інші властивості й особливості росту й розвитку рослин у спеціальних лізиметрах, що дозволяють ураховувати й контролювати пересування й баланс вологи й поживних речовин у природних умовах.

На відміну від вегетаційних дослідів, лізиметричні проводять у полі за природних умов освітленості, температури тощо. Технічно лізиметри являють собою різноманітні конструкції із цегли, бетону, металу та інших матеріалів, у яких ґрунт для вирощування рослин відгороджений з усіх боків від навколишнього ґрунту. Головним устаткуванням лізиметра є пристосування, що дозволяють вивчати й контролювати переміщення води й розчинених речовин. Потужність шару ґрунту в лізиметрі може варіювати від глибини орного шару до 1–2 м.

Залежно від способу наповнення ґрунтом розрізняють лізиметри з ґрунтом природної будови й лізиметри з насипним ґрунтом.

Лізиметричні досліди використовують для вивчення таких питань, як водний баланс під різними культурами, вимивання й переміщення поживних речовин ґрунту атмосферними опадами, визначення транспіраційних коефіцієнтів рослин у природних умовах та ін.

Слід мати на увазі, що цілковите відокремлення ґрунту в лізиметрі від нижче розташованих шарів створює для рослин інші водний, аераційний і поживний режими, ніж у звичайних польових умовах. Отже, лізиметричні досліди посідають проміжне місце між вегетаційними й польовими експериментами.

Використання такого типу експериментів в умовах школи є дуже складним і навряд чи здійсненним. Крім того, лізиметричні досліди найчастіше проводять з метою вивчення природних явищ, що протікають повільно, вони розраховані на багато років і навряд чи можуть становити інтерес як тип експериментів, що проводяться в школі.

Польовий сільськогосподарський дослід — це дослідження, що здійснюється в природних умовах поля на спеціально виділеній ділянці.

Головним завданням польового досліду є встановлення відмінностей між варіантами досліду для якісного та кількісного оцінювання впливу різних факторів на ті чи інші властивості рослин, насамперед на їхній урожай та його якість.

Польові досліди можуть сполучатися з вегетаційними, що утворює так званий вегетаційно-польовий дослід. При цьому рослини можуть певний час вирощуватися за контрольованих умов, а потім пересаджуватися в поле, або вирощуватися в полі в циліндричних чи квадратних посудинах без дна. Крім того, наразі існують стаціонарні й пересувні кліматичні вегетаційні камери й пересувні вегетаційні будиночки, які дозволяють на певний час створювати для рослин контрольовані кліматичні умови на різних етапах вегетації. Це дає експериментатору можливість оцінити вплив кожного кліматичного фактора на формування врожаю, що неможливо в природних умовах.

Як свідчить практика, саме більш або менш прості варіанти польового досліду є найпоширенішою формою дослідних робіт учнів. Проводити такі досліди в умовах пришкільних ділянок досить легко, і вони дозволяють учням самостійно виконувати всі етапи наукової праці.

З цієї причини надалі головні особливості проведення досліджень ми розглядатимемо саме на прикладі польових дослідів, хоча практично все сказане нижче стосується також інших типів біологічних експериментів і польових біологічних спостережень.

Розділ 3. ВИМОГИ ДО БІОЛОГІЧНИХ ЕКСПЕРИМЕНТІВ

Загальні зауваження

Незалежно від того, який тип експерименту було обрано для проведення дослідження, усі вони мають відповідати певним методичним вимогам. Дотримання цих вимог дозволяє проводити статистичне опрацювання результатів і не просто повідомляти про існування певної закономірності, але й надійно обґрунтувати виявлені зв'язки між залежними й незалежними змінними. Саме дотримання цих вимог дозволяє відрізнити справжній експеримент від наукоподібної роботи, що є марно витраченими часом і зусиллями.

Вимоги до польового досліду

Цінність польового досліду залежить від дотримання певних методичних вимог. Найважливішими з них є:

1. Відтворюваність.
2. Типовість.
3. Дотримання принципу єдиної відмінності.
4. Проведення досліду на спеціально виділеній ділянці.
5. Вірогідність за суттю.

Слід відзначити, що ці вимоги, крім четвертої, справедливі для будь-яких наукових експериментів, а не лише для польових.

Найважливіша вимога, якій має відповідати дослід, — це його відтворюваність. Кожен фахівець, відтворивши умови й методику проведення експерименту з тим самим об'єктом, що і його попередник, має отримати такі самі результати, принаймні, якісно. Невідтворюваність ставить під сумнів будь-які висновки, зроблені на основі подібних експериментів. До слова, саме неможливість отримати аналогічні результати в інших наукових центрах змусила американських учених відмовитися від відкриття хімічного елемента під № 108 і повідомити про помилку в експерименті з його одержання.

Під типовістю, або репрезентативністю, польового досліду розуміють відповідність умов його проведення ґрунтово-кліматичним і агротехнічним умовам району або зони, де планується впровадження результатів цього досліду. Будь-який польовий дослід має відповідати вимогам ґрунтово-кліматичної типовості. Цілком очевидно, що немає сенсу вивчати, наприклад, можливості боротьби з кореневими інфекціями в польовому досліді, закладеному на піщаних ґрунтах, якщо результати роботи передбачається використовувати на глинистих чорноземах. Що стосується відповідності досліду агротехнічним виробничим умовам, то ця вимога може бути застосована переважно до виробничих польових дослідів.

До поняття «типовість» польового досліду входить також вимога до проведення досліджень з типовими для певної зони культурами, з використанням районованих або перспективних сортів. Якщо проводяться агротехнічні дослідження, у яких використовуються екологічно не пристосовані до певної зони культури й сорти, то зазвичай результати подібної роботи не мають жодної цінності, оскільки районовані культури й сорти можуть зовсім інакше реагувати на досліджувані прийоми. З цієї причини не можна поширювати висновки подібних дослідів на звичайні виробничі умови. У зв'язку з цим неприпустимо, наприклад, вивчати захворювання, що трапляється на овочевих культурах у теплиці, на сортах відкритого ґрунту, і навпаки.

До типовості належить також вимога проведення польового досліду за загального високого рівня агротехніки та з насінням рослин високого класу. В іншому випадку польові досліди не мають практично ніякої цінності. Наприклад, дослід з добривами на неокультуреному ґрунті може справляти чимале враження, але він не відповідатиме практичним умовам звичайних ґрунтів зі старою ріллею. Подібно до цього можна досягти вражаючих результатів у передпосівній обробці насіння, якщо використати насіння з низькою схожістю.

Дотримання принципу єдиної відмінності означає, що всі інші умови, крім досліджуваного фактора, мають бути однаковими. Інакше цей принцип можна назвати принципом однаковості супутніх досліду умов. Це неодмінна вимога до проведення будь-якого наукового експерименту, що має суворо дотримуватися в дослідній роботі. Наприклад, якщо в досліді вивчається вплив дози фунгіциду на ураженість рослин захворюваннями, то єдиною відмінністю між варіантами досліду мають бути саме дози фунгіциду. Решта умов (ґрунт, сорти, попередні культури, добрива тощо) у всіх варіантах мають бути однаковими.

На практиці, однак, під час закладання польових дослідів далеко не завжди вдається дотримуватися цілковитої однаковості всіх умов, крім досліджуваної. Наприклад, можуть позначатися особливості мікрорельєфу дослідної ділянки, відмінності в складі ґрунту в різних місцях дослідної ділянки тощо. Ці проблеми розв'язуються шляхом використання кількох повторів кожного варіанта досліду та їх розміщення на дослідній ділянці спеціальними методами, наприклад методом цілком рендомізованих блоків. Про це йтиметься нижче.

Незважаючи на те що принцип єдиної відмінності — неодмінна умова будь-якого наукового експерименту, його не слід розуміти абсолютно. Наприклад, якщо порівнюються два сорти пшениці за стійкістю до певного захворювання й один з них через свої біологічні особливості потребує для оптимального розвитку густоти посіву, що відрізняється від такої в другого сорту, то було б нерозумно висівати їх однаковою нормою, тому що при цьому один із сортів виявився б у свідомо невигідних умовах. Точно так само під час випробування різних сортів збирання врожаю входить у досліджуваний комплекс, а не є супутньою умовою, яку потрібно вирівнювати. Час збирання врожаю пов'язаний з тривалістю вегетаційного періоду конкретного сорту, і було б неправильним збирати всі сорти одночасно, якщо тривалість їхньої вегетації відрізняється. У подібних випадках принцип однаковості слід розуміти творчо, виходячи з доцільності й оптимальності.

Вимога проведення польового досліду на спеціально виділеній ділянці з добре відомою господарською історією є обов'язковою для будь-якого польового досліду. Досліди, проведені не на спеціально виділеній і підготовленій ділянці (на пустирях, дачах тощо), не мають жодної наукової або практичної цінності, незалежно від завдань досліджень. Фактично ця вимога є логічним наслідком принципу єдиної відмінності. Не можна назвати польовим дослідом будь-які дослідження сортів або агротехніки, якщо вони проводяться на випадкових ділянках.

Під вірогідністю досліду по суті розуміють логічно правильно побудовану схему й методику проведення дослідів, їх відповідність поставленим перед дослідником завданням, правильний вибір об'єкта й умов. До невірогідності досліду по суті можна зарахувати використання неякісних матеріалів (реактивів), устаткування й неперевірених вимірювальних приладів. Досліди, недостовірні по суті, часто можуть призводити до неправильних висновків.

Розділ 4. ПЛАНУВАННЯ БІОЛОГІЧНИХ ЕКСПЕРИМЕНТІВ

З огляду на зазначені вище вимоги, дослідник насамперед планує експеримент. Під плануванням розуміють визначення цілей, завдань і об'єктів досліджень, розробку схеми експерименту, вибір оптимальної структури польового досліду й ділянки для його проведення. Планування — найважливіший елемент наукових досліджень, що становить низку послідовних етапів роботи. На жаль, плануванню дослідів не завжди приділяється належна увага. Але ж помилки, допущені під час планування, не можна виправити надалі ні ретельним проведенням дослідної роботи, ні наступним статистичним опрацюванням. Тут доречно навести цитату з книги О. О. Любіщева (1986), який, відзначаючи очевидну необхідність здійснення досліджень з мінімальною витратою сил і засобів, пише: «...чимало дослідників убачають особливу заслугу в тому, що зробили величезну кількість проб, дослідили величезну кількість об'єктів... Інші дослідники намагаються замінити чималу кількість об'єктів дослідження надмірною точністю спостережень, уважаючи, що зайва точність ніколи не зашкодить. Далеко не всі чітко усвідомлюють, що число досліджених об'єктів і точність мають випливати з конкретних умов дослідження. Якщо ж дослід організований неправильно, то педантична точність і значний обсяг матеріалу не запобіжать помилковим висновкам. Виходить, як висловлюється Р. Фішер, що не лише починають стріляти з гармат по горобцях, але, що ще сумніше, не поціляють у горобців» (с. 19).

У найбільш загальному вигляді структуру планування можна поділити на два етапи.

Перший етап містить:

- вибір теми;
- визначення актуальності досліджень;
- формулювання цілей і завдань досліджень;
- вибір об'єкта або об'єктів досліджень;
- збирання і критичний аналіз наявної інформації з досліджуваної проблеми;
- формулювання й висування робочих гіпотез (теоретичних моделей).

Другий етап планування досліду передбачає розробку програми досліджень. Найбільш важливими моментами цього етапу є такі:

- визначення розділів дослідної роботи;
- визначення місця і термінів їх виконання;
- складання схеми проведення дослідів за кожним розділом;
- складання календарного плану робіт з підготовки та проведення польового досліду, догляду за рослинами;
- складання плану фенологічних спостережень за розвитком рослин з конкретним зазначенням термінів і частоти проведення обліків і спостережень.

Успішне проведення досліду багато в чому залежить від вибору теми досліджень. При цьому слід уявляти рівень розробленості різних питань у галузі біології, їх перспективність, запити теорії та практики. Вибираючи тему, слід керуватися такими критеріями, як актуальність, новизна, перспективність. Проблема, для розв'язання якої здійснюється наукове дослідження, має бути чітко й однозначно визначена. Якщо вона не може бути сформульована, то вона не може бути й розв'язана.

Усвідомивши проблему, слід визначити задачі, відповіді на які приведуть нас до її розв'язання. Відповіді на них і дають експерименти, кожен з яких передбачає чітко сформульовану мету. Такою метою можуть бути гіпотези, які потрібно перевірити, або ефекти, що мають бути оцінені. Усупереч цьому цілі й завдання досліджень досить часто формулюються в загальній формі, що ускладнює їх виконання в межах одного досліду або дослідження. Досліднику слід обмежитися рамками саме того питання, на яке бажано отримати відповідь.

Вибираючи об'єкт досліджень, потрібно спиратися на поставлені цілі й завдання, господарське й наукове значення різних об'єктів, планованих методик і здорового глузду. Мабуть, досліднику, який працює в Україні, недоцільно вибирати як об'єкт вивчення кавове дерево або банан.

Вивчення наукової літератури є важливим етапом роботи, що дозволяє уникнути непотрібного дублювання й не витрачати часу на винахід велосипеда. Власне кажучи, формулювання теми, цілей і завдань дослідження неможливі без знання наявних у цій сфері наукових даних. Дослідник має добре знати, що саме в цій галузі науки є вивченим, а що залишається невивченим, і чітко уявляти, чому це невивчене важливо вивчити. Лише маючи ці знання, можна чітко сформулювати проблему, задля розв'язання якої проводитимуться експерименти. Крім того, дослідник ніколи не робить висновків лише на підставі свого експерименту, навіть якщо цей експеримент і було проведено кілька разів. До уваги береться інформація, наведена в наукових публікаціях інших фахівців. Звичайно, якщо робота ведеться в цілком новому напрямі, у якому немає ніякої інформації, то основою для висновків дослідника будуть лише його власні дані. Однак такі ситуації в науці трапляються нечасто.

Знання наукової літератури необхідне також для висування робочої гіпотези, тобто наукового припущення щодо розвитку явищ, на якому ґрунтується пояснення очікуваних у плановому досліді результатів. Робоча гіпотеза, висунута до проведення експерименту (апріорно), є найважливішим методологічним інструментом організації наукових досліджень. Вона є відправним пунктом для складання схеми досліду. Можна сказати так, що сам дослід проводиться для перевірки робочої гіпотези. Якщо результати експерименту не суперечать висунутій гіпотезі, то вона поглиблюється й розвивається далі. Якщо ж результати експерименту суперечать робочій гіпотезі, то вона видозмінюється з урахуванням нових даних.

Головні елементи методики й техніки польового досліду

Головними елементами методики польового досліду є: кількість варіантів і повторів, площа й форма ділянок, напрямок їх розміщення, метод розташування варіантів на земельній ділянці, система розміщення повторів, метод обліку врожаю, організація досліду в часі.

Кількість варіантів

Сукупність усіх варіантів досліду становить схему експерименту.

Кількість варіантів у схемі досліду залежить від теми досліджень і визначається метою та завданнями експерименту. Плануючи дослідження, завжди потрібно пам'ятати, що кількість варіантів впливає на точність досліду, оскільки зі збільшенням числа варіантів помилка досліду зростає. Це зумовлено тим, що зі збільшенням кількості варіантів збільшується площа під дослідом, зростає строкатість ґрунтової родючості й відстань між порівнюваними варіантами. За великої кількості варіантів важко помістити дослід або його окремі повтори в межах однорідної за ґрунтовою родючістю ділянки. Усе це збільшує дисперсію даних досліду і, відповідно, його помилку.

Слід мати на увазі, що на великих ділянках збільшення кількості варіантів значно сильніше збільшує помилку досліду, ніж на ділянках меншого розміру, і це необхідно враховувати в плануванні експерименту.

З іншого боку, якщо кількість варіантів занадто мала, площа дослідної ділянки буде використана нераціонально; крім того, незначна кількість варіантів не завжди може дати достатньо повну характеристику досліджуваних факторів. Крім іншого, необхідно раціонально використовувати вегетаційні періоди.

Як свідчить практика дослідної справи, слід прагнути того, щоб у досліді було 8–12, максимум 16 варіантів. Збільшення їх кількості вимагає, як правило, ускладнення методів постановки досліду.

Повторності й повтори

Використання кількох повторних ділянок для кожного варіанта досліду — найбільш дієвий засіб зниження помилки досліду. Власне кажучи, проведення досліду без повторів унеможливлює його оцінювання статистичними методами. Це припустимо лише в разі проведення попередніх, рекогносцирувальних і демонстраційних дослідів.

Повторністю досліду на території називають кількість одноіменних ділянок кожного варіанта, а повторністю в часі — число років проведення відповідного досліду.

Територіальна повторність дає можливість більш повно охопити строкатість земельної ділянки й отримати більш точні середні значення. Повторність у часі дозволяє встановити дію досліджуваного фактора за різних метеорологічних умов.

У разі збільшення повторності знижується помилка досліду (пропорційно кореню квадратному з числа повторень). Практика польових досліджень свідчить про те, що найбільш оптимальним є проведення польового досліду в чотири-шестиразовій повторності. Подальше збільшення числа повторень є недоцільним, оскільки при цьому зростає загальний обсяг роботи, але помилка знижується незначною мірою. Більша кількість повторень застосовується лише для доведення слабких ефектів впливів різних факторів.

Збільшення кількості повторних ділянок сильніше зменшує помилку досліду, ніж збільшення кожної ділянки за незмінної кількості повторень. Це особливо слід ураховувати під час проведення досліду на пришкільній ділянці, де площа під дослід завжди обмежена.

Результати польового експерименту значною мірою залежать від метеорологічних умов конкретного вегетаційного періоду (і року взагалі). У зв'язку з цим повторність у часі для короткострокового досліду необхідна для отримання достатньо об'єктивної характеристики досліджуваного прийому впродовж окремих років — сухих, нормальних, вологих тощо. Конкретна кількість повторень у часі залежить від завдань досліджень і від того, якими будуть метеорологічні умови. У кожному разі під час планування дослідів слід ураховувати, що навіть за сприятливого збігу обставин не можна розраховувати на отримання хоч трохи достовірної інформації менше, ніж за два-три роки.

Найчастіше польові досліди розміщують на площі методом так званих *організованих повторів*.

Організованим повтором польового досліду називають частину площі дослідної ділянки, що включає по одній ділянці (повторності) кожного варіанта досліду.

Частина земельної ділянки, що відводиться під повтор і містить усі варіанти досліду, має бути досить однорідною. Водночас між окремими повторами допускаються досить великі відмінності.

Розміщення повторів на дослідній ділянці може бути суцільним і розкиданим. За суцільного розміщення повтори розташовуються компактно й мають загальні межі. При цьому залежно від конфігурації ділянки повтори можуть розміщатися в один, два й більше ярусів. У кожному ярусі має бути ціле число повторів. За розкиданого розміщення повторів вони по одному або по кілька розташовуються в різних частинах дослідної ділянки або навіть на різних ділянках і не мають загальних меж. До такого методу розміщення повторів удаються зазвичай у тих випадках, коли немає достатньої за розміром земельної ділянки, на якій можна було б розмістити всі повтори в безпосередній близькості один від одного.

Розмір і форма дослідної ділянки

Польовий дослід ставлять на ділянках, що мають певну форму й розмір. Як відзначає Б. О. Доспєхов, у дослідній справі питанню розміру ділянки часто надається набагато більше значення, ніж воно на те заслуговує. Раніше спостерігалося захоплення великими ділянками (до 1 га й більше). Однак у всіх країнах практика дослідної роботи показала, що великі ділянки нічого не дають, крім збільшення витрат і зниження точності досліджень.

Загалом, розмір ділянки має забезпечувати таку кількість рослин, що компенсує індивідуальні генетичні відмінності між окремими рослинами. Тому чим більша за розміром рослина, яку висівають, тим більшою має бути площа ділянки, щоб її розмір дозволив помістити потрібну кількість рослин. Наприклад, для зернових культур уважається, що для виключення впливу мінливості окремих рослин на точність досліду на одній ділянці має бути як мінімум 80–100 рослин. За даними низки досліджень, часто буває достатньо й 40–50 рослин.

Визначаючи розмір ділянки, слід ураховувати також особливості агротехніки рослин і ступінь механізації польових робіт.

У практиці дослідної справи найбільш широко використовуються ділянки площею від 50 до 200 м^2. На початкових етапах дослідної роботи часто використовують ділянки площею від 10 до 50 м^2, однак чимало дослідів з успіхом проводять і на ділянках площею 0,5–2 м^2 (іноді польові досліди на ділянках такого розміру називають мікроділянковими дослідами). У будь-якому разі, навіть під час проведення виробничих дослідів, немає об'єктивних підстав до значного збільшення площі ділянок.

Форму й напрямок дослідних ділянок вибирають за результатами вивчення дослідної ділянки. Зазвичай витягнута форма ділянки гарантує меншу помилку досліду, оскільки більш повно охоплює строкатість земельного покриву.

Вірогідність досліду багато в чому залежить від орієнтації ділянок на дослідній ділянці. Довгі сторони ділянок потрібно розташовувати в тому напрямку, у якому найбільше змінюється родючість ґрунту, оскільки всі варіанти досліду слід поставити в однакові умови. За наявності полезахисних смуг ділянки мають у своєму розпорядженні довгу сторону перпендикулярно до них. Під час закладки досліду на однаковій за родючістю ділянці орієнтація ділянок не чинить впливу на помилку досліду.

По краях дослідних ділянок виділяють захисні смуги. Розрізняють бічні й окраїнні захисні смуги. Бічні захисні смуги виділяють уздовж довгих сторін ділянок для виключення впливу рослин сусідніх варіантів. Ширина бічної захисної смуги змінюється від 0,5 до 1,5 м. Іноді бічну захисну смугу просто просапують, залишаючи прохід між ділянками.

Окраїнні захисні смуги завширшки не менш ніж 2 м виділяють для захисту облікової частини ділянки від випадкових ушкоджень. Крім цього, часто всю дослідну ділянку або її частину з однойменними культурами засівають по периметру захисною смугою.

Методи розміщення варіантів у польовому досліді

Ділянки різних варіантів і повторів на дослідній ділянці можна розташовувати різними методами. Виокремлюють три основні групи методів розміщення варіантів на дослідній ділянці: стандартні, систематичні й рендомізовані, або випадкові (від англ. *random* — «випадковий», «обраний навмання»).

Стандартні методи характеризуються частим, зазвичай через один-два досліджувані варіанти, розміщенням контролю або стандарту. За систематичних методів розташування варіантів усередині кожного повтору залишається тим самим. У випадку рендомізованих методів порядок розташування варіантів визначається випадково.

Невипадкові методи розміщення варіантів

За стандартного методу один або два варіанти досліду чергуються з контролем або стандартом (*рис. 1а*). Отже, кожну ділянку досліджуваного варіанта порівнюють зі своїм контролем. Таке розміщення ґрунтується на ідеї, що родючість дослідної ділянки змінюється поступово й між урожаями сусідніх ділянок існує кореляція. Стандартний метод іноді здається дуже простим і надійним для зведення до мінімуму помилок експерименту. Уявляється, що контроль, розташований біля кожного досліджуваного варіанта, дасть найбільш точну оцінку ефективності варіанта. Однак практика застосування й порівняльного оцінювання стандартних методів виявила їхні істотні недоліки, головними з яких є такі:

1. Не завжди спостерігається тісна кореляційна залежність між урожаями поряд розташованих ділянок.
2. Дослідні варіанти, розташовані на значній відстані одна від одної, дуже складно порівнювати.
3. Стандартний метод дуже громіздкий, що пояснюється невиправдано великою кількістю контрольних ділянок.
4. Ненадійність статистичних оцінок під час опрацювання даних експерименту (точніше кажучи, за стандартного розміщення варіантів узагалі не існує жодних статистичних методів оцінювання результатів експерименту).

I						II						III						IV						
A St	B	C	A St	D	E	A St	B	C	A St	D	E	A St	B	C	A St	D	E	A St	B	C	A St	D	E	A St

а

I					II					III					IV				
A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E

б

I					II					III					IV				
C	A	D	B	E	B	D	A	E	C	D	B	E	C	A	A	C	E	B	D

в

Рис. 1. Методи розміщення п'яти варіантів (А–Е) на ділянках чотирьох повторів польового досліду: а — стандартний; б — систематичний; в — рендомізований. А — варіант, обраний як стандарт за стандартного методу, що для наочності позначено літерами St

За систематичного методу дослідні варіанти розташовують на ділянках усередині повторів у певній послідовності. Зазвичай у разі розміщення повторів в один ярус варіанти досліду розташовують однаково в кожному повторі. Наприклад, якщо в першому повторі варіанти розташовувалися в порядку A, B, C, D і E, то такий самий порядок зберігається й у решті повторів (*рис. 1б*). У разі розташування повторів у кілька ярусів зазвичай застосовують шаховий порядок, коли послідовність варіантів у повторах різних ярусів зсувається.

Упродовж тривалого часу в колишньому СРСР систематичне розташування варіантів було переважним. Але, незважаючи на гадану простоту, систематичний метод має недоліки й часто може призводити до дуже грубих помилок. Недоліки систематичного методу розміщення зумовлені двома причинами: непередбачуваним викривленням ефектів варіантів і відсутністю відповідних статистичних методів для оцінювання результатів досліджень.

Методи розміщення варіантів з допомогою випадкового вибору

У більшості випадків немає жодних причин використовувати стандартний або систематичний методи розміщення варіантів на ділянках дослідної площі. Завжди слід дотримуватися фундаментального принципу однаковості ймовірності будь-якого варіанта потрапити на будь-яку ділянку дослідної площі, тобто вдаватися до випадкового розміщення. Лише в такому випадку висновки, що робляться на підставі результатів досліду, будуть обґрунтованими, оскільки методи статистичного аналізу базуються саме на принципі випадкового відбору. Цього принципу дотримуються під час використання найпоширенішого у світовій практиці методу рендомізованих повторів (*рис. 1в*). Такий метод дозволяє зменшити помилку досліду в разі використання дисперсійного аналізу для статистичного опрацювання результатів польового досліду, оскільки дає можливість зменшити випадкове варіювання результатів на величину, зумовлену впливом повторів. У кожному повторі варіанти розподіляють по ділянках випадково.

У найпростішому випадку ділянки всіх варіантів і повторень розміщають цілком випадково, не об’єднуючи територіально повторення варіантів у компактні групи (повтори). Таке розміщення називають методом *неорганізованих повторів,* або *цілковитою рендомізацією* (*рис. 2*). Такий метод використовується лише в тих нечастих випадках, коли немає необхідності контролювати можливе закономірне варіювання ґрунтових умов експерименту, а польові досліди закладаються на добре вирівняних земельних ділянках.

Крім того, він може бути ефективним, якщо в досліді вивчається невелика кількість варіантів (два-чотири) і є підстави не контролювати можливе закономірне варіювання родючості ґрунту дослідної ділянки.

A	B	C	A	D
A	C	A	C	C
D	D	B	A	D
B	B	C	D	B

Рис. 2. Приклад цілком рендомізованого розміщення чотирьох варіантів (A–D) польового досліду, кожен з яких представлений у п'ятьох повтореннях

Однак найчастіше польові досліди розташовують на площі методом організованих повторів, про що було сказано вище. У кожному повторі варіанти розподіляють на ділянках випадково.

Випадковий метод розміщення варіантів часто називають також методом рендомізованих блоків, розуміючи під словом «блок» повтори.

Теоретичні засади необхідності рендомізації в польових сільськогосподарських дослідженнях було розроблено ще у 20–30-х рр. минулого століття найвидатнішим американським фахівцем Р. Фішером і надалі розвинені Д. Снедекором та ін.

Різновидом рендомізованого розміщення варіантів є методи латинського квадрата й латинського прямокутника. Використання латинського квадрата дозволяє значною мірою усунути з помилки досліду систематичну зміну родючості ґрунту у двох взаємно перпендикулярних напрямках.

Розташування досліду латинським квадратом передбачає, що число повторів має дорівнювати числу варіантів. На площі ділянки розташовуються ряди й стовпці, причому в кожному ряді й стовпці має бути повний набір усіх варіантів і при цьому жоден варіант не повинен повторюватися ні в ряді, ні у стовпці. Ділянки мають бути квадратної або майже квадратної форми (*рис. 3*). За видовженої форми ділянок латинський квадрат не має жодних переваг перед звичайним методом рендомізованих блоків.

У разі числа варіантів понад вісім замість латинського квадрата використовують латинський прямокутник, однак цей метод, як і метод розщеплених ділянок, вимагає досить глибокого розуміння методів статистичного опрацювання результатів такого експерименту і в пропонованому посібнику не розглядатиметься.

E	D	B	A	C
B	E	D	C	A
C	A	E	B	D
D	C	A	E	B
A	B	C	D	E

Рис. 3. Приклад розміщення п'яти варіантів досліду (A–E) методом латинського квадрата

Польові роботи на дослідній ділянці

Найважливішим правилом догляду за дослідом є одночасність виконання агротехнічних робіт на всій ділянці. Якщо з якихось причин це неможливо зробити одночасно, то впродовж одного дня необхідно завершити роботи на цілому числі повторів. Цю вимогу необхідно суворо виконувати. Її порушення, особливо багаторазове, призводить до невірогідності досліду по суті.

Навіть незначний розрив, наприклад, у термінах обробки або внесення добрив лише на 6–8 год (особливо якщо за цей час пройде дощ), часто призводить до істотних відмінностей у рості й розвитку рослин.

Друга загальна вимога — висока якість усіх виконуваних робіт. Агротехнічне тло на дослідній ділянці має бути оптимальним для прояву ефекту досліду.

Характер польових робіт, виконаних на дослідній ділянці, залежить від вирощуваної культури, завдань дослідження тощо. Обробка ґрунту, якщо вона сама не є досліджуваним чинником, має бути однаковою, одночасною та високоякісною на всіх ділянках досліду. Посів рослин на дослідній ділянці, як правило, слід проводити в один день. У всіх дослідах норми висіву бажано встановлювати за числом схожих насінин, а не за вагою.

Догляд за рослинами на дослідному полі не має відрізнятися від догляду за відповідними культурами у виробничих умовах. Прополку, міжрядну обробку, підгодівлю та інші заходи проводять однаково на всіх ділянках досліду й не розтягують у часі. До спеціальних робіт з догляду за дослідом зараховують створення і прочищення стежок, обрізку по шнуру кінців ділянок, а також створення захисних смуг. Сюди ж належить своєчасне розміщення кілочків, етикеток тощо.

На всій території дослідної ділянки підтримують чистоту, на ній не має залишатися виполотих рослин, решток бадилля, соломи та іншого сміття.

Ведення документації з досліду

Дуже часто веденню документації з досліду не надається особливого значення, особливо в разі проведення НДР школярами. Однак правильне ведення документації та звітності за дослідом є обов'язковим елементом експериментальної роботи. Об'єктивний аналіз і пояснення результатів досліджень можливі лише за умови обліку та своєчасного фіксування всіх виконуваних робіт, супровідних факторів і результатів спостережень і аналізів. Якість документації з досліду часто відбиває акуратність і якісність його проведення.

Документація з польового досліду має бути повною за змістом, об'єктивною, своєчасною й достовірною. Записи мають бути однотипними.

Обов'язковими документами польового досліду є:

1. Робочий план (програма).
2. Первинна поточна документація (щоденник польових робіт).
3. Допоміжні документи (робочі зошити або журнали).
4. Зведені документи (журнал польового досліду).
5. Звіт про проведення польового досліду.

Робочий план (програма) досліду включає: схему досліду; загальні умови проведення досліду (ґрунт, агротехніка та ін.); параметри польового досліду (площа ділянок, число повторень тощо); перелік і методику проведення обліків, спостережень і аналізів, необхідні для проведення досліду засоби, матеріали й устаткування; очікувані результати.

Найважливішою складовою робочої програми є календарний план з переліком усіх видів робіт, обліків, спостережень і аналізів із зазначенням їхніх обсягів і термінів виконання.

Щоденник польових робіт має бути належним чином оформлений. На перших сторінках указується тема експерименту, місце його проведення, прізвища виконавців, час проведення досліджень. Далі описується схема досліду, наводиться креслення з конкретним планом розміщення варіантів на дослідній ділянці.

До щоденника записуються в хронологічному порядку всі агротехнічні роботи, дані обліків і спостережень, погодні фактори, що супроводжують обліки, спостереження й агротехнічні роботи. Записи ведуть безпосередньо в полі або лабораторії під час виконання роботи або відразу ж після її завершення. Підчищення не допускаються, виправлення мають бути пояснені.

У допоміжних документах, починаючи з моменту вибору земельної ділянки під дослід і закінчуючи збиранням урожаю, слід робити докладні записи, що стосуються характеристики ґрунту, способів його обробки, добрив, підготовки насіння до посіву, догляду за посівами. Необхідно систематично реєструвати фенологічні спостереження за ростом і розвитком рослин. У документації слід відбивати відомості про всі явища, які можуть вплинути на точність досліду. Особливо необхідно звертати увагу на фактори погоди під час обприскування рослин різними препаратами для боротьби зі шкідниками та хворобами. Усі записи робляться за спеціальною формою або довільно, однак доцільно вести записи за певною схемою.

Журнал польового досліду є зведеним документом, що містить усі необхідні дані для подальших узагальнень і висновків. У журналі зосереджений весь головний матеріал з польового досліду (текст, таблиці, графіки), на підставі якого можна скласти звіт.

До журналу польового досліду обов'язково заносять такі відомості:

- назву теми досліду;
- терміни й місце проведення;
- прізвища виконавців і керівника;
- мету й завдання досліду;
- схему і план розміщення досліду;
- характеристику та історію дослідної ділянки;
- дані про особливості ґрунту й агротехніки;
- програму й методику досліджень;
- перелік усіх проведених робіт із зазначенням термінів і умов виконання;
- опрацьовані результати обліків, спостережень і аналізів у вигляді таблиць, графіків, діаграм;
- опрацьовані результати обліку врожаю в перерахуванні, за потреби, на 1 га;
- результати статистичного опрацювання даних.

Виправлення в журналі польового досліду неприпустимі, так само, як і записи олівцем.

Звіт про проведення польового досліду є завершальним етапом експериментальної роботи. До головних вимог до складання звіту належать чіткість побудови, логічність викладу матеріалу, переконливість аргументації, стислість і точність формулювань, що виключає можливість суб'єктивного й неоднозначного тлумачення, вірогідність і конкретність викладу результатів, доказовість висновків і обґрунтованість рекомендацій.

Звіт обов'язково має вичерпно повно відбивати зміст і результати проведеної роботи й мати доступну форму викладу. Завдяки цьому фахівець будь-якої категорії зможе отримати з нього потрібну інформацію.

Звіт зазвичай містить такі розділи:

- назву, виконавців і керівника теми;
- мету й завдання дослідження;
- стислу історію питання (огляд наукової літератури);
- схему, методику й умови проведення експерименту;
- результати досліджень та їх обговорення;
- висновки;
- рекомендації;
- перелік використаної літератури.

Наукові звіти підписуються виконавцем і керівником.

Розділ 5. ВСТУП ДО СТАТИСТИЧНОГО АНАЛІЗУ

Якщо нас цікавлять стабільні ознаки, величина яких є константою в усіх об'єктів певного виду, то жодне статистичне опрацювання результатів досліджень не потрібне. Наприклад, давно відомо, що в кожної мухи в нормі шість ніг, у кожної людини — дві руки, а в слона є довгий хобот. Статистичне опрацювання стає необхідним, коли ми аналізуємо ознаки, величини яких є змінними випадковими величинами. Тут для обґрунтованого висновку нам необхідно знати ступінь упевненості в спостережуваних відмінностях. Наприклад, нас може цікавити, чи відрізняється вага окремих особин деякого виду тварин у віддалених популяціях. Якщо ми в кожній популяції зважимо по одному екземпляру, чи можемо ми з упевненістю екстраполювати спостережувану різницю на популяції в цілому? А раптом у тій популяції, де тварини в середньому важчі, ніж в іншій, нам випадково трапився унікально легкий екземпляр? Зрозуміло, що потрібно зважити хоча б по кілька екземплярів у кожній з популяцій. А скільки? Яку різницю ми з упевненістю можемо вважати реальною, а яку — випадковою? Для отримання обґрунтованих відповідей на подібні запитання нам і потрібен статистичний аналіз даних (яким, на превеликий жаль, дуже часто нехтують під час виконання НДР школярами).

Нижче пропонується елементарний вступ до простих методів статистичного аналізу. Оскільки важливо *розуміти* суть цих методів, ми будемо детально розглядати як логіку, так і послідовність обчислень. Однак у практичній роботі настійливо рекомендується використовувати комп'ютерні програми для відповідних розрахунків, оскільки обчислення «у ручну», навіть з використанням мікрокалькулятора, можуть породжувати випадкові помилки. Така поширена комп'ютерна програма, як Microsoft Excel, має потужний пакет аналізу, що є надбудовою, яку варто підключити під час установки програми. Крім того, є спеціалізовані програмні продукти, наприклад STATISTICA, які дозволяють здійснювати будь-який вид статистичного аналізу — від найпростішого до найскладнішого.

У разі використання комп'ютерних програм достатньо не помилитися, уводячи дані. Якщо ж проводити довгі ланцюжки обчислень з багатьма проміжними результатами без комп'ютера, то існує значна небезпека припуститися помилок, які важко виявити.

Поняття про генеральну сукупність і вибірку

Коли в біології ми говоримо про середню величину окремої ознаки в певного виду живих організмів (наприклад, часто говорять про середній зріст людей, середню врожайність пшениці, середні надої молока та ін.), то під цим розуміють середню арифметичну величину цієї ознаки в усіх можливих організмів цього виду. Сукупність усіх організмів (або, у більш загальному випадку, усіх об'єктів певного виду), які існують або *могли б* існувати за певних умов, становить так звану *генеральну сукупність*. Інакше кажучи, генеральну сукупність можна визначити як фактичну або гіпотетичну (передбачувану) безліч *усіх* об'єктів, що цікавлять нас.

Уявімо собі, що на планету Земля прилетіли прибульці, які раніше тут не бували й про землян не мали жодного уявлення. Для того щоб таке уявлення мати, вони повинні одержати інформацію про ознаки, характерні для людей (уважатимемо, що прибульці налаштовані досить дружньо й вивчають людей лише з їхньої згоди). Для таких прибульців генеральною сукупністю будуть усі більш ніж 6 млрд людей, які живуть на Землі.

Якщо ознака, яка зацікавить прибульців, не варіює (тобто є константою), то один-два представники людства цілком і вичерпно охарактеризують усю генеральну сукупність усієї безлічі людей, що мешкають на нашій планеті. Наприклад, побачивши *одну* людину, прибульці зроблять цілком правильний висновок, що в *усіх* людей є по дві ноги, дві руки, два ока, дві вуха, один рот і т. д. (уважатимемо, що каліки прибульцям не трапляться).

Складнішими є справи з ознаками, які змінюють свою величину в разі переходу від однієї людини до іншої. Хоча величина такої ознаки зазвичай варіює лише в певних межах (наприклад, доросла людина не може мати зріст 20 см і менше або 3,5 м і більше), однак за одним або двома представниками складно отримати досить повне уявлення про середню величину ознаки в генеральній сукупності. Якщо прибульці захочуть з'ясувати середній зріст людей, то, вимірявши його величину в однієї або двох осіб, вони хоч трохи чіткого уявлення про цю ознаку не матимуть. Якщо зріст конкретної людини виявиться таким, що дорівнює 150 см, то звідси не випливатиме, що в *усіх* дорослих людей *у середньому* саме такий зріст.

Звичайно, прибульці дізналися б середній зріст цілком точно, якби виміряли його в усіх землян, однак навіть для просунутих у технічному плані прибульців таке завдання навряд чи виявиться під силу. У подібну ситуацію потрапляють не лише прибульці, але й учені та інші фахівці у своїй науковій і практичній діяльності, а також школярі під час проведення дослідів з біології.

Наприклад, у контролі якості продукції нас може цікавити частка бракованих виробів у партії товарів або їхня стійкість до ударів. Звичайно, ми могли б перевірити кожен виріб кожної випущеної партії, однак це може виявитися занадто складним технічно. Крім того, якщо ми, наприклад, кожну випущену лампочку перевірятимемо на максимальну тривалість горіння, кожен вироблений сірник — на легкість запалення, кожен складений комп'ютер — на стійкість до ударів, а кожне зерно пшениці у великому сховищі — на схожість, то після такої перевірки ми матимемо точну інформацію щодо якості партій товару, які нас цікавлять, однак самі ці партії як такі будуть цілком знищені в процесі перевірки.

Досить часто ми не лише практично, але й теоретично не можемо виміряти ознаку, яка цікавить нас, у кожного об'єкта дуже великої або нескінченно великої безлічі (як можна щось виміряти в нескінченно великій кількості об'єктів?).

Отже, підіб'ємо підсумок сказаного. Існує деяка велика або дуже велика генеральна сукупність об'єктів (наприклад, землян). У біології таку сукупність ми можемо найчастіше вважати нескінченною. У кожного об'єкта цієї сукупності є певна ознака, що варіює та є цікавою для нас (наприклад, зріст). Є також деяка середня величина цієї ознаки для всієї сукупності. У статистиці її називають *генеральним середнім арифметичним* і позначають літерою μ (велика грецька літера «мю»). Ми точно знатимемо, чому саме дорівнює генеральне середнє, якщо виміряємо цю ознаку в абсолютно всіх об'єктів (наприклад, зріст усіх людей Землі) і обчислимо середнє арифметичне. Однак зробити це з очевидних причин ми найчастіше не можемо (хто не згодний — виміряйте зріст кожного жителя нашої планети). Тож ця цілком конкретна середня величина зазвичай залишається не відомою нам. Водночас ми все-таки хочемо знати, чому вона може дорівнювати, хоча б приблизно. Як же нам учинити? Для того щоб приблизно *оцінити* величину μ, ми з нашої сукупності відбираємо певну кількість об'єктів, тобто робимо *вибірку*, у кожного члена вибірки визначаємо величину ознаки, що цікавить нас, і обчислюємо середнє арифметичне для об'єктів, які потрапили до вибірки. Це середнє арифметичне називається *вибірковою* та є найкращою *оцінкою*, яку ми можемо отримати для не відомої нам величини генерального середнього. Вибіркове середнє арифметичне найчастіше позначають як $\bar{X}$.

Число об'єктів, які потрапили до вибірки, називають *обсягами* вибірки, або *повторністю*, і позначають літерою n. Скільки ж об'єктів має бути у вибірці? Це залежить від багатьох обставин, зокрема від природи ознаки, що ми вивчаємо, від можливостей дослідника й від тієї величини, на яку ознака може відхилятися від свого середнього значення в окремих членів генеральної сукупності.

Остання обставина становить значний інтерес для аналізу отриманих даних. Зрозуміло, що чим більше можуть відхилятися значення змінної від середнього значення, тим більше вибіркове середнє може відхилятися від генерального середнього, якщо обсяг вибірки незначний.

Розглянемо, наприклад, два ряди цифр:

А: 4, 5, 5, 5, 5, 5, 5, 5, 5, 6.

Б: 0, 1, 2, 3, 4, 6, 7, 8, 9, 10.

І один, і другий ряд містять по 10 цифр і мають однакові середні арифметичні значення, що дорівнюють 5. Однак відхилення окремих значень від 5 між рядами цифр значною мірою відрізняються.

Уявімо, що кожен із цих рядів є дуже спрощеною моделлю генеральних сукупностей. Якщо з кожної сукупності ми зробимо випадкові вибірки з двох об'єктів, то для сукупності А найменше вибіркове середнє дорівнюватиме 4,5, а найбільше — 5,5. Отже, будь-яка вибірка буде порівняно непогано оцінювати генеральне середнє, що дорівнює 5 (багато вибірок із цієї сукупності, через слабке варіювання індивідуальних значень, матимуть середні, які дорівнюють 5, тобто генеральному середньому).

Водночас для сукупності Б найменше середнє у вибірці обсягом два об'єкти дорівнюватиме 0,5 (якщо до вибірки потраплять цифри 0 і 1); найбільше ж виявиться таким, що дорівнює 9,5 (якщо у вибірці опиняться цифри 9 і 10). Звичайно ж, такі значення набагато менш точно оцінюватимуть генеральне середнє, що дорівнює 5, ніж вибіркові середні із сукупності А.

Отже, мінливість окремих об'єктів сукупності відносно генерального середнього має велике значення для наших оцінок, і хотілося б мати можливість охарактеризувати її певною величиною, яка пояснювала б нам, наскільки значних відхилень ми можемо очікувати в індивідуальних членів генеральної сукупності. Як же можна виміряти цю величину?

Найпростіший підхід — обчислити різниці між найбільшим і найменшим значеннями, тобто визначити *розмах* значень. У випадку наших сукупностей А і Б він непогано відбиває суть справи, оскільки для першої сукупності дорівнює 2, а для другої — 10. Однак такий самий розмах значень матиме й наступний ряд цифр (назвемо його сукупністю В):

В: 0, 5, 5, 5, 5, 5, 5, 5, 5, 10.

У сукупності В середнє арифметичне значення, як і в сукупностях А і Б, також дорівнює 5, розмах між найменшим і найбільшим значеннями дорівнює 10, як і в сукупності Б, але все-таки варіювання окремих значень тут менше, ніж у цій сукупності, оскільки від середнього відрізняються лише два значення, а в сукупності Б — усі десять. І, наприклад, якщо із сукупності В ми складемо вибірки по два об'єкти, то найменше вибіркове середнє дорівнюватиме 2,5, а найбільше становитиме 7,5. Хоча такі середні менш точно оцінюватимуть генеральне середнє, що дорівнює 5, порівняно з вибіркою А, але оцінюватимуть її вони набагато точніше порівняно з вибіркою Б, незважаючи на однаковий розмах між мінімальним і максимальним значеннями.

Отже, розмах значень погано характеризує варіювання величини ознаки між окремими членами сукупності. Який же показник буде більш підходящим?

Якщо ми просто обчислимо відхилення кожного індивідуального значення від середнього, просумуємо їх і розділимо на число таких відхилень, то в результаті ми нічого не дістанемо (нічого — у розумінні нуль), оскільки однією з важливих властивостей середнього арифметичного є факт, що сума відхилень від неї дорівнює нулю (перевірте це для вибірок А, Б і В).

Значно кращим виявиться підхід, якщо ми просумуємо відхилення без урахування їхнього знака, тобто за їхнім модулем (абсолютною величиною). У цьому випадку ми матимемо так зване середнє абсолютне відхилення. Спробуймо обчислити його для наших модельних сукупностей.

Для А воно дорівнює:

$$\frac{\left\|4-5\right\|+\left|5-5\right|+\ldots+\left|6-5\right|}{10}=\frac{2}{10}=0{,}2.$$

Для Б воно становитиме:

$$\frac{\left|0-5\right|+\left|1-5\right|+\ldots+\left|10-5\right|}{10}=\frac{30}{10}=3.$$

Зрештою, для сукупності В ця величина дорівнюватиме:

$$\frac{\left|0-5\right|+\left|5-5\right|+\ldots+\left|10-5\right|}{10}=\frac{10}{10}=1.$$

Отже, логічні міркування показали нам, що варіювання цифр у сукупності А є найменшим, у сукупності В воно більше, а найбільшим воно є в сукупності Б. Середні абсолютні відхилення дорівнюють 0,2; 3 і 1 відповідно, тобто непогано характеризують те, що нам було потрібно. Проте, хоча середнє абсолютне відхилення є природним і зрозумілим показником і хоча воно добре характеризує величину варіювання ознаки серед членів генеральної сукупності, у біологічних дослідженнях воно не має застосування. Причини такої ситуації криються в складних питаннях, пов'язаних з теорією математичної статистики.

Оскільки обговорення цих питань виходить далеко за межі цієї книги, то залишимо кесареві кесареве, а фахівцям у математичній статистиці — теоретичне обґрунтування порушеного питання. Для нас воно не таке й важливе. Важливо інше — у біологічній статистиці ступінь варіювання ознаки оцінюють трохи по-іншому, а саме: обчислюють усі відхилення, підносять кожне з них у квадрат (і в такий спосіб позбуваються негативних величин), отримані квадрати складають, ділять на число відхилень і добувають квадратний корінь із отриманої цифри. Обчислене в такий спосіб усереднене відхилення називають *стандартним відхиленням*.

Спробуймо обчислити його для наших сукупностей А, Б і В.

Для А:

$$\sqrt{\frac{(4-5)^2+(5-5)+\ldots+(6-5)^2}{10}}=\sqrt{\frac{2}{10}}=\sqrt{0{,}2}\approx 0{,}45.$$

Для Б:

$$\sqrt{\frac{(0-5)^2+(1-5)+\ldots+(10-5)^2}{10}}=\sqrt{\frac{110}{10}}=\sqrt{11}\approx 3{,}32.$$

Для В:

$$\sqrt{\frac{(0-5)^2+(5-5)+\ldots+(10-5)^2}{10}}=\sqrt{\frac{50}{10}}=\sqrt{5}\approx 2{,}24.$$

Стандартні відхилення дозволили нам ранжувати наші сукупності за варіюванням вихідних ознак у тому самому порядку, що й середні абсолютні відхилення. Стандартне відхилення генеральної сукупності позначається літерою σ (грецька велика літера «сигма»). У статистиці самостійне значення відіграє також квадрат стандартного відхилення, що називається *дисперсією*. Отже, можна сказати, що дисперсія є середнім квадратом відхилення, а стандартне відхилення дорівнює кореню квадратному з дисперсії. Генеральне середнє арифметичне μ, генеральне стандартне відхилення σ і генеральну дисперсію σ^2 зазвичай називають параметрами генеральної сукупності, або генеральними параметрами.

Як вибіркове середнє арифметичне є оцінкою генерального середнього, так і вибіркове стандартне відхилення, позначуване літерою s, є оцінкою генерального вибіркового відхилення, а вибіркова дисперсія s^2 є оцінкою генеральної дисперсії.

Отже, знову підіб'ємо проміжний підсумок. Під час проведення біологічних наукових досліджень ми маємо генеральну сукупність об'єктів, у яких є параметр, що цікавить нас (наприклад, зріст людей, урожай ячменя, активність певного ферменту в лабораторної тварини або кількість плодів томату на одній рослині). Цей параметр у генеральній сукупності має середню величину μ і варіювання, що характеризується стандартним відхиленням σ. Найчастіше ні генеральне середнє, ні генеральне стандартне відхилення нам не відомі, і для того щоб оцінити їх, ми робимо вибірку з генеральної сукупності. До вибірки ми відбираємо n об'єктів, у яких вимірюємо параметр, що цікавить нас, і за вибіркою оцінюємо не відомі нам генеральні величини. Значення параметра, що цікавить нас, в окремих об'єктів *зручно* позначати як X_i, де X — чисельне значення параметра, а індекс i показує, у якого за ліком об'єкта в нашій вибірці параметр має величину x. Зрозуміло, що i може змінюватися від 1 для першого об'єкта до n в останнього члена нашої вибірки.

Вибіркове середнє є найзвичайнісіньким середнім арифметичним і обчислюється за очевидною формулою:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}. \qquad (1)$$

Слід нагадати, що позначення $\sum_{i=1}^{n} X_i$ показує, що ми підсумовуємо елементи X, при цьому кожен конкретний елемент ми позначаємо індексом i, що визначається порядковим номером елемента в сукупності. Коли пишемо X_i, то для першого елемента $i=1$, для другого $i=2$ і т. д. Під сигмою ми вказуємо, що підсумовування починається з першого елемента $(i=1)$. Для підсумовування домовилися, що після додавань до суми чергового значення величина індексу i зростає на одиницю, і так доти, поки не буде додано останній член сукупності, для якого $i=n$. Якщо потрібно просумувати лише якусь частину певного масиву величин, то над сигмою ставлять не n, а конкретну цифру, що дорівнює номеру члена сукупності, після додавання якого процедуру підсумовування слід завершити.

Дуже часто, коли підсумовуються всі елементи від першого до останнього, вираз n над та $i=1$ під знаком підсумовування (літерою сигма) опускають, і підсумовування позначають так: $\sum X_i$. Надалі подібне позначення буде використано й у цій книзі.

Для обчислення вибіркової дисперсії й вибіркового стандартного відхилення використовують таку формулу:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum\left(X_i - \bar{X}\right)}{n-1}}. \qquad (2)$$

Як бачимо з формули (2), для обчислення вибіркового стандартного відхилення ми від кожного індивідуального значення величини нашого параметра віднімаємо середнє значення, різницю підносимо до квадрата, підсумовуємо, ділимо на число об'єктів у вибірці та віднімаємо одиницю, а з отриманої величини вибіркової дисперсії добуваємо квадратний корінь.

Може викликати подив та обставина, що суму квадратів відхилень від вибіркового середнього ми ділимо не на число таких відхилень (тобто n), а на величину $(n-1)$. Величину $(n-1)$ зазвичай називають числом ступенів свободи, і для отримання вибіркової оцінки генерального стандартного відхилення суму квадратів відхилень краще ділити саме на неї, а не на число відхилень. За цієї умови ми дістаємо більш точну (незміщену) оцінку генерального стандартного відхилення σ. Якщо ж ми ділитимемо не на число ступенів свободи $(n-1)$, а на число відхилень від вибіркового середнього, що дорівнює обсягу вибірки (n), то дістанемо трохи занижену оцінку величини σ.

Незважаючи на можливе здивування, ці факти мають надійне теоретичне обґрунтування, і в усіх випадках для обчислення вибіркового стандартного відхилення потрібно використовувати число ступенів свободи, а не обсяг вибірки. Якщо трохи спростити ситуацію, то необхідність цього можна пояснити так. Якби нам була відома величина генерального середнього μ, то для обчислення оцінки генерального стандартного відхилення слід було б віднімати цю величину від кожного значення змінної в нашій вибірці та складати квадрати відхилень саме від генерального середнього. У такому випадку суму квадратів ми могли б з цілковитою підставою ділити на число таких відхилень n для отримання найбільш точної оцінки величини σ.

Однак найчастіше величина μ нам не відома, і ми користуємося її вибірковою оцінкою $\bar{X}$. Однак практично ніколи вибіркове середнє не дорівнює генеральному середньому (таке може трапитися лише випадково й дуже нечасто). Водночас сума відхилень від середнього арифметичного (і, відповідно, сума квадратів цих відхилень) завжди менша, ніж від будь-якого іншого числа. Отже, для об'єктів, що потрапили до вибірки, сума квадратів відхилень від вибіркового середнього завжди менша, ніж від генерального середнього, тобто є заниженою (або, як іще говорять, зміщеною). Щоб компенсувати цей зсув, суму квадратів від вибіркового середнього ділять не на n, а на трохи менше число (число ступенів свободи) — $(n-1)$. Повторимо ще раз, що цей підхід має чітке теоретичне обґрунтування.

Замість формули (2) краще користуватися тотожною формулою, що є більш зручною для обчислень і зазвичай більш точною:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}}{n-1}}. \qquad (3)$$

Формула (3) дозволяє отримати більш точні значення з тієї причини, що під час обчислення вибіркового середнього ми зазвичай удаємося до округлення. У разі використання формули (2) похибка округлення позначається в *кожній* різниці між індивідуальним значенням змінної X_i та вибірковим середнім $\bar{X}$, і якщо обсяг вибірки досить великий, то в сумі ці похибки можуть призвести до відносно великої неточності. Водночас у разі виконання обчислень за формулою (3) округляти доводиться не більш ніж двічі, тому вона й дає зазвичай більш точну оцінку.

Отже, ми навчилися обчислювати вибіркові оцінки генерального середнього й генерального стандартного відхилень.

Статистичне оцінювання

Вибіркові оцінки

Одним з питань статистичного оцінювання результатів дослідження є аналіз отриманих даних щодо того, як саме вони *оцінюють* генеральні параметри сукупності, з якої здійснено вибірку. Для розгляду конкретних даних повернімося до прибульців та уявімо собі, що їх зацікавив зріст чоловіків Євразії й Америки. За такої постановки питання є дві генеральні сукупності (перша — це дорослі чоловіки Євразії, друга — дорослі чоловіки Америки). Ці сукупності мають конкретні генеральні середні значення зросту, відповідно $\mu_{Є}$ і $\mu_{А}$, величини яких прибульцям не відомі. Крім того, ці самі сукупності характеризуються певною мінливістю величини зросту в окремих чоловіків, і ця мінливість характеризується генеральними стандартними відхиленнями $\sigma_{Є}$ та $\sigma_{А}$ для чоловіків Євразії й Америки відповідно. Значення генеральних стандартних відхилень також не відомі.

Для того щоб оцінити генеральні параметри, прибульці зробили вибірки по 10 осіб на кожному континенті й отримали такі результати вимірювання зросту в сантиметрах (це — УМОВНИЙ приклад).

Для чоловіків Євразії: 156, 156, 164, 167, 169, 171, 179, 181, 185, 192.

Для чоловіків Америки: 161, 164, 167, 169, 173, 180, 184, 189, 189, 194.

Які оцінки вони можуть отримати з цих даних? Допоможімо їм. Насамперед, звичайно, потрібно обчислити середні арифметичні.

Для Євразії вона становитиме:

$$\bar{X}_{\text{Є}} = \frac{156+156+\ldots+192}{10} = \frac{1720}{10} = 172.$$

Для Америки вона дорівнюватиме:

$$\bar{X}_{\text{А}} = \frac{161+164+\ldots+194}{10} = \frac{1770}{10} = 177.$$

Чи можна стверджувати на підставі цих даних, що середній зріст *усіх* чоловіків Євразії дорівнює 172 см, а Америки — 177 см? Логіка підказує, що за такою невеликою вибіркою, як 10 осіб, важко очікувати на дуже точну оцінку генерального середнього. Швидше за все, генеральні середні й у першому, і в другому випадках відрізняються від їхніх вибіркових оцінок. Однак наскільки та в який бік? На це питання, проводячи вибіркові дослідження, ми *ніколи* не знайдемо відповіді. Однак ми можемо дізнатися, чому *може дорівнювати* генеральне середнє. Для цього потрібно спочатку отримати вибіркові оцінки дисперсій і стандартних відхилень. Скористаймося формулою (3) і обчислімо ці вибіркові оцінки для обох вибірок (хоча практично завжди нам доведеться вдаватися до округлення, однак ми не користуватимемося знаком наближеної рівності — для вибіркових *оцінок* це безглуздо).

Для чоловіків Євразії:

$$s_{\text{Є}}^2 = \frac{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}}{n-1} = \frac{\left(156^2+156^2+\ldots+192^2\right) - \frac{(1720)^2}{10}}{9} =$$

$$= \frac{297150-295840}{9} = \frac{1310}{9} = 145{,}5556;$$

$$s_{\text{Є}} = \sqrt{s_{\text{Є}}^2} = \sqrt{145{,}5556} = 12{,}0646.$$

Для чоловіків Америки:

$$s_{\text{А}}^2 = \frac{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}}{n-1} = \frac{\left(161^2 + 164^2 + \ldots + 194^2\right) - \frac{(1770)^2}{10}}{9} =$$

$$= \frac{314530 - 313290}{9} = \frac{1240}{9} = 137{,}7778;$$

$$s_{\text{А}} = \sqrt{s_{\text{А}}^2} = \sqrt{137{,}7778} = 11{,}7379.$$

Урешті-решт, для обчислення можливих значень генерального середнього арифметичного нам знадобиться ще один показник, що залежить від стандартного відхилення й розміру вибірки. Цей показник називається стандартною помилкою, і він показує *точність* оцінки генерального середнього. Обчислюється стандартна помилка шляхом поділу стандартного відхилення на квадратний корінь із числа об'єктів у вибірці:

$$s_{\bar{X}} = \frac{s}{\sqrt{n}}. \qquad (4)$$

Стандартна помилка збільшується внаслідок зростання варіювання вихідних даних (тобто стандартного відхилення s) і зменшується в разі збільшення обсягу вибірки n. Загалом кажучи, це оцінка об'єктивного наявного показника, що характеризує саме точність оцінки генерального середнього. Не можна вважати, що вона є наслідком якихось помилок під час проведення досліджень (тут в оману може вводити власне слово «помилка», однак цей термін є загальноприйнятим). З цієї причини неправильною буде точка зору, що чим менша ця величина, тим краще. Як об'єктивно існує генеральне стандартне відхилення (σ), що має певну конкретну величину, незалежно від того, подобається нам це чи ні, так і існує не відоме нам відношення $\frac{\sigma}{\sqrt{n}}$ для вибірок обсягом n, і бажано, щоб наша вибірка давала якнайбільш точну його оцінку. Якщо це відношення є досить великим, то й наша вибіркова стандартна помилка теж має бути досить великою.

Для вибірок чоловіків Євразії й Америки стандартні помилки дорівнюватимуть:

для Євразії: $s_{\bar{X}_{\text{Є}}} = \frac{s_{\text{Е}}}{\sqrt{n}} = \frac{12{,}0646}{\sqrt{10}} = 3{,}81;$

для Америки: $s_{\bar{X}_{\text{А}}} = \frac{s_{\text{А}}}{\sqrt{n}} = \frac{11{,}7379}{\sqrt{10}} = 3{,}71.$

Стандартна помилка є надзвичайно корисною для обчислення так званих *довірчих інтервалів*, тобто інтервалів значень цікавої для нас змінної, які з відомою нам імовірністю включають невідому величину генерального середнього.

Для обчислення цих інтервалів ми спочатку множимо стандартну помилку $s_{\bar{X}}$ на певний коефіцієнт, а саме на значення так званого критерію t, або критерію Стьюдента (під таким псевдонімом писав свої роботи англійський математик Госсет). Далі отриманий добуток ми додаємо до значення вибіркового середнього й віднімаємо від нього. Отриманий інтервал власне й покриватиме не відоме нам значення генерального середнього:

$$\bar{X} - ts_{\bar{X}} \le \tfrac{1}{4} \le \bar{X} + ts_{\bar{X}}. \qquad (5)$$

Величини критерію t обчислені теоретично й наведені наприкінці книги в таблиці в додатку 1. Як користуватися цією таблицею? У першому стовпчику наведено число ступенів свободи. Нагадаємо, що для окремої вибірки воно дорівнює числу об'єктів у вибірці мінус одиниця, тобто $(n-1)$. У другому й третьому стовпчиках наведено значення критерію t для різних рівнів значущості. Що це таке? Коли ми висловлюємо статистичне судження на підставі даних вибірки, ми завжди враховуємо можливість того, що наша вибірка випадково виявилася невдалою й погано характеризує нашу сукупність. Для прикладу повернімося до простих модельних сукупностей А, Б і В, наведених на початку розділу. Якщо ми робитимемо з них випадкові вибірки по два об'єкти, то із сукупності Б, у якій значення значною мірою варіюють, найгіршими вибірками (щодо оцінки генерального середнього) будуть ті, що містять або два найменші, або два найбільші значення. Однак якщо вибір двох цифр здійснювати випадково, то ймовірність того, що до вибірки потраплять саме дві найменші або дві найбільші цифри, незначна. Проте, вона не дорівнює нулю (якщо говорити точно, вона дорівнює 0,044, або 4,4 %, або приблизно 1 шанс із 25). Умовно вважатимемо дві вибірки, що містять два найменші або два найбільші значення, «поганими», а решту — «гарними». У такому випадку, якщо ми після здійснення вибірки скажемо, що вона «гарна», то ми, швидше за все, не помилимося. Утім, усе-таки залишається ймовірність того, що ми саме помилилися, і така ймовірність називається рівнем значущості в статистиці. Часто в статистиці рівнем значущості називають не ймовірність помилки, а ймовірність безпомилкового твердження. Заплутатися тут неможливо, тому що в цьому випадку згадуються дуже різні цифри.

У наукових дослідженнях мінімально прийнятним рівнем значущості вважається ймовірність помилки менша за 5 %, або 0,05 у частках одиниці. За такого рівня значущості існує лише один шанс із 20, що ми помиляємося у своєму твердженні. Якщо ми хочемо мати ще більшу впевненість у безпомилковості наших тверджень, ми можемо прийняти рівень значущості 0,01 (за якого ймовірність того, що ми не помиляємося, дорівнює 99 %) або ще більш низький. У табл. 1 наведено значення критерію t для рівнів значущості 0,05 і 0,01. Розглянемо далі, як використовувати цю таблицю для обчислення довірчих інтервалів.

У кожній вибірці в нас 10 осіб. Виходить, число ступенів свободи — 9 $(10-1=9)$. Виберемо довірчий інтервал 0,05, або 5 % (або 95 %, що те саме, тільки в першому випадку ми говоримо про ймовірності помилки, а в другому — про ймовірності безпомилкового твердження). На рівні значення «9» у стовпчику кількості ступенів свободи знаходиться цифра 2,26 у стовпчику рівня значущості 0,05. Це і є критичне значення критерію t для 9 ступенів свободи й рівня значущості 0,05. Це значення і є тим коефіцієнтом, який ми можемо використати для обчислення довірчих інтервалів. Отже, для чоловіків Євразії вибіркове середнє дорівнює 172 см, а стандартна помилка — 3,81 см (стандартне відхилення й стандартна помилка вимірюються в тих самих одиницях, що й середнє арифметичне). Для $t=2{,}26$ маємо: $ts_{\bar{X}}=2{,}26\cdot3{,}81=8{,}6$ (використовувати більше число значущих цифр після коми не має жодного сенсу). За формулою (5): $172-8{,}6\le\mu\le172+8{,}6.$ Рівень значущості 0,05 означає, що на 95 % ми впевнені, що інтервал $(163{,}4\ldots180{,}6)$ см покриває генеральне середнє зросту чоловіків Євразії на підставі даних вибірки з 10 осіб.

Для вибірки чоловіків Америки значення критерію t буде тим самим (оскільки також є 9 ступенів свободи), а стандартна помилка дорівнюватиме 3,71. Використовуючи значення вибіркового середнього 177 см і формулу (5), можна обчислити 95%-й довірчий інтервал і для чоловіків Америки. Він дорівнюватиме $(168{,}6\ldots185{,}4)$ см.

Зверніть увагу на те, що довірчі інтервали *перекриваються*, тобто генеральні середні зросту чоловіків Євразії й Америки можуть дорівнювати одній і тій самій величині (нагадаємо, що цей приклад умовний і не ґрунтується на реальних даних).

Отже, узагальнюючи приклад, відзначимо, що на підставі вибірки ми можемо не лише отримати конкретну цифру — вибіркове середнє, що є оцінкою генерального середнього, але й, використовуючи стандартну помилку і значення критерію t, обчислити також інтервал значень, що з прийнятною для нас імовірністю покриватиме невідоме значення генерального середнього.

Порівняння двох вибірок

У попередньому розділі ми ознайомилися з тим, які оцінки можна зробити на підставі однієї вибірки з генеральної сукупності. Однак зазвичай нас цікавитиме не стільки окремо взята вибірка, скільки порівняння різних варіантів досліду.

Під час проведення експериментів найчастіше ми визначаємо, як незалежні змінні, що називаються також факторами, впливають на залежні змінні, наприклад, як доза добрив впливає на врожай рослин або як спосіб передпосівної обробки насіння впливає на його проростання й вагу проростків. Виявивши відмінності між варіантами нашого досліду, ми намагаємося вирішити, була така відмінність випадковою чи вона зумовлена впливом досліджуваного фактора. Урешті-решт, у цьому й полягає суть статистичного оцінювання результатів експерименту (таке оцінювання також часто називають статистичним опрацюванням результатів).

Для розуміння суті й логіки статистичного оцінювання повернімося до прикладу з прибульцями й уявімо собі, що їх зацікавило питання, чи розрізняються за зростом чоловіки, які проживають у Євразії й Америці. Прибульці зробили вибірки по 10 осіб на кожному континенті, виміряли їхній зріст і отримали первинні дані, наведені в попередньому розділі. Після цього можливі різні ситуації.

Якби вибіркові середні виявилися однаковими, то прибульці, імовірно, зробили б висновок, що зріст чоловіків на обох континентах не відрізняється. Інакше кажучи, висновок прибульців звучав би так: генеральні середні зросту чоловіків у Євразії $\left(\mu_{Є}\right)$ та Америці $\left(\mu_{А}\right)$, мабуть, однакові $\left(\mu_{Є} = \mu_{А}\right)$.

Однак, виявилося, що вибіркове середнє зросту чоловіків Євразії дорівнювало 172 см $\left(\bar{X}_{Є} = 172\text{ см}\right)$, а для чоловіків Америки — 177 см $\left(\bar{X}_{А} = 177\text{ см}\right)$. Як звучатиме висновок прибульців у цьому випадку? Чи почнуть вони вважати, що різниця в 5 см є випадковою і насправді різниці між зростом двох сукупностей чоловіків немає, чи ж вони мають вирішити, що ця різниця цілком доводить той факт, що чоловіки Америки *в середньому* на 5 см вищі за чоловіків Євразії?

Де та межа, за якою ми вже не можемо вважати спостережувану різницю випадковою? Відповідь на це питання дає статистичне оцінювання, у цьому випадку — використання критеріїв статистичної значущості. Для розглянутого прикладу цілком придатним виявиться вже знайомий нам критерій t.

Перед тим як ми розглянемо застосування цього критерію для порівняння двох вибірок, обговорімо загальний перебіг статистичних міркувань і загальну схему застосування статистичних критеріїв на прикладі з прибульцями й землянами.

Є дві генеральні сукупності (перша — це дорослі чоловіки Євразії, друга — дорослі чоловіки Америки). Ці сукупності мають конкретні генеральні середні значення зросту $\mu_{Є}$ та $\mu_{А}$, величини яких прибульцям не відомі (якби вони їх знали, то ніяких вибірок їм робити не довелося б). До проведення досліду щодо цих генеральних середніх висувається так звана *нульова* гіпотеза, тобто припускається, що вони однакові. Ця гіпотеза називається нульовою з тієї причини, що, якщо вона правильна, то різниця між генеральними середніми дорівнює нулю (це очевидно: якщо $\mu_{Є} = \mu_{А}$, то $\mu_{Є} - \mu_{А} = 0$).

Цій нульовій гіпотезі протиставляється *альтернативна* гіпотеза, відповідно до якої вибірки є вибірками із сукупностей, що мають різні генеральні середні, тобто $\mu_{Є} \neq \mu_{А}$ і, відповідно, $\mu_{Є} - \mu_{А} \neq 0$. Для розглянутого прикладу альтернативну гіпотезу можна сформулювати так: чоловіки Євразії й Америки відрізняються за зростом.

Отже, прибульці мають дві цілком зрозумілі статистичні гіпотези, і все, що їм потрібно зробити, — це, виходячи з результату вибіркового дослідження, перевірити, якій з гіпотез дані експерименту суперечать, а якій — ні.

Спостережувана різниця в 5 см доводить справедливість нульової гіпотези чи свідчить на користь альтернативної? На це питання не можна відповісти, якщо знати лише абсолютну величину різниці середніх, але не враховувати стандартну помилку різниці. Тактак, у різниці середніх є своя стандартна помилка, що показує, наскільки точно різниця вибіркових середніх оцінює різницю між генеральними середніми (немає значення, дорівнює ця різниця нулю чи іншому числу).

Для обчислення стандартної помилки різниці використовують дві формули, залежно від того, чи є однаковими обсяги наших вибірок.

Якщо помилку різниці двох вибіркових середніх позначити як s_d, то в цих двох випадках можна скористатися такими формулами:

а) Якщо $n_1 = n_2$, то:

$$s_d = \sqrt{\frac{s_1^2 + s_2^2}{n}}, \tag{6}$$

де s_1^2 і s_2^2 — дисперсії порівнюваних вибірок (які можна обчислити за формулою (3);

n — чисельність кожної вибірки.

б) Якщо $n_1 \neq n_2$, то:

$$s_d = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 n_2}}, \tag{7}$$

де s_1^2 і s_2^2 — дисперсії;

n_1 і n_2 — чисельності порівнюваних вибірок відповідно.

Обчисливши різницю вибіркових середніх та її помилку, далі ми обчислюємо величину критерію t:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_d}, \tag{8}$$

де t — фактичне значення критерію t, обчислене на підставі даних вибірок;

$\bar{X}_1$ і $\bar{X}_2$ — вибіркові середні;

s_d — стандартна помилка різниці вибіркових середніх.

З формули (8) випливає, що значення критерію t тим більше, чим більшою є різниця між вибірковими середніми й чим меншою — стандартна помилка різниці. Отже, критерій t завжди оцінює, наскільки різниця між середніми перевищує свою помилку. Щоби значення критерію t було додатним, завжди можна від більшого вибіркового середнього віднімати менше.

Тепер повернімося до нашого прикладу з чоловіками Євразії й Америки й порівняймо їх з допомогою цього критерію. Раніше ми обчислили, що вибіркова дисперсія для чоловіків Євразії дорівнює 145,5556, для чоловіків Америки вона становить 137,7778. Оскільки чисельності вибірок у нас однакові, то для обчислення помилки різниці можна використати формулу (6):

$$s_d = \sqrt{\frac{s_1^2 + s_2^2}{n}} = \sqrt{\frac{145{,}5556 + 137{,}7778}{10}} = 5{,}3.$$

Оскільки різниця в нас дорівнює 5 см, а її помилка — 5,3 см, то для цих двох вибірок величина критерію t дорівнюватиме:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_d} = \frac{5}{5{,}3} = 0{,}94.$$

Отже, про що свідчить ця величина? Чи доводить вона значущу відмінність між двома сукупностями, чи, навпаки, свідчить про те, що відмінність у 5 см отримано випадково, а насправді нульова гіпотеза про відсутність різниці між генеральними середніми є правильною? Для обґрунтованої відповіді на це питання ми маємо скористатися додатком 1, де наведено стандартні значення критерію t. У порівнянні двох вибірок число ступенів свободи дорівнює $n_1 + n_2 - 2$. Нульову гіпотезу про однаковість генеральних середніх відкидають, якщо фактична величина цього критерію дорівнює або перевершує стандартне (критичне) значення критерію, наведеного в таблиці, для відповідного рівня значущості й числа ступенів свободи. У випадку відхилення нульової гіпотези відмінності між двома вибірками називають значущими, або істотними. Іноді їх ще називають достовірними, але це не зовсім правильно, тому що не відповідає суворому визначенню поняття вірогідності.

У розглянутому нами прикладі для 18 ступенів свободи величина критичного значення критерію t для мінімального рівня значущості 0,05 дорівнює 2,1; ми ж отримали фактичне значення, що дорівнює 0,94. Отже, статистичне оцінювання показує, що ймовірність отримати різницю між вибірковими середніми в 5 см за умови однаковості генеральних середніх сукупностей, з яких зроблено дві вибірки, перевищує 5 % (фактично вона дорівнює 36 %). Звідси можна зробити висновок, що проведений експеримент не дозволяє зробити висновок про відмінність зросту між чоловіками Євразії й Америки. Інакше кажучи, нульова гіпотеза про однаковість середніх генеральних сукупностей не спростовується. Однак чи є вона істинною? Відповідь на це питання статистика не дає. Відповідь звучить саме так: ми не можемо стверджувати, що між двома вибірками чоловіків за їхнім зростом існує значуща відмінність (незважаючи на різницю між вибірковими середніми в 5 см). Такий висновок ми робимо на тій підставі, що обчислена за експериментальними даними величина критерію значущості t має високу ймовірність появи в разі, коли між генеральними середніми немає різниці, тобто є випадковою.

Можливо, якби прибульці зробили вибірки, що складаються не з 10, а, припустімо, зі 100 або навіть 1000 осіб, то вони виявили б цілком достовірні відмінності за їхнім зростом. У цьому випадку за тієї самої величини варіювання ознаки за рахунок збільшення числа повторів зменшилася б помилка різниці. Однак також імовірно, що для вибірок більшого розміру вибіркові середні практично не відрізнялися б.

У будь-якому разі існує загальна закономірність, яку слід ураховувати, плануючи експерименти: чим менша різниця між генеральними середніми, яку ми хочемо виявити з допомогою вибірок, і чим більше варіює ознака між окремими членами сукупності (тобто чим більшою є величина стандартного відхилення і, відповідно, дисперсії), тим більші вибірки нам доводиться використовувати.

Хотілося б відзначити одну обставину, якщо вона поки пройшла повз увагу читачів. Критерії значущості (наприклад, t) дозволяють нам зробити досить обґрунтований висновок про співвідношення між середніми *генеральних сукупностей*, з яких зроблено вибірки. Неправильно було б стверджувати, що вони дозволяють нам установити, відрізняються між собою *вибіркові середні* чи ні. У розглянутому прикладі (про зріст чоловіків Євразії й Америки) вибіркові середні відрізняються на 5 см, і жодні критерії для встановлення цього факту нам не потрібні. Але нас практично ніколи не цікавлять вибірки як такі; вони для нас важливі лише в тому плані, що дозволяють нам отримати *оцінки* генеральних параметрів.

Для того щоб ці оцінки були об'єктивними, вибірки мають бути правильно зроблені. Вибірковий метод являє собою досить самостійний і складний розділ статистики, і знайомство з його тонкощами не входить до завдання пропонованої книги. Відзначимо лише головну й безумовну вимогу до складання вибірок. Для того щоб вибірка справді характеризувала ту сукупність, з якої її зроблено, кожен член генеральної сукупності повинен мати *однакову ймовірність* потрапити до вибірки. Іншими словами цю головну вимогу можна сформулювати так: у членів генеральної сукупності не має бути відмітних ознак, які зробили б їхній вибір більш або менш імовірним. Лише тоді вибірка буде представницькою, або, як іще говорять статистики, *репрезентативною*.

Під час проведення наукових досліджень цієї вимоги дотримуватися нескладно. Важливо, щоб об'єкти потрапляли до вибірки *випадково*, без будь-якої упередженості. Випадковість можна забезпечити або з допомогою жереба, або з допомогою таблиць випадкових чисел, або в якийсь інший спосіб. Наприклад, якщо ми аналізуємо вплив біостимулятора на вкорінення черешків винограду і використовуємо в досліді по 50 черешків для кожного варіанта, то розподіл вихідних 100 черешків за варіантами слід зробити випадково. Так, ми можемо пронумерувати всі вихідні черешки (від 1 до 100), а потім «попросити» електронну таблицю Excel зробити випадковий вибір 50 цифр зі 100, а далі використати обрані черешки як контрольний варіант, а решту — як дослідні. Точно так само випадково слід розподіляти за варіантами досліду учнів під час вивчення різних прийомів навчання на засвоєння матеріалу, хворих для порівняння різних методів лікування, сорти рослин на ділянках повторів для порівняння врожайності тощо.

Якщо вибірки складено не випадково (або «на око»), то результати таких дослідів не мають жодної наукової або практичної цінності.

Наприкінці цього розділу розгляньмо також питання, пов'язане з двома типами помилок у разі використання статистичних методів. Критерій значущості t дозволяє оцінити ймовірність того, що конкретну його величину було отримано випадково, через вибіркове варіювання, а насправді відмінності між генеральними середніми немає. Якщо ця ймовірність дуже низька (наприклад, 0,00001 або ще менше), то ми можемо практично «на 100 %» бути впевнені, що генеральні середні справді відрізняються. Якщо ж вона значно більша, наприклад, перебуває в інтервалі 0,05–0,01, існує небезпека, що в нашому експерименті насправді сталася малоймовірна подія. У цьому випадку, хоча обчислена величина критерію й дозволяє відхилити нульову гіпотезу про відсутність різниці між генеральними середніми, однак це відхилення буде помилкою. Наприклад, ми можемо виявити, що сорт томатів А дає більший урожай, ніж сорт Б, і обчислений критерій t дозволить нам стверджувати це на рівні значущості 0,05. Однак імовірність помилки 5 % не така вже й мала, і цілком може виявитися, що у висновках за результатами нашого експерименту ми помилилися. Причина такої можливої помилки криється в самій сутності вибіркового методу, вона не пов'язана, наприклад, з неакуратним проведенням експерименту або з іншими подібними причинами. Така помилка в статистиці називається *помилкою першого роду*. Імовірність помилки першого роду дорівнює прийнятому рівню значущості: так, у додатку 1 для першого стовпчика критичних значень критерію t вона дорівнює 0,05, або 5 %, а для другого стовпчика — 0,01, або 1 %.

Під час проведення експериментів може виникнути бажання використати більш низькі рівні значущості, щоби зменшити ймовірність помилки першого роду. Наприклад, можна використати рівні значущості 0,001; 0,00001 і ще більш низькі. У цьому випадку ми й справді знижуємо ймовірність помилкового твердження про існування відмінностей між варіантами нашого досліду, тобто ймовірність помилкового відхилення нульової гіпотези про однаковість генеральних середніх, коли ця гіпотеза правильна.

Однак у разі використання надмірно низьких рівнів значущості на нас чатує інша помилка: ми не відхилятимемо нульову гіпотезу навіть у тих випадках, коли вона, безумовно, помилкова. Інакше кажучи, ми не помічатимемо відмінностей між варіантами досліду навіть тоді, коли вони фактично існують. Така помилка в статистиці називається *помилкою другого роду*. Її ймовірність обчислити не так просто, як імовірність помилки першого роду, однак слід мати на увазі, що вона стрімко зростає за необґрунтованого зниження рівня значущості.

Отже, у наукових дослідженнях слід дотримуватися балансу між обережністю й упевненістю, між острахом зробити помилку першого роду, тобто стверджувати наявність відмінності між варіантами, коли такої відмінності немає, і помилкою другого роду, тобто не помічати реальної різниці. Мінімальним загальноприйнятим рівнем значущості є ймовірність помилки першого роду 0,05.

Порівняння двох вибірок: метод пар

Іноді організація досліду вимагає іншого підходу до порівняння двох вибірок. Розгляньмо умовний приклад, що допоможе розібратися в подібних ситуаціях. Уявімо, що ми хочемо порівняти врожайність двох сортів ярового ячменя (А і Б). Ми висіяли ці сорти на дослідних ділянках у чотириразовому повторенні. Конфігурація дослідної ділянки дозволила розташувати ділянки у два ряди, і вздовж цих рядів спостерігається значна зміна ґрунтової родючості. На рисунку нижче показано розташування ділянок і отримані врожаї (у ц/га) (розташування ділянок у повторах обрано випадково, як і належить).

I повторність	II повторність	III повторність	IV повторність
А 32,6	Б 29,1	Б 26.1	А 25,8
Б 30,1	А 30.0	А 28,9	Б 23,0

→

Убутний градієнт родючості ґрунту

Перепишемо отримані дані у вигляді такої таблиці:

Повторність	Урожай, ц/га		Різниця
	Сорт А	Сорт Б	
I	32,6	30,1	2,5
II	30,0	29,1	0,9
III	28,9	26,1	2,8
IV	25,8	23,0	2,8
Середнє	29,3	27,1	2,3

Отже, сорт А має середню врожайність 29,3 ц/га, а сорт Б — 27,1 ц/га. Спробуймо застосувати метод попереднього розділу для оцінювання істотності цих відмінностей.

Сорт А: $\bar{X}_А = 29{,}3$; $s_А^2 = 7{,}929$.

Сорт Б: $\bar{X}_Б = 27{,}1$; $s_Б^2 = 10{,}269$.

Помилка різниці: $s_d = \sqrt{\dfrac{s_1^2 + s_2^2}{n}} = \sqrt{\dfrac{7{,}929 + 10{,}269}{4}} = 2{,}133.$

Тепер є всі дані, щоб обчислити величину критерію значущості t:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_d} = \frac{2{,}3}{2{,}133} = 1{,}08.$$

Для числа ступенів свободи, що дорівнює $(4+4-2=6)$, і рівня значущості 0,05 критичне значення критерію t становить 2,45. Оскільки обчислене нами значення набагато менше від цієї величини, то ми повинні зробити висновок, що в нас немає підстав відхиляти нульову гіпотезу щодо однаковості врожаїв сортів А і Б. Однак чи правильним буде такий висновок?

Зверніть увагу на те, що в *кожному* повторі врожай сорту А вищий, ніж урожай сорту Б. Уже одне це дозволяє досить обґрунтовано припустити, що перший сорт більш урожайний. Чому ж статистичне оцінювання не підтверджує цього припущення? Справа в тому, що в нас спостерігається систематична зміна родючості на дослідній ділянці, і в *кожного* сорту врожай знижується від першого повтору до четвертого. Ця мінливість вносить додатковий компонент у загальну дисперсію кожного сорту, і оскільки для обчислення стандартної помилки різниці ми використали саме індивідуальні дисперсії, то отримали завищену величину помилки. Фактично дисперсія врожаїв кожного сорту зумовлена сумою звичайної випадкової мінливості й систематичної зміни родючості ділянки. Для того щоб цей останній компонент виключити, більш правильним буде трохи інший підхід до оцінювання істотності різниці між варіантами.

Ми повинні оцінити, наскільки значущо відрізняється від нуля різниця в урожаї між сортами. Для цього ми аналізуємо останній стовпчик наведеної вище таблиці, що містить різниці між урожаями двох сортів за повторами. Середня величина різниці дорівнює 2,3 ц/га, стандартне відхилення — 0,911; стандартна помилка — 0,455. Виходячи з цих даних, величина критерію t становитиме:

$$t = \frac{\bar{X}_{\text{А}} - \bar{X}_{\text{Б}}}{s_d} = \frac{2{,}3}{0{,}455} = 5{,}055.$$

Оскільки обчислення вибіркових оцінок різниці ми проводили за чотирма цифрами, то число ступенів свободи в цьому випадку становить 3. Порівнюючи обчислену величину критерію з табличними значеннями для трьох ступенів свободи, ми можемо зробити висновок, що сорт А більш урожайний, ніж сорт Б, на рівні значущості 0,05 (це значення також майже «дотягує» до рівня значущості 0,01). Отже, застосувавши більш ефективний метод статистичного аналізу, ми цілком упевнено можемо стверджувати існування значущої різниці між сортами за врожайністю.

Метод парних порівнянь є дуже ефективним у статистичному аналізі даних експериментів подібного типу. Навіть більше, якщо є логічні підстави для складання пар значень (урожайність сусідніх ділянок у різних частинах дослідної ділянки, урожайність за роками, тварини одного виводку, порівняння близнюків тощо), то до такого методу варто вдаватися завжди, оскільки при цьому підвищується ефективність статистичних оцінок.

Однак хотілося б застерегти від помилкового складання довільних пар з різних повторів досліду. Порівнювані пари значень завжди вибираються *ДО* проведення дослідження, а не після того, як ми можемо побачити отримані значення.

Порівняння дискретних змінних

Наведені вище методи порівняння застосовуються для статистичного аналізу ознак, величина яких визначається безперервними змінними. У такому випадку розподіл конкретних величин ознаки в усіх об'єктів генеральної сукупності часто описується так званим нормальним розподілом, або розподілом Гаусса. На основі властивостей цього розподілу й був побудований критерій Стьюдента.

Однак у біологічних дослідженнях часто аналізуються ознаки, величини яких є дискретними змінними (наприклад, закономірності спадкування певних ознак, розподіл груп крові, кількість дітей у родинах тощо). Розподіл таких ознак у генеральній сукупності описується дискретними розподілами, наприклад біноміальним, розподілом Пуассона, геометричним та ін.

Розгляд особливостей імовірнісних розподілів значень безперервних і дискретних змінних у генеральних сукупностях виходить далеко за межі пропонованого посібника.

Тут же відзначимо, що критерій Стьюдента для порівняння дискретних даних можна застосовувати лише в окремих випадках; крім того, для дискретних даних змінюється порядок обчислення стандартної помилки.

Розгляньмо застосування методів аналізу дискретних даних для двох типів експериментів. Перший тип пов'язаний з перевіркою теорії, другий тип належить до порівняння вибірок.

Аналіз даних в експериментах першого типу розберемо на такому простому прикладі. Припустімо, що ми вирішили зіграти в «орлянку» й хочемо перевірити, чи справді запропонована для гри монета є правильною. Як ми це можемо зробити?

У загальному випадку ми можемо встановити це абсолютно точно, якщо підкидатимемо монету нескінченну кількість разів, щоб визначити, чи однаковими є ймовірності випадіння кожного з боків. Однак у такий спосіб правильність монети ми не доведемо, оскільки нескінченність ніколи не закінчується і наша перевірка ніколи не буде завершена. Отже, нам необхідно обмежитися якоюсь розумною кількістю підкидань, що дозволить зробити обґрунтований висновок.

Припустімо, ми підкинули монету десять разів. Якщо монета правильна, то чи можемо ми очікувати, що за *будь-яких* десяти підкидань рівно п'ять разів випаде цифра, і п'ять разів — герб? Звичайно ж, ми розуміємо, що ідеальні випадки трапляються нечасто, а співвідношення боків монети, що випали, є змінною випадковою величиною. Тому, якщо, наприклад, монета випаде шість разів одним боком і чотири — другим, то в нас немає особливих підстав уважати, що вона неправильна.

Імовірність того, скільки разів правильна монета за десяти підкидань (або за будь-якого іншого числа підкидань) випаде одним чи другим боком, обчислюється з використанням так званого біноміального розподілу. Цей розподіл описує альтернативні події (тобто події з двома результатами), коли ймовірність однієї події дорівнює p, а альтернативної події — q (у випадку правильної монети ймовірності випадіння боків однакові, тобто $p = q = 0{,}5$ (або $50\ \%$)).

Читачі, які зацікавилися цим, можуть знайти формули для обчислення відповідних імовірностей у рекомендованій літературі, тут вони не розглядатимуться. Автор наводитиме вже обчислені ймовірності.

Імовірність того, що правильна монета випаде кожним боком рівно п'ять разів, дорівнює приблизно $25\ \%$, а ймовірність того, що вона одним боком випаде чотири рази, а іншим — шість, помітно більша й дорівнює $41\ \%$.

Якого висновку з приводу правильності монети ми дійдемо, якщо вона тричі випаде одним боком, і сім разів — іншим? Чи можемо ми стверджувати в цьому випадку, що монета має деформацію або асиметрію? Якщо знову обчислити ймовірності біноміального розподілу, то виявиться, що три рази з десяти підкидань цифрою (або гербом) нагору правильна монета випадає не дуже й рідко — приблизно у $25\ \%$ випадків, тобто ймовірність такого результату така сама, що і ймовірність випадання однакових кількостей кожного з боків!

Звичайно, якщо монета всі десять разів випаде нагору одним боком, то ми матимемо цілковите право запідозрити щось нехороше (можливо, вона спаяна однаковими боками). Право на таку підозру дає нам незначна ймовірність подібного результату нашого експерименту, що становить близько $0{,}2\ \%$ у випадку справді правильної монети.

Досить підозрілим виглядатиме й випадок, коли співвідношення боків, що випали, дорівнює 9:1. Імовірність цього трохи менша за $2\ \%$.

Складніше зробити висновок щодо симетричності монети у випадку, якщо співвідношення боків, що випали, становитиме 8:2. Це не дуже нечаста подія для правильної монети, що має ймовірність близько $9\ \%$.

Можна сказати, що для того, щоб переконатися, чи правильною є монета, її потрібно підкинути не десять разів, а більше. Погодившись із цим, ми можемо підкинути монету, скажімо, сто разів. Однак і в цьому випадку ми навряд чи можемо очікувати, що правильна монета випаде рівно п'ятдесят разів одним боком і п'ятдесят разів — другим. Швидше за все, співвідношення боків, що випали, буде іншим (унаслідок того, що кількість гербів, які випали, або цифр є випадковою величиною). Цілком зрозуміло, що якщо співвідношення боків, які випали, становитиме 100:0 або 99:1, то ми, швидше за все, оголосимо монету неправильною, а якщо воно становитиме 49:51 або 48:52, тобто близько 50:50, ми вважатимемо її правильною. Однак де та межа, за якою не слід уважати монету правильною?

Загальний перебіг статистичних міркувань і загальна схема застосування статистичних критеріїв тут такі самі, які ми використали, розглядаючи використання критерію Стьюдента.

Ми маємо монету, щодо якої апріорно, *до досліду* (до наших підкидань) висуваємо *нульову* гіпотезу, що ця монета є правильною. У цьому випадку наша гіпотеза називається нульовою з такої причини: ми приймаємо, що існує реальна теоретична генеральна сукупність, яка складається з нескінченного числа сторазових підкидань абстрактної правильної монети. Якщо за випадкову величину прийняти число гербів, що випали, то, як неважко зрозуміти, середня величина цієї генеральної сукупності становить 50 гербів. Висуваючи припущення, що наша конкретна монета правильна, ми автоматично приймаємо, що ті сто підкидань, які ми збираємося зробити, є вибіркою з подібної генеральної сукупності з генеральної середньої, що також дорівнює 50 гербам. Інакше кажучи, відповідно до нульової гіпотези, різниця між середніми абстрактної генеральної сукупності правильних монет і фактичної генеральної сукупності, вибіркою з якої є наші 100 підкидань конкретної монети, дорівнює нулю $(50-50=0)$.

Цій нульовій гіпотезі протиставляється *альтернативна* гіпотеза, відповідно до якої 100 підкидань нашої монети є вибіркою з генеральної сукупності, у якій середня не дорівнює 50 гербам, а є більшою або меншою, і, відповідно, імовірності випадіння герба й цифри не однакові між собою. Простіше альтернативну гіпотезу можна сформулювати так: випробувана монета є неправильною.

Отже, ми маємо дві цілком зрозумілі статистичні гіпотези, і все, що нам потрібно зробити, — це провести експеримент і, виходячи з його результату, перевірити, якій з гіпотез дані експерименту суперечать, а якій — ні. Припустімо, ми провели наш експеримент, і виявилося, що герб випав 40 разів, а цифра — 60. Чи маємо ми вважати, що це цілком підтверджує або, навпаки, не підтверджує гіпотезу щодо правильності монети?

Для змістовної відповіді на це й подібні питання ми повинні мати певний критерій, що має вимірювати відхилення отриманої у вибірці величини, яка цікавить нас, від теоретичного значення. Відповідний критерій було введено в статистику в 1899 р. Карлом Пірсоном. Цей критерій називається критерієм χ^2 (читається «хі-квадрат»), або критерієм узгодження. Критерієм узгодження його названо через те, що його застосовують, зокрема, тоді, коли потрібно перевірити відповідність дискретних експериментальних даних установленій раніше теорії. Він може застосовуватися, звичайно ж, не лише для перевірки правильності монет, але й для таких випадків, як, наприклад, відповідність спадкування певної ознаки (за кількістю особин, що мають ознаку) законам Менделя, відповідність експериментальних даних окремому теоретичному розподілу, і для багатьох інших.

Обчислити конкретне чисельне значення цього критерію нескладно. У розглянутому нами прикладі за 100 підкидань монети випало 40 гербів і 60 цифр. Однак, відповідно до висунутого припущення щодо правильності монети, очікувалося, що кожен бік монети випаде однакову кількість разів — по 50. Відхилення спостережуваних значень від теоретичних тут такі: для гербів — $(40-50=-10)$, для цифр — $(60-50=10)$.

Значення хі-квадрат для цих даних обчислюється так:

$$\chi^2=\frac{(10)^2}{50}+\frac{(-10)^2}{50}=4.$$

З формули зрозуміло, що кожне відхилення підноситься у квадрат, кожен квадрат ділиться на гіпотетичне теоретичне значення, і частки складаються. Оскільки в знаменнику проставлено гіпотетичні чисельності (тобто вводиться обсяг вибірки), то очевидно, що χ^2 характеризує *відносні* значення відхилення, що є особливо важливим. Також очевидно, що якщо всі вибіркові чисельності дорівнюватимуть теоретичним, то хі-квадрат обернеться на нуль. Чим більше отримані дані відхиляються від теорії, тим більшою буде величина цього критерію.

У нашому прикладі, однак, поки ще не зовсім зрозуміло, чи є отримане значення хі-квадрат, що дорівнює 4, великим, помірним чи малим, тобто підтверджує чи спростовує нашу нульову гіпотезу щодо правильності монети. Тут перебіг аналізу такий самий, як і у випадку критерію Стьюдента. У додатку 2 наведено критичні значення критерію χ^2.

Нульову гіпотезу щодо однаковості генеральних середніх відкидають, якщо фактична величина цього критерію дорівнює або перевершує стандартне (критичне) значення критерію, наведеного в таблиці, для відповідного рівня значущості й числа ступенів свободи. У разі використання критерію хі-квадрат у дослідженнях, аналогічних розглянутому нами прикладу з монетою, число ступенів свободи буде на одиницю меншим від числа видів об'єктів (і, відповідно, числа доданків під час обчислення величини критерію), які ми аналізуємо. Оскільки в нас два види об'єктів (герб і цифра) і, відповідно, два доданки, число ступенів свободи дорівнює одиниці.

Для нашого прикладу з монетою критичне значення χ^2 для рівня значущості 0,05 дорівнює 3,84. Фактичне значення, що становить 4, перевершує критичне табличне значення; отже, з упевненістю щонайменше на 95 % ми можемо стверджувати, що монета, яка під час 100 підкидань 40 разів випала одним боком і 60 разів — другим, не є правильною. Відповідно, отримані нами в експерименті дані дозволяють відхилити нульову гіпотезу.

У загальному вигляді критерій хі-квадрат для перевірки теорії обчислюється в такий спосіб. Уявімо, що в нас є певна вибірка, яка містить об'єкти різного виду (наприклад, такі, у яких є і немає деяких ознак), причому число різних видів об'єктів дорівнює *n*. Позначимо для зручності чисельності об'єктів кожного виду через $f_1, f_2, f_3, \ldots, f_n$. Відповідні теоретичні чисельності для цих же об'єктів позначимо відповідно як $F_1, F_2, F_3, \ldots, F_n$. Відхилення фактичних чисельностей від теоретичних у цьому випадку дорівнюватимуть $f_1 - F_1$, $f_2 - F_2$ і т. д. Формула для обчислення хі-квадрат така:

$$\chi^2 = \sum \frac{\left(f_i - F_i\right)^2}{F_i}. \tag{9}$$

Для того щоб краще зрозуміти застосування критерію хі-квадрат, розгляньмо кілька прикладів.

В одному з моногібридних схрещувань Г. Мендель схрещував рослини гороху з пазушними квітками (домінантна ознака) і рослини з верхівковими квітками (рецесивна ознака). У F_2 було отримано 651 рослину з пазушними й 207 рослин з верхівковими квітками. Чи суперечать отримані дані нульовій гіпотезі щодо співвідношення домінантних ознак до рецесивних, яке дорівнює 3:1?

Виконаймо обчислення. Загальне число рослин нащадків другого покоління становить $(651+207=858)$. Теоретичні чисельності для пазушних і верхівкових квіток відповідно будуть такими: $\frac{3}{4}\cdot 858=643{,}5$ і $\frac{1}{4}\cdot 858=214{,}5$. Значення хі-квадрат дорівнює:

$$\chi^2=\sum\frac{(f_i-F_i)^2}{F_i}=\frac{(651-643{,}5)^2}{643{,}5}+\frac{(207-214{,}5)^2}{214{,}5}=\frac{7{,}5^2}{643{,}5}+\frac{(-7{,}5)^2}{214{,}5}\approx 0{,}349.$$

Число ступенів свободи тут дорівнює одиниці. Отримане значення критерію хі-квадрат цілком виразно свідчить про те, що в нас немає жодних причин відхиляти нульову гіпотезу, оскільки вона набагато менша від наведеного в додатку 2 критичного значення цього критерію.

За дигібридного схрещування рослин гороху, що відрізняються за формою й забарвленням насінин, Г. Мендель зібрав з рослин F_2 556 насінин, серед яких було:

гладеньких жовтих — 315; зморшкуватих жовтих — 101;
гладеньких зелених— 108; зморшкуватих зелених — 32.

Чи суперечать отримані дані гіпотезі про співвідношення фенотипів, яке дорівнює 9:3:3:1?

Відповідні теоретичні чисельності становитимуть:

- гладенькі жовті — $\frac{9}{16}\cdot 556=312{,}75$;
- зморшкуваті жовті — $\frac{3}{16}\cdot 556=104{,}25$;
- гладенькі зелені — $\frac{3}{16}\cdot 556=104{,}25$;
- зморшкуваті зелені — $\frac{1}{16}\cdot 556=34{,}75$.

Значення хі-квадрат становитиме:

$$\chi^2=\sum\frac{(f_i-F_i)^2}{F_i}=\frac{(315-312{,}75)^2}{312{,}75}+\frac{(101-104{,}25)^2}{104{,}24}+$$

$$+\frac{(108-104{,}25)^2}{104{,}25}+\frac{(32-34{,}75)^2}{34{,}75}\approx 0{,}469.$$

У цьому випадку число ступенів свободи дорівнює 3, оскільки для отримання величини хі-квадрат ми використовуємо чотири доданки. Отримане значення критерію дуже мале порівняно з табличним значенням для трьох ступенів свободи й рівня значущості 0,05, який становить 7,81. Тож дані експерименту Г. Менделя анітрошки не суперечать висунутій нульовій гіпотезі щодо співвідношення фенотипів у F_2, що дорівнює 9:3:3:1.

І ще один приклад. Соціологічна служба висунула прогноз, що на виборах мера в деякому невеликому місті, у якому проживають 10 000 виборців, їхні голоси розділяться нарівно між двома основними претендентами. Фактичний же поділ голосів після виборів виявився таким: 5100 проти 4900. Чи можна вважати, що результати голосування підтвердили прогноз соціологів про нічийний результат голосування?

Теоретичні чисельності виборців, які мали б віддати голоси одному й другому претенденту, становлять по 5000. Значення хі-квадрат у цьому випадку дорівнюватиме:

$$\chi^2 = \sum \frac{\left(f_i - F_i\right)^2}{F_i} = \frac{\left(4900 - 5000\right)^2}{5000} + \frac{\left(5100 - 5000\right)^2}{5000} = 4.$$

Число ступенів свободи дорівнює 1. Оскільки вибіркове значення хі-квадрат перевершує критичне табличне значення для одного ступеня свободи на рівні значущості 0,05, то слід зробити висновок, що отримане значення завелике для того, щоби прийняти нульову гіпотезу про розбивку голосів виборців на дві однакові групи. Імовірно, соціологічне опитування було проведено не зовсім коректно, або на вибір виборців вплинули якісь фактори, яких соціологи не врахували, формулюючи свій прогноз.

Критерій хі-квадрат використовується не лише для перевірки відповідності експериментальних даних певній теорії. Він також може використовуватися для порівняння двох або більше вибірок дискретних змінних. У цьому разі його застосовують особливим способом, після побудови таблиць спряженості ознак. Застосування таких таблиць є потужним і ефективним способом статистичного оцінювання, використовуваним і для множинних порівнянь кількох варіантів, і для перевірки однорідності кількох вибірок з дискретних генеральних сукупностей, і в багатьох інших випадках.

Що таке таблиця спряженості ознак? У загальному випадку це таблиці розміром $r \cdot c$ (рядки $\times$ стовпчики), у яких здійснено класифікацію дискретних даних за двома ознаками, кожна з яких може бути представлена двома або більше градаціями. Наприклад, людей можна поділити за статтю (одна ознака) і водночас за кольором очей (друга ознака). Однотипну продукцію можна поділити за придатністю (бракована вона чи ні) і водночас за заводами-виробниками.

Випадок, коли наявні дані можна представити у вигляді таблиці спряженості ознак, слід уважати досить загальним. У цьому розділі ми матимемо справу лише з таблицями розміром 2×2.

Розглянемо як приклад такі дані. Припустімо, ми випадково відібрали дві групи добровольців чисельністю 30 і 20 осіб. Членам першої групи ввели протигрипозну вакцину, а другій (контрольній) групі зробили нейтральну ін'єкцію (плацебо), при цьому самі учасники експерименту не знали, кому ввели справжню вакцину, а кому плацебо.

Після закінчення епідемії грипу виявилося, що в першій групі на грип занедужали 5 осіб із 30. У другій групі хворих виявилося 10 із 20. Отже, маємо частку захворілих у групі щеплених $\frac{5}{30} = 0{,}167$, або $16{,}7\ \%$, проти $\frac{10}{20} = 0{,}5$, або $50{,}0\ \%$, у контрольній групі. Під час статистичного оцінювання нульова перевірочна гіпотеза полягає в тому, що здійснені класифікації не залежать одна від одної, тобто що у вихідних сукупностях щеплених і не щеплених людей частки захворілих однакові. Під час роботи з вибірковими даними зазвичай виходить, що ці частки не дорівнюють одна одній, і питання полягає в тому, чи зумовлені ці відмінності лише випадковим варіюванням.

Складімо на підставі отриманих даних таблицю спряженості ознак. Біля фактичних даних у дужках наведено очікувані значення, які пояснюються нижче.

Добровольці	Число захворілих	Число не захворілих	Разом
Щеплені	5 (9)	25 (21)	30
Не щеплені	10 (6)	10 (14)	20
Разом	15	35	50

Для оцінювання представленого у вигляді таблиці спряженості ознак матеріалу найкращим способом статистичного оцінювання є обчислення величини χ^2, що в цьому випадку має бути мірою відмінності фактичних значень, розташованих у клітинках таблиці спряженості ознак, від тих очікуваних значень, які ми отримали б, якби обидва рядки були ідентичними (тобто якби щеплення не мало жодного впливу на захворюваність).

«Очікувані» значення для кожної клітинки таблиці обчислюються дуже просто. Для цього слід поділити добуток двох відповідних цій клітинці сум, записаних у стовпчику й рядку «Разом», на суму всіх чисел, записаних у всіх клітинках таблиці (крім стовпчика й рядка «Разом»). Так, очікуване число захворілих щеплених добровольців дорівнює $\frac{30\cdot 15}{50}=9$, а для захворілих не щеплених добровольців воно дорівнює $\frac{15\cdot 20}{50}=6$. Так само обчислено очікувані дані для всіх сполучень.

Маючи у своєму розпорядженні фактичні й теоретичні чисельності (обчислені в припущенні, що дві якісні ознаки, за якими здійснюються класифікації, не залежать одна від одної), ми можемо дуже просто обчислити величину критерію хі-квадрат (формула (9)). Для нашого прикладу вона дорівнюватиме:

$$\chi^2=\sum\frac{\left(f_i-F_i\right)^2}{F_i}=\frac{\left(5-9\right)^2}{9}+\frac{\left(10-6\right)^2}{6}+\frac{\left(25-21\right)^2}{21}+\frac{\left(10-14\right)^2}{14}\approx 6{,}349.$$

Обчисливши фактичне значення хі-квадрат для наших вибірок, ми, як звичайно, перевіряємо його на значущість шляхом зіставлення з критичними значеннями χ^2, наведеними в додатку 1, і приймаємо рішення щодо того, чи свідчить досягнутий рівень значущості на користь нульової гіпотези, а чи дозволяє відмовитися від неї. Яке тут число ступенів свободи? У разі використання таблиці спряженості ознак розміру $r\times c$ число ступенів свободи дорівнює $(r-1)\cdot(c-1)$. Для таблиці розміром 2×2 число ступенів свободи дорівнює 1.

Порівняння отриманого значення 6,349 з критичними значеннями χ^2 для одиничного ступеня свободи показує, що фактичне значення перевищує табличне для рівня значущості 0,05 і майже досягає рівня значущості 0,01. Тому ми відкидаємо висунуту нами нульову гіпотезу як неправдоподібну й погоджуємося з твердженням про наявність відмінностей у захворюваності на грип серед щеплених і не щеплених людей.

Для застосування критерію хі-квадрат для оцінювання зв’язку в таблицях спряженості ознак існує обмеження — жодне *очікуване* значення не має бути меншим від 5. В іншому випадку застосування описаного методу призводить до ненадійних результатів. Для такого випадку в таблиці 2×2 слід використовувати точний критерій Фішера, порядок обчислення якого міститься в пропонованому переліку літератури, зокрема у книзі Дж. В. Снедекора.

Якщо найменше очікуване значення в таблиці спряженості ознак — 5 або більше, для таблиць 2×2 немає необхідності обчислювати всі очікувані чисельності. Величину критерію χ^2 можна обчислити й без них. Запишемо таблицю 2×2 у загальному вигляді:

	Наявність ознаки A	Відсутність ознаки A	Разом
Наявність ознаки B	A	b	$a+b$
Відсутність ознаки B	C	d	$c+d$
Разом	$a+c$	$b+d$	n

Для такої таблиці фактичне значення критерію хі-квадрат можна обчислити за формулою:

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}. \quad (10)$$

Для порівняння двох вибірок, представлених дискретними змінними, можна використати також критерій Стьюдента. На початку розділу було відзначено, що це можна застосовувати в особливих випадках, а саме тоді, коли вибірки досить великі й мають не менш ніж 50 об'єктів кожна. Тут ми використовуємо *апроксимацію* біноміального розподілу нормальним розподілом. Крім того, у цьому випадку стандартна помилка різниці обчислюється інакше, ніж ми робили це в попередньому розділі, коли порівнювали безперервні величини.

Застосування критерію Стьюдента розглянемо на такому прикладі. Припустімо, що в нас є два сорти ячменя — A і B, і ми хочемо порівняти їхню стійкість до кореневої гнилизни за штучного зараження. З посівів було зроблено випадкові вибірки рослин цих сортів. До вибірки рослин сорту A потрапило 112 екземплярів, з яких 86 мали ознаки захворювання. До вибірки сорту B увійшло 138 рослин, серед яких 92 мали ознаки кореневої гнилизни. Чи є значущими відмінності між сортами в поширеності хвороби, що становить $\bar{X}_{\text{A}} = \frac{86}{112} \approx 0{,}768$, або $76{,}8\ \%$, для сорту A і $\bar{X}_{\text{B}} = \frac{92}{138} \approx 0{,}667$, або $66{,}7\ \%$, для сорту B?

Тут у нас кожна вибірка представлена значною кількістю об'єктів, і застосування критерію Стьюдента є виправданим.

Як нульову гіпотезу виберемо припущення про відсутність справжньої відмінності між величинами поширеності кореневої гнилизни. Відповідно до цієї гіпотези, частка рослин, уражених кореневою гнилизною, є однією й тією самою для обох сортів ячменя, а кожна вибірка дає окрему оцінку цій загальній частці; відмінності ж між вибірками зумовлені лише вибірковим варіюванням.

Альтернативною гіпотезою є та, що сорти фактично розрізняються за стійкістю до захворювання, і відмінності між вибірковими оцінками поширеності хвороби переважно й зумовлені цими відмінностями в стійкості.

Для перевірки нульової гіпотези ми насамперед оцінюємо загальну частку уражених рослин. Це узгоджується з прийнятою нульовою гіпотезою, і для обчислення загальної частки ми об'єднуємо дані двох вибірок. Загалом ознаки кореневої гнилизни спостерігаються у $86+92=178$ рослин із $112+138=250$ урахованих. Звідси для загальної частки уражених рослин дістаємо значення $\bar{X}=\frac{178}{250}=0{,}712$, або $71{,}2\ \%$. Відповідно, загальна частка неуражених рослин становить $1-0{,}712=0{,}288$, або $28{,}8\ \%$.

Тепер узагальнимо все сказане, щоби привести формулу обчислення критерію Стьюдента до загального вигляду.

Отже, ми маємо дві вибірки з n_1 і n_2 елементів, причому a_1 елементів з першої вибірки й a_2 елементів із другої мають ту саму визначену ознаку. Вибіркові середні (у частках одиниці) і середня частка для об'єднаної вибірки дорівнюватимуть:

$$\bar{X}_1=\frac{a_1}{n_1};\ \ \bar{X}_2=\frac{a_2}{n_2};\ \ \bar{X}=\frac{a_1+a_2}{n_1+n_2}.$$

Величину критерію t можна обчислити так:

$$t=\frac{\left|\bar{X}_1-\bar{X}_2\right|}{\sqrt{\bar{X}\left(1-\bar{X}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}.$$

Для розглянутого прикладу із сортами ячменя й кореневою гнилизною величини $\bar{X}_А$, $\bar{X}_В$ і $\bar{X}$ ми вже обчислили. Тепер обчислимо величину критерію Стьюдента:

$$t=\frac{\left|\bar{X}_А-\bar{X}_В\right|}{\sqrt{\bar{X}\left(1-\bar{X}\right)\left(\frac{1}{n_А}+\frac{1}{n_В}\right)}}=\frac{\left|0{,}768-0{,}667\right|}{\sqrt{0{,}712\left(1-0{,}712\right)\left(\frac{1}{112}+\frac{1}{138}\right)}}\approx\frac{0{,}101}{0{,}0576}=1{,}75.$$

Число ступенів свободи тут становить $112+138-2=248$. Критичне значення критерію Стьюдента для цього числа ступенів свободи й рівня значущості 0,05 дорівнює приблизно 1,97 (*див. додаток 1*). Оскільки фактично отримане значення менше від табличного, ми не маємо достатньо підстав для відхилення нульової гіпотези щодо відсутності відмінностей між сортами ячменя в поширеності захворювання.

Розгляньмо ще один подібний приклад. Аналізували ушкодження грушевою плодожеркою плодів двох сортів груші. З рослин кожного сорту випадково відібрали по 500 плодів. Число ушкоджених плодів відповідно становило 156 і 190 для сорту 1 і сорту 2. Чи є виявлена різниця в пошкодженні істотною?

Розмір вибірок у цьому експерименті достатньо великий, і з упевненістю можна використати наведені вище формули:

$$\bar{X}_1 = \frac{156}{500} = 0{,}312; \quad \bar{X}_2 = \frac{190}{500} = 0{,}380; \quad \bar{X} = \frac{156+190}{500+500} = \frac{346}{1000} = 0{,}346;$$

$$t = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\bar{X}\left(1-\bar{X}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\left|0{,}312-0{,}380\right|}{\sqrt{0{,}346\left(1-0{,}346\right)\left(\frac{1}{500}+\frac{1}{500}\right)}} = \frac{0{,}068}{0{,}030} = 2{,}267.$$

За $500+500-2=998$ ступенів свободи на рівні значущості 0,05 різниця, безумовно, є істотною.

Таблицю спряженості ознак можна будувати не лише для малих вибірок, але й для великих. Побудуємо її для останнього прикладу з ушкодженням плодів груш.

Сорт груші	Ушкоджених плодів	Неушкоджених плодів	Разом
Сорт 1	156	344	500
Сорт 2	190	310	500
Разом	346	654	1000

Значення хі-квадрат (за формулою (10)) для цих даних дорівнює 4,81. Якщо це значення зіставити з критичним значенням у додатку 2 для рівня значущості 0,05, то ми дійдемо висновку про відхилення нульової гіпотези щодо відсутності відмінностей в ушкодженні плодів між двома сортами груш.

Отже, у випадку великих вибірок як використання критерію Стьюдента, так і побудова таблиці спряженості ознак і використання критерію хі-квадрат дозволили нам зробити однаковий висновок. Це не випадково: для нескінченного числа ступенів свободи величина критерію Стьюдента дорівнює кореню квадратному величини критерію хі-квадрат для одиничного ступеня свободи. Яким критерієм користуватися в такому разі — справа смаку дослідника.

Однак використання критерію Стьюдента для вибірок дискретних даних, що мають чисельність менш ніж 50 об'єктів, призведе до помилкових висновків і є неприпустимим.

Кореляція

У навколишньому світі величини деяких кількісних ознак часто змінюються узгоджено, коли в разі зростання величини однієї ознаки збільшується або зменшується величина іншої. Наприклад, улітку спостерігається зворотний зв'язок між кількістю опадів і температурою: як правило, чим більше випадає опадів, тим нижча температура. Хоча з цього існує чимало винятків (трапляються теплі дощі й досить прохолодні ясні літні дні), проте в більшості випадків найбільш прохолодними днями влітку є найбільш дощові, а найбільш теплими — малохмарні. Подібні зв'язки виявляються не лише в метеорології. Наприклад, у людини виявляється сильний зв'язок між зростом і вагою. Хоча й бувають невисокі люди з великою вагою й досить сухорляві люди високого зросту, вага яких не дуже велика, однак *у середньому* очевидно, що чим більшого зросту людина, тим більшою є її вага. У рослин спостерігається зв'язок між кількістю опадів у різні періоди вегетаційного періоду й урожаєм: зазвичай урожай вищий за досить великої кількості опадів у періоди активного наростання вегетативної маси рослин, а також на початкових етапах формування плодів. Посуха в ці періоди може різко знизити продуктивність рослин, і, до певних меж, збільшення кількості опадів сприятиме збільшенню врожайності. Хоча цей зв'язок не є прямолінійним: надмірна вологість може призвести до зниження врожаю.

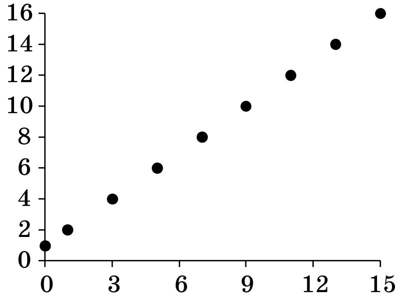
Наявність зв'язків між змінними, що варіюють, можна виявити на всіх рівнях організації живих організмів. У біології такий зв'язок називають кореляцією (від лат. *correlatio* — «співвідношення», «зв'язок»). Уперше цей термін використав Ж. Кюв'є в 1806 р. У біологічну статистику термін «кореляція» увів Ф. Гальтон у 1886 р.

Для аналізу таких зв'язків використовують *коефіцієнт кореляції*. Для того щоб краще зрозуміти його суть, розгляньмо умовний приклад. Уявімо, що люди полетіли на Марс, побудували там теплиці й почали вирощувати культурні рослини. Деякі марсіанські ґрунти багаті на Фосфор і Калій, але в них відсутній Нітроген (нагадаємо: цей приклад — УМОВНИЙ). Позначимо такий тип марсіанського ґрунту літерою А. В інших ґрунтах Нітроген наявний, але в недостатній кількості. Такий тип ґрунту позначимо Б. Урешті-решт, є марсіанські ґрунти, у яких Нітрогену цілком достатньо. Такому типу привласнимо літеру В. Далі припустімо, що марсіанські агрономи провели експеримент із ґрунтами згаданих трьох типів. У ґрунти додавали нітратне добриво в кількості від 0 до 15 г/кг ґрунту й потім вирощували на цих ґрунтах кукурудзу. Отримані врожаї (у т/га) наведено в таблиці нижче.

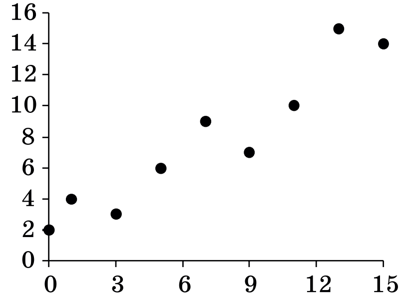
Більш наочно ці самі дані виглядатимуть, якщо представити їх у формі графіків, що й зроблено на рисунку нижче. Тут по осі X (горизонтальній) наведено кількості внесеного в ґрунт Нітрогену, а по осі Y — отримані врожаї.

Додано нітратного добрива, г/кг ґрунту	Урожаї зерна кукурудзи, т/га		
	Тип ґрунту		
	А	Б	В
0	1	2	9
1	2	4	9
3	4	3	5
5	6	6	7
7	8	9	11
9	10	7	6
11	12	10	9
13	14	15	5
15	16	14	10

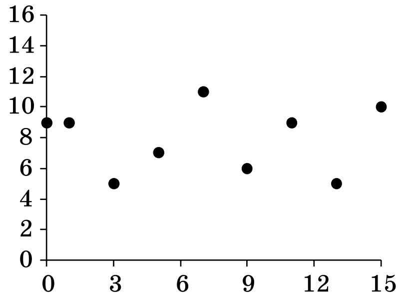
А



Б



В



З наведених даних добре помітно, що для ґрунту А ми спостерігаємо абсолютну прямо пропорційну залежність урожаю від кількості внесеного в ґрунт нітратного добрива: точки розташовані по прямій лінії. Прямо пропорційна залежність урожаю від кількості внесеного Нітрогену спостерігається й для ґрунту Б, хоча зв'язок не настільки сильний: точки розташовані більш вільно, не розміщаючись чітко на прямій лінії. Урешті-решт, для ґрунту В, імовірно, зв'язок між Нітрогеном і врожаєм відсутній — у разі зміни кількості внесеного в ґрунт нітратного добрива зміни врожайності не спостерігається.

Отже, на рисунку показано зв'язок між двома змінними (кількістю внесених добрив і врожаєм кукурудзи) різної *сили* — від цілковитої функціональної залежності однієї змінної від іншої до відсутності зв'язку між ними. Цілком зрозуміло, що хотілося б мати більш або менш точну *кількісну* оцінку сили цього зв'язку. Для цього використовують так званий коефіцієнт кореляції, що позначається літерою r. Для наведених на рисунку типів ґрунтів А, Б і В величини коефіцієнта кореляції дорівнюють 1; 0,95 і 0 відповідно. Як ці величини обчислюються?

Коли ми говоримо про кореляцію між двома змінними, ми маємо деяке число *пар* значень, у яких одне значення належить одній змінній, а друге — іншій змінній; у розглянутому вище прикладі такі пари утворювали дози внесених нітратних добрив (незалежна змінна) і відповідні цим дозам урожаї кукурудзи (залежна змінна). Якщо число пар значень двох змінних позначити як n, конкретні значення однієї змінної позначити як X_i, а другої як Y_i, і відповідні середні значення позначити як $\bar{X}$ і $\bar{Y}$, то коефіцієнт кореляції обчислюють за такою формулою:

$$r = \frac{\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum\left(X_i - \bar{X}\right)^2 \sum\left(Y_i - \bar{Y}\right)^2}}. \tag{11}$$

Для обчислень буває зручніше використовувати нижченаведену тотожну робочу формулу:

$$r = \frac{\sum X_i Y_\mathrm{i} - \dfrac{\sum X_i \sum Y_i}{n}}{\sqrt{\left(\sum X_i^2 - \dfrac{\left(\sum X_i\right)^2}{n}\right)\cdot\left(\sum Y_i^2 - \dfrac{\left(\sum Y_i\right)^2}{n}\right)}}. \tag{12}$$

Незважаючи на застрашливий вид формул (11) і (12), обчислення за ними виконувати нескладно, головне — не помилитися. Обчислена за цими формулами величина коефіцієнта кореляції може змінюватися від −1 до +1. Від'ємні значення коефіцієнта кореляції відповідають випадкам обернено пропорційного зв'язку між двома змінними, коли в разі збільшення одного значення друге зменшується; додатні значення пов'язані з прямо пропорційною залежністю.

Коефіцієнт кореляції за абсолютною величиною дорівнює одиниці у випадку, коли немає ніякого вибіркового варіювання, спостерігається абсолютна залежність однієї змінної від другої й усі значення розташовуються на прямій лінії. Саме такому випадку відповідають дані для типу ґрунту А, розглянуті в наведеному вище прикладі. У реальних біологічних дослідженнях такого не трапляється практично ніколи. Значення коефіцієнта кореляції, близькі до нуля, характеризують випадки відсутності зв'язку між змінними (тип ґрунту В). Випадком досить сильного зв'язку між ознаками є ґрунт типу Б — коефіцієнт кореляції через вибіркове варіювання не дорівнює одиниці, але є досить високим.

Для того щоб навчитися обчислювати коефіцієнт кореляції, розгляньмо наведені в книзі Дж. Снедекора дані про вплив ушкодження плодів яблуні личинкою молі на врожай яблук.

Номер дерева	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Відсоток ушкоджених плодів	59	58	56	53	50	45	43	42	39	38	30	27
Урожай одного дерева, сотні плодів	8	6	11	22	14	17	18	24	29	23	26	40

Швидкий погляд на дані таблиці дозволяє очікувати від'ємного коефіцієнта кореляції, тому що спостерігається явна зворотна залежність — за зниження відсотка ушкоджених плодів їхній урожай збільшується. Скористаймося тепер формулою (10) і проведімо необхідні обчислення. Найзручніше спочатку обчислити необхідні проміжні величини, а потім — величину r:

$$\sum X_i = 59 + 58 + \ldots + 27 = 540;\ \sum Y_i = 8 + 6 + \ldots + 40 = 238;$$

$$\sum X_i Y_i = 59 \cdot 8 + 58 \cdot 6 + \ldots + 27 \cdot 40 = 9714;$$

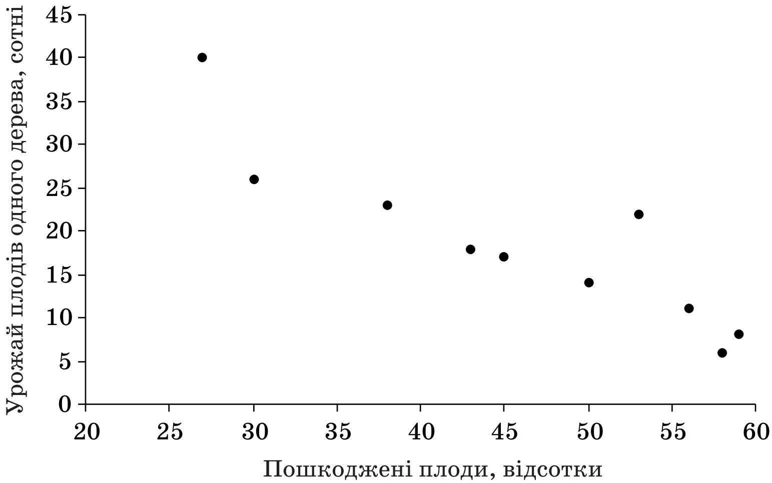
$$\sum X_i^2 = 59^2 + 58^2 + \ldots + 27^2 = 25522;\ \sum Y_i^2 = 8^2 + 6^2 + \ldots + 40^2 = 5736;$$

$$n = 12 \text{ (число пар значень);}$$

$$r = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left(\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right) \cdot \left(\sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}\right)}} = \frac{9714 - \frac{540 \cdot 238}{12}}{\sqrt{\left(25522 - \frac{540^2}{12}\right) \cdot \left(5736 - \frac{238^2}{12}\right)}} =$$

$$= \frac{-996}{1114} = -0{,}89.$$

Отже, спостерігається досить сильний обернено пропорційний зв'язок між відсотком ушкоджених яблуневою міллю плодів на дереві й урожаєм яблук із цього дерева. Найбільш наочно кореляція виявляється, коли дані представлені в графічному вигляді. Графік, що показує вплив ушкодження плодів на врожай, наведено нижче.



Певною мірою умовно, залежно від величини (абсолютної, без урахування знака) коефіцієнта кореляції силу зв'язку між двома змінними вважають високою, якщо величина r перебуває в інтервалі (0,66...0,99), середньою — для інтервалу (0,33...0,65) і слабкою, якщо $r < 0{,}33$. Якщо коефіцієнт кореляції дорівнює нулю, то зв'язок між змінними відсутній.

Обчислений на підставі даних вибірок коефіцієнт кореляції r, як і інші вибіркові показники, служить оцінкою генерального параметра, що зазвичай позначається літерою ρ (грецька буква «ро»). І аж ніяк не в усіх випадках, коли r не дорівнює нулю, ми можемо бути впевнені, що між двома змінними справді існує зв'язок. Достатньо навіть одного випадкового значення, що різко виділяється, щоб побачити зв'язок там, де його насправді немає, особливо коли в нашому розпорядженні — невеликі вибірки. З цієї причини після обчислення величини коефіцієнта кореляції проводиться визначення рівня значущості цієї кореляції. Визначаючи значущість, або істотність, вибіркового коефіцієнта r, ми зважаємо на те, що він є оцінкою генерального параметра ρ. Нульовою гіпотезою в цьому випадку буде припущення, що $\rho = 0$, тобто в генеральних сукупностях між змінними, що цікавлять нас, ніякого зв'язку немає.

Якщо нульова гіпотеза правильна, то обчислена на підставі вибірок величина r оцінює нуль. Альтернативною гіпотезою, як і звичайно, буде та, що генеральний параметр нулю не дорівнює, тобто між аналізованими змінними насправді існує зв'язок.

Для визначення істотності коефіцієнта кореляції використовують знайомий нам критерій Стьюдента (t), що в цьому випадку обчислюється так:

$$t = r\sqrt{\frac{n-2}{1-r^2}}. \tag{11}$$

Як і раніше, ми зіставляємо обчислену величину критерію Стьюдента з табличним значенням для відповідного рівня значущості й числа ступенів свободи, що дорівнює $(n-2)$. Істотною на конкретному рівні значущості вважається кореляція, для якої критерій t дорівнює або перевершує табличне значення. Перевірмо значущість кореляцій, обчислених для розглянутих у цьому розділі прикладів. Згадаймо, що величини коефіцієнта кореляції між дозою внесених нітратних добрив і врожаєм кукурудзи становили 1; 0,95 і 0 для ґрунтів А, Б і В відповідно. Величини критерію t для них будуть такі.

Для ґрунту А: $t = r\sqrt{\frac{n-2}{1-r^2}} = 1\cdot\sqrt{\frac{9-2}{1-1^2}} = ¥$.

Для ґрунту Б: $t = r\sqrt{\frac{n-2}{1-r^2}} = 0{,}95\cdot\sqrt{\frac{9-2}{1-0{,}95^2}} = 8{,}05.$

Для ґрунту В: $t = r\sqrt{\frac{n-2}{1-r^2}} = 0\cdot\sqrt{\frac{9-2}{1-0^2}} = 0.$

Для всіх типів ґрунтів число ступенів свободи дорівнює 7. Для ґрунту А величина критерію Стьюдента вийшла такою, що дорівнює нескінченності. Це не дивно, оскільки ми спостерігаємо абсолютну залежність між двома змінними, без будь-якого вибіркового варіювання. Отже, ми з упевненістю можемо відхилити нульову гіпотезу щодо відсутності кореляції між досліджуваними ознаками. Однак, як уже було сказано, такий абсолютний зв'язок у біологічних дослідженнях спостерігається рідко коли, якщо спостерігається взагалі.

Для ґрунту Б величина критерію Стьюдента перевершує табличне значення для рівня значущості 0,01 (фактичний рівень значущості для $t = 8{,}05$ і 7 ступенів свободи дорівнює 0,00009). Тож імовірність того, що для цього ґрунту високу кореляцію між дозою нітратного добрива й урожаєм отримано випадково, надзвичайно мала, і ми впевнено відхиляємо нульову гіпотезу щодо відсутності зв'язку між дозою й урожаєм.

Для ґрунту В критерій Стьюдента було обчислено лише для ілюстрації, він підтверджує раніше зроблений висновок про відсутність зв'язку між добривом і врожаєм. Інакше кажучи, ми приймаємо нульову гіпотезу про те, що генеральний параметр ρ дорівнює нулю.

Для даних про кореляцію між часткою ушкоджених плодів і врожаєм яблук з одного дерева маємо:

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0{,}89 \cdot \sqrt{\frac{12-2}{1-0{,}89^2}} = 6{,}17.$$

У цьому випадку число ступенів свободи дорівнює 10, і обчислена величина критерію Стьюдента також перевершує табличне значення для рівня значущості 0,01 (фактичний рівень значущості дорівнює 0,0001), що дозволяє нам обґрунтовано відхилити нульову гіпотезу щодо відсутності кореляції між змінними.

На завершення цього розділу хотілося б знову нагадати про обережність під час формулювання висновків за результатами експериментів або спостережень, пов'язаних з пошуком корелятивних зв'язків. Коли ми встановили наявність істотної кореляції між двома змінними величинами (неважливо, установили ми її під час проведення експерименту чи виявили в процесі спостереження), існує чимала спокуса приписати виявленій кореляції причинно-наслідковий зв'язок, тобто одну змінну розглядати як причину, а другу — уважати наслідком. Однак поспішний висновок не завжди принесе користь: установлення зв'язку ще не доводить причинно-наслідкової залежності.

Кілька вибірок з генеральних сукупностей: дисперсійний аналіз

У наукових експериментах ми рідко коли обмежуємося лише двома варіантами досліду. Якщо в досліді є більш ніж два варіанти, чи можемо ми використати статистичні методи, які були розглянуті вище, зокрема використати критерій Стьюдента для попарного порівняння всіх вибіркових середніх? На таке питання можна дати цілком певну відповідь: у жодному разі. Критерій Стьюдента призначений для порівняння лише *двох* варіантів досліду, і якщо в досліді є три або більше варіантів, цей критерій у такій ситуації «не працює».

У більшості випадків для статистичного опрацювання подібних даних найбільш підходящим є дисперсійний аналіз, теоретично обґрунтований на початку минулого століття Р. Фішером. Розвиток теорії дисперсійного аналізу сприяв значному прогресу в розробці статистичних методів для множинних порівнянь і надзвичайному розвитку теорії планування експерименту.

Для того щоб зрозуміти суть дисперсійного аналізу, розгляньмо спочатку дані умовного досліду, у якому зроблено чотири вибірки по п'ять повторень у кожній, з *однієї й тієї самої* генеральної сукупності, що має генеральне середнє, яке дорівнює 30, і генеральну дисперсію, що становить 100. Для визначеності вважатимемо, що ця генеральна сукупність являє собою врожайність (у ц/га) сорту ярового ячменя «Одеський 100», а кожна наша вибірка складається з чотирьох ділянок площею по 10 м^2, висіяних на однорідній за ґрунтовим покривом дослідній ділянці. У таблиці нижче наведено отримані в експерименті дані й результати деяких допоміжних обчислень, які нам знадобляться.

Урожай ярового ячменя сорту «Одеський 100»
(чотири вибірки з п'ятьох ділянок площею 10 м^2 кожна)

Урожай за повтореннями, ц/га	Вибірка 1	Вибірка 2	Вибірка 3	Вибірка 4	За дослідом у цілому
	38, 30, 40, 22, 35	29, 27, 20, 39, 45	11, 31, 17, 37, 39	21, 23, 28, 27, 21	
N	5	5	5	5	20
$\bar{X}$	33	32	27	24	29
$\sum\left(X_i - \bar{X}\right)^2$	208	396	616	44	1534

Отже, у нашому досліді є чотири вибірки по п'ять повторень у кожній, разом є 20 об'єктів $(4 \times 5 = 20)$.

Дані таблиці дозволяють отримати три вибіркові оцінки генеральної дисперсії $\sigma^2 = 100$.

Першу оцінку ми можемо дістати, використовуючи всі 20 об'єктів. Якщо розглядати їх як єдину вибірку, то вибіркову дисперсію ми можемо обчислити, використовуючи суму квадратів відхилень кожного значення від загального для всього досліду середнього (що дорівнює 29), поділивши її на число ступенів свободи, яке дорівнює загальній кількості об'єктів без одного (формула (2)). Суму квадратів відхилень для досліду наведено в правій нижній комірці таблиці й дорівнює 1534. Перша вибіркова оцінка генеральної дисперсії становить: $s^2 = \dfrac{1534}{19} = 80{,}7.$

Другу оцінку дисперсії генеральної сукупності можна отримати, виходячи із сум квадратів відхилень для кожної вибірки (нижній рядок таблиці). Тут ми можемо діяти двома способами.

Можна обчислити вибіркову дисперсію для кожної вибірки, а потім — середнє арифметичне (оскільки число повторень у кожній вибірці однакове, ми можемо зробити це без будь-яких застережень). Для вибірки 1 дисперсія дорівнює (формула ((2)): $\frac{208}{4} = 52$. Аналогічно, для вибірок 2, 3 і 4 вибіркові дисперсії дорівнюватимуть 99; 154 і 11 відповідно. Їх середнє арифметичне дорівнює: $\frac{52+99+154+11}{4} = 79$. Однак, з деяких формальних причин, краще вчинити інакше. Спочатку ми просумуємо (узагальнимо) суми квадратів відхилень для кожної вибірки: $209+396+616+44=1264$. Яке число ступенів свободи відповідатиме цій сумі квадратів? У кожній вибірці є п'ять повторів, тобто є чотири ступені свободи, і таких вибірок у нас чотири. Загалом, оскільки для кожної з чотирьох вибірок є чотири ступені свободи, сумі квадратів 1264 відповідатиме $4 \times 4 = 16$ ступеням свободи, а середній квадрат дорівнюватиме $\frac{1264}{16} = 79$. Результат, природно, такий самий, як і в разі обчислення вибіркової дисперсії для кожної вибірки та їх середнього арифметичного. Отже, 79 є другою вибірковою оцінкою σ^2.

Середні арифметичні вибірок приводять нас до третьої оцінки генеральної дисперсії. Якщо з генеральної сукупності зробити нескінченно багато вибірок обсягом n, то середні арифметичні цих вибірок матимуть розподіл із середнім, що дорівнює генеральній середній вихідної генеральної сукупності (μ), і дисперсією, яка дорівнює $\frac{\sigma^2}{n}$. З цієї причини в розглянутому експерименті середній квадрат відхилень вибіркових середніх від загального середнього буде оцінкою величини $\frac{\sigma^2}{n} = \frac{100}{5} = 20$.

Оскільки в нас чотири вибірки, то число ступенів свободи для вибірок дорівнюватиме 3. Середній квадрат відхилень вибіркових середніх від загального для всього досліду середнього арифметичного (що становить 29) складатиме:

$$\frac{(33-29)^2 + (32-29)^2 + (27-29)^2 + (24-29)^2}{3} = \frac{54}{3} = 18.$$

Оскільки 18 є оцінкою $\frac{\sigma^2}{5}$, то $18 \times 5 = 90$ буде третьою оцінкою генеральної дисперсії; ця оцінка ґрунтується на трьох ступенях свободи. Число 90 є оцінкою середнього квадрата; для обчислення *суми* квадратів, що відповідають вибіркам, його потрібно помножити на число ступенів свободи, тобто на 3. У результаті сума квадратів для вибірок становитиме $90 \times 3 = 270$.

Тепер у нас є всі дані для складання таблиці дисперсійного аналізу результатів визначення врожайності ячменя.

Дисперсійний аналіз для врожайності ярового ячменя за даними попередньої таблиці

Джерело варіювання	Число ступенів свободи	Сума квадратів відхилень	Середній квадрат
Вибіркові середні	3	270	90
Об'єкти всередині вибірок	16	1264	79
Загальне	19	1534	80,7

Ця таблиця висвітлює два чудові факти. По-перше, загальна сума квадратів поділяється на дві частини, зумовлені структурою нашого експерименту: загальне варіювання дорівнює сумі сум квадратів відхилень, зумовлених варіюванням вибіркових середніх і варіюванням об'єктів усередині вибірок $(1534 = 270 + 1264)$. По-друге, число ступенів свободи також поділяється на дві частини, а саме загальне число ступенів свободи дорівнює сумі ступенів свободи для вибіркових середніх і об'єктів усередині вибірок $(19 = 3 + 16)$. Оскільки всі обчислення виконувалися незалежно одне від одного, то вищесказане відзначає властивість складеності сум квадратів і ступенів свободи. Такий поділ загальної суми квадратів і загального числа ступенів свободи й називається дисперсійним аналізом.

Однак дисперсійний аналіз не обмежується лише розкладанням сум квадратів і числа ступенів свободи на складові. Як би елегантно це не виглядало, завданням дисперсійного аналізу є перевірка нульової гіпотези щодо однаковості середніх усіх вибірок, і для перевірки такої нульової гіпотези використовується відповідний критерій значущості. Крім того, під час проведення дисперсійного аналізу обчислення зазвичай виконуються трохи інакше: наведені вище докладні обчислення було зроблено лише з метою детального розгляду того, як отримують необхідні суми квадратів і ступені свободи й чому вони мають властивість складеності.

Повернімося до вихідної таблиці з урожаями ячменя та подивимося на неї з іншого боку. Забудьмо про те, що це умовний приклад, та уявімо, що це результати експерименту з порівняння врожаїв чотирьох різних сортів ячменя. Ми не маємо жодної попередньої інформації про ці сорти й після проведення експерименту повинні провести статистичний аналіз його результатів.

Оскільки в нас є чотири варіанти (чотири сорти), то нульова гіпотеза для дисперсійного аналізу звучатиме так: середні чотирьох генеральних сукупностей, з яких зроблено вибірки, однакові, тобто $\mu_1 = \mu_2 = \mu_3 = \mu_4$. Якщо нульова гіпотеза правильна, то всі чотири вибіркові середні є оцінками однієї й тієї самої величини. Альтернативна гіпотеза звучить так: принаймні одна генеральна середня не дорівнює решті.

Зробити вибір між цими гіпотезами нам дозволяє порівняння різних оцінок дисперсій, які були виконані нами вище. Оцінка загальної дисперсії, власне кажучи, особливого інтересу не становить, а от зіставлення оцінок дисперсій на підставі вибіркових середніх і всередині вибірок має виняткове значення для перевірки нульової гіпотези. Порівняння дисперсій здійснюється з допомогою критерію значущості F, запропонованого Дж. Снедекором і названого так на честь Р. Фішера. Критерій F у підручниках з біологічної статистики дуже часто називають критерієм Фішера, що є зовсім неправильним. Фішер для порівняння дисперсій запропонував трохи інший, менш зручний критерій значущості, який наразі не використовується.

Для обчислення величини критерію F ми ділимо дисперсію, оцінену за вибірковими середніми, на дисперсію, оцінену за варіюванням усередині вибірок, і порівнюємо отриману величину з табличним значенням для відповідного рівня значущості (точніше, ми ділимо не оцінки дисперсій, а середні квадрати). Нульова гіпотеза відхиляється в тому випадку, якщо обчислене значення дорівнює або перевершує табличне (як і для критерію Стьюдента). Однак таблиця критичних значень критерію F організована складніше, ніж таблиця значень критерію Стьюдента, оскільки в ній ураховуються дві групи ступенів свободи — для однієї та для другої дисперсій. Таблицю критичних значень критерію наведено в додатку 3.

Для розглянутого прикладу $F = \dfrac{90}{79} = 1{,}14$. Це значення дозволяє нам прийняти нульову гіпотезу щодо однаковості генеральних середніх сукупностей, з яких зроблено наші вибірки, оскільки воно набагато менше від табличного значення для рівня значущості 0,05 і числа ступенів свободи 3 для варіантів і 16 — для варіювання всередині вибірок. У такому висновку немає нічого дивного, адже ми знаємо, що чотири вибірки насправді було зроблено з однієї й тієї самої сукупності. Тож два середні квадрати є оцінкою однієї й тієї ж величини — генеральної дисперсії, і відмінності між ними — не більш ніж результат вибіркового варіювання, тобто є випадковими. З цієї причини ми не можемо вважати, що між чотирма вибірковими середніми є значущі відмінності, і хоча для першої вибірки середнє дорівнює 33 ц/га, а для четвертої — 24, відмінності між ними зумовлені випадковими причинами.

Наведену вище таблицю дисперсійного аналізу подано лише з метою ілюстрації розкладання загальної суми квадратів і загального числа ступенів свободи на складові. Зазвичай таблицю дисперсійного аналізу наводять у трохи іншому вигляді. Один з можливих виглядів подання аналізу дисперсій показано в таблиці нижче. У науковій літературі в таких таблицях число ступенів свободи часто позначають як df (degree freedom), суму квадратів як — SS (sum square), середній квадрат — як MS (medium square). Рівень значущості, як звичайно, позначають літерою P. У таблиці можна також навести критичне табличне значення (у таблиці нижче воно показано для рівня значущості 0,05), однак якщо наводити фактичний рівень значущості, то в цьому немає ніякої необхідності.

Таблиця дисперсійного аналізу даних за врожайністю ярового ячменя

Джерело варіювання	SS	df	MS	F	P	F критичне
Варіанти (сорти)	270	3	90	1,14	$>0{,}05$ (фактичне значення 0,36)	3,24
Залишкове (усередині груп)	1264	16	79			
Загальне	1534	19				

Отже, дисперсійний аналіз дозволив нам правильно прийняти нульову гіпотезу щодо однаковості генеральних середніх варіантів, з яких було зроблено вибірки. Однак що сталося б, якби дані вихідної таблиці про врожайність ми порівняли з використанням критерію Стьюдента? Порівняймо за цим критерієм, наприклад, урожаї вибірок 1 і 4. Скориставшись формулами (6) і (8), обчислимо величину критерію t. Для тренування читачам надається можливість виконати ці обчислення самостійно. У нас має вийти $t = 2{,}54$ (якщо в читачів вийшло інше значення, перевірте правильність обчислень). Для $5+5-2=8$ ступенів свободи отримане значення перевершує табличне для рівня значущості 0,05. Отже, якщо ми застосуємо неправильний метод аналізу, то можемо знайти відмінність, якої насправді не існує (вибірки зроблено з однієї й тієї самої сукупності). Використання дисперсійного аналізу дозволяє уникнути такої помилки.

Суть дисперсійного аналізу полягає в тому, що ми обчислюємо оцінку дисперсії за варіюванням окремих даних відносно вибіркових середніх (дисперсію, зумовлену вибірками, або залишкову дисперсію), а також оцінку дисперсії, зумовлену варіюванням вибіркових середніх відносно загального середнього арифметичного для досліду в цілому. Якщо ми маємо однакові середні генеральних сукупностей, з яких зроблено вибірки, тобто нульова гіпотеза є правильною, то ці дві величини служать оцінками однієї й тієї самої дисперсії. І відмінність між цими дисперсіями (оцінювана за критерієм F) зумовлена лише вибірковим варіюванням. У такому випадку величина обчисленого критерію F відрізняється від одиниці незначною мірою, і ймовірність того, що вона випадково перевершила табличне значення, незначна й визначається рівнем значущості.

Якщо ж генеральне середнє арифметичне хоча б однієї із сукупностей, з яких зроблено вибірки, відрізняється від інших, то в оцінку дисперсії з допомогою вибіркових середніх (тобто в середній квадрат варіантів) додатковий компонент, що приводить до її зростання відносно залишкової дисперсії, критерій F починає перевершувати табличне значення, і ми відхиляємо нульову гіпотезу.

У літературі залишкове варіювання часто називають помилкою, а середній квадрат залишкового варіювання — середнім квадратом помилки. Це цілком правильно, якщо лишень не забувати, що в поняття «помилка» тут вкладається статистичний зміст. Середній квадрат помилки справді є найкращою оцінкою загальної дисперсії, зумовленої варіюванням окремих значень усередині вибірок (варіантів), а корінь квадратний з цієї величини є найкращою оцінкою *стандартного відхилення* наших даних, спільного для всіх варіантів досліду, цілком еквівалентного стандартному відхиленню для однієї вибірки, що обчислюється за формулами (2) або (3). Використовуючи цю величину, ми можемо обчислити також звичайну стандартну помилку для кожного середнього, поділивши стандартне відхилення на корінь квадратний із числа повторів у кожній вибірці. Для прикладу з ячменем стандартна помилка дорівнюватиме: $\sqrt{\frac{79}{5}} = 3,97,$ і оскільки в кожному варіанті в нас є те саме число повторень (5), то стандартна помилка буде однаковою для кожного вибіркового середнього. Зверніть увагу на те, що було б украй неправильним підходом обчислювати стандартні помилки для кожного вибіркового середнього, виходячи з даних кожної вибірки окремо.

Якщо під час проведення досліду виявиться, що різні варіанти мають різне число повторень, то стандартні помилки вибіркових середніх можуть бути різними. Однак слід прагнути того, щоб у кожному варіанті було однакове число повторень — у цьому випадку полегшуються обчислення під час проведення множинних порівнянь, які ми розглянемо нижче.

Отже, озброївшись теоретичними відомостями про проведення дисперсійного аналізу, розгляньмо спочатку, як слід виконувати обчислення. У загальному випадку ми проводимо експеримент, що має l варіантів, кожен з яких представлений n_i числом повторень, що може бути однаковим для всіх варіантів і полегшує обчислення, а може бути й різним. Загальна схема подібного експерименту й деякі необхідні проміжні обчислення наведено в таблиці нижче (як X_{li} подано конкретні значення аналізованої змінної l-*го* варіанта i-го повторення).

Повторення	**Варіант досліду**				**У цілому за дослідом**
	A_1	**A_2**	**...**	**A_l**	
1	X_{11}	X_{21}		X_{l1}	
2	X_{12}	X_{22}		X_{l2}	
...					
n_i	X_{1i}	X_{2i}		X_{li}	
Суми за варіантами V	$\sum X_{1i}$	$\sum X_{2i}$		$\sum X_{li}$	$\sum V$
Вибіркові середні $\bar{X}_l$	$\frac{V_1}{n_1}$	$\frac{V_2}{n_2}$		$\frac{V_l}{n_l}$	
Квадрати сум за варіантами V^2	$\left(\sum X_{1i}\right)^2$	$\left(\sum X_{2i}\right)^2$		$\left(\sum X_{li}\right)^2$	
$\frac{V^2}{n}$	$\frac{V^2}{n_1}$	$\frac{V^2}{n_2}$		$\frac{V^2}{n_l}$	$\sum \frac{V^2}{n}$

Загальна сума квадратів відхилень зазвичай позначається як C_Y, сума квадратів між вибірками (варіантами досліду) — як C_V, а сума квадратів усередині вибірок (залишкова, помилки) — як C_Z. Для проведення дисперсійного аналізу нам слід розкласти загальну суму квадратів відповідно до формули: $C_Y = C_V + C_Z$.

Загальне число об'єктів, що зазвичай позначається як N, у нас дорівнює: $n_1 + n_2 + ... + n_l$. Якщо в кожному варіанті є однакове число повторень (що дорівнює n), то загальне число об'єктів просто дорівнює nl; якщо число повторень різне, то потрібно скласти їх. Загальне число ступенів свободи дорівнює $(N-1)$, число ступенів свободи для варіантів — $(l-1)$, а число ступенів свободи для залишкового варіювання — $(N-l)$. Як і загальна дисперсія, загальне число ступенів свободи також розчленовується на дві частини: $N-1=(l-1)+(N-l)$.

Обчислення, як правило, виконують у такій послідовності.

1) Обчислюють загальне число об'єктів: $N=n_1+n_2+...+n_l$, або у випадку однакового числа повторень у кожному варіанті: $N=nl$.
2) Обчислюють так званий коригувальний фактор (поправку), поділивши квадрат суми всіх значень на загальне число об'єктів: $C=\dfrac{\left(\sum X_{ij}\right)^2}{N}$.
3) Обчислюють загальну суму квадратів як різницю між сумою квадратів усіх значень і коригувальним фактором: $C_y=\left(\sum X_{ij}\right)^2-C$.
4) Обчислюють суму квадратів для варіантів як різницю між сумою квадратів сум для варіантів, ділених на число повторень у варіанті, і коригувальним фактором: $C_V=\sum\dfrac{V^2}{n}-C$.
5) Обчислюють залишкову суму квадратів як різницю між загальною сумою квадратів і сумою квадратів для варіантів: $C_Z=C_Y-C_V$.
6) Ділять C_V і C_Z на відповідні кількості ступенів свободи, обчислюючи середні квадрати (оцінки дисперсії).
7) Розраховують значення критерію F.
8) Заповнюють таблицю дисперсійного аналізу й роблять висновок щодо істотності відмінностей між варіантами.

Якщо критерій F виявився меншим за табличне значення, то після проведення дисперсійного аналізу роблять висновок про несуттєві відмінності між середніми значеннями (приймають нульову гіпотезу), і на цьому аналіз завершується. Якщо ж дисперсійний аналіз дозволив відхилити нульову гіпотезу, то наступне питання, що при цьому виникає, можна сформулювати так: а між якими саме середніми є істотні відмінності?

Відповісти на це питання дозволяють методи множинного порівняння вибіркових середніх. Таких методів є кілька, одним з найбільш уживаних є використання критерію найменшої істотної різниці (НІР). У пропонованому посібнику інші методи множинних порівнянь не розглядатимуться.

Як використання критерію НІР, так і проведення обчислень за наведеною вище схемою розглянемо на конкретному прикладі. У ньому ми використаємо ті самі дані про врожайність сортів ячменя, наведені на початку цього розділу, однак трохи змінимо їх. Уявімо, що вибірка 2 є вибіркою сорту ячменя, що має врожайність, яка перевищує врожайність інших сортів на 10 ц/га, тобто є вибіркою з генеральної сукупності, що має середнє, яке дорівнює 40 ц/га. Для того щоби внести цю зміну, до кожного повторення цієї вибірки додамо цифру 10. Решту вибірок залишимо без змін. Отже, у нашому досліді є чотири варіанти $(l=4)$, три з яких є вибірками з генеральних сукупностей, що мають однакові середні арифметичні, а одна являє собою сукупність, генеральне середнє якої більше за такі трьох інших. Для надання нашим даним наукового вигляду назвемо вибірки 1, 2, 3 і 4 сортами ярового ячменя А, Б, В і Г відповідно.

У зміненому вигляді таблиця вихідних даних набуде такого вигляду.

Урожаї чотирьох сортів ярового ячменя

Урожай за повтореннями, ц/га	Сорт А	Сорт Б	Сорт В	Сорт Г	За дослідом у цілому
	38, 30, 40, 22, 35	39, 37, 30, 49, 55	11, 31, 17, 37, 39	21, 23, 28, 27, 21	
N	5	5	5	5	
Суми за варіантами V	165	210	135	120	630
$\bar{X}$	33	42	27	24	
Квадрати сум за варіантами V^2	27 225	44 100	18 225	14 400	
$\frac{V^2}{n}$	5445	8820	3645	2880	20 790

Тепер виконаємо обчислення за наведеною схемою.

1) $N=nl=5\cdot4=20$;

2) коригувальний фактор:

$$C=\frac{\left(\sum X_{lj}\right)^2}{N}=\frac{(38+30+\ldots+21)^2}{N}=\frac{630^2}{20}=19845;$$

3) загальна сума квадратів:

$$C_Y=\sum X_{li}^2-C=\left(38^2+30^2+\ldots+21^2\right)-19845=22054-19845=2209;$$

4) сума квадратів для варіантів:

$$C_V=\sum\frac{V^2}{n}-C=20790-19845=945;$$

5) залишкова сума квадратів: $C_Z=C_Y-C_V=2209-945=1264.$

Подальші обчислення наведено в таблиці дисперсійного аналізу нижче. Відзначимо, що число ступенів свободи для загальної суми квадратів дорівнює $20-1=19$; для варіантів — $4-1=3$; для залишкової дисперсії — $19-3=16$.

Таблиця дисперсійного аналізу змінених даних про врожайність ярового ячменя

Джерело варіювання	SS	Df	MS	*F*	*P*	*F* критичне
Варіанти (сорти)	945	3	315	3,99	$<0{,}05$ (фактичне значення 0,027)	3,24
Залишкове	1264	16	79			
Загальне	2209	19				

Отже, у зміненому вигляді результати нашого експерименту приводять до відхилення нульової гіпотези щодо однаковості генеральних середніх усіх чотирьох сукупностей, які становлять сорти ярового ячменя А, Б, В і Г.

Перед тим як порівняти середні з допомогою критерію найменшої істотної різниці, зверніть увагу на один цікавий факт. Порівняймо таблиці дисперсійного аналізу для початкових даних і для змінених даних. Впадає в око, що сума квадратів залишкового варіювання не змінилася і залишилася такою, що дорівнює 1264 (а середній квадрат залишився таким, що дорівнює 79). Чому так сталося? Справа в тому, що варіювання даних усередині груп залишилося таким самим, яким і було. За зміни даних ми додали ту саму цифру до кожного числа, але при цьому на те саме число змінилося й вибіркове середнє. В остаточному підсумку ця сума квадратів не змінилася й середній квадрат помилки продовжує оцінювати генеральну дисперсію (однакову для всіх чотирьох варіантів). Водночас середній квадрат варіантів в останній таблиці оцінює суму генеральної дисперсії й додаткового компонента, що виникає через різницю між генеральними середніми різних варіантів та який привів до збільшення цього середнього квадрата.

Для використання критерію найменшої істотної різниці найбільш зручними є результати експериментів, у яких у кожному варіанті є однакове число повторень. У цьому випадку ми маємо *одну* величину НІР для порівняння всіх середніх за варіантами. Для обчислення цієї величини спочатку обчислюється узагальнена помилка різниці середніх, що дорівнює квадратному кореню з подвоєного середнього квадрата залишкової дисперсії, діленого на число повторів у одному варіанті. Оскільки середній квадрат залишкового варіювання є вибірковою оцінкою дисперсії, то ми цілком обґрунтовано можемо позначити його, як і раніше, s^2 (формула (2)). Помилку різниці середніх ми позначили як s_d (формула (6)). Отже, для наведених вище результатів дисперсійного аналізу:

$$s_d = \sqrt{\frac{2s^2}{n}} = \sqrt{\frac{2\cdot 79}{5}} = 5{,}6.$$

Далі ми множимо помилку різниці на критерій Стьюдента, значення якого вибирається з таблиці для прийнятого рівня значущості й числа ступенів свободи залишкової суми квадратів: $\text{НІР} = ts_d$. На прийнятому рівні значущості істотною вважається різниця між середніми, яка за абсолютною величиною дорівнює або перевершує НІР.

Оскільки у виконаному нами аналізі це число ступенів свободи дорівнює 16, то табличне значення критерію Стьюдента дорівнює 2,12 для рівня значущості 0,05. Тепер у нас є всі можливості для обчислення критерію найменшої істотної різниці: $\text{НІР}_{05} = 2{,}12\cdot 5{,}6 = {}$ $= 11{,}9$ (ц/га); позначення НІР_{05} показує, що ми обчислили найменшу істотну різницю саме для рівня значущості 0,05.

Якщо порівняти наші чотири вибіркові середні з допомогою цього критерію, то можна зробити висновок, що сорт Б перевершує за врожайністю сорти Г і В. Читачі можуть самостійно в цьому переконатися.

Цікавіші справи із сортом А. Різниця у врожаях між сортами А і Б становить $(42 - 33 = 9)$ ц/га. Ця різниця менша за величину найменшої істотної різниці, а отже, ми не можемо стверджувати, що між цими сортами є статистично істотна різниця. Однак цей приклад конструювали ми самі, тому ми знаємо, що ця різниця реально існує! Що ж сталося? Через примхливість випадкового вибіркового варіювання ми зробили помилку другого роду — не виявили реальної відмінності. Причому цю помилку ми зробили не через якісь неточності в проведенні досліду. Можливість такої помилки закладена в самому вибірковому методі.

Тут важливо зрозуміти таке: якщо статистичний аналіз не дозволив нам стверджувати про відмінності в урожайності сортів А і Б, то це не означає, що ми довели існування відсутності відмінності між генеральними середніми. А можливість помилок як першого, так і другого роду не є перешкодою для формулювання правильних висновків. У науковій роботі експеримент, як правило, виконується щонайменше двічі, беруться до уваги також дані інших дослідників, і лише на підставі такого комплексного аналізу робиться остаточний висновок.

У розглянутому прикладі з урожаями ячменя автор не вигадав цифри для кожного варіанта. Було зроблено випадкові вибірки з генеральної сукупності з $\mu = 30$ і $\sigma^2 = 100$, а потім до кожного значення вибірки 2 було додано цифру 10, щоб перетворити цю вибірку на вибірку із сукупності з $\mu = 40$.

Проведімо цей експеримент ще раз. Знову вважатимемо, що це врожаї ячменя чотирьох сортів, і потім до врожаїв сорту Б додамо цифру 10. Автор отримав такі дані:

Урожаї чотирьох сортів ярового ячменя — другий експеримент

	Сорт А	Сорт Б	Сорт В	Сорт Г
Урожай за повтореннями, ц/га	23, 26, 28, 33, 34	32, 40, 41, 44, 44	24, 25, 27, 28, 35	15, 29, 29, 35, 37
$\bar{X}$	28,8	40,2	27,8	29

Проведемо дисперсійний аналіз. Читачам надається можливість самостійно виконати всі обчислення. Має вийти така таблиця:

Джерело варіювання	SS	df	MS	*F*	*P*	*F* критичне
Варіанти (сорти)	514,55	3	171,5	4,95	<0,05 (фактичне значення 0,012)	3,24
Залишкове	554,4	16	34,65			
Загальне	1068,95	19				

Отже, як і раніше, величина критерію *F* дозволяє відхилити нульову гіпотезу щодо однаковості всіх середніх. Обчислимо величину найменшої істотної різниці:

$$s_d = \sqrt{\frac{2s^2}{n}} = \sqrt{\frac{2 \cdot 34,35}{5}} = 3,71;$$

$$\text{НІР} = ts_d = 2,12 \cdot 3,71 = 7,9 \ \text{(ц/га)}.$$

У повторному експерименті ми робимо абсолютно правильний висновок, що сорти А, В і Г за врожайністю не розрізняються, а сорт Б перевершує кожен з них. Такого ж висновку ми дійдемо, якщо об’єднаємо результати двох наших експериментів (тобто для кожного сорту матимемо по 10 повторень). Читачі можуть у цьому переконатися самостійно.

Отже, незважаючи на можливість помилкових висновків під час статистичного аналізу, дворазове повторення експерименту дозволило уникнути подібної помилки.

Розгляньмо ще один приклад, у якому ми, по-перше, закріпимо вміння виконувати обчислення під час проведення дисперсійного аналізу, а по-друге, навчимося порівнювати вибіркові середні за різного числа повторень у варіантах. Нижче наведено вплив трьох форм нітратних добрив на врожай вівсяниці лучної у вегетаційних дослідах (тут для нас не мають значення конкретні форми добрив, тому позначимо їх просто цифрами).

Урожаї вівсяниці (у г на посудину)

Урожай за повтореннями, ц/га	**Контроль (без добрив)**	**Форма нітратних добрив**			**За дослідом у цілому**
		1	**2**	**3**	
	16, 17, 14, 16	**29, 30, 30, 28**	**26, 29, 27, 27, 26, 28**	**20, 20, 21, 18, 21, 19**	
N	4	4	6	6	
Суми за варіантами V	63	117	163	119	462
$\bar{X}$	15,8	29,3	27,2	19,8	
Квадрати сум за варіантами V^2	3969	13 689	26 569	14 161	
$\frac{V^2}{n}$	992,25	3422,25	4428,17	2360,17	11202,84

Виконаємо обчислення за знайомою схемою.

1) $N = 4 + 4 + 6 + 6 = 20$ (складемо!);

2) коригувальний фактор:

$$C = \frac{\left(\sum X_{lj}\right)^2}{N} = \frac{\left(16 + 17 + \ldots + 19\right)^2}{N} = \frac{462^2}{20} = 10672{,}2;$$

3) загальна сума квадратів:

$$C_Y = \sum \mathrm{X}_{li}^2 - C = \left(16^2 + 17^2 + \ldots + 192\right) - 19845 =$$
$$= 11224 - 10672{,}2 = 551{,}8;$$

4) сума квадратів для варіантів:

$$C_V = \sum \frac{V^2}{n} - \mathrm{C} = 11202{,}84 - 10672{,}2 = 530{,}64;$$

5) залишкова сума квадратів:

$$C_Z = C_Y - C_V = 551{,}8 - 530{,}64 = 21{,}16.$$

Тепер заповнимо таблицю дисперсійного аналізу (читачі повинні самі зрозуміти, як обчислено ступені свободи для відповідних сум квадратів).

Таблиця дисперсійного аналізу змінених даних про врожайність ярового ячменя

Джерело варіювання	SS	df	MS	*F*	*P*	*F* критичне
Варіанти (форми Нітрогену)	530,64	3	176,88	134	$<0{,}01$ (фактичне значення $1{,}5\cdot 10^{-11}$)	3,24
Залишкове	21,16	16	1,32			
Загальне	551,8	19				

Проведений дисперсійний аналіз виявив вплив форм нітратних добрив на надзвичайно високому рівні значущості, що однозначно дозволяє відхилити нульову гіпотезу щодо відсутності різниці між варіантами досліду. Однак які різниці є істотними, а які — ні?

Оскільки в цьому досліді в нас різна кількість повторень у різних варіантах, то єдиний критерій НІР для порівняння всіх середніх ми обчислити не можемо. Залежно від порівнюваних значень критерії будуть трохи відрізнятися. Власне кажучи, ця відмінність буде зумовлена різними помилками різниці середніх; табличні значення критерію t залежать лише від числа ступенів свободи залишкової суми квадратів.

За порівняння середніх арифметичних варіантів, що мають по чотири повторення, помилка різниці обчислюється, як і раніше:

$$s_d=\sqrt{\frac{2s^2}{n}}=\sqrt{\frac{2\cdot 1{,}32}{4}}=0{,}81.$$

Абсолютно аналогічно обчислюється помилка різниці середніх у порівнянні варіантів, що мають по шість повторень:

$$s_d=\sqrt{\frac{2s^2}{n}}=\sqrt{\frac{2\cdot 1{,}32}{6}}=0{,}66.$$

А в разі порівняння варіантів, що мають чотири $(n_1=4)$ і шість $(n_2=6)$ повторень, для обчислення помилки різниці використовують таку формулу:

$$s_d=\sqrt{s^2\,\frac{n_1+n_2}{n_1 n_2}}=\sqrt{1{,}32\cdot\frac{4+6}{4\cdot 6}}=0{,}74.$$

Тепер нам ніщо не заважає обчислити величини НІР, придатні для оцінювання істотності різниці між середніми різних варіантів. Для порівняння за $n_1=n_2=4$ маємо $\text{НІР}_{05}=ts_d=2{,}12\cdot 0{,}81=1{,}7$. За $n_1=n_2=6$ маємо $\text{НІР}_{05}=ts_d=2{,}12\cdot 0{,}66=1{,}4$. І нарешті, за $n_1=4$ і $n_2=6$ маємо $\text{НІР}_{05}=ts_d=2{,}12\cdot 0{,}74=1{,}6$.

Застосування критерію найменшої істотної різниці до даних нашого досліду дозволяє зробити висновок, що в контрольному варіанті врожай вівсяниці нижчий, ніж у разі застосування будь-якої форми Нітрогену. Порівняння ж форм Нітрогену між собою показало, що найменше збільшення врожаю спостерігається в останньому варіанті (форма нітратного добрива 3). Урешті-решт, із двох форм Нітрогену, що дали найкращий результат, форма 2 істотно перевершує форму 3, хоча різниця між ними не дуже значна.

Щодо дисперсійного аналізу даних дискретної природи, таких, як схожість насіння, укорінення черешків, поширеність захворювань рослин тощо, схема дисперсійного аналізу не відрізняється від описаної вище, якщо в кожному повторенні використано досить велику кількість об'єктів (бажано не менш ніж 50, у крайньому випадку — не менш ніж 20).

Дискретні дані, загалом кажучи, можуть бути двох типів. В одному типі ми аналізуємо чисельності об'єктів, що трапляються або на певній площі, або за певний проміжок часу. Для прикладу проведемо дисперсійний аналіз результатів експерименту, у якому аналізували вплив чотирьох способів обробки ґрунту на чисельність бур'янів. Цей дослід проводили в чотирьох повтореннях. Результати наведено в таблиці нижче. Обчислення будуть виконані без додаткових пояснень.

Вплив чотирьох способів обробки ґрунту на кількість бур'янів, шт. на ділянку площею 100 м²

Число бур'янів за повтореннями, шт.	**Спосіб обробки ґрунту**				**За дослідом у цілому**
	1	**2**	**3**	**4**	
	210, 207, 288, 205	**169,162, 210, 165**	**160, 129, 103, 125**	**42, 75, 64, 88**	
n	4	4	4	4	
Суми за варіантами V	910	706	517	269	2402
$\bar{X}$	227,5	176,5	129,3	67,3	
Квадрати сум за варіантами V^2	828 100	498 436	267 289	72 361	
$\frac{V^2}{n}$	207 025	124 609	66822,25	18090,25	416546,5

1) $N = 4 \cdot 4 = 16$ (множимо!);

2) $C = \frac{\left(\sum X_{lj}\right)^2}{N} = \frac{(210 + 207 + \ldots + 88)^2}{N} = \frac{2403^2}{16} = 360600{,}25;$

3) $C_Y = \sum X_{li}^2 - C = \left(210^2 + 207^2 + \ldots + 882\right) - 360600{,}25 =$
$= 425752 - 360600{,}25 = 65151{,}75;$

4) $C_V = \sum \frac{V^2}{n} - \mathrm{C} = 416546{,}5 - 360600{,}25 = 55946{,}25;$

5) $C_Z = C_Y - C_V = 65151{,}75 - 55946{,}25 = 9205{,}5.$

Дисперсійний аналіз даних про вплив способів обробки ґрунту на кількість бур'янів

Джерело варіювання	SS	df	MS	*F*	*P*	*F* критичне
Варіанти (способи обробки ґрунту)	55946,25	3	18648,75	24,31	< 0,01 (фактичне значення 0,000022)	3,49
Залишкове	9205,5	12	767,125			
Загальне	65151,75	15				

Оскільки в нас однакова кількість повторень у кожному варіанті, обчислюємо загальну НІР для рівня значущості 0,05:

$$s_d = \sqrt{\frac{2s^2}{n}} = \sqrt{\frac{2 \cdot 767{,}125}{4}} = 19{,}6.$$

Для $P < 0{,}05$ $t = 2{,}18$; $\text{НІР}_{05} = 2{,}18 \cdot 19{,}6 = 42{,}7$.

Отже, між усіма способами обробки ґрунти є істотні відмінності. Найбільше бур'янів виростало на ділянках, які обробляли способом 1, а найменше — на ділянках, оброблених способом 4.

За іншого типу дискретних даних ми аналізуємо частки об'єктів, що мають ознаку, яка цікавить нас. Ці частки можуть бути виражені в частках одиниці або у відсотках. Для прикладу розглянемо дані про ураженість колосся восьми сортів пшениці ріжками. Рослини попередньо штучно заражали збудником шляхом нанесення спор ріжок на квітки. Із заражених рослин збирали насіння, яке потім висівали на дослідній ділянці, і проводили підрахунки здорових і уражених колосків. У кожному повторенні кожного варіанта аналізували по 500 колосків.

Ураженість колосся восьми сортів пшениці ріжками

Ураженість колосся ріжками за повтореннями, %	Сорт пшениці								За дослідом у цілому
	1	2	3	4	5	6	7	8	
	65; 67; 69	67; 64; 68	1; 0,4; 0,5	12; 11; 13	3; 1,5; 2,5	64; 62; 60	1; 1; 0,8	56; 55; 53	
n	3	3	3	3	3	3	3	3	
Суми за варіантами V	201	199	1,9	36	7	186	2,8	164	797,7

Ураженість колосся ріжками за повтореннями, %	Сорт пшениці								За дослідом у цілому
	1	2	3	4	5	6	7	8	
	65; 67; 69	67; 64; 68	1; 0,4; 0,5	12; 11; 13	3; 1,5; 2,5	64; 62; 60	1; 1; 0,8	56; 55; 53	
$\bar{X}$	67	66,33	0,63	12,00	2,33	62,00	0,93	54,67	
Квадрати сум за варіантами V^2	40401	39601	3,61	1296	49	34596	7,84	26896	
$\frac{V^2}{n}$	13467	13200,3	1,20	432	16,3	11532	2,6	8965,3	47616,7

1) $N = 8 \cdot 3 = 24;$

2) $C = \frac{\left(\sum X_{lj}\right)^2}{N} = 26513,55;$

3) $C_Y = \sum X_{li}^2 - C = 21136;$

4) $C_V = \sum \frac{V^2}{n} - \mathrm{C} = 416546,5 - 360600,25 = 21103,15;$

5) $C_Z = C_Y - C_V = 21136 - 26513,55 = 32,85.$

Дисперсійний аналіз даних про ураженість ріжками восьми сортів пшениці

Джерело варіювання	SS	df	MS	*F*	*P*	*F* критичне
Варіанти (способи обробки ґрунту)	21103,15	7	3014,74	1470,6	$<0,01$ (фактичне значення $3 \cdot 10^{-21}$)	2,66
Залишкове	32,85	16	2,05			
Загальне	21 136	23				

$$s_d = \sqrt{\frac{2s^2}{n}} = \sqrt{\frac{2 \cdot 2,05}{3}} = 1,17.$$

Для $P < 0,05$ $t = 2,12$; $\text{НІР}_{05} = 2,12 \cdot 1,17 = 2,48.$

Читачам надається можливість самостійно порівняти ураженість сортів пшениці за критерієм найменшої істотної різниці та зробити висновок про те, як групуються ці вісім сортів за стійкістю до ріжок.

Завершальні зауваження з приводу дисперсійного аналізу

Дисперсійний аналіз даних, приклади якого наведено вище, є найпростішим випадком однофакторного дисперсійного аналізу, оскільки в досліді ми аналізували вплив на певну ознаку лише одного фактора — сортів, форм нітратних добрив, способів обробки ґрунту тощо, хоча один фактор у кожному прикладі був представлений кількома варіантами (які часто називають градаціями фактора). Однак дисперсійний аналіз є надзвичайно ефективним і для дослідження дії на результативну ознаку кількох факторів (схеми так званого факторіального дисперсійного аналізу). Загальна логіка проведення дисперсійного аналізу в останньому випадку така сама, як і для проведення однофакторного аналізу, але обчислення при цьому ускладнюються. Ускладнюються й методи множинних порівнянь середніх. Для рівня шкільних дослідів (а також, між іншим, і для величезної кількості наукових експериментів у ВНЗ і НДІ), цілком достатньо однофакторного дисперсійного аналізу. У випадку ж, якщо читачів зацікавлять методи статистичного аналізу більш складних схем експериментів, рекомендуємо звернутися до спеціальної навчальної літератури зі статистики.

Крім цього, проведення дисперсійного аналізу є правомірним лише за дотримання певних умов, а саме: кожен варіант має являти собою випадкову вибірку з генеральної сукупності, що має нормальний розподіл, і дисперсії генеральних сукупностей, які представлені нашими варіантами, мають бути однаковими. Існують методи, що дозволяють здійснити перевірку дотримання цих умов для даних, які опрацьовуються з допомогою дисперсійного аналізу. Ці вимоги не дуже жорсткі, оскільки критерій F малочутливий до відхилення даних від нормального розподілу і трохи більш чутливий до неоднаковості дисперсій.

Під час проведення біологічних експериментів зазначені вимоги зазвичай виконуються, навіть без особливої турботи про їх дотримання. Однак є випадки й грубих порушень відзначених вимог. У такій ситуації допомагає, як правило, певне перетворення вихідних даних (наприклад, їх логарифмування, або витяг квадратного кореня, або інше). Особливо часто можуть порушуватися умови проведення дисперсійного аналізу тоді, коли ми працюємо з даними дискретного типу й наші вибірки невеликі за обсягом. У такому випадку їх також можна перетворити, зокрема, за методом Фішера, або будувати таблиці спряженості ознак.

Розділ 6. АНАЛІЗ РЕЗУЛЬТАТІВ ТА ОБҐРУНТУВАННЯ ВИСНОВКІВ

Загальні зауваження

Аналіз результатів становить головний науковий сенс будь-якої роботи. Він полягає, власне, в поясненні тих результатів, які були отримані в ході роботи. Часто для пояснення результатів автор висуває якусь гіпотезу. До речі, цю гіпотезу можна використовувати для планування подальших досліджень (навіть з іншими виконавцями) для її підтвердження або спростування.

Слід пам'ятати, що будь-яке наукове пояснення результатів має бути обґрунтованим. Це досягається завдяки залученню статистичної обробки результатів експерименту й логічному розгляду можливих альтернативних варіантів пояснення виявленого явища чи закономірності.

Для полегшення розуміння результатів аналізу даних дуже добре використовувати схеми, графіки, діаграми, фотографії та інші наочні матеріали. Це дозволяє значно легше сприймати аргументацію автора. Крім того, ілюстративні матеріали добре демонструють розуміння автором біологічного сенсу своїх досліджень. Як сам автор, так і сторонній спеціаліст із допомогою ілюстративного матеріалу може одразу побачити як переваги, так і недоліки роботи.

І ще одне важливе зауваження. Аналізу мають підлягати всі отримані матеріали. Часто в учнівських роботах частина даних виявляється немовби зайвою. Тобто показники, які були виміряні, у роботі є, але їхній аналіз не проводиться й у висновках вони не фігурують. Зрозуміло, що це викликає здивування. Навіщо було вимірювати показник, щоб потім ніяк його не використовувати? Звичайно, якщо цей показник не виявився суттєвим для формулювання основних висновків, то в них про нього можна й не згадувати. Але цей показник в аналізі згадати треба, адже саме він дає можливість виявити значення окремих досліджених показників.

На основі аналізу результатів автор роботи формулює висновки. З висновків видно, чи була досягнута мета, яка ставилася на початку роботи. Якщо мети не було досягнуто, то це потребує пояснення. Головним недоліком висновків учнівських робіт є введення до них зайвої інформації. Не треба вносити до висновків загальновідомих біологічних істин або тих фактів, які не були доведені результатами дослідження. Кількість висновків не має бути великою. Як показує практика, в учнівській роботі може бути в середньому від двох до п'яти основних висновків. Для отримання більшого їх числа масштаби дослідження мають бути надто вже грандіозними. Ще одна часта помилка у формулюванні висновків полягає в тому, що до них частково або навіть повністю не включають отримані учнями результати. Ця помилка безпосередньо пов'язана з попередньою. Бажання включити до висновків солідні біологічні відомості призводить до збільшення обсягу цього розділу, а отже, і до видалення з висновків власних результатів.

Аналіз добре виконаної дослідної роботи

Щоби проілюструвати згадані вище особливості аналізу результатів та обґрунтування висновків, розглянемо одну учнівську дослідну роботу. Вона є прикладом вдалої наукової роботи, бо у 2003 році посіла перше місце на всеукраїнському конкурсі МАН (секція лісового господарства) й отримала призове місце на міжнародній екологічній олімпіаді в м. Стамбулі (у складі спільного екологічного проекту, друга частина якого виконувалася в м. Луганську).

Тема роботи. Пошкодження рослин комахами-фітофагами в районі автошляхів з високою інтенсивністю руху транспорту

Виконала роботу учениця НВК № 46 ім. М. В. Ломоносова м. Харкова Олена Якунчикова.

Головною метою роботи було дослідження впливу забруднення, джерелом якого є автотранспорт, на деревні рослини й комах-фітофагів, що їх пошкоджують.

Об'єкти та методи досліджень

Основним об'єктом досліджень були листяні породи дерев, що ростуть у місті Харкові. Було обрано три парки, в яких у весняний (квітень) та літній (серпень) період здійснювалися дослідження. Збір матеріалу проводився в роки масового спалаху розмноження комах. У парках ім. Артема, ім. Маяковського та ім. Горького було обрано по три ділянки на відстані 25, 50 та 100 м від дороги — основного забруднюючого фактора, що досліджувався. На всіх деревах, що були обрані для дослідження, обиралося по одній модельній гілці (довжиною приблизно 1 м). На цій гілці підраховувалася кількість ушкодженого та неушкодженого комахами-фітофагами листя. На основі цих даних розраховувалися:

1) середній відсоток ураженого листя в кожного дерева та на кожній ділянці;
2) середній відсоток пошкодженого листя в різних порід дерев.

Листя, що було зібрано, висушувалося, і вимірювалася площа цілої та ушкодженої поверхні палеточним методом. На основі цих даних розраховувалися:

1) середній відсоток ураженої листової пластинки в кожного дерева та на кожній ділянці;
2) середній відсоток пошкодженої листової пластинки в різних порід дерев.

Для перевірки достовірності результатів було обрано двофакторний дисперсійний аналіз із повтореннями.

Коментар до методики проведення дослідження

Відносно методики проведення дослідження слід зазначити, що планування експерименту було здійснено з урахуванням того, що обробка результатів буде виконуватися з допомогою дисперсійного аналізу. Дисперсійний аналіз використовується в тих випадках, коли експерименти є більш складними й одночасно доводиться порівнювати кілька вибірок, що об'єднані в єдиний статистичний комплекс. У цьому випадку не можна використовувати t-критерій (критерій Стьюдента). Метод дисперсійного аналізу засновано на розділенні загальної дисперсії статистичного комплексу на компоненти, що її складають (звідси й походить назва методу). Порівнюючи окремі дисперсії між собою з допомогою F-критерію, можна визначити, яку долю загальної варіації ознаки обумовлює дія на неї окремих факторів.

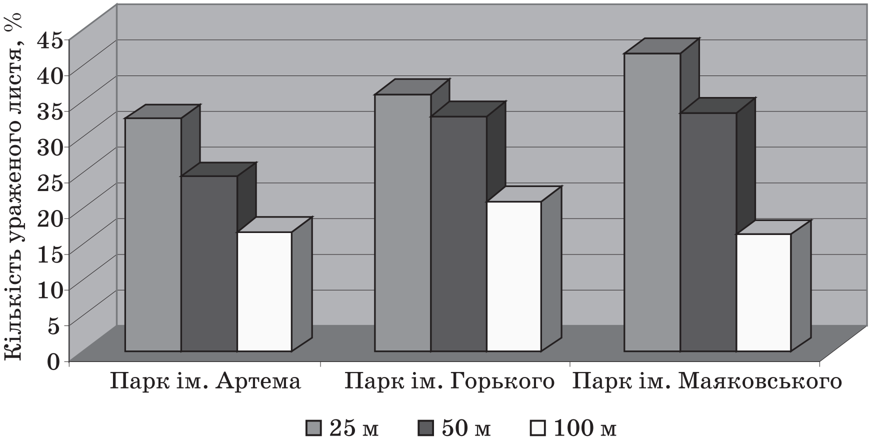
У цьому випадку досліджувалася дія на такий показник, як пошкодженість листя, двох факторів — віддалення від автомагістралі з інтенсивним рухом і територіальної локалізації (належності до того чи іншого парку).

Аналіз результатів

Як приклад аналізу результатів нижче наведемо кілька уривків з роботи.

Двофакторний дисперсійний аналіз із повтореннями показав, що в обох повтореннях дослідження (весною та влітку) різниця між ураженням листя на відстані 25, 50 та 100 м значуща ($p = 0,05$). Таким чином, результати дослідження доводять, що середній відсоток ураженого листя зменшується зі збільшенням відстані від дороги. Наприклад, у парку ім. Артема в середньому було уражено шкідниками на відстані 25 м — 43,63 %, на відстані 50 м — 28,93 %, на відстані 100 м — 11,34 %.

Залежність кількості ураженого листя від відстані до автомагістралей



Для того щоб оцінити кількісне значення ураженості листя, вимірювалася площа ураженої листової пластинки. Виявилося, що відсоток уражень зменшувався зі збільшенням відстані до дороги ($p = 0{,}05$). Наприклад, у парку ім. Артема середній відсоток ураженої поверхні весною на відстані 25 м від дороги дорівнює 2,97 %, на відстані 50 м — 2,03 %, на відстані 100 м — 0,62.

З метою оцінити стійкість до комах-фітофагів в умовах забруднення повітря отруйними газами різних порід дерев обчислювався ступінь ураження листя в кожної породи, що досліджувалася. Виявилося, що на відстані 25 м більш уражаються клен гостролистий, клен ясенолистий, клен польовий і липа, а менш уражається в'яз гладкий. Виходячи з цього, можна зробити припущення, що клен гостролистий, клен ясенолистий, клен польовий і липа менш стійкі до забруднення з автошляхів, а в'яз гладкий більш стійкий. На відстані 50 м від дороги комахи-фітофаги найчастіше поїдають горобину звичайну й гіркокаштан звичайний, які, таким чином, є найменш стійкими до забруднення. Значно менше поїдаються клен гостролистий, клен ясенолистий, липа і в'яз гладкий. Порівняно стійкими до комах і, таким чином, до забруднення є тополя і клен сріблястолистий. На відстані 100 м більш уражаються клен ясенолистий і липа, а менш — клен гостролистий і гіркокаштан звичайний.

Причиною цього явища може бути погіршення стану дерев під впливом забруднення. Ослаблення дерева викликає зменшення його здатності протистояти шкідникам. Воно може синтезувати менше захисних речовин, таких, як, наприклад, дубильні речовини. Найбільш ослабленими є дерева, що зазнають сильнішого впливу забруднюючого фактора, які знаходяться найближче до автомагістралей. Комахи-фітофаги зазвичай віддають перевагу ослабленим деревам, унаслідок чого й зосереджуються на деревах, що ростуть поблизу дороги. Якби забруднення сильніше впливало на комах, то тоді б найбільше пошкодження листя спостерігалося на середній відстані, де дерева вже є частково ослабленими, але вплив забруднення вже менший.

Коментар до аналізу роботи

Наведений аналіз у цілому відповідає вимогам, які наводились у загальних зауваженнях до цього розділу. Було проаналізовано всі параметри, що вивчалися в процесі роботи й висунуто обґрунтоване пояснення знайдених закономірностей. Отримані результати статистичної обробки свідчать про високу значимість отриманих результатів, що вказує на правильне планування й проведення досліду. Наведені діаграми показують, як можна ілюструвати результати дослідження.

Висновки роботи

Висновки роботи було сформульовано так:

Оскільки ураженість рослин комахами-фітофагами зменшується зі збільшенням відстані від дороги, то можна зробити висновок, що рослини більше страждають від забруднення дорогами, ніж комахи. У них знижується здатність опиратися ураженню комахами, знижується стійкість до фітофагів, а також до різних хвороб.

Комахи страждають від забруднення дорогами значно менше. Можливо, це обумовлено значно меншою тривалістю їх життя, міцністю покривів, природною здатністю акумулювати, переробляти чужорідні речовини. Щодо зміни стійкості рослин до комах-фітофагів під впливом забруднення дорогами можна зробити висновок, що найбільш стійкими до комах, а також, як виявилося, і до забруднення є такі породи рослин: на відстані 25 м — в'яз гладкий, на відстані 50 м — тополя та клен сріблястолистий, на відстані 100 м — клен гостролистий і гіркокаштан звичайний.

Розділ 7. ТИПОВІ ПОМИЛКИ В НАУКОВО-ДОСЛІДНИХ РОБОТАХ УЧНІВ

Для того щоби краще застосувати в практичній роботі все, про що йшла мова в попередніх розділах, ми пропонуємо вам розглянути кілька реальних дослідних робіт учнів з точки зору їх планування, проведення й аналізу. Для цього з кожної роботи наведено тему, мету, результати й висновки (зі збереженням орфографії оригіналу), а також усі наявні дані щодо методики планування й проведення дослідження.

Роботи добиралися за принципом наявності в них типових помилок, про які йдеться в наступному розділі.

Загальні зауваження

Як уже було зазначено на початку цієї книги, не буває науки шкільної, так, як не буває науки студентської, аспірантської або якоїсь іншої.

Є наука й ненаука, тобто будь-яке експериментальне дослідження або є науковою роботою, або лише має зовнішній вигляд наукової роботи, хоча до сфери науки, власне кажучи, не належить. Виникає питання: чи можна очікувати проведення справжнього наукового дослідження від учнів? Відповідь на це питання є позитивною. Незважаючи на відсутність досвіду наукової роботи в самого учня-виконавця, а часто і в учителя-керівника, обмежені методичні можливості й відсутність хоч якогось сучасного обладнання, науково-дослідні роботи учнів можуть бути надзвичайно ефективними. Особливо це стосується таких галузей, як спостереження за популяціями тварин і рослин, первинний скринінг різноманітних речовин з біологічною активністю (наприклад, тих, що стимулюють ріст, розвиток і врожайність рослин) тощо.

Проте щоб робота не була цікавою лише як ілюстрація бажання конкретного учня займатися науковою діяльністю, вона має відповідати певним вимогам. І більша частина таких вимог є абсолютними. Аналіз робіт, у яких були допущені типові відхилення від цих вимог, наведено нижче. А зараз хотілося б зробити кілька загальних зауважень відносно проведення дослідів.

По-перше, хоча розміри шкільної ділянки частіше за все не дозволяють провести польовий дослід за загальноприйнятими правилами, слід намагатися використовувати в досліді кілька повторностей, краще за все — чотири. Якщо доводиться вибирати між розмірами ділянки для окремої повторності й кількістю повторностей, то краще використовувати ділянки меншого розміру, але в кожному з варіантів мати більше повторностей, ніж проводити досліди на великих ділянках, але у двократній повторності чи взагалі без неї.

Скоріше за все, на шкільній ділянці не вдасться організувати проведення польового досліду з розміщенням повторностей у рендомізованих блоках. Тому під час розміщення конкретних ділянок конкретних варіантів слід використовувати схему повної рендомізації, тобто розміщувати ділянки випадково. Будь-якого систематичного розташування варіантів (хоч прямого, хоч зворотного, хоч із зсувом) слід уникати.

Польові досліди мають цінність лише у випадку їхнього проведення протягом кількох років, як мінімум трьох. Лише в цьому випадку можна чекати, що всі варіанти відчують на собі вплив різних погодних умов. Під час проведення однорічних дослідів може виявитися, що різниця між варіантами виникла випадково, через вплив метеорологічних умов. Особливо справедливим це є для дослідів, у яких перевіряють різні терміни посіву культури. Одного року більш пізні посіви, наприклад, можуть уникнути весняних заморозків і виявитися більш оптимальними. Але іншого року посів цієї ж культури в ті самі терміни може призвести, наприклад, до того, що в ранній період розвитку вони потраплять під вплив посухи, а більш ранній посів опиниться в кращих умовах. Тут, звичайно, може виникнути проблема у зв’язку з тим, що один учень часто не має змоги провести дво- або трирічні досліди. Але ж подібна робота може виконуватись і кількома учнями (не одночасно, а послідовно, протягом кількох років). Так, наприклад, у США й Англії проводять відомі тривалі польові досліди, розпочаті в середині XIX століття. У цьому випадку кожен учень може пред’являти свою частину роботи, а той, хто виконує завершальний етап, буде посилатися на співвиконавців.

У зв’язку з цими ж причинами документацію досліду слід доповнювати метеорологічними даними (іноді в роботах така інформація трапляється, але, на жаль, досить рідко), а аналіз результатів проводити з урахуванням, як мінімум, температури й опадів. Такого роду відомості можна знайти в районних службах або отримати самостійно в процесі досліду. Бажаним є співставлення погодних умов проведення досліду за середньодобовою температурою та кількістю опадів з багаторічними даними по декадах або хоча б по місяцях.

Крім того, польові досліди мають цінність лише з урахуванням якості ґрунту, на якому вони проводились. Особливо це стосується дослідів, метою яких є аналіз застосування добрив або різних біостимуляторів, а також порівняння різних сортів рослин. На бідних ґрунтах можна отримати вражаючі результати, які насправді можуть не мати ніякої цінності. Краще за все для розв'язання цієї проблеми раз на кілька років проводити в агрохімлабораторії аналіз ґрунту шкільної ділянки. У будь-якому випадку є дуже бажаною хоча б елементарна характеристика ґрунту за типом і фізичними властивостями, а також за величиною pH, умістом гумусу, Нітрогену, Фосфору й Калію.

У дослідах з рослинами, особливо в польових, для правильних висновків важливо проводити ретельні фенологічні спостереження для кожного з варіантів, бо динаміка росту рослин є одним з найважливіших факторів, що визначають кінцевий урожай.

Нарешті, результати досліду треба правильно статистично обробити, частіше за все з використанням дисперсійного аналізу. Хоча учень може й не мати відповідних математичних знань, інтуїтивне розуміння різниці між випадковими й закономірними відмінностями зазвичай притаманне навіть молодшим школярам, а провести відповідні розрахунки можна навіть без допомоги комп'ютера. Головне завдання вчителя в цій ситуації — надати учню необхідний алгоритм розрахунків.

Аналіз окремих не дуже вдалих дослідних робіт

Дослідна робота № 1

Вплив способу посіву на врожайність кукурудзи

Об'єкт: кукурудза сорту «Дніпровська — 315».

Мета досліду: виявити кращий спосіб посіву кукурудзи.

Площа ділянки: 100 м^2.

Місце проведення досліду: навчально-дослідна ділянка.

Схема досліду

Варіант 1 Посів 70×70 по дві-три рослини пунктирним способом		Варіант 3 Посів 60×60 по одній рослині
Варіант 2 Посів 70×20 по одній рослині		Варіант 2 Посів 70×20 по одній рослині
Варіант 3 Посів 60×60 по одній рослині		Варіант 1 Посів 70×70 по дві-три рослини пунктирним способом

Фенологічні дослідження відмінності між варіантами не виявили.

Урожай

Варіант 1 4,8 кг (48 ц/га)		Варіант 3 5,05 кг (50,5 ц/га)
Варіант 2 4,5 кг (45 ц/га)		Варіант 2 4,5 кг (45 ц/га)
Варіант 3 5 кг (50 ц/га)		Варіант 1 4,9 кг (49 ц/га)

Висновок

Найбільший урожай кукурудзи одержано з ділянки, де кукурудза посіяна за схемою 60×60 по одній рослині,— це 50,5 ц/га.

Коментар до роботи

Дослід проведено у двох повторностях, які розташовано систематично. Зауваження щодо цього викладено в загальній частині. Крім того, незрозуміло, чому було обрано саме ці три варіанти (в обґрунтуванні роботи про це вказано не було). У роботі відсутня статистична обробка, хоча значимість відмінностей між усіма варіантами підтверджується після проведення дисперсійного аналізу. Судячи з тексту, висновок зроблено лише на основі даних одного повторення (врожай 50,5 ц/га було отримано лише на одній ділянці), що, звичайно, є неправильним. Насправді максимальний урожай у середньому за даними двох повторностей дорівнює 50,25 ц/га. На жаль, у роботі відсутні дані, які є необхідними для отримання важливих висновків. Наведено лише врожай (скоріше за все зерна кукурудзи, але прямої вказівки на це немає), але відсутні такі показники, як висота рослин, середня кількість початків на одну рослину, кількість зернин у початку, маса 1 000 зернин тощо, без яких неможливо отримати відповідь на питання: завдяки чому врожаї різні в різних варіантів?

Дослідна робота № 2
Вплив якості насіння на врожай цибулі

Об'єкт: цибуля сорту «Золотиста».

Мета досліду: з'ясувати, як впливає якість насіння (величина, міцність зародка) на врожай цибулі.

Площа ділянки: 36 м^2.

Місце проведення досліду: навчально-дослідна ділянка. Попередник — бобові.

Схема досліду

Контрольна ділянка 9 м^2. Висівалося насіння різної величини	Дослідна ділянка 9 м^2. Висівалося насіння великого розміру	Дослідна ділянка 9 м^2. Висівалося насіння великого розміру	Дослідна ділянка 9 м^2. Висівалося насіння великого розміру

Фенологічні спостереження виявили різницю в датах полягання листя: 01.08 та 10.08 (поодиноке); 15.08 та 25.08 (масове). Яких ділянок стосуються такі цифри, у роботі не вказано. Різниця в строках появи сходів і стрілкуванні в роботі не наведена.

Урожай

Контрольна ділянка 115 ц/га. Цибулини різної величини	Дослідна ділянка 141 ц/га. Цибулини великого розміру	Дослідна ділянка 146 ц/га. Цибулини великого розміру	Дослідна ділянка 150 ц/га. Цибулини великого розміру

Висновок

Дослід показав, що на дослідних ділянках урожай зібрано вищий, якість цибулини краща. Отже, урожай цибулі залежить від якості насінного матеріалу. Рослини, вирощені з великого насіння, швидше ростуть і розвиваються. Вони раніше достигають і є високоврожайними.

Але, на жаль, насіння буває різноякісним не тільки за величиною, а й за вагою, навіть в одному плоді насіння буває різноякісним.

Коментар до роботи

Судячи зі змісту роботи мета її сформульована неправильно. Насправді автори аналізували вплив не якості, а розміру насінин на врожай, що, звичайно, не одне й те саме. Крім того, з тексту роботи не було зрозуміло, що саме висівалося — насінини чи цибулинки. Контрольний варіант у роботі підібрано дуже невдало. Якщо в дослідному варіанті висівалися великі насінини, то в контролі слід було висівати маленькі. У результаті виявилося, що в досліді взагалі немає контролю. Якщо під контролем розумілося висівання звичайної суміші насінин різного розміру, то його треба розглядати як окремий варіант. Слід також було навести критерії, згідно з якими насінини вважали великими або маленькими. До того ж так званий контроль представлено однією повторністю, що є неприпустимим. Через це неможливо провести статистичну обробку результатів. Незрозуміло також, що являли собою три ділянки з висіяним насінням великого розміру. Це три різні варіанти чи три повторності одного варіанта?

Дослідна робота № 3

Вплив підживлення на врожайність помідорів

Об'єкт: помідори сорту «Чудо землі».

Мета досліду: не вказана.

Площа ділянки: 28 м^2.

Місце проведення досліду: навчально-дослідна ділянка.

Схема досліду

Дослід проводився на двох грядках — дослідній і контрольній. На кожній грядці було розміщено 30 кущів помідорів у 2 рядки по 15 кущів. Довжина рядків — 10 метрів, ширина міжрядь — 0,7 м. Площа кожної грядки 14 м^2.

Контрольна грядка 14 м^2
Дослідна грядка 14 м^2

Урожай

Облік урожаю	Дослідна грядка	Контрольна грядка
Вага плодів 1 сорту	10,4 кг	6,2 кг
Вага плодів 2 сорту	12,8 кг	8,6 кг
Вага плодів 3 сорту	18,6 кг	8,4 кг
Вага плодів 4 сорту	14,0 кг	7,0 кг
Вага плодів 5 сорту	16,6 кг	12,8 кг
Загальна вага плодів	72,4 кг	43,0 кг
Урожай у перерахунку на 1 га	51,7 т	30,7 т

Висновок

Підживлення і підгортання має велике значення для підвищення врожайності помідорів.

Коментар до роботи

З тексту роботи (у висновку це вказано прямо) видно, що в ході досліду аналізувався вплив не лише підживлення, але й підгортання. Тобто в роботі досліджувалася дія двох факторів. Це потребувало й більш складної організації досліду. Також із тексту незрозуміло, у чому саме полягало підживлення і якої якості був ґрунт, де проводився дослід.

Кожен варіант досліду представлено однією повторністю, що повністю знецінює отримані результати. Крім того, отримані плоди краще було б розподілити не за сортами, а за градаціями маси або розміру. Результати такого розподілу краще представляти не в абсолютних значеннях, а у відсотках від загального врожаю.

Нарешті виникає й питання: Чи не можна було передбачити ще до початку досліду, що підживлення підвищує врожай рослин (та ще й на достатньо збіднілому ґрунті шкільної ділянки)?

Дослідна робота № 4
Вплив підживлення на врожай кукурудзи

Об'єкт: кукурудза сорту Дніпровський — 325.

Мета досліду: простежити вплив підживлення на врожай кукурудзи.

Площа ділянки: 10 м^2.

Місце проведення досліду: навчально-дослідна ділянка.

Схема досліду

Ділянка 1 (контрольна) — без підживлення, 5 м^2.

Ділянка 2 — дворазове підживлення, 5 м^2.

Фенологічні спостереження виявили затримку цвітіння та настання молочно-воскової стиглості рослин контрольної ділянки на чотири дні, а повної стиглості — на сім днів.

Урожай

Ділянка 1 (контрольна) — 59 ц/га.

Ділянка 2 — 69,9 ц/га.

Висновок

Під час проведення досліду було виявлено, що рослини з ділянки, де проводилося підживлення, розвивалися швидше й дали кращий урожай. Отже, підживлення кукурудзи сприяє покращенню врожаю. За висновками спеціалістів дослідної станції, рекомендовано проводити підживлення кукурудзи сорту Дніпровський — 325 у разі вирощування в місцевих господарствах.

Коментар до роботи

Дослід проведено без повторностей, що не дозволяє здійснити статистичну обробку, а отже, повністю його знецінює. Не вказано, за якою схемою висівалася кукурудза (а в роботі № 1, наприклад, показано, що різні схеми посіву призводять до різних результатів). Немає аналізу елементів структури врожаю (*див. зауваження до роботи № 1*), що не дозволяє пояснити отримані результати. З тексту роботи неможливо зрозуміти, чому підживлення проводили двічі. Може, з однократним підживленням результат був би кращим? І, як і в попередньому випадку, той факт, що підживлення підвищує врожай, є відомим для будь-якого селянина. Тим більше, що в роботі зазначено, що спеціалісти дослідної станції рекомендують проводити підживлення кукурудзи в підсобних господарствах. Тож чи треба проводити подібні досліди?

Дослідна робота № 5

Вплив субстрату на вкорінення живців чорної смородини

Об'єкт: чорна смородина сорту «Голубка».

Мета досліду: не вказана.

Площа ділянки: не вказана.

Місце проведення досліду: теплиця.

Схема досліду

Дослід у потрійній повторності (по 20 рослин у кожній повторності) у двох варіантах. Кожна повторність висаджувалася в окремий ящик.

Варіант 1: живці висаджувалися в пісок.

Варіант 2: живці висаджувались у субстрат (пісок з торфом у співвідношенні 1:1).

Результати досліду

Варіант 1 (пісок)		**Варіант 2 (пісок + торф)**	
посаджено	**укоренилося**	**посаджено**	**укоренилося**
20	12	20	16
20	10	20	15
20	11	20	17

Висновок

З досліду було встановлено, що живці чорної смородини, зрізані з однорічних здерев'янілих пагонів, добре вкорінювалися в піску, а ще краще в піску з торфом (1:1). Так, відсоток живців, що вкоренилися, у першому варіанті дорівнював 55 %, а в другому — 80 %.

Коментар до роботи

Назва роботи сформульована не зовсім вдало. Судячи зі змісту, тут проведено аналіз не впливу субстрату як такого (пісок — це теж субстрат), а вплив якості субстрату. Крім того, апріорно можна стверджувати, що між двома варіантами є різниця не лише за якістю субстрату, але й за його вологістю, бо вологоутримуюча здатність чистого піску набагато нижча, ніж у суміші піску з торфом. Цей фактор у роботі не проаналізовано, хоча він міг зробити суттєвий внесок у різницю, що спостерігалася. Хотілося б також побачити обґрунтування обраного співвідношення піску й торфу. Чому воно дорівнювало саме 1 : 1, а не 2 : 1 або 1 : 2?

У роботі також відсутня статистична обробка результатів. Треба зазначити, що тут аналізуються дискретні дані у двох варіантах. Кращим способом статистичної обробки для цього випадку є об'єднання всіх даних, складання таблиці зв'язаності й обчислення величини критерію хі-квадрат. Оскільки загальна кількість використаних живців у кожному з варіантів є достатньо великою (по 60), то прийнятним також було б порівняння об'єднаних даних за критерієм Стьюдента.

Дослідна робота № 6

Сортовивчення і сортовипробування томатів

Об'єкт: томати сортів «Грот», «Човник», «Прогрес».

Мета досліду: провести сортовивчення і сортовипробування томатів «Грот», «Човник», «Прогрес».

Площа ділянки: 22,05 м^2.

Місце проведення досліду: навчально-дослідна ділянка.

Схема досліду

Дослід проводився у потрійній повторності. У кожній повторності по сім рослин кожного з трьох сортів. Рослини висаджувалися 70×50 см.

Схеми розміщення рослин у кожній повторності та взаємного розташування варіантів не наведено. У роботі вказано кількість внесених на квадратний метр добрив та зазначено факт проведення двох обробок рослин проти пошкодження фітофторою. В описі сортів (літературні дані без посилання на джерело) наведено їх коротку характеристику й зазначено той факт, що вони є стійкими до фітофторозу.

Урожай

Сорт	Маса плоду, г	Урожай з одного куща, г	Кількість плодів з однієї рослини	Висота рослини	Врожайність з 1 м^2, г
Грот	30–40	600–700	20–25	40–50	1800
Човник	50–70	1000–1400	22–20	40–45	3000
Прогрес	40–50	800–1000	20–40	28–40	2400

Висновок

Із досліду було встановлено, що сорти томатів «Грот», «Човник» та «Прогрес» є стійкими до фітофторозу. Усі сорти ранні (від сходження до визрівання плодів проходить від 98 до 115 діб). У томатів сортів «Човник» та «Прогрес» кущі штамбові, компактні, не потребують підв'язування. Смакові якості всіх сортів доброго смаку, і їх використовують для вживання у свіжому вигляді та для засолювання.

Коментар до роботи

У назві роботи й у її меті є слова «сортовипробування» і «сортовивчення». Але з результатів роботи видно, що проводили лише порівняння сортів за врожаєм. Сортовипробування потребує достатньо регламентованого проведення польових дослідів за спеціальними схемами. У зв'язку з цим, до речі, під час вивчення сортів зазвичай не застосовують хімічного захисту від захворювань, бо це заважає виявленню генетично обумовленої стійкості сорту.

Незрозуміло, які величини наведено в таблиці. Судячи з усього, наведено ліміти (тобто найменше й найбільше значення) значень ознак. Хоча треба було наводити середні величини, тим більше що дослід проведено у трикратній повторності. Ще можна було б показати варіювання величини ознак, які аналізуються, й указати стандартні відхилення. Величини стандартних відхилень дозволять оцінити розмах варіювання ознаки, якщо це нас зацікавить. У роботі взагалі відсутня статистична обробка результатів досліду, що не дає змоги зробити насправді обґрунтовані висновки (хоча можливість зробити її в авторів була). У висновках указано, що всі сорти є стійкими до фітофторозу. Але за умови використання хімічних засобів захисту рослин важко робити висновки щодо власної стійкості сорту до конкретного збудника. Крім того, оцінка стійкості до захворювань потребує оцінки ураженості рослин, яку навряд чи учень зможе виконати самостійно без допомоги відповідного фахівця. Тому не слід писати у висновках про те, що в роботі не вивчалось.

У роботі відсутнім є і найголовніший висновок про те, який сорт (або сорти) є найкращим. Крім того, під час аналізу результатів слід було б пояснити, завдяки яким особливостям (вища середня маса плодів, більша кількість плодів на кущі тощо) урожай плодів у досліджених сортів відрізнявся.

ДОДАТКИ

Додаток 1

Критичні значення критерію *t* для рівнів значущості 0,05 і 0,01

Число ступенів свободи	Рівні значущості		Число ступенів свободи	Рівні значущості	
	0,05	0,01		0,05	0,01
1	12,71	63,66	14	2,14	2,98
2	4,30	9,92	15	2,13	2,95
3	3,18	5,84	16	2,12	2,92
4	2,78	4,60	17	2,11	2,90
5	2,57	4,03	18	2,10	2,88
6	2,45	3,71	19	2,09	2,86
7	2,36	3,50	20	2,09	2,85
8	2,31	3,36	25	2,06	2,79
9	2,26	3,25	30	2,04	2,75
10	2,23	3,17	40	2,02	2,70
11	2,20	3,11	50	2,01	2,68
12	2,18	3,05	100	1,98	2,63
13	2,16	3,01	1000+	1,96	2,58

Додаток 2

Критичні значення критерію χ^2 для рівнів значущості 0,05 і 0,01

Число ступенів свободи	Рівні значущості		Число ступенів свободи	Рівні значущості	
	0,05	0,01		0,05	0,01
1	3,84	6,63	11	19,67	24,72
2	5,99	9,21	12	21,03	26,22
3	7,81	11,34	13	22,36	27,69
4	9,49	13,28	14	23,68	29,14
5	11,07	15,09	15	25,00	30,58
6	12,59	16,81	16	26,30	32,00
7	14,07	18,47	17	27,59	33,41
8	15,51	20,09	18	28,87	34,80
9	16,92	21,67	19	30,14	36,19
10	18,31	23,21	20	31,41	37,57

Додаток 3а

Критичні значення критерію *F* для рівня значущості 0,05

Число ступенів свободи для меншої дисперсії	Число ступенів свободи для більшої дисперсії																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0	243,9	244,7	245,4	245,9	246,5	246,9	247,3	247,7	248,0	248,3	248,6	248,8	249,0	249,26
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,42	19,42	19,43	19,43	19,44	19,44	19,44	19,45	19,45	19,45	19,45	19,45	19,46
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70	8,69	8,68	8,67	8,67	8,66	8,65	8,65	8,64	8,64	8,63
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86	5,84	5,83	5,82	5,81	5,80	5,79	5,79	5,78	5,77	5,77
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,60	4,59	4,58	4,57	4,56	4,55	4,54	4,53	4,53	4,52
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,92	3,91	3,90	3,88	3,87	3,86	3,86	3,85	3,84	3,83
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51	3,49	3,48	3,47	3,46	3,44	3,43	3,43	3,42	3,41	3,40
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22	3,20	3,19	3,17	3,16	3,15	3,14	3,13	3,12	3,12	3,11
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,99	2,97	2,96	2,95	2,94	2,93	2,92	2,91	2,90	2,89
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77	2,76	2,75	2,75	2,74	2,73
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,69	2,67	2,66	2,65	2,64	2,63	2,62	2,61	2,60
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,60	2,58	2,57	2,56	2,54	2,53	2,52	2,51	2,51	2,50
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53	2,51	2,50	2,48	2,47	2,46	2,45	2,44	2,43	2,42	2,41
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,44	2,43	2,41	2,40	2,39	2,38	2,37	2,36	2,35	2,34
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,38	2,37	2,35	2,34	2,33	2,32	2,31	2,30	2,29	2,28
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35	2,33	2,32	2,30	2,29	2,28	2,26	2,25	2,24	2,24	2,23
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31	2,29	2,27	2,26	2,24	2,23	2,22	2,21	2,20	2,19	2,18
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27	2,25	2,23	2,22	2,20	2,19	2,18	2,17	2,16	2,15	2,14
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,17	2,16	2,14	2,13	2,12	2,11	2,11
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,18	2,17	2,15	2,14	2,12	2,11	2,10	2,09	2,08	2,07
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18	2,16	2,14	2,12	2,11	2,10	2,08	2,07	2,06	2,05	2,05
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15	2,13	2,11	2,10	2,08	2,07	2,06	2,05	2,04	2,03	2,02
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,18	2,15	2,13	2,11	2,09	2,08	2,06	2,05	2,04	2,02	2,01	2,01	2,00
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15	2,13	2,11	2,09	2,07	2,05	2,04	2,03	2,01	2,00	1,99	1,98	1,97
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09	2,07	2,05	2,04	2,02	2,01	2,00	1,98	1,97	1,96	1,96

Додаток 3б

Критичні значення критерію *F* для рівня значущості 0,01

Число ступенів свободи для меншої дисперсії	Число ступенів свободи для більшої дисперсії																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6083	6106	6126	6143	6157	6170	6181	6192	6201	6209	6216	6223	6229	6235	6240
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,41	99,42	99,42	99,43	99,43	99,44	99,44	99,44	99,45	99,45	99,45	99,45	99,46	99,46	99,46
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,13	27,05	26,98	26,92	26,87	26,83	26,79	26,75	26,72	26,69	26,66	26,64	26,62	26,60	26,58
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45	14,37	14,31	14,25	14,20	14,15	14,11	14,08	14,05	14,02	13,99	13,97	13,95	13,93	13,91
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,96	9,89	9,82	9,77	9,72	9,68	9,64	9,61	9,58	9,55	9,53	9,51	9,49	9,47	9,45
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56	7,52	7,48	7,45	7,42	7,40	7,37	7,35	7,33	7,31	7,30
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,36	6,31	6,28	6,24	6,21	6,18	6,16	6,13	6,11	6,09	6,07	6,06
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61	5,56	5,52	5,48	5,44	5,41	5,38	5,36	5,34	5,32	5,30	5,28	5,26
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,01	4,96	4,92	4,89	4,86	4,83	4,81	4,79	4,77	4,75	4,73	4,71
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65	4,60	4,56	4,52	4,49	4,46	4,43	4,41	4,38	4,36	4,34	4,33	4,31
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25	4,21	4,18	4,15	4,12	4,10	4,08	4,06	4,04	4,02	4,01
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01	3,97	3,94	3,91	3,88	3,86	3,84	3,82	3,80	3,78	3,76
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,91	3,86	3,82	3,78	3,75	3,72	3,69	3,66	3,64	3,62	3,60	3,59	3,57
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66	3,62	3,59	3,56	3,53	3,51	3,48	3,46	3,44	3,43	3,41
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52	3,49	3,45	3,42	3,40	3,37	3,35	3,33	3,31	3,29	3,28
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50	3,45	3,41	3,37	3,34	3,31	3,28	3,26	3,24	3,22	3,20	3,18	3,16
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40	3,35	3,31	3,27	3,24	3,21	3,19	3,16	3,14	3,12	3,10	3,08	3,07
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32	3,27	3,23	3,19	3,16	3,13	3,10	3,08	3,05	3,03	3,02	3,00	2,98
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24	3,19	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,92	2,91
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18	3,13	3,09	3,05	3,02	2,99	2,96	2,94	2,92	2,90	2,88	2,86	2,84
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17	3,12	3,07	3,03	2,99	2,96	2,93	2,90	2,88	2,86	2,84	2,82	2,80	2,79
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,07	3,02	2,98	2,94	2,91	2,88	2,85	2,83	2,81	2,78	2,77	2,75	2,73
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	3,02	2,97	2,93	2,89	2,86	2,83	2,80	2,78	2,76	2,74	2,72	2,70	2,69
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98	2,93	2,89	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,68	2,66	2,64
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,06	2,99	2,94	2,89	2,85	2,81	2,78	2,75	2,72	2,70	2,68	2,66	2,64	2,62	2,60

ЛІТЕРАТУРА

1. *Бейли Н.* Статистические методы в биологии.— М.: Мир, 1963.— 272 с.
2. *Гланц С.* Медико-биологическая статистика.— М.: Практика, 1998.— 459 с.
3. *Доспехов Б. А.* Методика полевого опыта.— М.: Колос, 1985.
4. *Лакин Г. Ф.* Биометрия.— М.: Высшая школа, 1990.— 352 с.
5. *Любищев А. А.* Дисперсионный анализ в биологии.— М.: Изд-во Моск. ун-та, 1986.— 200 с.
6. *Снедекор Дж. У.* Статистические методы в применении к исследованиям в сельском хозяйстве и биологии.— М.: Изд-во сельскохозяйственной литературы, журналов и плакатов, 1961.— 504 с.
7. *Эренберг А.* Анализ и интерпретация статистических данных.— М.: Финансы и статистика, 1981.— 406 с.

ЗМІСТ