Міністерство освіти та науки України
Національний університет „Чернігівський колегіум"
імені Т.Г.Шевченка

МАТЕМАТИЧНІ МЕТОДИ В БІОЛОГІЇ

Методичні рекомендації для студентів природничих спеціальностей

Чернігів 2020

Математичні методи в біології: методичні рекомендації для студентів природничих спеціальностей / Укладачі О. Б. Мехед, О. В. Ткаченко.- Чернігів, НУЧК, 2020. – 93 с.

Рецензенти Ткач Ю. М. - завідувач кафедри кібербезпеки та математичного моделювання Національного університету "Чернігівська політехніка", доктор педагогічних наук, професор

Третяк О. П. - декан природничо-математичного факультету, професор кафедри біології Національного університету "Чернігівський колегіум" імені Т. Г. Шевченка, кандидат біологічних наук, професор кафедри біології

Затверджено на засіданні кафедри
протокол № 2 від 21 вересня 2020 р.

ЗМІСТ

Тема I. Математичні методи в біології

1. Поняття варіаційної статистики

2. Значення математичних методів обробки матеріалу в біологічному дослідженні

3. З історії використання математичних методів опрацювання біологічного експерименту

1. Поняття варіаційної статистики

Предметом є статистична сукупність, або безліч індивідуально помітних, але біологічно однорідних одиниць, які об'єднуються (групуються) за деякими загальними умовами для вивчення.

Окремі одиниці статистичної сукупності називаються її членами, або варіантами. Загальна кількість варіант, що складають дану сукупність, називається її обсягом.

Варіанти - це окремі значення ознаки, за якими утворюється статистична сукупність (врожайність, вміст ліпідів в організмі або в сім'ї, активність ферментних систем, успішність, розумові здібності).

Характерні риси або предмети, за якими один об'єкт відрізняється від іншого, називаються ознаками. Статистична сукупність може бути утворена і за кількома ознаками, складатися з декількох однорідних груп, що об'єднуються в єдиний комплекс щодо прийнятих в досліді умов. У таких випадках вона називається статистичним комплексом.

Яку б форму і зміст не приймала статистична сукупність, вона являє собою явище масове, для якого характерна наявність індивідуальності у його компонентів. Аналіз масових явищ вимагає від дослідника певних знань, які складають зміст біометрії: вміння правильно узагальнювати і аналізувати результати масових спостережень; робити науково обґрунтовані висновки.

2. Значення математичних методів обробки матеріалу в біологічному дослідженні

Знання про природу набуваються шляхом спостереження, порівняння та досвіду. При цьому як об'єкти, так і результати спостереження можуть бути і одиничними і масовими.

Аналіз причинно-наслідкових відносин показує, що результати одиничних спостережень, як правило, не збігаються один з одним, варіюють від випадку до випадку в певних межах. Варіювання, або внутрішньогрупова мінливість - явище універсальне, характерне для всіх живих систем.

Характерною рисою біометрії є те, що вона розглядає поодинокі і масові явища в їх єдності на основі існуючого між ними внутрішнього зв'язку. Внутрішній зв'язок виражається статистичними закономірностями, які діють тільки в сфері масових явищ і не піддаються виявленню на одиничних спостереженнях.

Біометрія не займається питаннями складних математичних рішень, обґрунтуванням формул і рівнянь. Біометрія використовує готові математичні висновки і застосовує їх для вирішення біологічних задач. Тому питання технічного додатка математичних висновків і формул займають в курсі біометрії велике місце.

3. З історії використання математичних методів опрацювання біологічного експерименту

Біометрія має історію, яка своїм корінням сягає в сиву давнину. Біометрія на початку характеризувалася додатком кількісних методів до дослідження живої природи.

У 1768 французький гіпполог Буржела видає книгу «Екстер'єр коня», де пропонується програма вимірювання для визначення пригодності коня до тієї чи іншої служби. В цей же час розвивається військова антропологія з метою відбору чоловіків, найбільш придатних для несення військової служби.

У середині XVII ст. були покладені початок теорії ймовірностей, що виникла на ґрунті азартних ігор. Статистика виникла з потреб держави, а біометрія - в процесі розвитку біології. Першим, хто вдало поєднав емпіричні методи антропології і соціальної статистики з математичною теорією ймовірностей був Кетле. У 1835 вийшла його книга «Про людину та розвиток її здібностей і досвід соціальної фізики».

У XX ст. з'явилися класичні праці Госсета, які друкувалися під псевдонімом "Стьюдент", Фішера та інших представників англійської школи біометрії. З ім'ям Стьюдента пов'язано обґрунтування так званої "теорії малої вибірки", що відкрила нову сторінку в історії біометрії. Фішер розробив метод дисперсійного аналізу, який знайшов застосування не тільки в біології, але і в техніці.

Тема II. Угруповання результатів спостережень

1. Таблиці та ряди розподілу.
2. Класифікація ознак.
3. Побудова варіаційних рядів.
4. Графіки розподілу.

1. Таблиці та ряди розподілу

Результати дослідження фіксуються зазвичай в первинних документах - протоколах дослідів, польових щоденниках, робочих журналах і т.п. Зібраний фактичний матеріал потім піддається статистичній обробці. Мета обробки - витяг з маси фактів укладеної в них інформації, отримання на підставі проведеного дослідження об'єктивних і переконливих висновків.

Перший крок на шляху статистичної обробки - угруповання зібраних даних відповідно до завдань дослідження і тими умовами, в яких воно проводилося. Найбільш раціональною формою угруповання служать статистичні таблиці. У них зазвичай зводяться результати масових спостережень. Статистичні таблиці бувають складні і прості, і їх будова залежить від того, за якими ознаками і за якою їх кількістю групується матеріал, а також від завдань, які вирішуються угрупованням зібраного матеріалу. Приклад порівняно простий угруповання: результати П.М. Константинова (1955), отримані в досвіді по випробуванню врожайності ячменю і вівса в умовах нечорноземної смуги Російської Федерації:

Таблиця 1

Культура	Врожайність зерна по рокам досвіду (*ц/га*)							Середня урожайність
	1928	1929	1930	1931	1932	1933	1934	
Ячмінь	7,70	9,00	9,40	7,40	7,40	10,90	8,0	8,54
Овес	8,26	7,22	8,43	5,57	6,35	8,00	9,13	7,57

Прикладом складних таблиць, що ілюструють залежність однієї з варіюючих ознак від змін іншої, служать кореляційні таблиці, а також таблиці дисперсійних комплексів.

Найбільш просту форму статистичного угруповання представляють ряди розподілу, які будуються на основі операції ранжирування, тобто шляхом розташування варіант (окремих числових значень варіюючої ознаки) в зростаючому або спадному порядку. Наприклад, є такий сукупність 20 вимірювань ознаки: 2, 5, 3, 6, 4, 7, 4, 5, 6, 6, 5, 9, 5, 6, 10, 8,12, 9, 7, 6.

Видно, що ознака варіює від 2 до 12 одиниць. Розташуємо цю сукупність у зростаючому порядку:

2, 3, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 8, 9, 9, 10, 12.

Вийде ранжирований ряд значень ознак.

При розподілі членів сукупності в ряд переслідуються певні цілі. Одна з них - розкриття закономірності варіювання досліджуваної ознаки. Тому до рядів розподілу пред'являються певні вимоги:

1) вони повинні бути легкодоступні для огляду;

2) добре ілюструвати закономірність варіювання. Ранжируваний нами ряд сам по собі погано задовольняє цим вимогам. Якщо, ті ж варіанти розташувати у вигляді подвійного ряду, враховуючи їх повторюваність в загальному строю, сукупність розподілиться таким чином:

Таблиця 2

Варіанти	2	3	4	5	6	7	8	9	10	11	12
Повторюваність варіант (p)	1	1	2	4	5	2	1	2	1	0	1

Такий упорядкований ряд розподілу, в якому вказана повторюваність варіант, що належать до даної сукупності, називається варіаційним рядом. Повторюваність варіант в сукупності називають вагами або частотами.

У біометрії ознаки позначають прописними буквами *X, Y, Z* ... їх числові значення - малими x_p, x_2, x_3, ... або y_p, y_2 y_3 ... Їх частоти позначаються латинською буквою p. Загальна кількість варіант, що входять до складу даної сукупності, називають обсягом сукупності і позначають буквами п або N.

$$\sum \mathrm{p} = n \text{ [1].}$$

Відносна частота або частість обчислюється як частка від ділення *p / n* і виражається в частках або відсотках.

$$\sum \frac{\mathrm{p}}{\mathrm{n}} = 1 \text{ [2].}$$

Заміна абсолютних значень ознаки (частот) частостей полегшує зіставлення одного варіаційного ряду з іншим. І робить більш виразними характерні риси варіювання.

2. Класифікація ознак

Біологічні ознаки діляться на якісні та кількісні. До якісних належать такі ознаки, як забарвлення листя і квіток, смак і запах продуктів. Якщо мова йде про вимірювані або обчислювальні величини - це будуть кількісні ознаки.

У варіаційні ряди розподіляються тільки кількісні ознаки, а якісні ознаки зазвичай розглядають в альтернативній формі. Кількісні ознаки можуть бути рахунковими (варіюють дискретно) та мірними (варіюють безперервно).

3. Побудова варіаційних рядів

Відомо два види варіаційних рядів: безінтервальні та інтервальні. Прикладом безінтервального варіаційного ряду може служити розподіл американських вугрів (n = 863) за кількістю хребців (по Бергу, 1924):

Таблиця 3

Число хребців (x)	103	104	105	106	107	108	109	110	111
Число особин (p)	1	8	45	183	274	221	96	31	3

В даному випадку ознака варіює слабо. Але багато ознак варіюють у дуже широких межах і розподіл їх в безінтервальний ряд не досягає мети: ряди виходять занадто розтягнутими, погано доступним для огляду, що не відображають чітко закономірності варіювання. Наприклад, число зерен ячменю в 50 колосках:

Таблиця 4

Число зерен (x)	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Кількість колосків (p)	1	1	1	1	1	1	6	5	6	5	3	2	5	5	2	3	1	0	1

У таких випадках найкращий результат виходить від розподілу сукупності в інтервальний варіаційний ряд. Для цього вся варіація ознаки від мінімуму до максимуму варіанти розбивається на рівні інтервали (від і до) або класи. Потім всі варіанти розподіляють по класах і частоти *p* будуть частотами класів. Для обчислення середніх величин на таких дискретних варіаційних рядах використовуються не класові інтервали, а їх середні, рівні півсумі верхньої і нижньої меж класу.

Число класів залежить від завдання дослідження і характеру зібраного матеріалу. Ширина класового інтервалу позначається не тільки на характері розподілу варіант по класах, а й на точності середніх характеристик. Установка широких класових інтервалів спотворює типові риси варіювання і веде до зниження точності числових характеристик ряду. При виборі надмірно вузьких інтервалів точність узагальнюючих числових змінних підвищується, але ряд виходить занадто розтягнутим і не дає чіткої картини варіювання. Щоб визначити величину класового інтервалу для побудови доступного для огляду варіаційного ряду Г.А. Стерджес (Sturges, 1926) рекомендує наступну формулу:

$$i = \frac{x_{max} - x_{min}}{1 + 3{,}32lg_n} \text{ [4] или } i = \frac{x_{max} - x_{min}}{5\text{x}lg_n}$$

Формулою [5] рекомендується користуватися при наявності в сукупності великого числа членів (n> 100).

Отже, розглянемо побудову варіаційного ряду для випадку з колосками ячменю: x_{min} = 9, x_{max} = 27; величина класового інтервалу буде наступною - i = 3 [4] і i = 2 [5]. Візьмемо i = 3. При розбивці варіації на класи межі першого класу встановлюємо так, щоб мінімальна варіанту потрапила приблизно в середину цього класу. Якщо нижня (мінімальна) межа 1-го класу дорівнює 8, то вийде 8 класів з нижніми межами, рівними 8, 11, 14, 17, 20, 23, 26, 29. Щоб однозначно вирішити питання про приналежність варіанти до якогось класу, його верхню межу зменшують на 0,1 або 0,01, що і дає необхідне розмежування класів. наприклад, таблиця 5

Таблиця 5

Класи по числу зерен в колосках ячменю	Середнє значения класів (x)	Частоти *(p)*
8-10,9	9,5	2
11-13,9	12,5	3
14-16,9	15,5	12
17-19,9	18,5	14
20-22,9	21,5	12
23-25,9	24,5	6
26-28,9	27,5	1
сума	-	50

В результаті виходить інтервальний варіаційний ряд з переривчастим варіюванням. При побудові варіаційного ряду не допускається подвійний облік однієї і тієї ж варіанти.

4. Графіки розподілу

Щоб надати більшу наочність закономірності варіювання ознак, варіаційні ряди прийнято зображати графічно у вигляді гістограми, або полігону, а також у вигляді кумуляти або огіви.

Графік, званий гістограмою розподілу частот, виходить, якщо в системі координат по осі абсцис відкласти межу класів, а по осі ординат - їх частоти.

У випадку з розподілом класів ячменю за кількістю зерен в колосі (розглянутий раніше) гістограма буде виглядати, як показано на рис. 1.

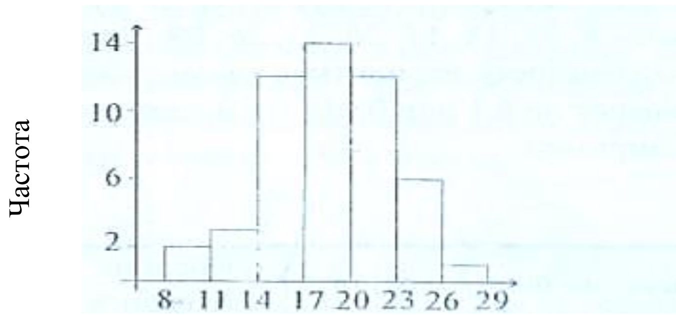

Рис . 1. Гістограма розподілу класів ячменю за кількістю зерен

Гістограма зображує закономірності розподілу варіант по класах варіаційного ряду, тобто при безперервному варіюванні ознаки. Прямокутники відповідають класам, а їх висота - частотам варіаційного ряду.

Якщо з середніх точок прямокутників гістограми опустити перпендикуляри на вісь абсцис, а самі точки з'єднати між собою, вийде графік дискретного варіювання, званий полігоном розподілу (рис. 2).

Полігон розподілу можна побудувати незалежно від гістограми, завдаючи на вісь абсцис середні значення класів. А коли виникає необхідність, можна полігон перетворити в гістограму.

В інших випадках графік варіаційного ряду будується в вигляді кумуляти (від слова cumulo - накопичувати). Для побудови кумуляти розподілу класів ячменю за кількістю зерен скористаємося даними 1 та 3 стовпців табл. 5. При цьому по осі абсцис відкладають значення класових варіант, а по осі ординат - накопичені частоти (рис. 3).

Таблиця 6

Середнє значення класів	Частоти (Р)	Накопичення частоти
9,5	2 1	2
12,5	3	5
15,5	12	17
18,5	14	31
21,5	12	43
24,5	6	49
27,5	1	50
Сума	50	50

Рис. 2. Полігон розподілу класів ячменю за кількістю зерен

Рис. 3. Кумулята розподілу класів ячменю за кількістю зерен

Якщо ряд накопичених частот нанести на вісь абсцис, а значення варіант розташувати по осі ординат і побудувати графік, виходить огива.

Значення графіків полягає в їх наочності. Але вони не дають точної характеристики варіюючої ознаки, так як залежать від прийнятих масштабів. Точну характеристику варіюючих ознак дають статистичні показники, про які йтиметься далі.

Тема III. Закономірності розподілу

1. Характерні риси варіювання.
2. Імовірність і її властивості.
3. Поняття випадкової величини.
4. Нормальний розподіл і його параметри.

1. Характерні риси варіювання

У розподілі емпіричних сукупностей кидається в очі одна важлива особливість - переважне накопичення варіант в центральних класах і поступове убування їх числа в міру віддалення від серединної точки варіаційного ряду. Ця особливість, яка становить одну з характерних рис варіювання біологічних ознак, - факт дуже важливий, що має широке поширення в природі (рис. 4). У самих різних випадках проявляється одна і та ж закономірність: в масі відносно однорідних одиниць (варіант) переважна більшість складають варіанти середнього розміру, і чим далі вони відхиляються від середнього рівня ознаки, тим рідше зустрічаються в даній сукупності. Отже, між значеннями ознаки та їх зустрічаємістю існує певний зв'язок. Наочним виразом зв'язку з цим служить варіаційний ряд і його графік - варіаційна крива.

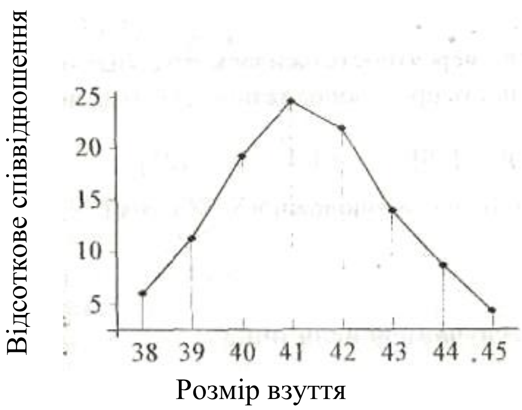

Рис. 4. Полігон процентного співвідношення в розподілі чоловічого взуття серед населення центральних областей Російської Федерації

2. Імовірність і її властивості

Припустимо, що в пологовому будинку народилося 208 хлопчиків і 200 дівчаток, всього 408 дітей. Число 208 - це абсолютна частота народжених хлопчиків, а число 200 абсолютна частота народжених дівчаток. Якщо число хлопчиків і дівчаток, які народилися в цьому пологовому будинку, віднести до загальної кількості новонароджених дітей, вийдуть відносні частоти або частости цих подій:

208/408 = 0,51 відносна частота народжених хлопчиків,

200/408 = 0,49 відносна частота народжених дівчаток.

Однак теоретично повинно народжуватися порівну хлопчиків і дівчаток. Теоретичне значення відносної частоти очікуваної події називається його ймовірністю. Подія - це той результат, який виходить при кожному випробуванні. Якщо при кожному випробуванні подія неминуче настає, вона називається достовірною. Якщо немає - неможливою. Якщо і може наступити і не може - це випадкова подія. Події, які при випробуванні в постійних умовах повторюються багаторазово, називаються масовими.

Ймовірністю називається відношення числа випадків або випадків m сприяють настанню очікуваної події A, до числа всіх можливих і несумісних в даному випробуванні результатів n, тобто

$$P(A) = \frac{m}{n} \ [19]$$

Для зручності, ймовірність очікуваної події прийнято позначати через p, а ймовірність протилежної події через q, тоді

$P(A) = p$ [20], $P(\bar{A}) = q$ [21], $p + q = 1$ [22],

де $(\bar{A})$ - ймовірність протилежної події (*не A*).

3. Поняття випадкової величини

Все, що можна підрахувати або виміряти, називається величиною. Величини (або "ознаки") діляться на постійні і змінні. Постійною називається величина, яка в заданих умовах не змінює свого значення. Змінна - це така величина, яка в даних умовах здатна приймати різні числові значення.

Змінна величина називається випадковою, якщо в заданих умовах вона може приймати то одні, то інші значення.

Випадкова величина в *N* повторних випробуваннях може приймати самі різні значення, але в кожному окремому випробуванні вона приймає завжди тільки одне з можливих значень. Яке значення прийме випадкова величина в результаті кожного випробування, заздалегідь сказати неможливо. Тому, характеризувати випадкову величину можна лише з певною ймовірністю, тобто вказуючи ймовірність її можливих значень.

4. Нормальний розподіл і його параметри

При відкритті закону розподілу змінної випадкової величини була виведена формула залежності, графік функції якої називається кривою нормального розподілу. Її площа дорівнює 1. Чим далі від середньої арифметичної стоїть значення, тим менше ймовірність (частота) появи даного значення.

Нормальний розподіл характеризується математичним очікуванням (відповідає поняттю середньої арифметичної) і дисперсією (мірою варіювання) випадкової величини. Дисперсія служить мірою відхилення можливих значень випадкової величини *X* від її математичного очікування *M*. Випадкова величина, що належить до даної сукупності, може відхилятися від середньої арифметичної до трьох ст.

Тема IV. Середні величини

план:

1. Види середніх і їх значення.

2. Середня арифметична:

а) проста середня арифметична;

б) зважена середня.

3. Скорочений спосіб обчислення середньої арифметичної (спосіб умовної середньої).

1. Види середніх і їх значення

Однією з найважливіших узагальнюючих характеристик варіюючих ознак є середня величина. Характеризуючи ту чи іншу популяцію, кажуть, наприклад, про середню продуктивність тварин або рослин, середню успішність, про середню швидкість біохімічної реакції і т.д. Значення середніх полягає в їх властивості нівелювати індивідуальні відмінності, в результаті чого виступає більш-менш стійка числова характеристика ознаки - не окремих представників, а цілої групи статистичних одиниць.

2. Середня арифметична

а) Проста середня арифметична

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x_1}{n} \text{ [6]}$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x_1}{n} \text{ [6]}$$

б) Зважена середня

Якщо в сукупності спостережень окремі варіанти повторюються p раз, то середня арифметична обчислюється за формулою [7] з урахуванням повторюваності (або "ваг") окремих варіант.

$$\bar{\mathrm{x}} = \frac{\mathrm{x}_1 p_1 + \mathrm{x}_2 p_2 + \mathrm{x}_3 p_3 + \cdots + \mathrm{x}_n p_n}{p_1 + p_2 + p_3 + \ \dots p_n} = \frac{\sum x_1 p_1}{p_1} \ [7]$$

Так як середня обчислюється з урахуванням частот або «ваг» окремих варіант, то вона називається зваженою середньою.

Таблиця 7

Кількість зерен в колоску (варіанти) (*x*)	7	8	9	10	11	12	13
Число колосків (повторюваність) (*p*):	1	1	2	7	3	3	1

x = (7x1+ 8x1+9x2+10x7+11x3+12x3+13x1)/18 = 185/18 =10,28 (зерен)

Аналогічним способом розраховується і загальна середня $(x ̅)$ з суми приватних середніх $((x\_1) ̅)$.

$$\bar{\mathrm{x}} = \frac{\mathrm{x}_1 n_1 + \mathrm{x}_2 n_2 + \mathrm{x}_3 n_3 + \cdots + \mathrm{x}_k n_k}{n_1 + n_2 + n_3 + \ \dots n_n} = \frac{\sum x_1 n_1}{n_1} \ [8]$$

3. Скорочений спосіб обчислення середньої арифметичної (спосіб умовної середньої)

Обчислення середньої арифметичної способом зваженої середньої не завжди зручно, особливо на сукупностях великого обсягу і при наявності багатозначних чисел, коли обчислювальна робота стає особливо трудомісткою. У таких випадках простіше розрахувати середню арифметичну спрощеним способом - способом умовної середньої.

Одна з варіант, все одно якась, умовно приймається за середню величину, що позначається через *A*. Зазвичай в якості умовної середньої береться варіанти (або клас) з найбільшою або близькою до неї частотою, хоча це не обов'язково. Потім розраховують відхилення варіант (або класів) від цієї умовної середньої і знаходять середню арифметичну за такою формулою:

$$\bar{x} = A + \frac{\sum pa}{n} \ [9]$$

де А - умовна середня, $a = x\text{-}A$ - відхилення варіанти від умовної середньої.

Наприклад, скористаємося наведеної формулою [9] для визначення середнього числа зерен в 18 колосках озимого жита:

Таблиця 8

Варіанта (x):	7	8	9	10	11	12	13
Повторюваність (p):	1	1	2	7	3	3	1
a:	-4	-3	-2	-1	0	+1	+2
pa:	-4	-3	-4	-7	0	+3	+2

$$\sum \text{pa} = +5 - 18 = -13$$

$$\bar{x} = 11 + \frac{-13}{18} = 11 - 0{,}72 = 10{,}28 \text{ (зерен)}$$

Таким же способом можна розрахувати середнє число хребців у американських вугрів. Якщо ж сукупність досить велика, то дані розрахунки простіше проводити за допомогою таблиць. Наприклад, при вивченні вмісту кальцію (в мг%) в сироватці 100 павіанів-гамадрилів отримали значення від 8,99 до 14,7. Розбивши варіацію на класи відповідно до формул [4] або [5], побудуємо таблицю і рознесемо відповідні дані експерименту (табл. 9).

Таблиця 9

Середні значення класів або класові варіанти	Частоти	Відхилення класових варіант від умовної середньої	Похідні відхилень на частоти	$a' = \frac{x - A}{i}$	a'
(x)	*(P)*	*(a=x- A)*	*(pa)*		
8,9	2	-2,8	-5,6	-4	-8
9,6	3	-2,1	-6,3	-3	-9
10,3	9	-1,4	-12,6	-2	-18
11,0	17	-0,7	-11,9	-1	-17
Разом	-	-	-36,4	-	-52
A = 11,7	25	0	0	0	0
12,4	23	+0,7	+ 16,1	+ 1	+23
13,1	10	+1,4	+ 14,0	+2	+20
13,8	7	+2,1	+14,72	+3	+21
14,5	4	+2,8	11,2	+4	+ 16
Разом	-	-	+56.0	-	+80
Сума	100	-	+ 19,6	-	+28

$$\bar{x} = A + \frac{\sum pa}{n} = 11{,}7 + \frac{19{,}6}{100} = 11{,}896 \approx 11{,}90 \text{ (мг \%)}$$

Якщо замість, *a* = *x*-*A* використати $\text{a}' = \frac{\text{x}-\text{A}}{\text{i}}$, де i - величина класового інтервалу (*i* = 0,7) і формулу

$$\bar{x} = A + \text{i}\frac{\sum p\text{a}'}{n} \text{ [10]}$$

то результат обчислень виявиться таким же:

$$\bar{x} = 11{,}7 + 0{,}7\left(\frac{28}{100}\right) = 11{,}7 + 0{,}196 = 11{,}896 \approx 10{,}90 \text{ (мг \%)}$$

Порівнюючи перший і другий спосіб розрахунку середньої, бачимо, що другий спосіб набагато простіше.

Тема V. Показники варіації

план:

1. Ліміти і розмах варіації.

2. Дисперсія і середнє квадратичне відхилення.

3. Обчислення середнього квадратичного відхилення (спосіб умовної середньої).

4. Коефіцієнт варіації.

5. Нормоване відхилення.

1. Ліміти і розмах варіації

Середня арифметична - найважливіша статистична характеристика, але вона нічого не говорить про величину варіювання ознаки. Одним з показників варіації служать ліміти, що показують min і max варіанти сукупності, а також розмах варіації (R), що представляє собою різницю між максимальною і мінімальною варіантами сукупності, тобто R = -x_{max}-x_{min}. Але вони не здатні характеризувати істотні риси варіювання, а, крім того, можуть сильно змінювати своє значення. Прикладом можуть служити наступні дві вибірки:

Таблиця 10

(x)	<u>10</u>	15	20	25	30	35	40	45	<u>50</u>	$\bar{x} = 30$
(y)	<u>10</u>	28	30	30	30	30	32	32	<u>50</u>	$\bar{y} = 30$

2. Дисперсія і середнє квадратичне відхилення

Найбільш підходящою мірою варіювання служить середній квадрат відхилень або дисперсія, що позначається символом ст.

$$\sigma^2 = \frac{\sum(x_1 - \bar{x})^2}{n-1} \ [11]$$

З огляду на повторюваність відхилень, можна висловити цю формулу так:

$$\sigma^2 = \frac{\sum pd^2}{n-1} \ [12]$$

де $d=(x_1\text{-}\bar{x})$

Середнє квадратичне відхилення або a обчислюється за такою формулою:

$$\sigma = \sqrt{\frac{\sum(x_1 - \bar{x})^2}{n-1}} \ [13] \text{ або } \sigma = \sqrt{\frac{\sum pd^2}{n-1}} \ [13а]$$

Величина n-1 носить назву числа ступенів свободи. Середнє квадратичне відхилення (стандартне) - величина іменована і виражається в тих же одиницях виміру, що і ознака. Чим сильніше варіює ознака, тим більше і величина середнього квадратичного відхилення і навпаки. Обчислимо середнє відхилення для розглянутих в 1 розділі даної теми рядів *x* і *y*.

Алгоритм розрахунку середнього квадратичного відхилення для ряду *x*

Таблиця 11

(x):	<u>10</u>	15	20	25	30	35	40	45	<u>50</u>	$\bar{x} = 30$
(a):	20	15	10	5	0	5	10	15	20	
(a^2):	400	225	100	25	0	25	100	225	400	$\sum a^2$=1500

$$\sigma_x = \sqrt{\frac{1500}{9-1}} = \sqrt{187{,}5} = 13{,}7$$

Алгоритм розрахунку середнього квадратичного відхилення для ряду *y*

Таблиця 12

(y):	<u>10</u>	28	28	30	30	30	32	32	<u>50</u>	$\bar{y} = 30$
(a):	20	2	2	0	0	0	2	2	20	
(a^2):	400	4	4	0	0	0	4	4	400	$\sum a^2$=816

Середні квадратичні відхилення для рядів *x* і *y* відрізняються один від одного і характеризують специфіку варіювання ознаки. При великій кількості спостережень різниця між *n* і *n-1* істотно не позначається на величині показника варіації, тому при *n> 30* замість *n-1* можна брати значення *n*.

Середня арифметична і середнє відхилення дають повну кількісну характеристику будь-якій емпіричній сукупності, що розподіляється по нормальному закону. Середня арифметична відображає дію на ознаку основних причин, що визначають типовий для популяції рівень його розвитку, тоді як середнє відхилення характеризує варіювання значень цієї ознаки навколо центру розподілу, тобто середньої арифметичної. Середнє квадратичне відхилення є мірою ступеня впливу на ознаку різних другорядних причин, що викликають його варіювання.

Таким чином, ці показники, хоча і відображають різні сторони варіюючих ознак, тісно пов'язані між собою.

3. Обчислення середнього квадратичного відхилення (спосіб умовної середньої)

Формула розрахунку середнього квадратичного відхилення методом умовної середньої набуває такого вигляду:

$$\sigma = \sqrt{\frac{\sum pa^2}{n} - \left(\frac{\sum \text{pa}}{n}\right)^2}\ [14],$$

де, *a*=*x*-*A*

Покажемо застосування цієї формули на прикладі розподілу кальцію (мг%) в сироватці крові павіанів-гамадрилів використовуючи наступний алгоритм (див. Табл. 13).

Таблиця 13

Класові варіанти *(x)*	Частоти *(p)*	Відхилення класових варіант від умовної середньої *(a = x - A)*	Похідні відхилень на частоти *(pa)*	pa^2
8,9	2	-2,8	-5,6	15,86
9,6	3	-2,1	-6,3	13,23
10,3	9	-1,4	-12,6	17,64
11,0	17	-0,7	-11,9	8,33
Итого	-	-	-36,4	
A = 11,7	25	0	0	0
12,4	23	+0,7	+ 16,1	11,27
13,1	10	+ 1,4	+ 14,0	19,60

13,8	7	+2,1	+14,7	30,87
14,5	4	+2,8		31,36
Итого	-	-	56,0	-
Сумма	100	-	+19,6	147,98

Підставивши отримані значення з таблиці в формулу [14] отримаємо значення середнього квадратичного відхилення

$$\sigma = \sqrt{\frac{147{,}98}{100} - \left(\frac{19{,}6}{100}\right)} = \sqrt{1{,}48 - 0{,}04} = \sqrt{1{,}44} = 1{,}20 \text{ (мг\%)}$$

Якщо відхилення класових варіант від умовної середньої *A* віднести до величини класового інтервалу, тобто замість, $a = x\text{-}A$ взяти $\mathrm{a}' = \frac{x_i - A}{i}$, то розрахунки за формулою

$$\sigma = \sqrt[i]{\frac{\sum pa'}{n} - (\frac{\sum pa'}{n})^2} \ [15]$$

значно спростяться, що видно з наступного прикладу:

Таблиця 14

Класові варіанти (x)	Частоти (p)	$\mathrm{a}' = \frac{\mathrm{x} - \mathrm{A}}{\mathrm{i}}$	(pa’)	pa'^2
8,9	2	-4	-8	32
9,6	3	-3	-9	27
10,3	9	-2	-18	36
11,0	17	-1	-17	17
Разом	-	-	-52	-
A = 11,7	25	0	0	0
12,4	23	+ 1	+23	23
13,1	10	+2	+20	40
13,8	7	+3	+21	63
14,5	4	+4	+16	64
Разом	-	-	+80	-
Сума	100	-	+28	302

На вибірках невеликого обсягу, особливо коли вони не згруповані в варіаційний ряд, середньоквадратичне відхилення обчислюється тим же способом за такою формулою:

$$\sigma = \sqrt{\frac{n}{n-1}[\frac{\sum a^2}{n} - (\frac{\sum a}{n})]}\ [16],$$

де *a=x-A.*

Наприклад, вміст гемоглобіну в крові 9 одновікових осіб:

Таблиця 15

x:	74	70	77	72	79	88	76	80	86	
a:	0	-4	+3	-2	+5	+14	+2	+6	+12	$\sum a = 36$
a^2:	0	16	9	4	25	196	4	36	144	$\sum a^2 = 434$

звідки

$$\sigma = \sqrt{\frac{9}{9-1}[\frac{\sum 434}{9} - (\frac{\sum 36}{9})^2]} = \sqrt{\frac{9}{5}(32,22)} = \sqrt{36.25} = 6,02$$

4. Коефіцієнт варіації

Середнє квадратичне відхилення є основним мірилом варіабельності ознак. Цей показник не залежить від числа спостережень, і тому може використовуватися для порівняльної оцінки варіювання однорідних ознак. Разом з тим широкого використання середнього квадратичного відхилення в якості запобіжного порівняння варіабельності ознак заважає те, що цей показник є величиною іменованою. Знаючи, що середнє відхилення ряду розподілу кальцію в сироватці крові павіанів-гамадрилів одно 1,20 мг%, а варіювання датських вугрів за кількістю хребців характеризується значенням середнього квадратичного відхилення 1,35 хребців не можна сказати, що данські вугри більш мінливі за ознакою, ніж павіани-гамадрили.

Щоб середньоквадратичне відхилення могло бути використано в якості запобіжного порівняння варіабельності ознак, воно повинно бути неіменованим. Для цього використовують коефіцієнт варіації:

$$CV = \frac{\sigma}{\bar{x}} \times 100\% \quad \text{або} \quad CV = \frac{\sigma}{\bar{x}}\ [17]$$

Так,

$$CV_{гамадрилів} = \frac{1,20}{11,9} \times 100\% = 10,1\%$$

$$CV_{вугрів} = \frac{1,35}{114,74} \times 100\% = 1,2\%$$

Із зіставлення цих показників видно, що перша ознака варіює сильніше, ніж друга. При нормальному розподілі коефіцієнт варіації зазвичай не перевищує 45-50% і часто буває набагато нижче цього рівня. У випадках же асиметричних розподілів він може бути досить високим, що досягає іноді 100% і вище.

5. Нормоване відхилення

В області біометричного аналізу чільне місце належить ще одному показнику, званому нормованим відхиленням (*t*). Нормоване відхилення показує, на скільки *a* та чи інша варіанта відхиляється від середнього рівня варіюючої ознаки. У найпростішому вигляді нормоване відхилення виглядає так:

$$t = \frac{x_i - \bar{x}}{\sigma} \quad [18]$$

Наприклад, при обстеженні фізичного розвитку учнів ремісничих шкіл Москви в 1957/58 н. р. показало, що середній зріст юнаків у віці 15 років дорівнює 164,8 см при *a* = 5,8 см. Оцінимо, наскільки зріст юнака (171,2 см) відрізняється від середнього:

$$t = \frac{171,2 - 164,8}{5,8} = +1,1$$

Оскільки будь-яка варіанта, що належить до сукупності, що розподіляється по нормальному закону, може відхилятися від середньої до трьох *a*, знайдена величина вказує на незначне збільшення зросту у юнака в порівнянні із середнім рівнем для даної групи.

Тема VI. Вибірковий метод

1. Вибірка і її репрезентативність.

2. Репрезентативність вибіркових показників:

а) помилка середньої арифметичної;

б) властивості середньої помилки. Закон великих чисел;

в) помилка при різних способах відбору варіант з генеральної сукупності;

г) помилки інших вибіркових показників;

д) показник точності оцінки параметрів.

3.Статистична перевірка гіпотез:

а) нульова гіпотеза;

б) рівні значущості і довірчі ймовірності;

в) довірчий інтервал і його межі;

г) критерії оцінки.

4. Малі вибірки:

а) t-розподіл Стьюдента,

а-1) випадки незалежних вибірок,

а-2) випадки залежних вибірок;

б) порівняння дисперсій. F-розподіл Фішера.

1. Вибірка і її репрезентативність

Щоб отримати вичерпну інформацію про стан тієї чи іншої статистичної сукупності, потрібно врахувати весь її склад без винятку. Однак в силу різних обставин, не завжди доводиться вдаватися до суцільного обстеження досліджуваних сукупностей. По-перше, тому, що ця робота пов'язана з великими витратами праці, часу, матеріальних засобів. По-друге, маючи на увазі практичну неможливість або недоцільність повного врахування всіх членів сукупності. Природні популяції, як правило, недоступні суцільному статистичному опису (фітопланктон, схожість насіння і т.д.). Внаслідок цього замість суцільного обліку всіх членів досліджуваної сукупності аналізу піддається якась її частина і по ній судять про стан сукупності в цілому.

Сукупність, з якої відбираються варіанти для спільного вивчення, називається генеральною, а відібрана з генеральної сукупності частина її членів носить назву вибірки (N і n відповідно).

Сутність вибіркового методу полягає в тому, щоб за властивостями частини (вибірки) можна судити про численні характеристики цілої (генеральної сукупності). Основу вибіркового методу складає той внутрішній зв'язок, який існує в популяціях між одиничним і загальним, частиною і цілим. Досвід показує, що правильно вироблена вибірка досить добре уявляє або репрезентує структуру і стан генеральної сукупності. Репрезентативність вибірки залежить, перш за все, від способу відбору варіант. У будь-якому випадку вибірка повинна бути типовою і цілком об'єктивною. І цьому сприяє випадковий відбір варіант або принцип рандомізації (за принципом лотереї та існуючої для цих цілей таблиці випадкових чисел).

2. Репрезентативність вибіркових показників

Характеристики генеральної сукупності, такі як середня величина (M), дисперсія (σ^2) і середньоквадратичне відхилення (σ) - являють собою величини постійні. По відношенню до них відповідні вибіркові характеристики, що оцінюють генеральні параметри ($\bar{x}$, σ^2 *и* σ), є величинами випадковими: вони можуть збігатися і не збігатися з величиною генеральних параметрів. Тому виникає питання про репрезентативність вибіркових показників. Можливі відхилення вибіркових показників від генеральних називаються помилками репрезентативності. Ці помилки не технічні, а статистичні, що виникли в результаті недостатньої точності, з якою вибірка репрезентує генеральну сукупність. Розміри вибіркових помилок залежать головним чином від обсягу вибірки і від розміру варіювання ознаки. На розмірах вибіркових помилок позначаються також і способи відбору варіант з генеральної сукупності.

а). Помилка середньої арифметичної

Групові середні завжди будуть відрізнятися одна від одної, тому що є випадковими величинами. Але в той же час вони будуть варіювати навколо одного і того ж центру розподілу - генеральної середньої (М), яка є величиною постійною. А так як вибіркові середні варіюють в $\sqrt{n}$ разів менше, ніж окремі варіанти однієї і тієї ж генеральної сукупності, то вибіркова помилка середньої або помилка репрезентативності обчислюється за формулами:

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \quad \text{або} \quad m_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}, \quad [23]$$

При $n<30$ користуються наступними формулами:

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{n-1}}, \quad \text{або} \quad m_{\bar{x}} = \sqrt{\frac{\sum (x_1 - \bar{x})^2}{n(n-1)}} \quad [24]$$

або

$$m_{\bar{x}} = \sqrt{\frac{1}{n-1}\frac{\sum x^2}{n} - \bar{x}^2} \quad [25]$$

або

$$m_{\bar{x}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n(n-1)}} \quad [26]$$

Для пркладу візьмемо наступні вісім варіант 2, 4, 3, 7, 5, 6, 4, 5.

Таблиця 16

(x):	2	4	3	7	5	6	4	5	$\sum x = 36$
(x^2):	4	16	9	49	25	36	16	25	$\sum x^2 = 180$

$$\bar{x} = \frac{36}{8} = 4{,}5 \quad \bar{x}^2 = 4{,}5^2 = 20{,}25$$, звідки по [25]

$$m_{\bar{x}}^2 = \frac{1}{7}(\frac{180}{8} - 20{,}25) = 0{,}32, \quad m_{\bar{x}} = \sqrt{0{,}32} = 0{,}57$$

Якщо середня арифметична обчислюється у спрощений спосіб, і її помилка визначається тим же способом за формулою:

$$m_{\bar{x}} = \sqrt{\frac{1}{n-1}[\frac{\sum a^2}{n} - (\frac{\sum a}{n})^2]} \quad [27]$$

де $a=x-A$ (відхилення від умовної середньої).

Застосуємо цю формулу до того ж прикладу:

Таблиця 17

(x):	2	4	3	7	5	6	4	5	A=4
(a):	-2	0	-1	+3	+1	+2	0	+1	$\sum a = +4$
(a^2):	4	0	1	9	1	4	0	1	$\sum a^2 = 20$

Знаходимо середню: $\bar{x} = A + \frac{\sum a}{n} = 4 + \frac{4}{8}4,5$ і її помилку:

$$m_{\bar{x}}^2 = \frac{1}{7}\left[\frac{20}{8} - (\frac{4}{8})^2\right] = 0,32, \quad \text{або} \quad m_{\bar{x}} = \sqrt{0,32} = 0,57$$

Вибіркова помилка виражається в тих же одиницях виміру, що і супроводжувані нею показники. Вона має два знака: плюс і мінус і характеризує відхилення вибіркових показників як в сторону великих (+), так і в бік менших (-) значень по відношенню до генерального параметру (M).

Середня арифметична і її помилка записується так: $x \pm m_s$, в даному прикладі цей запис виглядає у вигляді $\bar{x} \pm m_{\bar{x}}$=4,5 $\pm$ 0,57.

б). Властивості середньої помилки. Закон великих чисел

Вибіркова помилка характеризує варіювання вибіркових показників навколо їх генеральних параметрів; вона має ті ж властивості, що і середнє відхилення. Лише одна властивість специфічна для вибіркової помилки: вона зменшується при збільшенні числа спостережень (n). Це властивість вибіркової помилки обумовлена дією статистичного закону великих чисел. В цьому законі виражається внутрішній зв'язок між числом випробувань і наближенням вибіркової середньої до свого генерального параметру - математичного сподівання.

Значення вибіркової помилки: вона вказує на точність, з якою визначається середня величина.

Величина середньої помилки залежить не тільки від обсягу вибірки, а й від розмаху варіювання ознаки: чим більше розмах варіації, тим більше буде і величина вибіркової помилки і навпаки.

Поряд із зазначеними причинами на величині середньої помилки позначається і спосіб відбору варіант з генеральної сукупності.

в) Помилка при різних способах відбору варіант з генеральної сукупності.

Залежно від характеру і методики дослідження відбір варіант з генеральної сукупності може проводитися по-різному. Існує два основних способи відбору: повторний і бесповторний випадковий відбір. Повторний відбір проводиться, коли відібрані варіанти повертаються в ту ж сукупність, з якої вже були взяті. Тому вони можуть бути відібрані повторно.

Якщо ж відбираються варіанти назад в генеральну сукупність не повертаються, відбір називається бесповторного або безповоротним. У першому випадку відбір варіант не впливає на склад генеральної сукупності, і ймовірність кожної варіанти потрапити до вибірки не змінюється. У другому випадку кожен попередній відбір, впливає на результат подальшого відбору, тому що змінюється склад генеральної сукупності і ймовірність варіант потрапити до складу вибірки. Тому, помилка середньої арифметичної обчислюється за формулою:

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}} = \sqrt{\frac{\sigma^2}{n}(1-\frac{n}{N})} \quad [28]$$

Наприклад, із загального числа чоловіків (5000), які підлягають призову на військову службу, методом випадкового бесповторного відбору виміряно 500 осіб. Середній зріст призовників виявився рівним 170 см, вибіркова дисперсія σ^2 = 66,3 см. Звідси вибіркова помилка середньої арифметичної зросту призовника дорівнює

$$m_{\bar{x}} = \sqrt{\frac{66{,}3}{500}(1 - \frac{50}{5000})} = \sqrt{0{,}119} = 0{,}33 см$$

Якби відбір варіант проводився з цієї сукупності повторним випадковим способом, то помилка вибіркової середньої [23] була б рівною:

$$m_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{66{,}3}{500}} = \sqrt{0{,}13} = 0{,}36$$

Повторний і бесповторний випадковий відбір може проводитися по-різному, в залежності від того, як організовується спостереження над досліджуваним об'єктом. Звідси випливають різновиди відбору:

а) типовий або груповий відбір, який називається також районованим, може бути:

1) пропорційним,

2) непропорційним,

б) серійний або гніздовий,

в) механічний.

Всі ці види відбору спрямовані на підвищення репрезентативності вибірки, хоча вони і порушують принцип рандомізації (випадковості), тому що проводяться по заздалегідь наміченою схемою.

У разі типового відбору генеральна сукупність розчленовується на окремі і однакові (при пропорційному відборі) за складом групи або райони, з яких випадковим способом проводиться відбір якоїсь кількості варіант. Потім відібрані з кожної групи варіанти об'єднуються в одну вибіркову сукупність і піддаються спільній статистичній обробці. Вибіркова помилка середньої арифметичної у випадках типового пропорційного повторного відбору визначається за формулою $m_{\bar{x}} = \sqrt{\frac{\bar{\sigma}^2}{n}}$ [29], а при бесповторному – за формулою $m_{\bar{x}} = \sqrt{\frac{\bar{\sigma}^2}{n}(1-\frac{n}{N})}$ [30], де σ^2 середня зважена з вибіркових дисперсій. Детальніше про це читайте. Лакіна, 1973, стор.89.

При серійному відборі, генеральна сукупність ділиться на групи, звані серіями, або гніздами, які і відбираються в необхідній (намічуваному дослідником) кількості для спільної обробки. Серії можуть бути рівночисельними і неоднаковими за кількістю складових варіант. Наприклад, при вивченні фізичного розвитку учнів сільських шкіл з 50 рівновеликих груп підлітків у віці від 14 до 16 років обстеження піддалися 6 груп. Результати вимірювання окружності грудей виявилися наступними.

Таблиця 18

N групи (серії)	1	2	3	4	5	6
середнє груп ($\bar{x}_1$)	71,9	72,8	73,5	76,1	73,2	74,8
дисперсії (σ^2)	16,8	18,3	20,8	17,1	23,9	22,6

Щоб визначити помилку середньої спочатку треба обчислити середню величину міжсерійної дисперсії:

$$\bar{\sigma}_i^2 = \frac{\sum(x_i - \bar{x})}{n_i} = \frac{(71{,}9-74{,}0)^2 + (72{,}8-74{,}0)^2 + \ldots + (74{,}8-74{,}0)^2}{6} = \frac{12{,}98}{6} = 2{,}16$$ см

Звідки по [30] $m_{\bar{x}} = \sqrt{\frac{2{,}16}{6}(\frac{50-6}{50-1})} = \sqrt{0{,}32} = 0{,}57$ см

Окружність грудей підлітків - 74,0±0,57 *см.*

При механічному відборі генеральна сукупність розбивається на кілька рівних частин або груп. Потім з кожної групи випадковим чином відбирають по одній одиниці. Механічний відбір може проводитися і за іншою схемою, коли в вибірку потрапляє кожна десята, сота і.п. одиниця генеральної сукупності. Наприклад, при проведенні ботанічних екскурсій можна реєструвати кожен п'ятий, десятий і т.п. екземпляр зустрінутих рослин даного виду.

г). Помилки інших вибіркових показників

У дослідницькій роботі може виникнути необхідність об'єднати ряд вибіркових середніх з їх помилками або знайти похідні середніх, супроводжуваних помилками і т.д. У таких випадках поступають таким чином.

1. Коли визначається середня арифметична з декількох незалежних середніх з їх помилками, то вибіркова помилка такого результату обчислюється за такою формулою:

$$m_{\bar{x}s} = \frac{1}{n_1}\sqrt{m_1{}^2 + m_2{}^2 + \dots m_n{}^2} \quad [31]$$

Наприклад, на трьох рівновеликих вибірках отримані середні:

$\bar{x}_1 = 10{,}2 \pm 0{,}12$ $\bar{x}_2 = 11{,}5 \pm 0{,}18$ $\bar{x}_3 = 13{,}1 \pm 0{,}09$

$$\overline{x_s} = \frac{10.2 + 11.5 + 13.1}{3} = 11.6$$

Знаходимо помилку цього результату:

$$m_{\bar{x}_s} = \frac{1}{3}\sqrt{0{,}12^2 + 0{,}18^2 + 0{,}09^2} = \frac{1}{3}\sqrt{0{,}045} = \frac{0{,}21}{3} = 0{,}07;$$

$$\bar{x}_s \pm m_{\bar{x}_s} = 11{,}6 \pm 0{,}07.$$

2. Якщо обчислювальна сума декількох середніх арифметичних супроводжуються з їх помилками, то вибіркова помилка суми обчислюється за формулою:

$$m_{\bar{x}_s} = \sqrt{m_1^2 + m_2^2 + \ldots + m_n^2} \quad [32].$$

На тому ж прикладі:

$$\sum \bar{x} = 10{,}2 + 11{,}5 + 13{,}1 = 34{,}8;$$

$$m_{\bar{x}} = \sqrt{0{,}12^2 + 0{,}18^2 + 0{,}09^2} = \sqrt{0{,}045} = 0{,}21.$$

3. Помилка похідних двох вибіркових середніх з їх помилками визначається за формулою:

$$m_{n} = \bar{x}_1 \times \bar{x}_2 \sqrt{\left(\frac{m_1}{\bar{x}_1}\right)^2 + \left(\frac{m_2}{\bar{x}_2}\right)^2} \qquad [33].$$

Наприклад.

$$\bar{x}_1 = 10{,}3 \pm 0{,}11\,;\ \bar{x}_2 = 8{,}2 \pm 0{,}12\,.$$

$$m_{n} = 10{,}3 \times 8{,}2\sqrt{\left(\frac{0{,}11}{10{,}3}\right)^2 + \left(\frac{0{,}12}{8{,}2}\right)^2} = 84{,}46 \times 0{,}18 = 15{,}2.$$

4. Помилка часткового від ділення середніх арифметичних з їх помилками визначається за такою формулою:

$$m_{ch} = \frac{\bar{x}_1}{\bar{x}_2}\sqrt{\left(\frac{m_1}{\bar{x}_1}\right)^2 + \left(\frac{m_2}{\bar{x}_2}\right)^2} \qquad [34].$$

Для розглянутого раніше прикладу:

$$m_{ch} = \frac{10{,}3}{8{,}2}\sqrt{\left(\frac{0{,}11}{10{,}3}\right)^2 + \left(\frac{0{,}12}{8{,}2}\right)^2} = 1{,}26 \times 0{,}18 = 0{,}22.$$

5. Помилка різниці вибіркових середніх $\overline{x_1} - \overline{x_2} = D\overline{x_1} - \overline{x_2} = D$ двох незалежних, але нерівновеликих вибірок (тобто n1 = n2) дорівнює:

$$m_D = \sqrt{\sigma_S^2 \times \frac{n_1 + n_2}{n_1 \times n_2}} = \sqrt{\sigma_S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad [36],$$

где $\sigma_S^2 = \dfrac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2} = \dfrac{n_1(n_1 - 1)m_1^2 + n_2(n_2 - 1)m_2^2}{n_1 + n_2 - 2} = \dfrac{\sum a_1^2 + \sum a_2^2}{n_1 + n_2 - 2}$

и $a = (x - \bar{x})$. Так, что

$$m_D = \sqrt{\frac{\sum a_1^2 + \sum a_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{\frac{\sum a_1^2 + \sum a_2^2}{n_1 + n_2 - 2} \times \frac{n_1 + n_2}{n_1 \times n_2}} \qquad [37].$$

7. Вибіркова помилка різниці середніх $(\overline{x_1} - \overline{x_2} = D)$ сполучених розподілів, тобто які знаходяться в залежності один від одного або від якоїсь загальної причини, обчислюється за формулою [35] з поправкою на спряженість, тобто

$$m_{1+2} = \sqrt{m_1^2 + m_2^2 - 2rm_1m_2} \qquad [38].$$

Тут r - коефіцієнт кореляції, що показує ступінь спряженості двох рядів розподілів (про нього будемо говорити пізніше).

Але вибіркову помилку різниці середніх арифметичних сполучених розподілів можна обчислити, не використовуючи коефіцієнта кореляції за такими аналогічним формулами:

$$m_d = \sqrt{\frac{1}{n-1}\left(\frac{\sum d^2}{n} - \bar{d}^2\right)} = \sqrt{\frac{\sum\left(d-\bar{d}\right)^2}{n(n-1)}} \quad [39];$$

$$m_d = \sqrt{\frac{\sum d^2 - \left(\frac{\sum d}{n}\right)^2}{n(n-1)}} \quad [39a].$$

тут d- різниця між відповідними варіантами сполучених рядів X і $У$, тобто d = x-y, d - середня різниця, тобто $\frac{\sum d}{n} = \bar{d};\ \bar{x} - \bar{y} = D_y$, де

n - загальне число парних спостережень.

д) Показник точності оцінки параметрів

Сама по собі абсолютна величина вибіркової помилки як показник іменований мало придатна для випадків порівняльної оцінки точності, з якою визначені середні результати спостережень по відношенню їх до генеральних параметрів.

Наприклад, є середні $\bar{x}_1$ =86,1$\pm$0,7 см, $\bar{x}_2$=17,4±0,2 г.

За абсолютною величиною їх помилок важко сказати, яка середня визначена більш точно, оскільки середні з їх помилками виражені різними одиницями виміру. Щоб отримати певне уявлення про точність, з якою визначено той чи інший середній результат, прийнято використовувати так званий показник точності Cs:

$$Cs = \frac{m_{\bar{x}}}{\bar{x}} \times 100\% \quad [40].$$

Коли відомо значення коефіцієнта варіації (CV), показник точності можна визначити за такою формулою:

$$Cs = \frac{CV}{\sqrt{n}} \quad [40a].$$

Під точністю визначення вибіркової середньої розуміється ступінь наближення її до середньої генеральної сукупності. Чим точніше визначено середній результат, тим менше буде *Cs*, і, навпаки, при менш точному середньому результаті показник *Cs* виявиться більше. Точність вважається достатньою, якщо *Cs* не перевищує 3-5%. Так, для наведених раніше середніх, показники точності в обох випадках будуть наступними:

$$Cs_1 = \frac{0,7}{86,1} \times 100\% = 0,81\%;$$

$$Cs_2 = \frac{0,2}{17,4} \times 100\% = 1,15\%.$$

Звідси також видно, що перша середня визначена більш точно, ніж друга.

3. Статистична перевірка гіпотез

а). нульова гіпотеза

Майже у всіх випадках вибіркового спостереження параметри генеральної сукупності залишаються невідомими. Про них доводиться судити за вибірковими даними, які є величинами випадковими. Для оцінки величини генеральних параметрів за вибірковими показниками використовується так звана нульова гіпотеза. Тобто припущення про те, що генеральні параметри, про які судять за вибірковими даними, не відрізняються один від одного, і що різниця, яка спостерігається між вибірковими показниками, носить не систематичний, а виключно випадковий характер. Висуваючи нульову гіпотезу, експериментатор виходить з припущення, що спостережувана мінливість ознаки залежить не від дії організованого фактора, а визначається другорядними, нерегульованими в досвіді випадковими причинами.

б). Рівні значущості і довірчі ймовірності

Сформульована гіпотеза потребує перевірки. Щоб її прийняти або відкинути, потрібні підстави. Підстави дає теорія ймовірностей, що дозволяє пов'язувати статистичні гіпотези з певною ймовірністю. Дотримуючись закону нормального розподілу, можна стверджувати, що в 95% випадків вибіркова середня ($\bar{x}$) не відхиляється від середньої (М) генеральної сукупності більше, ніж на *2t*, де $t = \frac{\bar{x} - M}{\sigma}$. І тільки в 5% випадків, вважаючи відхилення в + і - напрямках від (М), вибіркова середня вийде за ці межі. Це означає, що ймовірність отримати у вибірці середній результат, який відхилиться від генерального параметра на *2t,* дорівнює лише 0,05. Якщо ж мова йде про відхилення від (М) тільки в одну сторону, ймовірність буде вдвічі менше (Р = 0,025). Так ось, відсоток таких малоймовірних випадків, які суперечать прийнятій гіпотезі, ставить її під сумнів, називається рівнем значущості гіпотези. У біологічних дослідженнях зазвичай приймається 5% -й рівень значимості, якому відповідав би ймовірність $P_1 = 0{,}05$. У більш відповідальних випадках, коли висновки повинні бути особливо суворими, приймається 1% -й або 0,1% -й рівні значущості, яким відповідає $P_2 = 0.01$ і $P_3 = 0.001$.

Таким чином, ймовірність, якій вирішено знехтувати при оцінці генеральних параметрів поданням вибіркових спостережень, виражається прийнятим рівнем значущості.

Імовірність же зворотних випадків, коли гіпотеза заслуговує довіри, називається довірчою ймовірністю. Зазвичай в дослідницькій практиці приймаються три порога довірчої ймовірності: $P_1 = 0{,}95$; $P_2 = 0{,}99$; $P_3 = 0{,}999$.

Кожен поріг, або рівень довірчої ймовірності, зв'язується з певною величиною нормованого відхилення:

Довірна ймовірність *(P)*	Нормоване відхилення (t)
0,95	1,96
0,99	2,58
0,999	3,29

Величина довірчої ймовірності або рівень значущості при перевірці гіпотез встановлюється самим дослідником в залежності від ступеня точності, з якою проводиться дослідження і відповідальності висновків, що випливають з нього. Якщо $P \geq 0{,}05$ або ж $P < 0{,}95$, то відкидати нульову гіпотезу немає підстав. Коли ж $P < 0{,}05$ або $P > 0{,}95$ нульова гіпотеза відкидається.

в) Довірчий інтервал і його межі

Межі, в яких з тією чи іншою ймовірністю знаходиться параметр генеральної сукупності, називаються довірчими, а інтервал, укладений між цими межами, носить назву довірчого. У загальній формі можна наступним чином встановити довірчий інтервал для невідомого параметра М генеральної сукупності:

$$-t \le \frac{x - M}{\sigma} \le +t,$$

так як ймовірність відхилення кожен вид від центру розподілу визначається функцією нормованого відхилення. Перетворивши цей вираз, отримуємо

$$x - t\sigma \le M \le x + t\sigma .$$

Це і є довірчий інтервал, в якому знаходиться величина генерального параметра М. Тут $x - t\sigma$ і $x + t\sigma$- довірчі кордону, t - нормоване відхилення, яке визначається порогом довірчої ймовірності.

Так з ймовірністю $P = 0{,}95$ (відповідає $t = 1{,}96$) можна стверджувати, що невідомий генеральний параметр М нормального розподіляється сукупності знаходиться в інтервалі:

$$x - 1{,}96\sigma \le M \le x + 1{,}96\sigma .$$

г). Критерії оцінки

Визначення можливих значень генеральних параметрів за величиною вибіркових показників носить загальну назву оцінки генеральних параметрів. Критерієм оцінки служить стандартна величина нормованих відносини (t_{st}) з якою порівнюється фактичне значення цього критерію ($t_ф$). Відносно генеральної середньої М цей критерій виражається наступними аналогічними відносинами:

$$t_ф = \frac{\bar{x} - M}{m_{\bar{x}}}, \quad \text{или} \quad t_ф = \frac{\bar{x} - M}{\sigma}\sqrt{n} \quad [41].$$

При $t_ф < t_{st}$ нульова гіпотеза зберігається. Якщо ж $t_ф \ge t_{st}$ нульову гіпотезу слід відкинути. Наприклад, в одній з кошар овцесовхозу на 95 особин середній настриг вовни на 1 вівцю склав 6,2 кг при $\sigma = 0{,}43$кг. Чи можна на підставі цього результату зробити висновок, що настриг вовни на цій кошарі достовірно знижений у порівнянні із середнім настригом вовни по радгоспу, рівному 6,4 кг / вівцю! Нормуючи відомі величини, знаходимо:

$$t_{\phi} = \frac{6,2-6,4}{0,43} \times \sqrt{95} = -4.5.$$

Для довірчої ймовірності P = 0,99 t_s = 2,58. Так як $t_{\phi} \geq t_{st}$ нульова гіпотеза відкидається.

Можна також оцінити достовірність (тобто не випадковість) відмінностей, що спостерігаються між середніми $\overline{x_1}$ и $\overline{x_2}$ (див. Лакіна, 1973, стор.101).

При порівнянні статистичних показників один з одним слід враховувати, на яких сумах - залежних або незалежних - вони отримані. Якщо варіанти однієї ознаки X розподіляються незалежно від значень іншої ознаки Y, вони називаються незалежними. Якщо ж значення однієї ознаки в тій чи іншій мірі пов'язані з відповідними значеннями іншої ознаки, вони залежні один від одного.

4. Малі вибірки

а) Т-розподіл Стьюдента

У багатьох випадках обсяг вибіркової сукупності не перевищує 20-30 спостережень. Такі вибірки називають малими.

Коли вибірки незалежні, різниця між генеральними пар-метрами оцінюється по різниці вибіркових середніх ($\overline{x_1} - \overline{x_2} = D$). Число ступенів свободи в таких випадках визначається за формулою:

$$k = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2.$$

Якщо ж порівнювані вибірки залежні одна від одної, то різницю між параметрами слід обчислювати не по різниці вибіркових середніх, а за середньою різниці між парними варіантами *(x - y = d)* сполучених розподілів. У цьому випадку число ступенів свободи визначаються за формулою:

$$k = n - 1$$ або ($k = n - 2$)

$$t_ф = \frac{\bar{x}_1 - \bar{x}_2}{m_D} \geq t_{st}$$

Нульова гіпотеза відкидається при

для відповідних *P* і *к*.

Переходимо до розгляду відповідних прикладів, на яких легше засвоїти значення критерію *t* в оцінці генеральних параметрів за даними вибіркових спостережень.

а-1) Випадки незалежних вибірок

Розберемо на наступному прикладі. Вивчався вплив кобальту на збільшення живої ваги кроликів. Дослід проводився на двох групах тварин - дослідної та контрольної. Вік кроликів в обох групах коливався від 1,5 до 2-х місяців. Вихідна вага особин не виходив за межі 500-600 г. Дослід проводився півтора місяці. Обидві групи тварин містилися на одному і тому ж кормовому раціоні. Досвідчені кролики щодня отримували у вигляді водного розчину по 0,06 г хлористого кобальту на 1 кг живої ваги. За час досвіду тварини дали наступні збільшення у вазі:

Таблиця 19

контрольні:	504	560	580	600	420	530	490	580	470	n=9
дослідні:	580	692	700	621	640	561	680	630	-	n=8

Відразу ж потрібно відзначити, що дані величини варіюють незалежно: кожна величина приймає те чи інше значення незалежно від значення іншої величини. Зводимо результати досвіду в таблицю.

Таблиця 20

Прирости (г)		Відхилення від $\overline{x}$		Квадрати відхилень	
дослід	контроль	дослід	контроль	дослід	контроль
580	504	58	22	3364	484
692	560	54	34	2916	1156
700	420	62	106	3844	11236
621	600	17	74	289	5476
640	580	2	54	4	2916
561	530	77	4	5929	16
680	490	42	36	1764	1269
630	580	8	54	64	2916
-	470	-	54	-	3136
∑=5104	∑=7434	-	-	∑=18174	∑=28632
$\overline{x}_1$ =638	x, =526	-	-	∑=46806	

Різниця в контролі і досвіді 638-526 = 112 г. По формулі [37] визначаємо помилку цієї різниці.

$$m_D = \sqrt{\frac{\Sigma a_1^2 + \Sigma a_2^2}{n_1 + n_2 - 2} \times \frac{n_1 + n_2}{n_1 \times n_2}} = \sqrt{\frac{46806}{8+7} \times \frac{9+8}{9 \times 8}} = \sqrt{736,8} = 27,13 \text{ г}.$$

Критерій достовірності

$$t_\phi = \frac{\overline{x}_1 - \overline{x}_2}{m_D} = \frac{112}{27,13} = 4,1$$

Для рівня значущості P = 0,01 і для числа ступенів свободи k = 9 + 8- 2 = 15 знаходимо (по таблиці стандартних значень t-критерію Стьюдента) значення t_{st}, рівне 2,95, $t_ф > t_{st}$, (фактичне значення значно перевершує критичне значення), отже, нульову гіпотезу потрібно відкинути і визнати статистично достовірної різницю в приростах кроликів в досвіді і контролі.

Розглянемо ще один аналогічний приклад. На двох групах лабораторних мишей - дослідної та контрольної - з'ясовувалося дію хіміко-терапевтичного препарату на розвиток організму тварин. В результаті місячних випробувань виявилися наступні відмінності у вазі тварин (в грамах):

Таблиця 21

контрольні:	70	78	60	80	60	60	68	$\bar{x}_1$=68 г
дослідні:	80	75	62	70	68	71	-	$\bar{x}_2$=71 г

D=71 г-68 г=3,0 г

Визначимо помилку цієї різниці. Спочатку розрахуємо суми квадратів відхилень варіант від їх середніх за формулою:

$$\Sigma a^2 = \Sigma(x_i - \bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n} \quad [42].$$

Контроль: ь: $\Sigma a_1^2 = (70^2 + 78^2 + 80^2 + ... + 68^2) - \frac{476^2}{7} = 32808 - 32368 = 440.$

Експеримент: $\Sigma a_2^2 = (80^2 + 75^2 + 62^2 + ... + 71^2) - \frac{426^2}{6} = 30434 - 30246 = 188.$

Знаходимо об'єднаний середній квадрат відхилень:

$$\sigma_x^2 = \frac{440 + 188}{7 + 6 - 2} = \frac{628}{11} = 57,1,$$

звідки помилка різниці середніх обчислюється так:

$$m_D = \sqrt{\sigma_x^2 \times \frac{n_1 + n_2}{n_1 \times n_2}} \quad [43].$$

$$m_D = \sqrt{57,1 \times \frac{13}{42}} = \sqrt{17,7} = 4,2г.$$

Критерій достовірності відмінностей $t_ф$ = *3,0/4,2=0,71..* По таблиці стандартних значень критерію *t*-Стьюдента для рівня значущості *P* = 0,05 і *k* = 11 знаходимо t_{st} = *2,2,* $t_ф < t_{st}$ отже, нульова гіпотеза зберігається, тобто різницю між генеральними середніми цих груп статистично недостовірна і, отже, дія хіміко-терапевтичного препарату статистично недостовірно.

Коли відома генеральна середня (М), то різниця між нею і вибіркової середньої ($\bar{x}$) оцінюється помилково вибіркової середньої, тому що генеральна середня помилки не має:

$$t = \frac{\bar{x} - M}{m_{\bar{x}}} \quad \text{или} \quad t = \frac{\bar{x} - M}{\sigma}\sqrt{n} \quad [44].$$

Наприклад, методом селекції на підвищення жирномолочності створена лінія великої рогатої худоби загальною чисельністю 12 тварин із середнім відсотком жиру в молоці 4,16 ± 0,025%. Вихідна порода характеризується середньою жирномолочністю 4,09%. Чи правдива різниця в жирності молока у новоствореної лінії тварин?

$$t = \frac{4,16 - 4,09}{0,025} = \frac{0,07}{0,025} = 3,2.$$

Для рівня значущості *P = 0,01* и *k = n-1 = 12-1 -11* t_{st} *=3,11.* $t_ф < t_{st}$ нульова гіпотеза відкидається. Відбір на жирномолочність виявився ефективним.

А-2). Випадки залежних вибірок

При оцінці вибірок, значення яких варіюють в певній залежності один від одного (а це часто буває пов'язано з самим характером досвіду), розглянутий раніше спосіб оцінки генеральних параметрів виявляється неточним. Розглянемо наступний приклад. Вивчався вплив чорного і квітневого пара на урожай жита. Досвід тривав протягом 6 років. Враховувалася вага 1000 зернин в грамах. Результати досліду виявилися наступними (по Сапегін, 1937):

Таблиця 22

Рік посіву	1898	1899	1901	1902	1903	1904	
По чорному пару	31,3	24,0	24,6	28,6	29,1	30,1	$\bar{x}_1$=27,9 г
По квітневому	31,6	24,2	24,8	29,1	31,0	31,0	$\bar{x}_2$=28,4 г

Видно, що урожай жита по квітневому пару трохи вище, ніж по чорному $D = \bar{x}_1 - \bar{x}_2 = 0,5$ *г.* Чи можна покластися на цю різницю, чи достовірна вона?

Якщо використовувати вже розглянутий нами підхід до незалежних вибірках то $t_ф$*= 1,6.* Для *k = 10* и *P = 0,05* t_{st}*= 2,23.* $t_ф < t_{st}$,, отже, нульову гіпотезу відкинути не можна. Якщо ж порівнювати ні середні, а варіанти, тобто оцінювати генеральні параметри по середній різниці варіант з урахуванням пов'язаності між ними, вийде наступний результат.

Таблиця 23

посів	Вага 1000 зерен в грамах по роках						середнє
	1898	1899	1901	1902	1903	1904	
По чорному пару	31,1	24,0	24,6	28,6	29,1	30,1	27,9
По квітневому	31,6	24,2	24,8	29,8	29,9	31,0	28,4
Різниця (d)	0,5	0,2	0,2	0,5	0,8	0,9	-
Квадрат різниці (d^2)	0,25	0,04	0,04	0,25	0,64	0,81	$\sum d^2=2.03$

Помилка середньої різниці, яка визначається за формулою [39] дорівнює:

$$d = \bar{x}_1 - \bar{x}_2 = 28,4 - 27.9 = 0.5.$$

$$m_d = \sqrt{\frac{1}{n-1}\left(\frac{\sum d^2}{n} - \bar{d}^2\right)} = \sqrt{\frac{1}{5}\left(\frac{2,03}{6} - 0,5^2\right)} = \sqrt{0,018} = 0,13 \text{ г}.$$

Критерій $t_ф = \frac{\bar{x}_1 - \bar{x}_2}{m_d} = \frac{\bar{d}}{m_d} = \frac{0,5}{0.13} = 3,85.$

Для P = 0,05, k = 6-1= 5, t_{st}= 2,57, $t_ф > t_{st}$, отже, нульова гіпотеза відхиляється, і різниця визнається статистично достовірною.

Наведений приклад служить наочним підтвердженням того, що статистичні методи можна застосовувати огульно, не погодившись з вмістом експериментального матеріалу.

Як критерій достовірності середньої різниці може служити також і відношення:

$$t = \frac{\bar{d}}{\sigma_d}\sqrt{n} \qquad [45],$$

$$\text{де} \quad \bar{d} = \frac{1}{n}\Sigma d_i \ , \quad \text{а} \quad \sigma_d = \sqrt{\frac{1}{n-1}\left(\Sigma\left(d_i - \bar{d}\right)^2\right)},$$

Яке оцінюється по таблиці Стьюдента.

б) Порівняння дисперсії. F-розподіл Фішера

Для роботи на малих вибірках Фішер запропонував певні підходи, які привели до показника, названого критерієм Фішера, який розраховується за формулою:

$$F = \frac{\sigma_1^2}{\sigma_2^2} \quad [46], \text{ причому } \sigma_1^2 > \sigma_2^2, \text{ т.к. } F \geq 1.$$

Критерій Фішера функціонально пов'язаний з ймовірністю. Він залежить тільки від числа ступенів свободи $k_1=n_1-1$ и $k_2=n_2-1$ порівнюючих дисперсій. Характерним для F виявляється те, що він повністю визначається вибірковими дисперсіями і не залежить від генеральних параметрів, тому що передбачається, що обидві дисперсії з однієї і тієї ж генеральної сукупності. Для визначення довірчих меж для критерію Фішера існують таблиці стандартних значень. У цій таблиці ступеня свободи для більшої дисперсії беруться по горизонталі, для меншої дисперсії - по вертикалі. На перетині цих стовпців дані стандартні значення критерію Фішера для P = 0,05 і P = 0,01.

Нульова гіпотеза виходить з визнання рівності дисперсій. Якщо емпіричні значення ($F_ф$) менше теоретичних (F_{st}) для відповідного рівня значущості і ступенів свободи k_1 і k_2, різниця розглядається як випадкова. Якщо ж $F_ф \geq F_{st}$ нульова гіпотеза відкидається, різниця між порівнюваними величинами визнається статистично достовірною. Наприклад, порівнюються дві вибірки насіння квасолі за вагою:

одна з посівного матеріалу: $n_1 = 100$, σ_1 - 58,3 мг;

інша - з урожаю: $n_2 = 200$, $\sigma_2 = 59.3$ мг.

Чи правдива різниця?

Обчислюємо значення критерію Фішера:

$$F_{ф} = \frac{(59{,}3)^2}{(58{,}3)^2} = \frac{3516}{3398} = 1{,}03$$

По таблиці Фішера для Р = 0,05 и k = 200 - 1 - 199 и k_2 = 100 - 1 - = 99, знаходимо F_{st}= 1,31. $F_{ф} < F_{st}$, нульова гіпотеза зберігається, розбіжності між вибірками за цією ознакою виявляються статистично недостовірними.

Тема VII. Оцінка законів розподілу

1. Оцінка вискакуючих варіант.

2. Наближені оцінки закону розподілу:

а) обчислення асиметрії і ексцесу за способом умовної середньої;

б) оцінка показників асиметрії та ексцесу;

в) про причини асиметричних розподілів.

3. Критерій відповідності емпіричних і теоретичних розподілів. Критерій χ2.

4. Поняття трансгресії.

1. Оцінка вискакуючих варіант

Бувають випадки, коли окремі крайні варіанти сильно відхиляються від сусідніх з ними варіант варіаційного ряду. Тоді виникає сумнів у тому, чи належать вони до даної генеральної сукупності. Причини таких явищ можуть бути різними: по-перше, можливі технічні помилки, допущені при утворенні вибіркової сукупності, по-друге, "вискакування" варіант може бути наслідком сильної варіабельності ознаки, тобто явищем цілком нормальним.

Критерієм оцінки вискакуючих варіант служить нормоване відхилення:

$$t = \frac{x_i - \bar{x}}{\sigma} \quad [47]$$

Нульова гіпотеза в даному випадку говорить про те, що "вискакуюча" варіантаналежить до тієї ж генеральної сукупності і відкидається, коли $t_ф < t_{st}$.

Наприклад, на п'яти дослідних ділянках було отримано (в перерахунку в ц / га) наступний урожай:

8,3 7,9 9,1 6,8 і 12,1 ц / га.

Викликає сумнів варіанту 12,1, сильно ухиляється від середнього врожаю x = 8,84, σ = 2,0 ц / га. Оцінимо її "вискакування" на достовірність. Знайдемо нормоване відхилення t $= \frac{12{,}1-8{,}84}{2{,}0} = \frac{3{,}26}{2{,}0} = 1{,}63.$ По таблиці

"Критичні значення величини нормованого відхилення при оцінці сумнівних варіант", знаходимо значення t_{st} для P = 0,95 и n = 5, t_{st}= 1,92. Так як $t_{st}> t_{ф}$ нульова гіпотеза зберігається, відкидати цю варіанту при розрахунку середнього врожаю не можна.

2. Наближені оцінки закону розподілу

Середня арифметична, *сг* і *о* самі по собі не містять інформації про закон розподілу. Не всі ознаки розподіляються по нормальному закону: деякі виявляють явну асиметрію, можливі й інші випадки відхилення від нормального закону розподілу. Тому, перш ніж використовувати той чи інший критерій оцінки генеральних параметрів, слід скласти уявлення про закон розподілу досліджуваної ознаки. Наближена оцінка закону розподілу може бути отримана за допомогою коефіцієнтів асиметрії та ексцесу.

а) Обчислення асиметрії і ексцесу за способом умовної середньої

Варіаційні ряди можуть бути:

а) скошеними (позитивна і негативна асиметрія),

б) гостро-і плосковерхівними (позитивний і негативний ексцес).

Мірою скошеності рядів розподілу служить коефіцієнт асиметрії, позначається символом As.

$$As = \frac{\sum p(x_i - \bar{x})^2}{n\sigma^2} = \frac{\sum pa^2}{n\sigma^2} \quad [48].$$

При строго симетричному розподілі As=0, т.к.$\sum(x_1-\bar{x})^3= 0$. Коефіцієнт асиметрії величина відносна; він коливається від 0 до 1. Якщо As $\leq$ 0,2 асиметрія вважається незначною, при As> 0,5 скошенність розподілу виявляється вже сильною. Робоча формула для обчислення коефіцієнта асиметрії за способом умовної середньої наступна:

$$As = \frac{\frac{\Sigma pa^3}{n} - 3b\frac{\Sigma pa^2}{n} + 2b^3}{\sigma^3} \quad [49],$$

де $b = \frac{\Sigma pa}{n}$ - умовний момент 1 порядку,

$\frac{\Sigma pa^2}{n}$ - умовний момент 2 порядку,

$\frac{\Sigma pa^3}{n}$ - умовний момент 3 порядку.

де $a = (x_i - A) / i$.

Обчислимо за цією формулою коефіцієнт асиметрії для розподілу Са в сироватці крові павіанів-гамадрилів.

Таблиця 24

Класові варіанти (x)	Частоти (P)	A=(x_i-A)/i	pa	Pa^2	pa^3	pa^4
8,9	2	-4	-8	32	-128	512
9,6	3	-3	-9	27	-81	243'
10,3	9	-2	-18	36	-72	144
11,0	17	-1	-17	17	-17	17
Разом	-	-	-52	-	-298	-
A = 11,7	25	0	0	0	0	0
12,4	23	+ 1	+23	23	+23	23
13,1	10	+2	+20	40	+80	160
13,8	7	+3	+21	63	+ 189	567
14,5	4	+4	+ 16	64	+256	1024
Разом	-	-	+80	-	+458	-
Сума	100	-	+28	302	+250	2690

$$b = \frac{+28}{100} = 0,28$$ - умовний момент 1 порядку,

$$\frac{\Sigma pa^2}{n} = \frac{302}{100} = 3,02$$ - умовний момент 2 порядку,

$$\frac{\Sigma pa^3}{n} = \frac{250}{100} = 2,50$$ - умовний момент 3 порядку.

$$3b=0,84;\ 3b\frac{\Sigma pa^2}{n} = 2,537;\ b^3 = 0.0222;\ 2b^3=0.044;\ \sigma^3 = 5,0$$

Підставивши в формулу [49], отримуємо As = (2,50 - 2,54 + 0,04) / 5,0 я «0,01. Отримана величина настільки мала, що дає вагому підставу вважати цей розподіл симетричним.

Величина ексцесу вимірюється за допомогою коефіцієнта ексцесу *Ex*, який обчислюється за формулою:

$$Ex = \frac{\Sigma p(x_i - \bar{x})^4}{n\sigma^4} - 3 = \frac{\Sigma pa^4}{n\sigma^4} - 3 \quad [50].$$

Для строго симетричних розподілів As ~ 0. позитивний ексцес має знак (+), а негативний (-). Гранична межа негативного ексцесу -2, а позитивний ексцес може бути будь-якої величини. Позитивний ексцес вважається незначним, якщо Ex <0,5.

Для розрахунку ексцесу за способом умовної середньої скористаємося формулою:

$$Ex = \frac{\frac{\Sigma pa^4}{n} - 4b\frac{\Sigma pa^3}{n} + 6b^2\frac{\Sigma pa^2}{n} - 3b^4}{\sigma^4} - 3 \quad [50a].$$

де $a=(x_i - A)/i$, $b = \frac{\Sigma p(x_i - A)}{n}$.

Скористаємося таблицею для утримання Ca^{++} в крові павіанів-гамадрилів, отримаємо:

$$\frac{\Sigma pa^4}{n} = \frac{2690}{100} = 26{,}90; \qquad \frac{\Sigma pa^3}{n} = \frac{250}{100} = 2{,}50; \qquad \frac{\Sigma pa^2}{n} = \frac{302}{100} = 3{,}02;$$

$$b = \frac{\Sigma pa}{n} = \frac{+28}{100} = 0{,}28; \quad 4b = 1{,}12; \quad 6b^2 = 0{,}47; \quad 3b^4 = 0{,}005;$$

$$4b\frac{\Sigma pa^3}{n} = 2{,}8; \quad 6b^2\frac{\Sigma pa^2}{n} = 1{,}42; \quad \sigma = 1{,}715; \quad \sigma^4 = 8{,}64.$$

Підставивши значення, отримаємо

$$Ex = [(26{,}90 - 2{,}8 + 1{,}42 - 0{,}005)/8{,}64] - 3 = 2{,}95 - 3 = -0{,}05.$$

Виходить мізерна величина, що дозволяє вважати відсутність ексцесу у даного розподілу.

б). Оцінка показників асиметрії та ексцесу

Як і інші вибіркові показники, коефіцієнти асиметрії та ексцесу є величинами випадковими. Щоб відрізнити уявну асиметрію і ексцес від дійсних, необхідна статистична оцінка достовірності вибіркових показників асиметрії та ексцесу.

в). Про причини асиметричних розподілів

Асиметрія може виникати від причин двоякого роду:

- Чисто технічних - внаслідок угруповання матеріалу у варіаційні ряди (помилкова асиметрія).

- Від біологічних, пов'язаних головним чином з перемішуванням генетично константних чистих ліній в процесі їх схрещування один з одним.

3. Критерій відповідності емпіричних і теоретичних розподілів. критерій χ2

Статистична оцінка розбіжностей, які спостерігаються між емпіричними і теоретичними (або очікуваними) частотами варіаційного ряду проводиться за допомогою особливих критеріїв відповідності. Один з них, званий критерієм χ^2, запропонований Пірсоном в 1901 р.

$$\chi 2 = \sum \frac{(p - p')^2}{p'} \quad [51]$$

де p - емпірична частота, p'- відповідна теоретична частота. Якщо p-p'=d, тоді

$$\chi 2 = \sum \frac{d^2}{p'} \quad [51a]$$

Для знаходження величини критерію χ^2 необхідно:

1. За кожним класом варіаційного ряду знайти d.

2. Звести різницю в квадрат, і розділити її на p '.

3. Підсумувати отримані відношення для всіх класів варіаційного ряду.

Отримана величина і буде значенням χ^2. Вона завжди позитивна. Якщо Σ (p-p ') = 0, то $\chi^2 = 0$. отже, має місце повна відповідність фактичних частот очікуваним (або обчисленим). Якщо ж $\chi^2 \neq 0$, то необхідно оцінити достовірність відмінностей, порівнявши зі стандартним значенням χ^2. Якщо $\chi^2{}_1 \geq \chi^2{}_{st}$- нульова гіпотеза відкидається.

Продемонструємо застосування критерію χ2 на прикладі розподілу датських вугрів за кількістю хребців, маємо на увазі гіпотезу про нормальність цього розподілу. Зведемо дані в таблицю.

Таблиця 25

Варіанти (x)	Частоти		Різниця (d)	Квадрат різниці (d^2)	d^2/p'
	емпіричні (p)	вирахувані (p')			
111	3	1.7	1.3	1,69	0,99
112	9	10.0	1.0	1,00	0,10
1 13	31	34,3	-1	10,89	0,32
114	71	67,9	j, j	9,61	0,01
115	82	77,6	3,1	19,36	0,25
116	46	51,1	4.4	26,01	0,51
117	19	19,5	5.1 ,5	0,25	0,01
118	6	4.8	1.2	1,44	0,30
Сума	267	267,2	-	-	2,49

Звідки χ^2 = 2,49. При оцінці емпіричних розподілів, які слідують нормальному закону, число ступенів свободи k = n-3 (з урахуванням трьох обмежень свободи варіації: п. .7, А). По таблиці стандартних значень χ^2 для числа ступенів свободи k=n-3=8-3=5 $\chi^2{}_{st}$ =11,1. $\chi^2{}_{ф} < \chi^2{}_{st}$ - нульова гіпотеза зберігається, розбіжності між емпіричними частотами і обчисленими за нормальним законом слід визнати випадковими.

Наступний приклад ілюструє застосування критерію χ^2 для з'ясування достовірності закономірностей розщеплення при схрещуванні. В результаті схрещування гібридів першого покоління між собою вийшло таке потомство гороху: 787 високих і 277 низьких рослин.

Оцінимо достовірність цього розщеплення як моногібридного з повним домінуванням (3: 1). Для цього занесемо дані в таблицю:

Таблиця 26

	Значення ознаки		Всього
	високі	низькі	
Емпіричні(p)	787	277	1064
Вирахувані (p’)	1064 x 3/4 = 798	1064 x 1/4 = 266	1064
d=p-p'	-11	+ 11	0
d^2	121	121	-
d^2-p'	0,15	0,44	χ^2=0,59

Звідки χ^2 = 0,59. По таблиці стандартних значень χ^2 для числа ступенів свободи k=n-1=2-1 = 1 χ^2_{st} =3,84. $\chi^2_{ф}< \chi^2_{st}$- нульова гіпотеза зберігається, розбіжності між емпіричними частотами і обчисленими за законом моногібридного схрещування слід визнати випадковими.

4. Поняття трансгресії

При розподілі незалежних вибірок в варіаційні ряди нерідко доводиться спостерігати, що частина варіант цих вибірок розподіляються по одним і тим же класах, хоча між середніми арифметичними цих рядів існує статистично достовірна різниця.

Ряди, у яких частина класів виявляється загальною, а між середніми арифметичними виявляється статистично достовірна різниця, називаються трансгресуючими рядами. Сам же факт неповного розмежування варіаційних рядів носить назву трансгресії.

Ступінь трансгресії може бути дуже різною. Вимірювання величини трансгресії є важливим елементом біометричного аналізу (Лакин, 1973, стор. 144), тому що дає відповідь на питання про те, чи належать або не належать розглянуті вибірки однієї і тієї ж генеральної сукупності.

Тема VIII. Кореляційний аналіз

1. Поняття кореляції і основні завдання кореляційного аналізу.
2. Коефіцієнт кореляції.
3. Основні властивості коефіцієнта кореляції.
4. Довірча оцінка коефіцієнта кореляції.
5. Метод Z.
6. Мінімальна кількість спостережень для планованої точності коефіцієнта кореляції.
7. Оцінка різниці між коефіцієнтами кореляції.
8. Обчислення коефіцієнта кореляції на малих вибірках.
9. Кореляційне відношення.

1. Поняття кореляції і завдання кореляційного аналізу

Наявність існуючого в живій природі зв'язку необхідно виміряти. Залежність між змінними випадковими величинами X і Y, при якій кожному значенню однієї з них відповідає не якесь конкретне значення, а певна групова середня іншої величини, тобто $y_x = f(x_i)$ или $x_y = f(y_i)$ називається кореляційною або просто кореляцією. Математичний аналіз зв'язків, існуючих між випадковими величинами, становить зміст кореляційного аналізу.

У напрямку кореляція буває позитивною або прямою, і негативною, або зворотною, а за формою - лінійною (прямолінійною) або нелінійною, або криволінійною.

При позитивній кореляції, групові середні однієї ознаки зростають зі збільшенням значень іншої ознаки. При негативній кореляції групові середні однієї ознаки зменшуються при збільшенні значень іншої ознаки. Наприклад, зі збільшенням ваги молочних корів помічається зростання їх удою, в той же час, жирномолочні корови, як правило, дають менше молока в порівнянні, з коровами, що володіють низьким відсотком жиру в молоці.

Кореляція називається лінійною, коли напрямок зв'язку між ознаками X і У графічно та аналітично виражається прямою лінією і навпаки.

У всіх випадках завдання кореляційного аналізу залишаються одні і ті ж: встановлення форми і напрямки зв'язку, що існує між варіюючими ознаками, вимір її сили або тісноти з подальшою оцінкою достовірності емпіричних показників зв'язку.

2. Коефіцієнт кореляції

Щоб виміряти ступінь спряженості між ознаками X і Y, необхідно зіставити відповідним чином їх значення один з одним. Коефіцієнт кореляції обчислюємо за формулою:

$$r = \frac{\sum(x_i - \bar{x})^2(y_i - \bar{y})}{n\sigma_x\sigma_y} \quad [52]$$

Коефіцієнт кореляції величина відносна і виражається в частках одиниці. Формула

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \quad [53]$$

звільняє дослідника від необхідності обчислювати середнє відхилення, що помітно полегшує розрахунок коефіцієнта кореляції.

3. Основні властивості коефіцієнта кореляції

Коефіцієнт кореляції служить для вимірювання сили або тісноти лінійного зв'язку між значеннями ознак X і Y. Коефіцієнт кореляції лежить в межах -1 <r <+1. При наявності позитивної зв'язку між варіюючими ознаками коефіцієнт кореляції має знак (+), при наявності зворотного або негативного зв'язку r має знак (-). Коли r = 0, це означає відсутність кореляції, при r = 1 очевидний функціональний зв'язок між ознаками. Таким чином, при r> 0 цей показник характеризує не тільки наявність, але і ступінь пов'язаності між значеннями варіюючих ознак. Чим сильніше спряженість, тим вище коефіцієнт кореляції і навпаки. А знак при r дозволяє визначати напрямок зв'язку.

Зазвичай вважається, що $r < 0,3$ вказує на слабкий зв'язок, при $0,3 < r < 0,5$ зв'язок визнається помірною. Якщо ж $0,5 < r < 0,7$ - кореляція вважається значною, при $0,7 < r < 0,9$ - сильною і при $r > 0,9$ дуже сильною, близькою до функціонального зв'язку. Зрозуміло, це чисто умовні підрозділи, а не загальноприйнятий стандарт при оцінці ступеня пов'язаності між варїруючими ознаками.

4. Довірча оцінка коефіцієнта кореляції

Так як вибірковий коефіцієнт кореляції є величиною випадковою, і може виявитися відмінним від 0 навіть при незначному варіюванні ознак. Звідси виникає необхідність розглядати коефіцієнт кореляції в якості оцінки генерального параметра (ρ). Нульова гіпотеза стосовно оцінки генерального ρ величиною емпіричного коефіцієнта кореляції (r) полягає в припущенні, що ρ = 0, тобто між випадковими величинами X і Y кореляція відсутня. Для перевірки нульової гіпотези служить t-критерій Стьюдента.

Помилка коефіцієнта кореляції обчислюється за формулою (для $n < 100$):

$$t = \frac{r\sqrt{n-2}}{1-r^2} \text{ [54]}$$

Нульова гіпотеза відхиляється, якщо $t_ф > t_{st}$ для $k = n-2$ і взятого рівня значущості (P). Це означає, що в генеральній сукупності $\rho \neq 0$ і, отже, вибірковий коефіцієнт кореляції достовірно відрізняється від 0, і між X і Y існує кореляційний зв'язок. При $t_ф < t_{st}$ нульова гіпотеза зберігається, відхилення вибіркового коефіцієнта кореляції від 0 вважається чисто випадковим.

Наприклад, на вибірці n = 36 отриманий r - 0,46. Потрібно оцінити достовірність цієї величини. критерій достовірності

$$t_{\phi} = \frac{0,46\sqrt{36-2}}{\sqrt{1-0,46^2}} = \frac{2,682}{0,888} = 3,0.$$

По таблиці Стьюдента для k = 36 - 2 = 34 і P = 0,01 знаходимо t_{st} = 2,58, $t_{ф} > t_{st}$. Нульова гіпотеза відхиляється.

5. Метод Z

На нечисленних вибірках оцінка коефіцієнта кореляції описаним способом може виявитися недостатньо точною. Обійти труднощі дозволяє запропонований Фішером метод "зет". Фішер запропонував замість коефіцієнта кореляції для його оцінки використовувати пов'язану з ним допоміжну величину Z:

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r} \quad \text{або} \quad Z = 1,15129\lg\frac{1+r}{1-r} \quad [55]$$

Перетворити коефіцієнт кореляції в показник "зет" можна за допомогою спеціальної таблиці, складеної Фішером. У цій таблиці вказані значення Z, що відповідають різним величинам коефіцієнта кореляції. Критерієм достовірності показника служить такий вираз:

$$t_Z \frac{Z}{\sigma_Z} = Z\sqrt{n-3} \quad [56]$$

Цей критерій придатний як для малих, так і для великих вибірок; він використовується у всіх випадках, коли замість коефіцієнта кореляції береться відповідне йому значення 2.

Для оцінки достовірності і встановлення довірчого інтервалу, по якому з достатньою ймовірністю можна судити про величину коефіцієнта кореляції генеральної сукупності, надходять у такий спосіб: за значенням емпіричного коефіцієнта кореляції в таблиці знаходимо значення Z. Потім визначаємо величину помилки показника 7 :

$$\sigma_Z = \frac{1}{\sqrt{n-3}} \quad [57]$$

За формулою [56] знаходимо емпіричне значення, яке порівнюємо зі стандартним по таблиці Стьюдента для прийнятого Р і до - п - 2.

Наприклад, на вибірці n = 28 отримуємо r = 0,52. По таблиці знаходимо Z = 0,576. Обчислюємо помилку:

$$t_z = Z\sqrt{n-3} = 0,576 \times \sqrt{25} = 2,88.$$

По таблиці Стьюдента для А = 28 - 2 = 26 і Р = 0,05 знаходимо t_{st} = 2,06. Оскільки $t_ф > t_{st}$ нульова гіпотеза не зберігається.

$$\sigma_z = \frac{1}{\sqrt{28-3}} = \frac{1}{5} = 0,20.$$

За величиною $\Delta Z = t\sigma_z$ = 1,96x0,20 = 0,392 знаходимо межу довірчого інтервалу для показника Z:

нижня межа = Z - ΔZ = 0,576 - 0,392 = 0,184;

верхня межа = Z + ΔZ = 0,576 + 0,392 = 0,968.

Користуючись таблицею, переводимо значення Z в коефіцієнт кореляції і знаходимо його довірчі межі:

нижня = 0,18; верхня = 0,74.

Це означає, що величина коефіцієнта кореляції в генеральній сукупності знаходиться між межами 0,18 <r <0,74. Можна сказати, що емпіричний коефіцієнт кореляції r = 0,52 визначений з достатньою точністю.

6. Мінімальна кількість спостережень для планованої точності коефіцієнта кореляції

Статистична недостатність емпіричного коефіцієнта кореляції ще не доводить, що зв'язку між варіюючими ознаками немає. При достатній кількості спостережень цей зв'язок може виявитися достовірно. Розрахувати необхідний обсяг вибірки для планованої точності коефіцієнта кореляції можна за формулою:

$$n = \frac{t^2}{Z^2} = 3 \quad [58],$$

де n - бажаєма чисельність спостережень, t - величина, задана за прийнятим порогом довірчої ймовірності (краще для P = 0,99).

Наприклад, для rr=0,25 и n=20, Z=0,2554, $t_z = 0{,}2554\sqrt{17} = 1{,}05$. P = 0,05, k = 20 - 2 = 18 , $t_{st} = 2{,}10$. Нульову гіпотезу відкинути не можна. Відомо, що довірчої ймовірності P = 0,95 відповідає нормоване відхилення t = 1,96, тому

$$n = \frac{1{,}96^2}{0{,}2554^2} + 3 = \frac{3{,}842}{0{,}065} + 3 = 59 + 3 = 62$$

Отже, щоб задовольнити поставленому завданню, необхідно провести не менше 62 спостережень.

7. Оцінка різниці між коефіцієнтами кореляції

Метод Z дозволяє оцінити достовірність різниці між емпіричними коефіцієнтами кореляції, обчисленими на незалежних вибіркових сукупностях. Помилка різниці Z - Z визначається за формулою:

$$m_{D_z} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \qquad [59];$$

критерій достовірності різниці за формулою:

$$t_D = \frac{Z_1 - Z_2}{m_{D_z}} \qquad [60].$$

Наприклад, вимірюючи залежність між довжиною класів (см) і кількістю зерен, що міститься у озимого жита в одному випадку на вибірці n = 50, r = +0,56, а в іншому - на вибірці n = 44, r = + 0,48. Різниця $r_1 - r_2 = 0{,}56 - 0{,}48 = 0{,}08$. З'ясуємо, чи випадкова ця розбіжність. По таблиці знаходимо значення $Z_1 = 0{,}633$, $Z_2 = 0{,}523$. критерій достовірності

$$t_D = \frac{0{,}633 - 0{,}523}{\sqrt{\frac{1}{50-3} + \frac{1}{44-3}}} = \frac{0{,}11}{\sqrt{0{,}0456}} = 0{,}52.$$

Так як для P=0,95 $t_{st} = 1{,}96$. На основі того, що $t_D = 0{,}52$, т. е. $t_D < t_{st}$, робимо висновок про випадковість різниці між коефіцієнтами кореляції.

8. Обчислення коефіцієнта кореляції на малих вибірках

На вибірках невеликого обсягу коефіцієнт кореляції обчислюють, не вдаючись до розподілу вибіркового матеріалу в варіаційні ряди і угруповання його в кореляційну таблицю. Для прикладу проаналізуємо дані про вагу 20 новонароджених павіанів-гамадрилів і вазі їхніх матерів, виміряних на початку вагітності (див. Табл. 27). Передбачається, що між вагою матерів і вагою їх приплоду існує прямолінійний позитивний зв'язок. Обчислимо для цих даних коефіцієнт кореляції. Для цього можна скористатися робочими формулами для малих вибірок:

$$r = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{\sqrt{n\Sigma x_i^2 - (\Sigma x_i)^2} \times \sqrt{n\Sigma y_i^2 - (\Sigma y_i)^2}} \quad [61];$$

$$r = \frac{D_x + D_y - D_d}{2\sqrt{D_x \times D_y}} \quad [62];$$

де $D_x = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$; $D_y = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}$; $D_d = \Sigma d_i^2 - \frac{(\Sigma d_i)^2}{n}$;

$d = x_i - y_i$; n - число парних спостережень

Щоб використовувати ту чи іншу формулу необхідно попередньо знайти допоміжні значення. Їх отримуємо, складаючи робочу таблицю (табл. 27).

Беремо з даної таблиці підсумкові цифри і знаходимо середнє арифметичне

$$\bar{x} = \frac{237,4}{20} = 11,87\ кг; \qquad \bar{y} = \frac{14,06}{20} = 0,703\ кг.$$

Визначаємо суми квадратів відхилень:

$$D_x = 2861,60 - \frac{237,4^2}{20} = 43,662; \qquad D_y = 9,9598 - \frac{14,06^2}{20} = 0,0756;$$

$$D_d = 2535,7218 - \frac{223,34^2}{20} = 41,6840.$$

Таблиця 27

Вага матері X	Вага приплоду Y	X x Y	X^2	$У^2$	x-y=d	d^2
10,0	0,70	7,000	100,00	0,4900	9.30	86.4900
10,8	0,73	7,884	116,64	0,5329	10.07	101,4049
11,3	0.75	8.475	127,69	0,5625	10,55	111,3025
10,0	0,70	7,000	100,00	0,4900	9,30	86,4900
10,1	0.65	6.565	102,01	0Д22Г1	9,45	89,3025
11,1	0.65	7,215	123,21	0.4225	10,45	109,2025
11,3	0.70	7.910	127,69	0,4900	10.60	112.3600
10.2	0,61	6,222	104,04	0,3721	9,59	91.9681
13.5	0.70	9,545	182,25	0,4900	12,80	163,8400
12,3	0,63	7/749	151.29	0,3969	11,67	136,1889
14,5	0,70	10,150	210.25	0.4900	13,80	190,4400
11,0	0.65	7,150	121.00	0,4225	10,35	107,1225
12,0	0,72	8,640	144,00	0,5184	11,28	127.2384
11,8	0,69	8.142	139,24	0,4761	11.11	123,432
13,4	0.78	10,452	179.56	0,6084	12.62	159,2644
11.4	0,70	7,980	129,96	0,4900	10.70	114,4900
12.0	0,60	7.200	244.00	0,3600	11.40	129.9600
15,6	085		1243,36	0/7225	14,75	1217,5625
13.0	0,80	10,400	169.00	0,6400	12,20	148,8400
12,1	0,85	9.075	146,41	0,5625	1L35	128,8225
∑ 237.4	14,06	167,939	2861,60	9,9598	223,34	2535,7218

Підставляємо знайдені значення в будь-яку з формул [61-62] і визначаємо коефіцієнт кореляції:

$$r = \frac{20 \times 167,939 - 237,4 \times 14,06}{\sqrt{20 \times 2861,60 - (237,4)^2} \times \sqrt{20 \times 9,9598 - (14,06)^2}} =$$

$$= \frac{20,936}{\sqrt{873,24} \times \sqrt{1,51}} = \frac{20,936}{36.346} \approx +0,58;$$

$$r = \frac{43.662 + 0,0756 - 41.6840}{2\sqrt{43.662 \times 0,0756}} = \frac{2,054}{2\sqrt{3.309}} \approx \frac{2,054}{3,616} = +0,58.$$

Отримана величина (r = 0,58) вказує на наявність значного позитивного зв'язку між вагою матері і вагою новонароджених дитинчат у павіанів-гамадрилів, Оцінимо достовірність отриманої величини. Для r = 0,58 знаходимо по таблиці значення Z = 0,663. Звідси критерій

$$t_z = Z\sqrt{n-3} = 0,663\sqrt{20-3} = 0,663 \times 4,12 = 2,73$$

По таблиці Стьюдента для P = 0,05 і Z = 20-2 = 18 і знаходимо t_{st} = 2,10. Оскільки t_Z> t_{st} нульова гіпотеза відкидається, величина r = 0,58 виявляється достовірною.

9. Кореляційне відношення

Для вимірювання криволінійної залежності між змінними величинами X і Y коефіцієнт кореляції непридатний. У таких випадках використовується інший показник - кореляційне відношення, що позначається грецькою буквою η (ета). На відміну від коефіцієнта кореляції, який характеризує залежність між X і К з точки зору прямої пропорційності, кореляційне відношення описує її двосторонньо. Розберемо на прикладі. Візьмемо кілька парних значень двох змінних величин V і Y:

Таблиця 28

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

Ранжируємо цю сукупність за X

Таблиця 28.1

X:	2	2	4	4	6	6	6	8
Y:	4	6	4	8	10	8	12	7
$\bar{Y}$:	5		6		10			7

З урахуванням повторюваності

Таблиця 28.2

X:	2	4	6	8
$\bar{Y}_x$:	5	6	10	7

де $\bar{Y}_x$ - приватні чи групові середні, відповідні однаковим значенням X. Якщо ж ранжувати сукупність по У, то вийде наступне:

Таблицяа 27.3

У:	4	4	6	7	8	8	10	12
X	2	4	2	8	6	4	6	6
X:	3		2	8	5		6	6

З урахуванням повторюваності

Таблица 28.4

У:	4	6	7	8	10	12
$\overline{X_y}$	3	2	8	5	6	6

Таким чином, з'ясовується, що залежність між змінними X і Y виражається по різному в залежності від того, за значеннями який з них ранжирується сукупність по X або У. Тому кореляційне відношення виражається не одним, а двома показниками $\eta_{y/x}$ и $\eta_{x/y}$. Вони обчислюються за такими формулами:

$$\eta_{x/y} = \sqrt{\frac{\sigma_{xy}^2}{\sigma_y^2}} \text{ и } \eta_{x/y} = \sqrt{\frac{\sigma_{xy}^2}{\sigma_x^2}} \text{ [63]}$$

де $\sigma_{yx}^2 = \frac{\sum \overline{y}_x - \overline{y}}{т}$ - середній квадрат відхилень приватних або групових середніх ($\overline{y_x}$) від загальної середньої ($\bar{y}$), тобто приватна дисперсія,

$\sigma_y^2 = \frac{\sum y_i - \bar{y})^2}{n}$ –загальна дисперсія сукупності.

Відповідно:

$$\sigma_{xy}^2 = \frac{\sum(\bar{x}_y - \bar{x})^2}{n} \text{ и } \sigma_x^2 = \frac{\sum x_i - \bar{x})^2}{n}$$

Як і коефіцієнт кореляції, кореляційне відношення - величина відносна, η має значення від 0 до 1: чим сильніше зв'язок, тим вище значення η. При відсутності кореляції η = 0. При цьому кореляційне відношення - величина завжди позитивна.

Показники кореляційного відношення зазвичай не рівні між собою, тобто . $\eta_{y/x} \neq \eta_{x/y}$.. Лише при строго лінійного зв'язку між X і У здійснюється рівність $\eta_{y/x} = \eta_{x/y}$.. Ця особливість кореляційного відношення дозволяє характеризувати будь-яку кореляційну залежність між варїруючими ознаками - і лінійну і криволінійну. Чим ближче зв'язок між ознаками наближається до прямолінійного функціонального зв'язку, тим ближче за абсолютною величиною показники кореляційного відношення один до одного.

На малих вибірках кореляційне відношення обчислюється за розглянутою вище формулою без угруповання в варіаційні ряди і в кореляційні таблиці.

Розглянемо обчислення кореляційного відношення на прикладі: з'ясовувалася наявність кореляційного зв'язку між вагою самок павіанів-гамадрилів (У) і тим віком (V), коли у них настає перший статевий цикл. Обчислимо для цих даних кореляційне відношення.

Таблиця 29

Вага (у)	Вік (х)	$\bar{x}_y$	$\bar{x}_y$-$\bar{x}$	$(\bar{x}_y-\bar{x})^2$	x_i-$\bar{x}$	$(x_i-\bar{x})^2$
5,0	24,0	26,25	4,55	20,70	6,8	46,24
6,2	26,0	26,25	4,55	20,70	2,3	5,29
5,4	26,0	32,00	1,20	1,44	1,2	1,44
5,0	28,5	28,50	2,30	5,29	4,8	23,04
6,7	28,5	28,50	2,30	5,29	0,2	0,04
8,0	29,0	32,50	1,70	2,89	0,7	0,49
5,7	31,0	32,50	1,70	2,89	2,7	7,29
5,4	31,0	32,50	1,70	2,89	1,7	2,89
6,1	31,0	31,00	0,20	0,04	0,2	0,04
5,5	31,5	32,00	1,20	1,44	1,2	1,44
5,8	32,0	33,50	2,70	7,29	2,7	7,29
5,3	32,0	31,00	0,20	0,04	0,2	0,04
6,8	32,5	26,00	4,80	23,01	4,8	23,04
5,6	32,5	34,00	3,20	10,24	3,2	10,24
6,4	32,5	32,50	1,70	2,89	1,7	2,89
7,5	33,0	28,50	2,30	5,29	2,3	5,29
5,5	33,5	33,25	2,45	6,00	3,2	10,24
6,0	33,5	33,25	2,45	6,00	1,7	2,89
6,3	34,0	33,00	2,20	4,84	2 2	4,84
6,8	34,0	29,00	1,80	3,24	1,8	3,24
Сума	-	-	-	132,44	-	158,20

$\bar{x}$=616/20=30,8 мес $\quad \bar{y} = \frac{121,0}{20} = 6,05$

визначаємо г/ ваги по віку

$$\eta_{y/x} = \sqrt{\frac{132,44}{158,20}} = \sqrt{0,84} = 0,92.$$

Таким же способом визначаємо кореляційне відношення віку за вагою г / (див. табл, 30), Підставивши в формулу [63] значення з таблиці отримаємо:

$$\eta_{y/x} = \sqrt{\frac{9,17}{12,57}} = \sqrt{0,73} = 0,85.$$

Таблиця 30

Вага (y)	Вік (x)	$\bar{y}_y$	$\bar{y}_y$-$\bar{y}$	$(\bar{y}_y-\bar{y})^2$	y_i-$\bar{y}$	$(y_i-\bar{y})^2$
5,0	24,0	5,00	1,05	1,1025	1,05	1,1025
6,2	26,0	5,80	0,25	0,0625	0,15	0,0225
5,4	26,0	5,80	0,25	0,0625	0,65	0,4225
5,0	28,5	5,85	0,20	0,0400	1,05	1,1025
6,7	28,5	5,85	0,20	0,0400	0,65	0,4225
8,0	29,0	8,00	1,95	3,8025	1,95	3,8025
5,7	31,0	5,73	0,32	0,1024	0,35	0,1225
5,4	31,0	5,73	0,32	0,1024	0,65	0,4225
6,1	31,0	5,73	0,32	0,1024	0,05	0,0025
5,5	31,5	5,50	0,55	0,3025	0,55	0,3025
5,8	32,0	5,55	0,50	0,2500	0,25	0,0625
5,3	32,0	5,55	0,50	0,2500	0,75	0,5625
6,8	32,5	6,27	0,22	0,0484	0,75	0,5625 ,
5,6	32,5	6,27	0,22	0,0484	0,45	0,2025
6,4	32,5	6,27	0,22	0,0484	0,65	0,4225
7,5	33,0	7,50	1,45	2,1025	1,45	2,1025
5,5	33,5	5,73	0,32	0,1024	0,55	0,3025
6,0	33,5	5,73	0,32	0,1024	0,05	0,0025
6,3	34,0	6,55	0,50	0,2500	0,25	0,0625
6,8	34,0	6,55	0,50	0,2500	0,75	0,5625
Сума	-	-		9,1722	-	12,5700

$\eta_{y/x}$=0,85, $\eta_{x/y}$=0,92, Ці показники говорять про досить сильну залежності, що існує між вагою тіла і тим віком павіанів-гамадрилів, в якому у самок наступає перший статевий цикл. Оцінимо достовірність отриманих величин за критерієм:

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

для прийнятого рівня значущості (P) і відповідного числа ступенів свободи (k = n = 2), Якщо $t_ф \geq t_{st}$ нульова гіпотеза відхиляється

$$t_{y/x} = 0,85\sqrt{\frac{20-2}{1-0,85^2}} = 0,85\sqrt{66,7} = 6,94;$$

$$t_{x/y} = 0,92\sqrt{\frac{20-2}{1-0,92^2}} = 0,92\sqrt{112,5} = 9,71.$$

По таблиці Стьюдента для $P = 0,01$ і $k = 18$, $t_{st} = 2,28$, $t_ф \geq t_{st}$ нульова гіпотеза відхиляється, отримані значення $\eta_{y/x}$, $\eta_{x/y}$ статистично достовірні.

Тема IX. Регресійний аналіз

1. Поняття регресії.

2, Вирівнювання емпіричних рядів регресії.

3, Лінійна регресія.

1. Поняття регресії

Коефіцієнт кореляції та кореляційне відношення дозволяють вимірювати ступінь спряженості між ознаками, визначати напрямок і форму існуючого між ними зв'язку. Але вони не дають уявлення про те, наскільки в середньому може змінитися варіююча ознака при зміні на одиницю виміру іншої, пов'язаної з ним ознаки. Функція, що дозволяє за величиною однієї ознаки (x) знаходити середні (очікувані) значення іншої ознаки ($\overline{(y_x)}$), пов'язаної з X кореляційно, називається регресією. Статистичний аналіз регресії отримав назву регресійного аналізу.

Показники регресії - величини іменовані: вони характеризують залежність між змінними X та У за їх абсолютними значеннями.

2. Вирівнювання емпіричних рядів регресії

Під вирівнюванням мається на увазі спосіб заміни ламаної лінії або ряду - регресії, динаміки, розподілу на плавно поточну, згладжену лінію, або звільнений від вагаючих значень чисельний ряд. Існують різні способи вирівнювання рядів.

а) Графічний спосіб. Даний спосіб не вимагає обчислювальної роботи. Після нанесення емпіричного ряду на графік, на око визначають середні точки лінії регресії, які потім з'єднуються. Недоліком способу є те, що він не виключає вплив індивідуальних властивостей дослідника на результат вирівнювання.

б) Спосіб ковзаючої середньої. Більш точний результат дає вирівнювання емпіричних рядів послідовним визначенням середніх арифметичних з двох або трьох сусідніх значень ряду.

Наприклад, є дані про вікові зміни ваги дитинчат гамадрилів.

Таблиця 31

Вік (,wee,):	0	1	2	1 J	4	5	6
Середня ваг (кг):	0,7	1,0	1,6	1,4	1,9	2,0	2,6

Знаходимо сумми та середні арифметичні

3.3/3=1,1

4.0/3 1,3

4.9/3=1 6

5.3/3=1,8

6,5/3=2,2

Отримуємо усереднені значення ряду: 1,1; 1,3; 1,6; 1,8; 2,2. Спосіб ковзної середньої простий і особливо зручний в тих випадках, коли емпіричний ряд представлений багатьом числом членів і втрата двох з них (крайніх) помітно не позначається на його загальній структурі. Цінність цього методу в тому, що він дозволяє себе модернізувати: усереднені величини можна отримувати з двох, трьох і більшого числа членів емпіричного ряду.

3. Лінійна регресія

Для математичного виразу зв'язку між змінними X і У служить рівняння загального вигляду Y = f(x), де символом f(x) позначається підбираєма форма рівняння, яка більш-менш повно виражає функціональну залежність середньої величини однієї змінної від значень іншої змінної величини X. Такого роду математичні рівняння називаються кореляційними або регресійний (Ф.Гальтон).

Залежність між біологічними ознаками може бути найрізноманітнішою. В більшості випадків емпіричні регресії виражаються простим рівнянням лінійної залежності:

$$\overline{y_x} = a = bx$$

Тут $\overline{y_x}$ групова середня арифметична, або очікуване значення змінної Y, відповідне заданому значенню змінної X; a й b - параметри рівняння; a - служить вільним членом, b - є показником пропорційності, який називають коефіцієнтом регресії.

Розрахунок емпіричного рівняння регресії проведемо за формулами:

$$b_{x/y} = \frac{\sum xy - n\bar{x}\bar{y}}{\sum y^2 - n\bar{y}^2} \quad \text{або} \quad b_{x/y} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \qquad [64];$$

$$b_{y/x} = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} \quad \text{або} \quad b_{y/x} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \qquad [65].$$

Для з'ясування залежності між окружністю грудей (У) і довжиною тіла (X) чоловіків попередньо розрахуємо допоміжні величини

Таблиця 32

Ріст X	Окруж-ність грудей У	XУ	X^2	Y^2	y_x	$\bar{Y}_x - Y$	$(\bar{Y}_x - Y)^2$
148,5	81,5	12102,75	22052Д5	6642,25	80,7	3,56	12,6736
150,5	82,0	1234 1,00	22650,25	6724,00	812	3,06	9,3636
152,5	80,7	12306,75	23256,25	6512,49	81,7	2,56	6,5536
154,5	81,0	12514,50	2387025	6561,00	822	2,06	4,2436
156,5	82,1	12848,65	24492Д5	6740,41	82,6	1,66	2,7556
158,5	83,8	13282,30	2512225	7022,44	83,3	0,96	0,9216
160,5	83,9	13465,95	2576025	7039,21	83,6	0,66	0,4356
162,5	83,9	13633,75	26406,25	7039,21	84,3	0,04	0,0016
164,5	85,0	13982,50	2706025	7225,00	84,8	0,54	0,2916
166,5	86,1	14335,65	2772275	7413,21	852	0,94	0,8836
168 J	86,7	14608,95	2839225	7516,89	85,7	1,44	2,0736
170 5	86,1	14680,05	2907025	7413,21	86,3	2,04	4,1616
172,5	86,4	14904,00	2975625	7464,96	87,0	2,74	7,5076
174,5	85,9	14989,55	3045025	7378,81	87,3	3,04	9,2416
176 5	88,8	15673,20	31152,25	7885,44	88,0	3,74	13,9876
2437,5	1263,9	205669,55	397213,75	106758,53	1263,9	29,04	75,0960

$n - 15,\ \bar{x} = 162{,}5;\ \bar{y}$=84,26

За підсумковими даними таблиці за формулами знаходимо:

$$b_{y/x} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{15 \times 205669{,}55 - 2437{,}5 \times 1263{,}9}{15 \times 397213{,}75 - (2437{,}5)^2} = \frac{4287}{16800} = 0{,}255;$$

$$a - \bar{y} - b\bar{x} - 84{,}26 - 0{,}255 \times 162{,}5 - 42{,}79;$$

$$\bar{y}_x - 0{,}255x + 42{,}793.$$

Додатково до графічного зображення регресії можна визначити міру лінійності:

$$\gamma = \eta^2 - r^2 \quad [66].$$

Коефіцієнт кореляції визначаємо за формулою:

$$r - \frac{\sum xy - \frac{\sum x \times \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \times \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}} \quad [67].$$

$$\sum(y_i - \bar{y})^2 - \sum y^2 - \frac{(\sum y)^2}{n} = 106578{,}53 - \frac{1263{,}9^2}{15} - 82{,}0;$$

$$\sum(x_i - \bar{x})^2 - \sum x^2 - \frac{(\sum x)^2}{n} - 397213{,}75 - \frac{2437{,}5^2}{15} = 1120{,}0;$$

$$r - \frac{205669{,}55 - \frac{1}{15}(2437{,}5 \times 1263{,}9)}{\sqrt{1120 \times 82}} - 0{,}943;$$

$$r^2 - 0{,}889;\ \eta^2_{y/x} = \frac{\sum(\bar{y}_x - \bar{y})^2}{\sum(y_i - \bar{y})} = \frac{75{,}0960}{82{,}00} - 0{,}914.$$

Визначимо міру лінійності ($\gamma=\eta^2-r^2$):

γ =0,914-0,889 = 0,025. Вибіркова помилка дорівнює:

$$m_\gamma = \frac{\sqrt[2]{0{,}025 - 0{,}025^2(2 - 0{,}914 - 0{,}889)}}{\sqrt{15}} = \frac{\sqrt[2]{0{,}025 - 0{,}025^2 \times 0{,}197}}{\sqrt{15}} = 0{,}087$$

Критерій $t_\gamma = \frac{\gamma}{m_\gamma}$ [69]. При $t_\gamma < 3$ (у відповідних випадках $t_\gamma < 2{,}5$) кореляція між ознаками оцінюється практично прямолінійною.

$t_{\gamma} = \frac{0,025}{0,087} = 0,28$, т.е $t_{\gamma} < 1$.

Отже, і графічно та аналітично підтверджується початкове припущення про лінійність регресії окружності грудей у чоловіків по їх зростанню.

Тема X. Дисперсійний аналіз

1. Суть методу і його основні завдання.
2. Основні поняття і терміни.
3. Дисперсійний аналіз однофакторних комплексів малих груп.

1. Суть методу і його основні завдання

Дисперсійний аналіз призначений для з'ясування причинно-наслідкових відносин. Р.Л. Фішер запропонував оцінювати результати дослідів не по середнім арифметичним, а шляхом порівняння вибіркових дисперсією, вірніше їх відносин з критичним значенням критерію F. Звідси і сам метод, розроблений Р.А Фішером (1925) отримав назву дисперсійного аналізу. В випадках комплексної оцінки результатів спостережень цей метод виявився більш "економним" і досить ефективним в порівнянні з іншими біометричними методами.

Дисперсійний аналіз проводиться як на малих, так і на великих вибірках, на однорідному і біологічно неоднорідному матеріалі, коли в одному і тому ж комплексі об'єднуються результати спостережень, проведених на особинах різної статі, віку, видової або расової приналежності і т.д. Цей метод є одним з найбільш потужних і ефективних методів біометричного аналізу; він дозволяє вирішувати найрізноманітніші завдання, які постають перед дослідниками.

2. Основні поняття і терміни

Ознаки, що змінюються під впливом тих чи інших причин, називаються результативними. А причини, що діють на результативні ознаки, прийнято називати факторами. Наприклад, зріст, вага, фізичний стан організму або цілої популяції - це ознаки. А такі засоби впливу, як дози лікарських або токсичних речовин, дози внесених в грунт добрив, раціони годівлі тварин або норми харчування людей і т.п. відносяться до категорії факторів. Зазвичай кожен із чинників представлений деякою кількістю груп, званих градаціями.

Вибіркова сукупність, організована певним чином для вивчення ефективності дії організованих факторів на результативну ознаку, називається статистичним, або дисперсійним комплексом. Структура такого комплексу визначається числом градацій, на які поділяються організовані фактори і враховується в досвіді ознака. Форма комплексу дається таблицею, в якій число стовпців дорівнює числу градацій одного або декількох організованих факторів, а по рядках відкладаються градації результативної ознаки, залежно від числа чинників, що враховуються, розрізняють одно- дво- і багатофакторні дисперсійні комплекси.

Оцінка достовірності впливу організованого фактора на результативну ознаку проводиться за критерієм Фішера (F), який потім оцінюється по таблиці Фішера для відповідних ступенів свободи к / і до, і прийнятого рівня значущості P ~ 0,05 або P = 0,01, Нульова гіпотеза, тобто, припущення про відсутність впливу організованого фактора на результативну ознаку, відкидається за умови, якщо $F > F_r$ якщо ж $F_ф < Fs_l$ спостерігаються між груповими середніми розбіжності визнаютьсятстатистично недостовірними.

3. Дисперсійний аналіз однофакторних комплексів малих груп

Дисперсійний аналіз однофакторних комплексів малих груп проводиться за такою приблизною схемою:

1. Знаходять середні величини, х всього комплексу і приватні чи групові середні х, по градаціях фактора А.

2. Визначають загальну суму квадратів відхилень (D):

$$D_y = \sum(x_i - \bar{x})^2 \quad [70].$$

3. Обчислюють міжгрупову суму квадратів (D):

$$D \ n\sum(\bar{x}_i - \bar{x})^2 \quad [71].$$

при різних числах варіант в градаціях фактора:

$$D_x = \sum[n_i(\bar{x}_i - \bar{x})^2] \quad [71a].$$

4. Знаходять внутрішньогрупову суму квадратів (D_):

$$D_z = \sum\sum(x_i - \bar{x}_i)^2 \quad [72].$$

Розрахунки спрощуються якщо використовувати такі робочі формули:

$$D_y = \Sigma x^2 - \frac{(\Sigma x)^2}{N} \quad [73]; \qquad D_x = \Sigma\frac{(\Sigma x_i)^2}{n_A} - \frac{(\Sigma x)^2}{N} \quad [74];$$

$$D_z = \Sigma x^2 - \Sigma\frac{(\Sigma x_i)^2}{n_i} \quad [75],$$

де x - варіанти, що входять до складу дисперсійного комплексу;

x_i - варіанти, що входять до складу градацій;

$\bar{x}$ - загальна середня арифметична;

$\bar{x}_A$ и $\bar{x}_i$ - групові або приватні середні арифметичні;

N - загальне число варіант;

n_A і n_i - числа варіант по градаціях і в групах комплексу, Так як D_v = Dt + D, то D = D - D_x,

5. Визначивши суми квадратів відхилень, встановлюють числа ступенів свободи (k):

для загальної дисперсії K_y = N-1,

для груповий дисперсії Kx = a-1,

для внутрішньогрупової дисперсії Kz = N-a,

де а - число градацій фактора А.

6. Обчислюємо загальну дисперсію:

$$\sigma_Y^2 = \frac{D_y}{N-1} = \frac{D_y}{K_y} \quad [76];$$

міжгрупову:

$$\sigma_X^2 = \frac{D_x}{a-1} = \frac{D_x}{K_x} \quad [77];$$

внутрішньогрупову:

$$\sigma_z^2 = \frac{D_z}{N-a} = \frac{D_z}{K_z} \quad [78].$$

7. Оцінимо достовірність:

$$F_{ф} = \frac{\sigma_y^2}{\sigma_z^2} \quad [79].$$

Розглянемо застосування даного алгоритму на наступному прикладі: на одній з дослідних станцій відчувався вплив доз мінеральних добрив на врожайність озимого жита. Результати випробувань виявилися наступні:

Таблиця 33

Дози добрив (кг/га)	Врожай (и/га) по повторюваностям						n_i	Сума	Середній врожай ($\bar{x}_i$)
	1	2	3	4	5	6			
15	8,0	8,4	9,0	8,6			4	34,0	8,5
20	8,2	9,0	10,0	10,2	9,2	10,0	6	56,4	9,4
25	11,0	13,0		12,0			3	36,0	12,0
30	7,5	8,5					2	16,0	8,0
Сума	-	-	-	-	-	-	15	142,4	9,5

$$\bar{x} = 9{,}5\frac{\text{ц}}{\text{га}}, \sum n = N = 15, \qquad \sum x = 142{,}4$$

При групировці даних в таблицю розрахунки полегшуються:

Таблиця 34

Показники	Градації фактора А (дози добрив)				Сума
	1 (15)	2(20)	3(25)	4(30)	
Врожай по повторюваностям (X)	8,0; 8,4 9,0; 8,6	8,2; 9,0 10,0;10,0 9,2; 10,0	11,0 13,0 12,0	7,5 8,5	a= 4
n_A	4	6	3	2	$\sum n_A$ = N=15
$\sum x_i$	34,0	56,4	36,0	16,0	$\sum\sum x_i = \sum x$= 142,4
$(\sum x_i)^2$	1156,00	3180,96	1296,00	256,00	-
$(\sum x_i)^2/n_A$	289,00	530,16	432,00	128,00	$\sum(\sum x_i)^2/n_A$=1379,16
$\sum x_i^2$	289,52	532,88	434,00	128,50	$\sum x^2$ =13804,9

$$D_y = \sum x^2 - \sum\frac{(\sum x)^2}{n} = 1384{,}9 - \frac{142{,}4^2}{15} = 1384{,}9 - 1351{,}85 = 33{,}05$$

$$D_y = \sum\frac{(\sum x_i)^2}{n_A} - \frac{(\sum x)^2}{N} = 1379{,}16 - 1351{,}85 = 27{,}31$$

$$D_Z = \sum x^2 - \sum \frac{(\sum x_i)^2}{n_A} = 1384{,}9 - 1379{,}16 = 5{,}74$$

$$K_y = 15 - 1 = 14, K_x = 4 - 1 = 3, K_Z = 14 - 3 = 11,$$

$$\sigma_x^2 = \frac{D_x}{K_x} = \frac{27{,}31 = 9{,}1}{3}; \quad \sigma_Z^2 = \frac{D_Z}{K_Z} = \frac{5{,}74}{11} = 0{,}52; \quad F_ф = \frac{\sigma_х^2}{\sigma_Z^2} = \frac{9{,}1}{0{,}52} = 17{,}7$$

Визначаємо по таблиці Фішера значення F_{st}: $k_x = 3$ (по горизонталі), $k_Z = 11$ (по вертикалі), $F_{st} = 6{,}2$, $F_ф > F_{st}$ нульова гіпотеза відкидається, статистично достовірно, що різні дози мінеральних добрив з різною силою впливають на урожай озимого жита.

ДОДАТКИ

Таблиця І. Критичні точки t-критерію Стьюдента при різних рівнях значущості α

Число ступенів свободи *k*	*α, %*			Число ступенів свободи *k*	*α, %*		
	5	1	0,1		5	1	0,1
1	12,71	63,66	64,60	18	2,10	2,88	3,92
2	4,30	9,92	31,60	19	2,09	2,86	3,88
3	3,18	5,84	12,92	20	2,09	2,85	3,85
4	2,78	4,60	8,61	21	2,08	2,83	3,82
5	2,57	4,03	6,87	22	2,07	2,82	3,79
6	2,45	3,71	5,96	23	2,07	2,81	3,77
7	2,37	3,50	5,41	24	2,06	2,80	3,75
8	2,31	3,36	5,04	25	2,06	2,79	3,73
9	2,26	3,25	4,78	26	2,06	2,78	3,71
10	2,23	3,17	4,59	27	2,05	2,77	3,69
11	2,20	3,11	4,44	28	2,05	2,76	3,67
12	2,18	3,05	4,32	29	2,05	2,76	3,66
13	2,16	3,01	4,22	30	2,04	2,75	3,65
14	2,14	2,98	4,14	40	2,02	2,70	3,55
15	2,13	2,95	4,07	60	2,00	2,66	3,46
16	2,12	2,92	4,02	120	1,98	2,62	3,37
17	2,11	2,90	3,97	∞	1,96	2,58	3,29
P	0,05	0,01	0,001	-	0,05	0,01	0,001

Таблиця II. Критичні значення величини нормованого відхилення при оцінці сумнівних варіант з урахуванням обсягу вибірки *n* і рівнів значущості α

n	α,%		*n*	α,%		*n*	α,%	
	5	1		5	1		5	1
4	1.71	1,73	13	2,56	2,81	23	2,84	3.16
5	1.92	1.97	14	2,60	2,86	24	2,86	3,18
6	2.07	2.16	15	2,64	2.90	25	2.88	3.20
7	2,18	2.31	16	2,67	2 95	26	2,90	3,22
8	2.27	2.43	17	2,70	2.98	27	2.91	3,24
9	2.35	2,53	18	2,73	3,02	28	2,93	3,26
10	2.41	2,62	19	2.75	3.05	29	2.94	3,28
11	2.47	2,69	20	2.78	3,08	30	2.96	3.29
12	2,52	2,75	21	2,80	3.11			
P	0.05	0,01	–	0,05	0.01		0,05	0.01

Таблиця III. Значення F-критерію Фішера при рівнях значимості α = 5% (верхній рядок) і α = 1% (нижня рядок)

k_2	k_1 --- ступінь свободи для великої дисперсії							
	1	2	3	4	5	6	7	8
1	161 4052	200 4999	216 5403	225 5625	230 5764	234 5859	237 5928	239 5982
2	18.51 98,50	19,00 99,00	19,16 99.17	19,25 99,25	19,30 99,30	19,33 99,33	19.35 99.36	19,37 99.37
3	10.13 34,12	9.55 30,82	9,28 29.16	9,12 28,71	9.01 28,42	8.94 27.91	8,89 27.67	8,85 27,49
4	7.71 21,20	6.94 18.00	6,59 16,89	6.39 15.98	6.26 15,52	6,16 15.21	6,04 14.98	6.00 14,80
5	6.61 16,26	5,79 13,27	5,41 12,06	5.19 11.39	5.05 10,97	4.95 10,67	4.88 10.46	4.82 10,29
6	5.99 13,75	5.14 10,92	4,76 9,78	4.53 9.15	4,39 8,75	4,28 8.47	4.21 8,26	4.15 8,10
7	5.59 12.25	4.74 9.55	4,35 8.47	4.12 7.85	3,97 7,46	3,87 7,19	3.79 6,99	3.73 6.84
8	5.32 11,26	4.46 8.65	4,07 7.59	3,84 7,01	3,69 6.63	3,58 6,37	3.50 6,18	3.44 6,03
9	5.12 10,56	4,26 8.02	3,86 6,99	3,63 6,42	3,48 6,06	3.37 5.80	3,29 5.61	3,23 5,47
10	4,96 10.04	4.10 7,56	3.71 6.55	3,48 5,99	3.33 5,64	3,22 5.39	3.14 5,20	3,07 5,06
11	4.84 9.65	3.98 7.21	3,59 6,22	3.36 5,67	3,20 5,32	3,09 5.07	3.01 4,89	2,95 4,74
12	4.75 9.33	3.80 6,93	3.49 5,95	3,26 5,41	3,11 5.06	3,00 4.82	2,91 4,64	2,85 4.50
13	4,67 9.07	3.80 6,70	3,41 5.74	3.18 5,21	3.02 4.86	2,92 4,62	2.83 4.44	2,77 4,30

Продовження табл. III

k_2	k_1 --- ступінь свободи для великої дисперсії						
	9	10	12	15	20	30	∞
1	241 6022	242 6056	244 6106	246 6157	248 6209	250 6261	254 6366
2	19,38 99,39	19,40 99,40	19,41 99,42	19,43 99,43	19,45 99,45	19,46 99,47	19,50 99,50
3	8,81 27,35	8,79 27,23	8,74 27,05	8,70 26,87	8,66 26,69	8,62 26,50	8,53 26,13
4	5,94 14,66	5,94 14,55	5,91 14,37	5,86 14,20	5,80 14,02	5,75 13,84	5,63 13,46
5	4,77 10,16	4,74 10,05	4,68 9,89	4,62 9,72	4,56 9,55	4,50 9,38	4,36 9,02
6	4,10 7,98	4,06 7,87	4,00 7,72	3,94 7,56	3,87 7,40	3,81 7,23	3,67 6,88
7	3,68 6,72	3,64 6,62	3,57 6,47	3,51 6,31	3,44 6,16	3,38 5,99	3,23 5,65
8	3,39 5,91	3,35 5,81	3,28 5,67	3,22 5,52	3,15 5,36	3,08 5,20	2,93 4,86
9	3,18 5,35	3,14 5,26	3,07 5,11	3,01 4,96	2,94 4,81	2,86 4,65	2,71 4,31
10	3,02 4,94	2,98 4,85	2,91 4,71	2,85 4,56	2,77 4,41	2,70 4,25	2,54 3,91
11	2,90 4,63	2,85 4,54	2,79 4,40	2,72 4,25	2,65 4,10	2,57 3,94	2,40 3,60
12	2,80 4,39	2,75 4,30	2,69 4,16	2,62 4,01	2,54 3,86	2,47 3,70	2,30 3,36
13	2,71 4,19	2,67 4,10	2,60 3,96	2,53 3,82	2,46 3,66	2,38 3,51	2,21 3,16

Продовження табл. III

k_2	k_1 --- ступінь свободи для великої дисперсії							
	1	2	3	4	5	6	7	8
14	4,60 8,86	3.74 6,51	3.34 5,56	3,11 5,04	2,96 4,69	2,85 4,46	2.76 4,28	2,70 4.14
15	4,54 8,68	3.68 6,36	3,29 5,42	3.06 4,89	2.90 4,56	2,79 4,32	2.71 4,14	2.64 4,00
16	4,49 8,53	3,63 6,23	3.24 5,29	3,01 4,77	2,85 4,44	2,74 4,20	2,66 4.03	2.59 3,89
17	4,45 8,40	3.59 6,11	3.20 5,18	2,96 4.67	2,81 4,34	2,70 3,93	2,61 3.79	2.55 3,68
18	4,41 8,29	3.55 6.01	3,16 5.09	2.93 4,58	2,77 4,25	2,66 4,01	2.58 3 84	2,51 3.71
19	4,38 8,18	3.52 5,93	3.13 5,01	2.90 4.50	2.74 4,17	2,63 3,94	2.54 3.77	2.48 3.61
20	4,35 8,10	3,49 5,85	3.10 4 94	2,87 4,43	2,71 4,10	2,60 3,87	2.51 3,70	2.45 3.56
21	4.32 8,02	3,47 5.78	3,07 4,87	2,84 4.37	2,68 4,04	2,57 3.81	2,49 3.64	2,42 3,51
22	4,30 7,95	3,44 5,72	3.05 4,82	2,82 4,31	2.66 3,99	2.55 3,76	2.46 3.59	2,40 3,45
23	4,28 7,88	3,42 5,66	3,03 4,76	2,80 4,26	2,64 3,94	2,53 3,71	2.44 3,54	2.37 3,41
24	4.26 7.82	3,40 5,61	3,01 4.72	2,78 4,22	2.62 3,90	2,51 3,67	2.42 3.50	2,36 3,36
26	4,24 7.77	3,39 5,57	2,99 4,68	2,76 4,18	2,60 3,85	2,49 3,63	2,40 3,46	2,34 3,32
26	4,22 7,72	3,37 5,53	2,98 4,64	2,74 4,14	2.59 3,82	2,47 3,59	2,39 3.42	2,32 3,29
27	4,21 7,68	3,35 5,49	2,96 4,60	2,73 4,11	2,57 3,78	2,46 3,56	2,37 3.39	2,31 3.26
28	4,20 7,64	3,34 5,45	2,95 4,57	2,71 4,07	2,56 3.75	2,45 3,53	2,36 3,36	2.29 3.23

Продовження табл. III

k_2	k_1 --- ступінь свободи для великої дисперсії						
	9	10	12	15	20	30	∞
14	2,65 4,03	2,60 3,94	2,53 3,80	2,46 3,66	2,39 3,51	2,31 3,35	2,13 3,00
15	2,59 3,89	2,54 3,80	2,48 3,67	2,40 3,52	2,33 3,37	2,25 3,21	2,07 2,87
16	2,54 3,78	2,49 3,69	2,42 3,55	2,35 3,41	2,28 3,26	2,19 3,10	2,01 2,75
17	2,49 3,68	2,45 3,59	2,38 3,46	2,31 3,31	2,23 3,16	2,15 3,00	1,96 2,65
18	2,42 3,60	2,38 3,51	2,31 3,37	2,23 3,23	2,16 3,08	2,07 2,92	1,88 2,57
18	2,42 3,52	2,38 3,43	2,31 3,30	2,23 3,15	2,16 3,00	2,07 2,84	1,88 2,49
20	2,39 3,46	2,35 3,37	2,28 3,23	2,20 3,09	2,12 2,94	2,04 2,78	1,84 2,42
21	2,37 3,40	2,32 3,31	2,25 3,17	2,18 3,03	2,10 2,88	2,01 2,72	1,81 2,36
22	2,34 3,35	2,30 3,26	2,23 3,12	2,15 2,98	2,07 2,83	1,98 2,67	1,78 2,31
23	2,32 3,30	2,27 3,21	2,20 3,07	2,13 2,93	2,05 2,78	1,96 2,62	1,76 2,26
24	2,30 3,26	2,25 3,17	2,18 3,03	2,11 2,89	2,03 2,74	1,94 2,58	1,73 2,21
26	2,28 3,22	2,24 3,13	2,16 2,99	2,09 2,85	2,01 2,70	1,92 2,54	1,71 2,17
26	2,27 3,18	2,22 3,09	2,15 2,96	2,07 2,81	1,99 2,66	1,90 2,50	1,69 2,13
27	2,25 3,15	2,20 3,06	2,13 2,93	2,06 2,78	1,97 2,63	1,88 2,47	1,67 2,10
28	2,24 3,12	2,19 3,03	2,12 2,90	2,04 2,75	1,96 2,60	1,87 2,44	1,65 2,06

Продовження табл. III

k_2	k_1 --- ступінь свободи для великої дисперсії							
	1	2	3	4	5	6	7	8
29	4,18 7,60	3,33 5,42	2,93 4,54	2,70 4,04	2,55 3,73	2,43 3,50	2,35 3,33	2,28 3,20
30	4,17 7,56	3,32 5,39	2,92 4,51	2,69 4,02	2,53 3,70	2,42 3,47	2,33 3,30	2,27 3,17
32	4,15 7,50	3,29 5,34	2,90 4,46	2,67 3,97	2,51 3,65	2,40 3,43	2,31 3,25	2,24 3,13
34	4,13 7,44	3,28 5,29	2,88 4,42	2,65 3,93	2,49 3,61	2,38 3,39	2,29 3,22	2,23 3,09
36	4,11 7,40	3,26 5,25	2,87 4,38	2,63 3,89	2,48 3,57	2,36 3,35	2,28 3,18	2,21 3,05
38	4,10 7,35	3,24 5,21	2,85 4,34	2,62 3,86	2,46 3,54	2,35 3,32	2,26 3,15	2,19 3,02
40	4,08 7,31	3,23 5,18	2,84 4,31	2,61 3,85	2,45 3,51	2,34 3,29	2,25 3,12	2,18 2,99
42	4,07 7,28	3,22 5,15	2,83 4,29	2,59 3,80	2,44 3,49	2,32 3,27	2,24 3,10	2,17 2,97
44	4,06 7,25	3,21 5,12	2,82 4,26	2,58 3,78	2,43 3,47	2,31 3,24	2,23 3,08	2,16 2,95
46	4,05 7,22	3,20 5,10	2,81 4,24	2,57 3,76	2,42 3,44	2,30 3,22	2,22 3,06	2,15 2,93
48	4,04 7,20	3,19 5,08	2,80 4,22	2,57 3,74	2,41 3,43	2,30 3,20	2,21 3,04	2,14 2,91
50	4,03 7,17	3,18 5,06	2,79 4,20	2,56 3,72	2,40 3,41	2,29 3,19	2,20 3,02	2,13 2,89
60	4,00 7,08	3,15 4,98	2,76 4,13	2,53 3,63	2,37 3,34	2,25 3,12	2,17 2,95	2,10 2,82
70	3,98 7,01	3,13 4,92	2,74 4,08	2,50 3,60	2,35 3,29	2,23 3,07	2,14 2,91	2,07 2,78
80	3,96 6,96	3,11 4,88	2,72 4,04	2,49 3,56	2,33 3,26	2,21 3,04	2,13 2,87	2,06 2,74
100	3,94 6,90	3,09 4,82	2,70 3,98	2,46 3,51	2,31 3,21	2,19 2,99	2,10 2,82	2,03 2,69
150	3,90 6,81	3,06 4,75	2,66 3,92	2,43 3,45	2,27 3,14	2,16 2,92	2,07 2,76	2,00 2,63
200	3,89 6,76	3,04 4,71	2,65 3,88	2,42 3,41	2,26 3,11	2,14 2,89	2,06 2,73	1,98 2,60
∞	3,64 6,63	3,00 4,61	2,60 3,78	2,37 3,32	2,21 3,02	2,10 2,80	2,01 2,64	1,94 2,51

Продовження табл. III

k_2	k_1 --- ступінь свободи для великої дисперсії								
	9	10	11	12	14	16	20	30	∞
29	2,22 3,09	2,18 3,00	2,14 2,93	2,10 2,87	2,05 2,77	2,01 2,69	1,94 2,57	1,83 2,41	1,64 2,03
30	2,21 3,00	2,16 2,17	2,13 2,90	2,09 2,84	2,04 2,74	1,99 2,66	1,93 2,55	1,84 2,38	1,62 2,01
32	2,19 3,02	2,14 2,93	2,10 2,86	2,07 2,80	2,01 2,70	1,97 2,62	1,91 2,50	1,82 2,34	1,39 1,96
34	2,17 2,98	2,12 2,89	2,08 2,28	2,05 2,76	1,99 2,66	1,95 2,58	1,89 2,46	1,80 2,30	1,57 1,91
36	2,15 2,95	2,11 2,86	2,07 2,76	2,03 2,72	1,98 2,62	1,93 2,54	1,87 2,43	1,78 2,26	1,55 1,87
38	2,14 2,95	2,09 2,82	2,05 2,75	2,02 2,69	1,96 2,59	1,92 2,51	1,85 2,40	1,76 2,23	1,53 1,84
40	2,12 2,89	2,08 2,80	2,04 2,73	2,00 2,66	1,95 2,56	1,90 2,48	1,84 2,37	1,74 2,20	1,51 1,80
42	2,11 2,86	2,06 2,78	2,03 2,70	1,99 2,64	1,93 2,54	1,89 2,46	1,83 2,34	1,73 2,18	1,49 1,78
44	2,10 2,84	2,05 2,75	2,01 2,68	1,98 2,62	1,92 2,52	1,88 2,44	1,81 2,32	1,72 2,15	1,48 1,75
46	2,09 2,82	2,04 2,73	2,00 2,66	1,97 2,60	1,91 2,50	1,87 2,42	1,80 2,30	1,71 2,13	1,46 1,73
48	2,08 2,80	2,03 2,72	1,99 2,64	1,96 2,58	1,90 2,48	1,86 2,40	1,79 2,28	1,70 2,12	1,45 1,70
50	2,07 2,79	2,03 2,70	1,99 2,63	1,95 2,56	1,89 2,46	1,85 2,38	1,78 2,26	1,69 2,10	1,44 1,68
60	2,04 2,72	1,99 2,63	1,95 2,56	1,92 2,50	1,86 2,39	1,82 2,31	1,75 2,20	1,65 2,03	1,39 1,60
70	2,02 2,67	1,97 2,59	1,93 2,51	1,89 2,45	1,84 2,35	1,79 2,27	1,72 2,15	1,62 1,98	1,35 1,53
80	2,00 2,64	1,95 2,55	1,91 2,48	1,88 2,42	1,82 2,31	1,77 2,23	1,70 2,12	1,60 1,94	1,32 1,49
100	1,97 2,59	1,93 2,50	1,89 2,43	1,85 2,37	1,79 2,26	1,75 2,19	1,68 2,06	1,57 1,89	1,28 1,43
150	1,94 2,53	1,89 2,44	1,85 2,37	1,82 2,31	1,76 2,20	1,71 2,12	1,64 2,00	1,53 1,83	1,22 1,33
200	1,93 2,50	1,88 2,41	1,84 2,34	1,80 2,27	1,74 2,17	1,69 2,09	1,62 1,97	1,52 1,79	1,19 1,28
∞	1,88 2,41	1,83 2,32	1,79 2,25	1,76 2,18	1,69 2,08	1,64 2,00	1,57 1,88	1,46 1,70	1,00 1,00

Таблиця IV. χ^2-Розподіл. Критичні (процентні) точки для різних значень ймовірності P і числа ступенів свободи k

k	α, %					P, %				
	5	2,5	1	0,5	0.1	99,9	99,5	99,0	97,5	95,0
1	3,84	5,02	6,64	7,88	10,83	–	–	–	–	–
2	5,99	7,38	9,21	10,60	13,82	·	0,01	0,02	0,05	0,10
3	7,82	9.35	11,34	12,84	16,27	0.02	0,07	0.12	0.22	0.35
4	9,49	11.14	13.28	14,86	18,47	0.09	0,21	0.30	0.48	0.71
5	11,07	12,83	15,09	16,75	20,52	0.21	0,41	0.55	0,83	1,14
6	12,59	14.15	16,81	18,55	22,46	0.38	0,68	0,87	1,24	1,64
7	14,07	16,01	18,48	20,28	24,32	0.60	0,99	1,24	1,69	2,17
8	15,51	17,54	20,09	21,96	26,12	0.86	1,44	1,65	2.18	2,73
9	16,92	19,02	21,67	23,59	27,88	1.15	1,54	2,09	2,70	3,32
10	18,31	20 48	23,21	25,19	29,59	1.48	2,65	2.56	3,25	3,94
11	19,68	21.92	24,72	26,76	31.26	1.83	2,31	3,05	3,82	4.58
12	21,03	23.34	26,22	28,30	32,91	2,21	3.47	3,57	4.40	5,23
13	22.36	24,74	27,69	29,82	34,53	2,62	3,56	4,11	5.01	5,89
14	23,68	26,12	29,14	31,32	36.12	3,04	4,57	4,66	5.63	6,57
15	25.00	27.49	30,58	32,80	37,70	3,48	4.11	5,23	6,26	7,26
16	26,30	28,84	32,00	34,27	39,25	3.94	5,24	5,81	6,91	7,96
17	27.59	30,19	33,41	35,72	40,79	4.42 ,	5,80	6,41	7,56	8,67
18	28,87	31,53	34,80	37,16	42,31	4.91	6,56	7.02	8,23	9,39
19	30,14	32,85	36.19	38,18	43,82	5.41	6.45	7,63	8,91	10,12
20	31,41	34.17	37,57	40,00	45,32	5,92	7,43	8,27	9,59	10.85
21	32,67	35.48	38.93	41,40	46,80	6.45	8,43	8.90	10.28	11,59
22	33,92	36,78	40.29	42,80	48,27	6 98	8,35	9,54	10,98	12,34
23	35,17	38,08	41,64	44,18	49,73	7,53	9,06	10,20	11,69	13,09
24	36,42	39,36	42,98	45,56	51.18	8 09	9,69	10,86	12,40	13,85
25	37,65	40,65	44,31	46,93	52,62	8.65	10.02	11,52	13,12	14,61
26	38.88	41,92	45,64	48,29	54.05	9.22	11,06	12,20	13,84	15.38
27	40.11	43,19	46,96	49,64	55,48	9.80	11.81	12.88	14,57	16,15
28	41,34	44,46	48,28	50,99	56,89	10.39	12,46	13,56	15,31	16,93
29	42,56	45,72	49,59	52,34	58,30	10.99	13,12	14,25	16,05	17,71
30	43,77	46,98	50,89	53,67	59,70	11.59	13,79	14,95	16,79	18,49
31	44,93	48,23	52,19	55,00	61,10	12,20	14,46	15,66	17,54	19,28
32	46,19	49,48	53,49	56,33	62,49	12,81	15,13	16,36	18,29	20,07
33	47,40	50.72	54.78	57,65	63,87	13.43	15,82	17,07	19,05	20,88
34	48,60	51,97	56,06	58,96	65,25	14,06	16,50	17,79	19,81	21,66
35	49,80	53,20	57,34	60.28	66,62	14,69	17,19	18,51	20,57	22,46
36	51,00	54,44	58,62	61,58	67,98	15.32	17,89	19,23	21,34	23,27
37	52.19	55,67	59,89	62,88	69,35	15.94	18,59	19.96	22,11	24,08
38	53,38	56.90	61,18	64,18	70,70	16,61	19,29	20.69	22,88	24.88
39	54,57	58,12	62,43	65,48	72,06	17,26	20,00	21.43	23,65	25,70
40	55,76	59,34	63,69	66,77	73,40	17.92	20,71	22,16	24,43	26,51
41	56,94	60,56	64.95	68,05	74.74	18,58	21,42	22,91	25.22	27,33
42	58,12	61.78	66,21	69,34	76.08	19,24	22,14	23,65	26,00	28.14
43	59,30	62,99	67.46	70,62	77,42	19,91	22,86	24,40	26,78	28,97
44	60.48	64.20	68,71	71,89	78.75	20,58	23.58	25,15	27,58	29,79
45	61,66	65,41	69,96	73,17	80,08	21,25	24.31	25.90	28.37	30,61
46	62,83	66,62	71,20	74,44	81,40	21.93	25,04	26.66	29,16	31,44
47	64,00	67.82	72,44	75,70	82,72	22,61	25.78	27,42	29.96	32,27
48	65.17	69,02	73,68	76,97	84,04	23,30	26.51	28.18	30,76	33,10
49	66,34	70,22	74.92	78,23	85,35	23,98	27,25	28.94	31,56	33,93
50	67,51	71,42	76,15	79,49	86.66	24.67	27.99	29.71	32.36	34,76
51	68.67	72,62	77,39	80,75	87.97	25.37	28,74	30.48	33,16	35,60

Продовження табл. IV

k	α, %					P, %				
	5	2,5	1	0,5	0,1	99,9	99,5	99,0	97,5	95,0
52	69,83	73,81	78,62	82,00	89,27	26,06	29,48	31,25	33,97	36,44
53	70,99	75,00	79,84	83,25	90,57	26,76	30,23	32,02	34,78	37,28
54	72,15	76,19	81,07	84,50	91,87	27,47	30,98	32,79	35,59	38,12
55	73,31	77,38	82,29	85,75	93,17	28,17	31,74	33,57	36,40	38,96
56	74,47	78,57	83,51	86,99	94,46	28,88	32,49	34,35	37,21	39,80
57	75,62	79,75	84,73	88,24	95,75	29,59	33,25	35,13	38,03	40,65
58	76,78	80,94	85,95	89,48	97,04	30,30	34,01	35,91	38,84	41,49
59	77,93	82,12	87,17	90,72	98,32	31,02	34,77	36,70	39,66	42,34
60	79,08	83,30	88,38	91,95	99,61	31,74	35,54	37,48	40,48	43,19
61	80,23	84,48	89,59	93,19	100,89	32,46	36,30	38,27	41,30	44,04
62	81,38	85,65	90,80	94,42	102,17	33,18	37,07	39,06	42,13	44,89
63	82,53	86,83	92,01	95,65	103,44	33,91	37,84	39,86	42,95	45,74
64	83,68	88,00	93,22	96,88	104,72	34,63	38,61	40,65	43,78	46,60
65	84,82	89,18	94,42	98,11	105,99	35,36	39,38	41,44	44,60	47,45
66	85,97	90,35	95,63	99,33	107,26	36,09	40,16	42,24	45,43	48,30
67	87,11	91,52	96,83	100,6	108,53	36,83	40,94	43,04	46,26	49,16
68	88,25	92,69	98,03	101,8	109,79	37,56	41,71	43,84	47,09	50,02
69	89,39	93,86	99,23	103,0	111,06	38,30	42,49	44,64	47,92	50,88
70	90,53	95,02	100,4	104,2	112,32	39,04	43,28	45,44	48,76	51,74
71	91,67	96,19	101,6	105,4	113,58	39,78	44,06	46,25	49,59	52,60
72	92,81	97,35	102,8	106,6	114,84	40,52	44,84	47,05	50,43	53,46
73	93,94	98,52	104,0	107,8	116,09	41,26	45,63	47,86	51,26	54,32
74	95,08	99,68	105,2	109,1	117,35	42,01	46,42	48,67	52,10	55,19
75	96,22	100,8	106,4	110,3	118,60	42,76	47,21	49,48	52,94	56,05
76	97,35	102,00	107,58	111,50	119,85	43,51	48,00	50,29	53,78	56,92
77	98,48	103,16	108,77	112,70	121,10	44,26	48,79	51,10	54,62	57,79
78	99,62	104,32	109,96	113,91	122,35	45,01	49,58	51,91	55,47	58,65
79	100,75	105,47	111,14	115,12	123,59	45,76	50,38	52,72	56,31	59,52
80	101,88	106,63	112,33	116,32	124,84	46,52	51,17	53,54	57,15	60,39
81	103,01	107,78	113,51	117,52	126,08	47,28	51,97	54,36	58,00	61,26
82	104,14	108,94	114,70	118,73	127,32	48,04	52,77	55,17	58,84	62,13
83	105,27	110,09	115,88	119,93	128,56	48,80	53,57	55,99	59,69	63,00
84	106,40	111,24	117,06	121,13	129,80	49,56	54,37	56,81	60,54	63,88
85	107,52	112,39	118,24	122,32	131,04	50,32	55,17	57,63	61,39	64,75
86	108,65	113,54	119,41	123,52	139,28	51,08	55,97	58,46	62,24	65,62
87	109,77	114,69	120,59	124,72	133,51	51,85	56,78	59,28	63,09	66,50
88	110,90	115,84	121,77	125,91	134,74	52,62	57,58	60,10	63,94	67,37
89	112,02	117,00	122,94	127,11	135,98	53,39	58,39	60,93	64,79	68,25
90	113,14	118,14	124,12	128,30	137,21	54,16	59,20	61,75	65,65	69,13
91	114,27	119,28	125,29	129,49	138,44	54,93	60,00	62,58	66,50	70,00
92	115,39	120,43	126,46	130,68	139,67	55,70	60,82	63,41	67,36	70,88
93	116,51	121,57	127,63	131,87	140,89	56,47	61,62	64,24	68,21	71,76
94	117,63	122,72	128,80	133,06	142,12	57,25	62,44	65,07	69,07	72,64
95	118,75	123,86	129,97	134,25	143,34	58,02	63,25	65,90	69,92	73,52
96	119,87	125,00	131,14	135,43	144,57	58,80	64,06	66,73	70,78	74,40
97	120,99	126,14	132,31	136,62	145,79	59,58	64,88	67,56	71,64	75,28
98	122,11	127,29	133,48	137,80	147,01	60,36	65,69	68,40	72,50	76,16
99	123,22	128,42	134,64	138,99	148,23	61,14	66,51	69,23	73,36	77,05
100	124,34	129,56	135,81	140,17	149,45	61,92	67,33	70,06	74,22	77,93
P	0,05	0,025	0,01	0,005	0,001	0,999	0,995	0,990	0,975	0,950

Таблиця V. Значення z, що відповідають значенням вибіркового коефіцієнта кореляції r_{xy}.

r_{xy}	*Соті дози коефіцієнта кореляції*									
	0	1	2	3	4	5	6	7	8	9
0.0	0.000	0,010	0.020	0.030	0.040	0,050	0,060	0,070	0,080	0.090
0,1	0,100	0,110	0.121	0,131	0.141	0,151	0.161	0.172	0,182	0,192
0.2	0.203	0.213	0.224	0.234	0.245	0,255	0,266	0,277	0,288	0,299
0.3	0,310	0.321	0,332	0.343	0.354	0,365	0.377	0.388	0.400	0.412
0,4	0.424	0,436	0,448	0.460	0.472	0.485	0,497	0.510	0.523	0.536
0.5	0.549	0,563	0,576	0,590	0,604	0,618	0.633	0.648	0,663	0,678
0.6	0,693	0,709	0.725	0.741	0.758	0,775	0.793	0,811	0,829	0.848
0.7	0.867	0,887	0.908	0,929	0,951	0 973	0.996	1,020	1.045	1,071
0,8	1.099	1.127	1,157	1,188	1.221	-1,256	1.293	1.333	1.376	1.422
0.9	1,472	1.528	1,589	1,658	1,738	1.832	1.946	2.092	2.298	2,647
0,99	2.647	2.700	2.759	2,826	2,903	2,995	3.106	3,250	3.453	3,800

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Атраментова Л. О. Біометрія : підруч. для студ. вищ. навч. закладів / Л. О. Атраментова, О. М. Утєвська. – Харків : Ранок, 2007. – 176 с.
2. Глазов Н. М. Статистический метод в таксации и лесоустройстве / Н. М. Глазов. – М. : Лесн. пром-сть, 1976. – 143с.
3. Гланц С. Медико-биологическая статистика. –М.: Мир, 1999.–652 с.
4. Гмурман В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. – М. : Высшая школа, 1977. – 477 с.
5. Горкавий В. К. Статистика : підручник / В. К. Горкавий. – К. : Аграрна освіта, 2009. – 511 с.
6. Горошко М.П. Біометрія / М.П. Горошко, С.І. Миклуш, П.Г. Хомюк. – Львів, Камула, 2004. – 285 с.
7. Гумецький Р.Я., Паляниця Б.М., Чабан М.Є. Математичні методи в біології : Теоретичні відомості, програмований практикум, комп’ютерні тести / Навч. посібник.– Львів: ЛНУ, 2004. – 112 с.
8. Дика М.В., Тарновська М.М., Яремчук М.М., Генега А.Б., Санагурський Д.І. Біометрія: теоретичні відомості та лабораторний практикум / Навч. посібник. – Львів: ЛНУ, 2016. – 100 с.
9. Калінін М.І. Біометрія: Підручник для студентів вузів біологічних і екологічних напрямків./ Калінін М.І., Єлісєєв В.В. - Миколаїв: Вид-во МФ НаУКМА, 2000. - 204 с.
10. Лакин Г.Ф. Биометрия : Учеб. пособие для биол. специальностей вузов / 4-е изд.– М.: Высш. школа, 1980. – 293 с.
11. Чепур С.С. Біометрія: Методичний посібник. – Ужгород: Видавництво УжНУ «Говерла», 2015. – 40 с.
12. S. P. Otto, T. Day. A biologist's guide to mathematical modeling. Princeton University Press. 2007