МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ»

БІОМЕТРІЯ

Навчальний посібник

Ужгород – 2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ»

БІОМЕТРІЯ

Навчальний посібник

Ужгород – 2023

УДК 57.087.1(075.8)
Б 63

Біометрія: навчальний посібник / Укладач: С.С. Чепур. Ужгород: Вид-во УжНУ «Говерла», 2023. 196 с.
ISBN 978-617-7825-90-5

Посібник включає основи планування дослідження і методи статистичної обробки результатів спостереження.

Для студентів лісогосподарських, біологічних, екологічних факультетів вищих навчальних закладів. Може бути корисним для аспірантів, науковців, практиків, спеціалістів сільської, лісової та суміжних галузей.

Рецензенти:

Савіна О.І., доктор сільськогосподарських наук, професор кафедри плодоовочівництва і виноградарства ДВНЗ «УжНУ»;

Шишканинець І.Ф., кандидат сільськогосподарських наук, заст. директора з наукової роботи НПП "Зачарований край".

Рекомендовано до друку:

Вченою радою ДВНЗ «УжНУ»
(протокол №2 від 21 лютого 2023 року)

Редакційно-видавничою радою ДВНЗ «УжНУ»
(протокол №1 від 17 лютого 2023 року)

ISBN 978-617-7825-90-5

ЗМІСТ

ВСТУП

Біометрія представляє сукупність математичних методів, що використовуються у біології, має свою специфіку, свої відмінні риси і займає визначене місце в системі біологічних наук. Сучасна біометрія – це розділ біології, змістом якого є планування спостережень і статистична обробка їхніх результатів.

Без математично-статистичної обробки сьогодні немислиме проведення досліджень. До таких досліджень відносяться порівняння вибіркових груп по досліджуваних показниках і визначення достовірності результатів такого порівняння з заданою імовірністю безпомилкових прогнозів, визначення достатньої кількості досліджуваних об’єктів, з’ясування впливу різних факторів на біологічні явища та процеси, вивчення залежності між ознаками при вивченні спадковості і повторюваності господарсько важливих ознак, і, розробка алгоритмів для діагностування, прогнозування в біології, в лісовому і сільському господарствах, медицині за допомогою комп’ютерних програм.

Багато підручників по біометрії, які є загальнодоступними, видані в 60-80 роки минулого століття російською мовою. Матеріал є застарілим, потребує оновлення, доповнення. В той же час в галузі математичної статистики та біометрії є напрацювання українських дослідників останніх десятиліть, наприклад праці М.П. Горошко, С.І. Міклуша, П.Г. Хомюка, М.І. Калініна та ін.

В силу сказаного видання сучасного посібника із застосування біометрії в лісовому і сільському господарстві, в якому були б використані загально-методичні розробки в галузі математичної статистики Плохицького, Лакіна, Рокицького та ін., та доповнені новими напрацюваннями є актуальним. Є необхідність у посібниках по біометрії різного рівня складності, щоб у студента чи практика був вибір доступного йому матеріалу за рівнем підготовки.

Навчальний посібник розрахований на формування у студентів базового комплексу знань з методології проведення досліджень, чисельного опису, статистичного опрацювання результатів спостережень та математичного моделювання об'єктів і явищ, як предметів діяльності фахівців різних галузей, в т. ч. сільського і лісового господарства.

Посібник складається з 10 тем, відповідно до тематики лекційних занять, що визначаються навчальною програмою з біометрії.

Під час підготовки посібника враховано досвід проведення занять з біометрії, а також математико-статистичної обробки та аналізу вихідних матеріалів при проведенні автором наукових досліджень.

У посібнику увагу зосереджено на доступності викладу матеріалу для розуміння його студентами, підкріплений прикладами і алгоритмами для застосування біометричних методів у лісовому господарстві і суміжних галузях.

В доступній формі наводяться основні математично-статистичні методи, які використовуються в біологічних дослідженнях. Викладено закономірності розподілу випадкових величин, методику побудови варіаційних рядів, техніку обчислення та оцінку вибіркових показників, кореляційний, регресійний, дисперсійний аналіз та інші питання біометрії. Наведено приклади, користуючись якими навіть слабо підготовлений з математики фахівець зможе застосувати у роботі рекомендовані способи обробки матеріалів. Наведено матеріали, необхідні для обчислення біометричних показників.

Даний навчальний посібник буде корисним не тільки студентам, які навчаються за галузями знань «Аграрні науки і продовольство», «Природничі науки», «Біологія», а й аспірантам, викладачам, науковцям.

РОЗДІЛ 1
ПРЕДМЕТ І ЗНАЧЕННЯ БІОМЕТРІЇ

1.1. Специфіка біометрії, її місце в системі біологічних наук.
1.2. Значення біометрії в дослідницькій роботі та професійній підготовці спеціалістів сільського та лісового господарства.
1.3. Короткі історичні відомості виникнення і розвитку біометрії.
1.4. Лісова біометрія як частина загальної біометрії.

1.1. Специфіка біометрії, її місце в системі біологічних наук

Сучасна біометрія – це розділ біології, змістом якої є планування спостережень і статистична обробка їх результатів.

Якщо математична статистика і теорія ймовірностей – розділи математики, теоретичні, фундаментальні науки, які розглядають масові явища без врахування специфіки складових їх елементів, то біометрія – прикладна наука, що досліджує конкретні біологічні об'єкти із застосуванням математичних методів.

Предметом *біометрії* є будь-який біологічний об'єкт, при вивченні якого використовується розрахунок або вимірювання, тобто з кількісної сторони більш-менш точна оцінка його якісного стану. При цьому, досліджуються не одиничні, а групові випадки та масові явища, до яких можна використовувати статистичні закони. Наприклад, лісівник виміряв діаметр одного дерева – це поодинокий випадок. Якщо ж лісівник виміряв діаметри у кількох дерев, або ж вимірював діаметр одного дерева, але щорічно протягом певного періоду (5-10 років), – це масове явище, незалежно від того, яким був об'єкт спостереження – одиничним або груповим.

Біологія вивчає загальні і локальні закони і закономірності, які притаманні живій матерії в усій її різноманітності, формах існування і розвитку протягом усіх періодів існування планети. Це одна з найбільш давніх наук. З часом з біології деталізувались ботаніка і зоологія. Згодом ботаніка і зоологія подрібнюються на окремі розділи як спеціальні біологічні науки – ембріологія, гістологія, палеологія, мікробіологія і таке інше. Винахід мікроскопу спрямував розвиток біології декількома шляхами – поглибленням вивчення внутрішньої будови живої матерії; вивчення біохімічних процесів, що відбуваються в живій матерії (біохімія, фізіологія), вивчення організації живої речовини з визначенням структурних рівнів цієї організації.

Біометрія за призначенням і специфікою займає місце на стику біологічних і математичних наук. За своєю сутністю і очікуваними результатами досліджень біометрія відноситься до біологічних наук, за методологією – це самостійний розділ біології. Вона не займається питаннями одержання суто математичних рішень, обґрунтуванням математичних формул і рівнянь. Але вона використовує готові математичні висновки і застосовує їх для рішення біологічних задач. Наука, яка дала обґрунтування доцільності і можливості застосування математичних методів дослідження в області біології була названа біометрією. Біометрія відпрацювала методологію і концептуальну можливість математичного аналізу і визначила конкретні математичні аспекти щодо встановлення доцільності застосування методів математичного аналізу в дослідженнях тих або інших аспектів біологічної природи.

На особливості біологічних об'єктів в плані зміни їх морфології, фізіології, біохімії, генетики впливають екологічні фактори. Розвиток біології як науки неможливий без врахування характеру таких змін, виявлення зовнішніх причин, що їх обумовлюють, визначення ступеню і напрямку цих змін. Це можливо лише за умови відповідних досліджень біологічних об'єктів шляхом вимірювання певних морфологічних ознак або особливостей фізіологічних і біохімічних процесів.

Біометричні дослідження (біо – життя, метрія – вимірювання) – це дослідження, результатом яких є визначення відповідних характеристик біологічних об'єктів. Окремі частини (елементи) біологічних об'єктів, які можуть бути визначені шляхом вимірювання або співставлення, доцільно характеризувати як *біометричні елементи*.

З формальної точки зору біометрія представляє сукупність математичних методів, які застосовуються в біології і запозичені головним чином з області математичної статистики і теорії імовірностей. Найбільш тісно біометрія пов'язана з математичною статистикою, висновками якої вона переважно користується. Однак біометрія має свою специфіку, свої відмінні риси і займає певне місце в системі біологічних наук.

Біометрія виникла з потреб біології. Крім біометрії, між біологією і математикою склалися і інші напрямки математичної біології. Кожен напрямок має свої завдання і стосовно них використовує відповідні математичні методи. Загальним для всіх напрямків математичної біології є дедуктивний підхід до вирішення конкретних завдань, коли на перше місце висуваються математичні моделі з подальшою перевіркою їх досвідом. Біометрія же спирається переважно на індуктивний метод, відштовхується від конкретних даних, отриманих імперичним (набутим, отриманим в результаті досвіду) шляхом, які аналізує за допомогою математичних методів.

Перш ніж застосовувати математичні методи аналізу, біолог чи лісівник повинен з'ясувати який аспект, чи характеристику цього об'єкту він хоче одержати за допомогою математичного аналізу і чи є доцільним його застосування. Слід уникати математичного обробітку експериментальних даних у випадках, коли їх результати дають чітку відповідь на завдання дослідження.

Доцільно застосувати математичні методи для з'ясування таких аспектів:

1. Особливості розподілу маси однотипних біологічних об'єктів за будь-якою ознакою.

2. Математичне моделювання і прогнозування зміни будь-якої характеристики біологічного об'єкту при зміні якогось фактору, що впливає на цю характеристику.

3. Особливості і характер зв'язку між окремими ознаками біологічних об'єктів у відповідності до їх чисельності.

4. Виявлення характеру і ступеню впливу якогось фактору на зміну відповідної характеристики біологічного об'єкту шляхом співставлення середніх значень і визначення достовірності різниці.

5. Визначення достовірності одержаних результатів експериментальних досліджень.

6. Методика досліджень біологічних об'єктів для одержання достовірних даних при мінімальній кількості досліджень.

Наведені питання вирішуються в основному із застосуванням: кореляційного аналізу; регресивного аналізу; законів розподілу випадкових величин.

1.2. Значення біометрії в дослідницькій роботі та професійній підготовці спеціалістів сільського та лісового господарства

На даний час важко назвати галузь знань, де б не використовували математичні методи.

Біометричні методи використовують як при вивченні анатомії людини, так і при дослідженні взаємозв'язку між діаметром та висотою дерева в лісовому господарстві. Біометрія необхідна і при вивченні успадкування і повторюваності господарсько-важливих ознак, вимірюванні зв'язку між ними і в багатьох інших випадках. Застосування біометрії виявилося корисним у багатьох областях прикладної біології.

Так, завдяки біометричному аналізу масових антропологічних вимірювань антропологам вдалося підійти до досить точного обґрунтування принципів розкрою і стандартизації взуття та одягу, що виготовляється для масового використання. Біометричні показники лягли в основу кількісної оцінки фізичного розвитку людини, його спортивних і трудових досягнень.

Існують дослідження, які не мають відношення до біометрії. Це описові методи в біології. Але, якщо в дослідженнях присутні розрахунки чи заміри, то використання біометричних методів є вкрай необхідним. У таких випадках нехтування методами біометрії або неправильне їх застосування призводить до невиправданих витрат праці і часу, а головне – до мало переконливих, а нерідко і помилкових висновків.

Дослідники різних галузей (біологи, агрономи, лісівники) повсякденно змушені вирішувати питання про те, як реагує живий організм на зміну того чи іншого фактору зовнішнього середовища. Наприклад, зміна температури середовища впливає на швидкість росту організму, швидкість біохімічних процесів, що в ньому відбуваються, та інше. Для того, щоб вивчити це питання, необхідно поставити серію спостережень, тобто зафіксувати, що при даній температурі швидкість росту одна, при збільшенні температури вона прийняла таке-то значення, при подальшому збільшенні температури – таке-то. Опрацювання одержаних даних методами біометричного аналізу дозволить одержати математичну апроксимацію залежності організму від даного фактору оточуючого середовища. Знання природи живої матерії, форм і особливостей її існування, закономірностей розвитку одержуються шляхом проведення відповідних спостережень і порівняння їх результатів.

Разом з цим, в комплекс факторів зовнішнього середовища, під впливом яких формуються організми в онтогенезі, входить значна частина різних значень цих факторів. Наприклад, на ріст, формування, цвітіння, утворення плодів дерев і чагарників впливають хімічний склад ґрунту, умови зволоження, температура ґрунту, температура і вологість повітря, інтенсивність сонячного опромінювання і таке інше. Комплекс цих факторів обумовлює процес онтогенезу кожної рослини. Але кожен з факторів може в цей комплекс входити, маючи різне номінальне значення. Так, наявність поживних речовин (водорозчинних сполук фосфору, калію, азоту та інше) може бути в різноманітних дозах від мінімальної кількості, що ледве забезпечує життя рослини, до максимальної, яка викликає отруєння рослини, те ж саме з наявністю вологи і рівнем присутності інших факторів. Отже, кількість можливих співвідношень факторів та їх градацій в комплексі, під впливом якого відбувається онтогенез рослин, за відомим законом теорії множин може досягати десятків мільйонів. Кожний такий комплекс накладає певний відбиток на ту чи іншу морфологічну ознаку: інтенсивність сонячного опромінювання на довжину і ширину листка, вологість під час росту на масу рослини, а під час формування плоду – на його форму, масу, хімічний склад та інше.

Між різними значеннями будь-яких морфологічних ознак живих організмів існує відповідний взаємозв’язок. Наприклад, чим більший діаметр стовбура дерева, тим, як правило, більша його висота. Але часто-густо у дерев з однаковими діаметрами висоти можуть бути різними. Отже, для того, щоб визначити й описати характер взаємозв'язку між цими морфологічними елементами дерев, необхідно провести не одне, а певну кількість замірів, які після опрацювання їх методами біометричного аналізу дадуть достовірну відповідь на поставлене питання.

Для отримання обґрунтованих і достовірних відповідей на вищенаведені питання та на багато інших, які постійно постають перед дослідниками необхідні вивчення біологічних закономірностей із застосуванням біометричних методів.

Працівники лісового господарства постійно стикаються з роботою, де необхідно вимірювати і розраховувати. Наприклад, середній діаметр, середня висота, повнота, запас деревини і т.п. Тому знання використовувати біометричні методи в їх дослідженнях є необхідним.

1.3. Короткі історичні відомості виникнення і розвитку біометрії

Термін "біометрія" був введений Ф. Гальтоном у 1889 році для позначення кількісних методів, що застосовувались в біології.

Слово "біометрія" від лат. bios – життя і metron – міра. На даному періоді застосування цього терміну означає сукупність математично-статистичних методів, що застосовувались в біологічних дослідженнях. Біометрія має свою історію розвитку.

Біометрія як відносно самостійна наукова дисципліна склалася в другій половині 19 століття. Однак її витоки сягають більш раннього періоду в історії природознавства: на той час, коли вимірювання біологічних об’єктів стали розглядати як метод наукового пізнання.

В міру свого розвитку суспільство потребувало точних знань про природу. Так, в свій час стає актуальним афоризм Г. Галілея (1564-1642): «Виміряй все те, що вимірюється і зроби те, що не вимірюється тим, що вимірюється».

У 1614 р з'явилася книга Санторіо (1561-1636 рр.) «Про статичну медицину». У 1680 році вийшла в світ книга Бореллі

(1608-1679 рр.) «Про рух тварин». У 1768 році французький гіполог Буржеля видав свою працю «Екстер'єр коня». У цій книзі наведено набір вимірів, необхідних для визначення придатності коней до тієї чи іншої служби. У 1718 році в Лондоні вийшла у світ книга французького математика А. де Муавра (1667-1754 рр.) «Вчення про випадки». Вимірявши ріст у 1375 дорослих жінок і, розташувавши результати вимірювань в ряд, він виявив закономірність, що відповідає відомому в теорії імовірностей закону нормального розподілу. Виникла необхідність інтеграції методів біології з методами теорії імовірностей і математичної статистики.

Теорія ймовірностей і математична статистика виникли в середині XVII ст. незалежно одна від одної. Стимулом до обґрунтування теорії імовірностей послужив розвиток грошових відносин в суспільстві. Відому роль при цьому зіграли азартні ігри – метання монет, гральних кісток, ігри в карти, – які виявилися простими моделями, що дозволили помітити закономірність у поведінці випадкових подій масового характеру. Біля витоків теорії імовірностей стояли французькі вчені П. Ферма (1601-1665) і Б. Паскаль (1623-1662), а також голландський математик і натураліст Х. Гюйгенс (1629-1696). Вагомий внесок у становлення теорії імовірностей внесли Я. Бернуллі (1654-1705) і А. де Муавр. Однак найбільш істотний розвиток отримала ця теорія в працях таких видатних математиків, як П. Лаплас (1749-1827), К. Гаусс (1777-1855), С. Пуассон (1781-1840).

Розвиток математичної статистики пов'язаний з проблемами державознавства. В середині XVII століття в економічно розвинених країнах Європи накопичилася така кількість відомостей щодо демографічної та страхової ситуації, а також в галузі торгівлі, охорони здоров'я та інших галузях господарства, що розбиратися в них за допомогою способів описової статистики стало майже неможливим. Назріла гостра необхідність пошуку нових методів аналізу статистичних даних, їх теоретичного обґрунтування. Завдання зводилось до того, щоб по частині судити про стан цілого, тобто за вибіркою робити висновок про всю сукупність суспільних явищ в цілому, повний опис яких ставав справою дуже трудомісткою і дорогою. Розробка теорії вибіркового методу зближувала математичну статистику з висновками теорії імовірностей, що стало важливою віхою на шляху до виникнення біометрії.

Першим, хто вдало поєднав методи антропології і соціальної статистики з висновками теорії ймовірностей і математичної статистики, був бельгійський антрополог і статистик А. Кетле (1796-1874). З робіт Кетле випливало, що задача статистики полягає не в одному лише зборі та класифікації статистичних даних, а в їх аналізі, метою якого має бути відкриття закономірностей, що діють у сфері масових явищ. Знання цих закономірностей і мало перетворити статистику в джерело наукового пізнання соціальних і біологічних явищ.

А. Кетле заклав основи біометрії. Створення ж математичного апарату цієї науки належить англійській школі біометриків XIX ст., на чолі якої стояли Ф. Гальтон (1822-1911) і К. Пірсон (1857-1936). Ця школа виникла під впливом праць Ч. Дарвіна (1809-1882).

Гальтону належить перша спроба застосувати статистичні методи до розв'язання проблеми спадковості і мінливості організмів. Гідним продовжувачем досліджень Гальтона був його учень К. Пірсон, який зайнявся вивченням проблеми спадковості і мінливості організмів. Він створив математичний апарат біометрії; розвинув вчення про різні типи кривих розподілу, розробив метод моментів (1894) і критерій «хі-квадрат» (1990). Пірсон ввів в біометрію такі показники, як середнє відхилення (1894) і коефіцієнт варіації (1896). Йому належить удосконалення методів кореляції і регресії Гальтона (1896, 1898).

На помилку Гальтона і Пірсона, які вважали, що зовнішня подібність між родичами говорить про ступінь їх спорідненості вказав датський вчений В. Йогансен (1857-1927). У дослідах з квасолею Йогансен прийшов до важливого висновку про те, що біологічні проблеми повинні вирішуватися за допомогою математики, але не як математичні задачі. «Статистиці, – писав Йогансен, – завжди повинен передувати біологічний аналіз, інакше результати можуть бути «статистичною брехнею». Це був новий, реалістичний підхід до оцінки ролі математичних методів в біологічних дослідженнях.

Значення біометрії в дослідницькій роботі біологів стало очевидним вже тоді, коли були відкрито статистичні закони, що діють у сфері масових явищ. Оскільки статистичні методи базувалися на великих кількостях спостережень, і вимагали великої обчислювальної роботи, до чого у біологів, які звикли до роботи на нечисленних вибірках, не було досвіду, тому біологи не відразу оцінили всю важливість цих відкриттів.

Положення стало змінюватися після того, як була обґрунтована теорія малої вибірки, піонером якої був учень К. Пірсона В. Госсет (1876-1937), псевдонімом «Стьюдент». Подальший розвиток теорія малої вибірки отримала в працях Пірсона і особливо Р. Фішера (1890-1962), який зробив величезний внесок в біометрію, збагативши її новими методами статистичного аналізу. Крім того, Фішер заклав основи планування експериментів, ввів в біометрію цілий ряд нових термінів і понять і переконливо показав, що планування експериментів і обробка їх результатів – це два нерозривно пов'язані завдання статистичного аналізу.

Початок визнання біометрії як обов'язкового механізму біометричних досліджень, що може забезпечувати високу достовірність одержаних результатів відносять до вивчення питань про розмір військового одягу, яке провів американський лікар А. Нетле у 1835 році. Він розставив десять тисяч солдат американської армії по росту і вперше з'ясував, що розміри рекрутів мають дивовижну закономірність розподілу: найменша кількість рекрутів мала самі маленькі розміри, найбільша кількість рекрутів – середні розміри, а кількість рекрутів із збільшенням розміру знов зменшувалась. Такий характер розподілу виявився діючим по відношенню до всіх біологічних об'єктів.

1.4. Лісова біометрія як частина загальної біометрії

Лісова біометрія – це розділ біометрії, змістом якого є планування і організація кількісних експериментів в лісознавстві, лісівництві, лісовій таксації та в інших лісових дисциплінах, обробка та аналіз отриманих результатів, використовуючи методи математичної статистики.

Лісова біометрія досить розвинена наука. Для опису лісу, лісових біогеоценозів, окремих елементів лісового насадження використовуються і дають важливі практичні результати багато математичні підходи, що застосовуються в загальній біометрії.

Методи лісової біометрії широко використовуються при аналізі процесів в природних явищах, властивих деревам і деревостанам, а також окремим ділянкам лісу. Як приклад можна привести аналіз успадкування ознак дерев і деревостанів, оцінка ефективності різних лісогосподарських заходів.

Лісова біометрія – необхідний методичний інструмент для проведення наукових досліджень в лісі, при розробці різних нормативних і довідкових матеріалів. Наприклад, всі лісівники користуються сортиментними і об'ємними таблицями для обліку лісу на корені, застосовують спеціальні таблиці для обліку готової лісо продукції, заміріючи довжину сортименту і його діаметр у верхньому відрізі т. д. В цих таблицях наведені усереднені величини, отримані на основі обробки методами біометрії великого експериментального (його ще називають первинним або вихідним) матеріалу. Далеко не кожна колода певної довжини і діаметру має об'єм, що точно співпадає з тим, що наведений у довіднику для параметрів довжини і діаметра, які необхідно знайти. Але, користуючись методами лісової біометрії, вчені отримали об'єми з достатнім ступенем наближення до реальності, і точність визначення збільшується при зростанні числа вимірів.

У лісовій біометрії широко застосовуються методи математичного аналізу, математичне моделювання. В даний час найбільш раціональним є застосування біометричних методів з використанням комп'ютерів. Практично сьогодні всі розрахунки проводяться тільки на комп'ютерах.

У той же час, щоб правильно використовувати методи лісової біометрії, необхідно добре знати лісівництво, лісову таксацію, лісові культури та інші дисципліни, де необхідне застосовування цих методів. Використання біометричних методів не гарантують отримання коректного результату без знань в галузі. Механічне, бездумне оперування цифрами, зведеними в моделі, нехай навіть і обробленими статистичними методами, неприпустимо, і може привести до помилкових висновків.

Перш ніж виводити модель, треба уявити загальну біологічну закономірність. Наприклад, відомо, що дерево росте вгору, діаметр дерева не може зменшуватися і т. д. Це найпростіші приклади, де все очевидно. Але є багато випадків, коли бездумне застосування статистичних моделей призводить навіть при аналізі росту дерев до негативних величин. Відносно наведених прикладів помилку виявити легко, але далеко не всі закономірності настільки очевидні. Саме тому, починаючи роботу з матеріалом, який буде оброблений біометричними методами, треба дуже добре знати свій предмет: біологію, лісівництво, таксацію, лісову селекцію і т. д. Надалі при вивченні різних методів біометрії ми це розглянемо більш докладно на конкретних прикладах.

Запитання для самоконтролю:

1. Що є предметом біометрії? Що таке біометричні показники і біометричні елементи?
2. На чому побудована біометрія? Яке значення біометрії в дослідницькій діяльності?
3. Наведіть приклади з вашого побуту, де використовуються біометричні методи.
4. Назвіть основні етапи розвитку біометрії, як науки.

РОЗДІЛ 2
СТАТИСТИЧНА СУКУПНІСТЬ ВИПАДКОВИХ ПОДІЙ ТА ЇЇ ОСОБЛИВОСТІ

2.1. Статистичні сукупності і статистичні спостереження. Статистичні вибірки.

2.2. Генеральна і вибіркова сукупність і їх обсяг.

2.3. Методи збору і обробки інформації в лісовій біометрії.

2.1. Статистичні сукупності і статистичні спостереження. Статистичні вибірки

Вивчення біологічних явищ проводиться не по окремим спостереженням, які можуть виявитися випадковими, нетиповими, такими які неповно виражають сутність даного явища, а на безлічі однорідних спостережень, що дає більш повну інформацію про досліджуваний об'єкт. Наприклад, вивчаючи ліс, скажімо ялинник в певних умовах місця зростання, ми не обмежуємося вимірами 1-2 або 5-10 дерев, а проводимо заміри на всій пробної площі, де росте не менше 200 ялинових стовбурів, тобто вивчаємо ту їх сукупність, яка вже складає ліс.

Застосування біометричних методів дослідження можливе лише при обстеженні значної кількості об'єктів. Кожен такий об'єкт або його окрема частина і ознака вимірюється і позначається певним значенням. Разом ці значення складають статистичну сукупність або множину.

Статистична сукупність або множина індивідуально різних, але біологічно однорідних одиниць, які об'єднуються (в сукупність) для сумісного (групового) вивчення є предметом біометрії.

При цьому абсолютно не обов'язково, щоб сукупність складалася з безлічі особин одного виду і віку, наприклад з однорідних дерев сосни. Вона може бути утворена і в результаті численних випробувань, тобто проб, спостережень і т. п., що проводяться на одному і тому ж індивідуумі.

Наприклад, статистичною сукупністю будуть дані спостереження зміни густоти травостою, фенологічні спостереження за одним або кількома деревами дуба і т. д.

Таким чином, сукупність об'єднує якесь число однорідних спостережень або реєстрацій. Сукупностями є дерева в деревостані, популяції комах, рослини на дослідних ділянках. Поняття сукупності можна застосувати не тільки до тварин або рослин. Такими ж сумами є молекули газу в деякому обсязі, населення міста і т. д.

До складу сукупності входять різні члени, або одиниці: для популяції тварин – кожна окрема тварина, для стада корів одиницею є кожна корова, для сукупності насіння сосни – кожне зернятко, для деревостану – дерево, при вивченні дерева – його клітини і т. д.

Ознака – це загальна властивість досліджуваного предмета. Так, при вивченні лісового насадження ознаками будуть вид дерева, вік дерев і деревостанів, розмір, форма або колір насіння, особливості вегетації (у дуба є рання і пізня форми) і т. д.

Біологічні ознаки можуть вивчатись як в кількісному, так і в якісному їх визначенні. В кількісному визначенні вивчаються розміри організмів та їх морфологічних органів, маса, об’єм, число пелюсток в квітах або листків на гілці, кількість зернин в колосі, маса врожаю або окремих насінин, приріст організмів за масою, довжиною, об’ємом та інше.

Кількісні показники можуть бути рахівними або мірними.

Ознаки, які існують дискретно (листки, зерна, пагони, окремі організми тощо) називають рахівними.

Ознаки, які не можуть бути представлені дискретно, тобто відокремлено, називають мірними. В якісному визначенні вивчаються такі ознаки, як смак, запах або забарвлення органу. Вони можуть бути представлені в їх номінальному значенні (світло-коричневий, коричневий, темно-коричневий) або в альтернативній формі (високий – низький; важкий – легкий; світлий – темний).

Число одиниць сукупності називають об'ємом сукупності і позначають латинською літерою N. Одиниця сукупності може характеризуватися певними ознаками, наприклад: молекули газу – швидкостями їх руху; насіння сосни – формою, кольором, вагою; дерева – товщиною, висотою і т. д.

Кожна досліджувана ознака приймає різні значення у різних одиниць сукупності, вона змінюється в своєму значенні від однієї одиниці сукупності до іншої. Ця різниця між одиницями сукупності називається варіацією або дисперсією, тобто розсіюванням. Ми говоримо – "ознака варіює". Це означає, що вона приймає різні значення у різних членів сукупності, наприклад насіння одного виду, дерев одного виду і т. п.

Елементи, що входять до складу сукупності, називаються її членами, або *варіантами*. Останній термін походить від латинського varians – змінюється. Варіанти – це окремі спостереження або деякі числові значення ознаки. Так, якщо позначити ознака через X (велика), то його значення або варіанти будуть позначатися через x (мала), тобто як x_1, x_2, x_3, x_4, ... x_k.

Загальна кількість варіант, що входять до складу даної сукупності, називається її *об'ємом* і позначається буквою N або Σn. Саму ж варіюючу величину, тобто величину, що змінюється під впливом багатьох випадкових причин і спроможну приймати різні значення, називають випадковою змінною n_i.

Варіанти є її числовими значеннями. Тут значок і – порядковий номер варіанти.

У той же час, не дивлячись на відмінності між варіантами, що входять в сукупність, остання має внутрішню однорідність. Члени сукупності подібні за рядом важливих ознак. білячі шкурки неоднакові за розмірами, якістю хутра, фарбування, але всі вони – шкурки особин одного і того ж виду – білки звичайної. Зерна пшениці відрізняються один від одного за вагою та іншим хімічним і фізичним ознаками, але всі вони – зерна пшениці, а не ячменю, хоча обидві культури могли бути вирощені на одному полі. Дерева варіюють за висотою, але всі вони одного виду, наприклад, сосни. Жолуді дуба відрізняються за розмірами, вагою, кольором, але вони насіння одного деревного виду – дуба черешчатого і т. д.

Найчастіше, до складу сукупностей входять окремі особини. Так, наприклад, при характеристиці плодоносних дерев сосни на лісо насіннєвій плантації за одиницю сукупності можна взяти кожне дерево. Однак одиницею сукупностей може бути не саме дерево, а деяка його характеристика. В даному випадку допустимо взяти урожай шишок або насіння за певні роки. Тоді при загальній кількості дерев на лісонасіннєвій плантації, скажімо 200 штук, кількість варіантів, одержуваних за кілька років насіння, складе 600, 800 або іншу величину.

Можна вивчати варіацію тієї чи іншої ознаки в часі навіть на одному дереві. Як відомо, розмір насіння і їх вага величина мінлива в залежності від ботанічних і абіотичних факторів. Вивчаючи їх зміни на одному дереві за ряд років, теж отримуємо статистичну сукупність, яка вивчається методами біометрії. Такою ж сукупністю є час розпускання бруньок на одному і тому ж дереві, але в різні роки. Сукупністю буде довжина і вага голок сосни на одній гілці і т. д.

Таким чином, сума спостережень або вимірювань є теж сукупність. Кожне окреме спостереження, при якому встановлюється значення випадкової змінної, є одиницею цієї сукупності.

Сукупність може складатися з інших, більш приватних сукупностей. Так, сукупність з усіх диких тварин одного виду розпадається на приватні сукупності – окремі популяції. У межах однієї популяції можна виділити ще більш приватні сукупності, наприклад, нащадків певних самців або самок. Вивчаючи деревостани сосни, їх можна розділити по областях, лісгоспах або лісо рослинних районах.

У всіх випадках ми стикаємося з постійними відмінностями як всередині окремих приватних сукупностей, так і між ними.

Розрізняють генеральну і вибіркову (часткову) сукупність.

2.2. Генеральна і вибіркова сукупність і їх обсяг

Генеральною називають найбільш загальну сукупність. Це теоретично нескінченно велика або, у всякому разі, що наближається до нескінченності сукупність всіх одиниць або членів, які можуть бути до неї віднесені. *Генеральна сукупність* об'єднує всю кількість біологічних об'єктів, яка вивчається. Так, якби можна було описати всі особини окремого виду, наприклад всі дерева бука в карпатських лісах, то вони склали б генеральну сукупність.

Генеральна сукупність може складатися з такої великої кількості одиниць, що вивчити їх всіх немає можливості. Тому практично доводиться мати справу з порівняно невеликими вибірковими сукупностями. Так, лісівник закладаючи пробну площу, де всього-то 200 дерев, робить висновки про всю сукупність.

Наприклад, при вивченні поновлення бука лісового на площі лісу 10000 га генеральною сукупністю буде весь підріст на досліджуваній території. За умови заміру всіх рослин підросту одержані характеристики будуть мати абсолютний рівень достовірності. Однак неможливо заміряти відповідну досліджувану ознаку у кожної рослини підросту лісу. А, якщо і можливо, то займатиме багато часу, людської праці і затрат. Власне, в цьому нема необхідності, якщо застосувати методи біометричного аналізу. Ці методи дозволяють одержати середнє значення відповідної ознаки, замірявши лише незначну частину рослин підросту, а також одержати іншу необхідну інформацію з певним ступенем наближеності результатів досліджень до значення, яке притаманне генеральній сукупності.

Питання про те, в якій степені по вибірковій сукупності можна судити про генеральну, належить до числа найважливіших теоретичних і практичних питань в біологічній статистиці. Це питання розглянемо згодом.

Завданням вивчення будь-якої сукупності є отримання статистичних (біометричних) характеристик, або показників. Вони дозволяють судити про дану сукупність в цілому, про відмінності всередині неї і про відмінність її від інших, схожих з нею або близьких до неї сукупностей. Сукупність стає статистичною саме тоді, коли в її опис вноситься кількісний метод.

Застосування кількісного методу вивчення сукупності і дозволяє отримувати для неї ряд статистичних показників. З їх допомогою ми отримуємо основну інформацію про сукупності.

Щоб вибіркова сукупність якомога повніше відображала генеральну, необхідно враховувати такі основні положення.

1. Вибірка повинна бути цілком представницькою, або *типовою*, тобто щоб до її складу входили переважно ті варіанти, які найбільш повно відображають генеральну сукупність. Тому

перед тим як приступити до обробки вибіркових даних, їх уважно переглядають і видаляють явно нетипові варіанти. Наприклад, при вимірі довжини класів не можна включати до вибірки зіпсовані сажкою або обірвані колосся, оскільки вони нетипові для такого роду вибірки. При вивченні ходу росту дерев у висоту треба виключити дерева, зламані бурею, пошкоджені вогнем і т. д.

2. Вибірка повинна бути об'єктивною. При формуванні вибірки не можна чинити по сваволі, включати до її складу лише ті варіанти, які здаються типовими, а всі інші бракувати. Доброякісна вибірка проводиться без упереджених думок, за методом жеребкування або лотереї, коли жодна з варіант генеральної сукупності не має ніяких переваг перед іншими – потрапити або не потрапити в склад вибіркової сукупності. Іншими словами, вибірка повинна проводитися за принципом випадкового відбору, без яких би то не було суб'єктивних вилучень на її склад. Наприклад, якщо ми хочемо для визначення середньої висоти виміряти 20 дерев ялини, то не можна їх вибирати на свій смак, виключаючи, скажімо, низькі пригноблені стовбури.

Треба вимірювати висоту у кожного 10 або 20 і т. д. дерева ялини, при цьому треба враховувати обмеження, згадані вище в п.1, наприклад, виключати дерева, зламані вітром.

3. Вибірка повинна бути *якісно однорідною*. Не можна включати до складу однієї і тієї ж вибірки дані, отримані на особинах різної статі, виду, віку або фізіологічного стану, так як наперед відомо, що ці чинники по-різному позначаються на величині і функціональному стані ознак, за якими може бути утворена вибіркова сукупність. Неоднорідний за складом матеріал не дає вірної інформації про досліджувані явища. Наприклад, не можна об'єднувати в одну пробну площу

деревостани різних типів лісу, хоча б вони росли поруч, скажімо грабова і букова діброви.

Всі ці умови може дотримати тільки фахівець, який добре знає не тільки біометрію, а й предмет свого дослідження.

Емпіричні, або вибіркові, сукупності можуть мати самий різний обсяг. Залежно від числа спостережень прийнято розрізняти *малі вибірки, що містять не більше 30 варіант*, і вибірки більше > 30, що включають в свій склад до 100-200 одиниць сукупності і більше. Верхня межа тут не обмежена. Принципової різниці між великою і малою вибірками немає.

Розрізняти їх доводиться на тій підставі, що порівняльна оцінка біометричних показників, що обчислюються на малих вибірках, знаходиться в залежності від числа спостережень, про що буде розказано нижче.

Якісна однорідність сукупності визначається метою дослідження.

Таким чином, можна зробити висновок, що сукупність статистична – це сукупність предметів, явищ, речей, якісно однорідних і варіюють за обліковою (біометричною) ознакою: звідси назва – варіаційна статистика.

Генеральна статистична сукупність – певна філософська категорія, всеосяжна сукупність предметів, явищ, речей, яка може розглядатися як кінцевою, так і нескінченною, але обов'язково якісно однорідною і варіює за обліковою ознакою.

Вибіркова (часткова) статистична сукупність – це частина генеральної сукупності, яка задовольняє вимоги репрезентативності.

2.3. Методи збору і обробки інформації в лісовій біометрії

Вибір правильних методів збору і обробки інформації визначає результат дослідження.

Виділять три різні типи джерел інформації, які необхідні при проведенні досліджень: 1) біологічні об’єкти, як безпосередні джерела біологічної інформації; 2) наукові публікації, як джерела опосередкованої інформації про об’єкти; 3) бібліографія джерел наукової інформації.

Розглянемо як отримати інформацію з різних типів джерел більш детально. Для прикладу, яку інформацію дослідник отримує від біологічного об’єкту? Це потік даних, які сприймає людина органами відчуттів, або приладами та механізмами, які розширюють його можливості. Органами відчуттів сприймаються колір, вага, запах, звук, зволоженість, структура поверхні, рухи та вібрації і таке інше, що фіксується в пам’яті дослідника. Кількість та якість цієї інформації залежать від суб’єктивних можливостей органів відчуттів дослідника. Для запобігання втрати інформації чи її викривлення з часом використовують методи негайної фіксації: записи в польових щоденниках та лабораторних журналах, зарисовки, в останні роки фото- та відео фіксації.

У випадку використання під час досліджень допоміжних приладів і точних методів поле сприйняття людини штучно розширюється.

Сукупність даних, отриманих за допомогою приладів, значно точніше описує об’єкт. Але слід відзначити, що їх об’єктивність обмежується точністю вимірювальних приборів, надійністю методів вимірювання, випадковими факторами та ін., тобто практично ніколи не дорівнює 100 %.

Між безпосередніми спостереженнями та вимірюванням показників за допомогою приладів є різниця, яка полягає в тому, що людина водночас фіксує весь комплекс властивостей та характеристик об’єкту, на основі яких робить висновки. При цьому людина на основі інтуїції може розширити можливості отримання інформації чи утворити багаторівневі асоціації.

З іншого боку, спостереження за допомогою приладів дозволяє отримати порівняно вузький спектр даних, який залежить від можливостей обладнання.

Отож, не можна повністю замінити дослідження, які проводить дослідник з використанням приладів автоматичними спостереженнями за живими об'єктами без присутності людини.

Будь-який біологічний об'єкт дослідження завжди є частиною природної ієрархії. На її нижньому рівні знаходяться атоми хімічних елементів, а на верхньому – об'єкти рівня біосфери, яка включає в себе всі інші. Так, наприклад, дерево є складовим компонентом деревостану, який займає певне місце в своєму біогеоценозі. Далі по збільшенню масштабу йдуть біогеоценотичний комплекс-ландшафт, біом (наприклад, лісостеп), біогеографічна область (наприклад, палеоарктична – на північ від Гімалаїв), екосистема суходолу, біосфера Землі.

Дерево в свою чергу складається з надземної та підземної морфологічних частин, далі – по зменшенню масштабу: система органів (наприклад, асиміляційний апарат), органи (листя), тканини (наприклад, поглинаюча – хлоренхіма), клітини певного типу, органоїди клітини (хлоропласти), структурні одиниці органоїдів (квантосоми хлоропластів), молекули пігментів, білків та ліпідів, атоми хімічних елементів.

Всі складові цієї ієрархії безпосередньо або опосередковано пов'язані одне з одним. Структура цих зв'язків дуже складна. Реально при проведенні дослідження увага концентрується лише на певних властивостях, характеристиках та деяких зв'язках конкретного об'єкту.

Ще одним ускладнюючим фактором є те, що біологічний об'єкт – це завжди динамічна структура, яка змінює свої властивості протягом часу.

Таким чином, об'єм даних про біологічний об'єкт та його зв'язки, який може бути отримано під час дослідження, настільки великий, що майже не може бути виміряний. Він обмежується особливостями сприйняття людини, точністю вимірювальних приладів та діапазоном досліджуваних характеристик, а також метою дослідження.

Дані, отримані під час досліджень біологічного об'єкту далі аналізуються і обробляються різними методами, в тому числі біометричними, та становляться базою для висновків, підтвердження або відхилення попередніх гіпотез і т. ін.

Для науковця, якщо він сам не проводив даного дослідження, сума цих опосередкованих даних становиться значимою та досяжною після їх публікації. Основними типами наукових публікацій є колективні та авторські монографії, узагальнюючі праці, наукові статті, звіти, автореферати, стенограми матеріалів наукових конференцій, семінарів і т. ін. Це також можуть бути і аудіовізуальні та цифрові записи. В наш час все більшого значення набувають публікації в Інтернеті та інших інформаційних мережах.

Всі публікації групуються за напрямками досліджень. В кожної з публікацій розглядається певне коло властивостей об'єкту або групи об'єктів. Таким чином, в порівнянні з потоком даних від об'єкту, об'єм даних в публікаціях значно менший. Але це не свідчить про те, що він малий або недостатній. Загальна кількість наукових біологічних публікацій на даний час рахується на мільйони.

Їх перевага перед джерелами першого типу полягає в більшій досяжності (наприклад, науковець може отримати необхідну для порівняння інформацію про географічно-віддалений об'єкт), в більшій структурованості та упорядкованості, в зниженні витрат на отримання інформації.

Джерела цього типу виконують функцію збереження фактографічних даних. Вони акумулюють в собі досягнення біологічної науки та забезпечують їх передачу в часі та поширення в просторі. Цей механізм став одним з головних чинників швидкого розвитку наукової думки в XX сторіччі. Вивчення існуючих публікацій за напрямком дослідження є обов'язковою умовою його ефективності, що запобігає зайвій витраті зусиль на пошуки вже знайденого.

Джерела цього типу насичені спеціальною термінологією, більша частина якої майже ніколи не зустрічається в повсякденній мові. В біологічних публікаціях традиційно використовується велика кількість латинських слів для позначення органів, частин тіла, назв видів та ін.

У зв'язку з наведеними особливостями наукової біологічної інформації, а також з метою підвищення структурованості джерел в них часто розміщуються предметні та алфавітні покажчики, покажчики латинських назв.

Окремо видаються словники біологічної термінології. Ці особливості обмежують використання біологічних наукових джерел широким колом читачів, роблять цю інформацію менш досяжною.

Інформація цього типу застаріває, тобто втрачає з часом адекватність біологічному об'єктові, який вивчається. Це майже не стосується публікацій, в яких йдеться про загальні, фундаментальні закони та принципи, і безпосередньо торкається публікацій, які описують властивості конкретних об'єктів. Застарілості інформації сприяє згадана вище динамічність всіх біологічних структур, їх схильність до зміни властивостей. Цього можна уникнути, вивчаючи динаміку біологічних структур в часі, періодичність їх змін.

Об'єктивність наукових публікацій гарантується автором, інститутом рецензування, авторитетними редакційними колегіями. Але це не виключає водночас суб'єктивного тлумачення авторами даних, що покликано співіснуванням в науці різних ідей, напрямків, шкіл, течій. Наукові публікації відрізняються вагомістю та науковою новизною.

Бібліографування має величезне значення для будь-якого наукового дослідження. Сьогодні бібліографія грає навігаційну роль, орієнтує дослідника в потоках публікацій, є засобом ідентифікації документа в масиві йому подібних за бібліографічним описом. Масиви бібліографічних описів, як і масиви джерел упорядковуються за допомогою різноманітних класифікацій, які побудовані на основі виділення розділів людського знання. На цей розподіл значною мірою впливають тенденції розвитку науки.

Сьогодні основними класифікаціями є десятинна класифікація Дьюї, універсальна десятинна класифікація (УДК), бібліотечно-бібліографічна класифікація (ББК) і інші.

За допомогою класифікацій формуються індекси. Вони показують належність джерела до конкретного підрозділу знання, вказують на місце бібліографічного запису в упорядкованому масиві, наприклад, в систематичному каталозі наукової бібліотеки або в комп'ютерній базі даних.

Певне наукове або виробниче завдання, яке ставиться у відповідності з черговими потребами і планами розвитку галузі, може бути вирішене на основі накопиченої інформації або для її рішення може знадобитися збір (повний або частковий) нової інформації. В обох випадках в більшості лісівничих завдань на різних етапах використовують нормативно-довідкову інформацію; іноді частка її в загальній кількості інформації і вплив на кінцеві результати вельми істотні.

Якщо наявної інформації достатньо, то на її основі формулюють відповідну гіпотезу і розробляють модель, яку перевіряють емпірично за допомогою статистичних методів. Якщо ж такої інформації недостатньо, то приймають рішення про конкретні шляхи дослідження, які визначаються метою роботи, фінансовими і трудовими можливостями, наявними в наявності засобами збору і обробки інформації.

Одним з основних моментів при організації спостереження є розробка правильної методики. Треба пам'ятати, що помилки, допущені в методиці збору первинного матеріалу, не можуть бути потім виправлені ніякою камеральною обробкою і виникнуть неточності або помилки у висновках.

Як відомо, існує два основних способи спостережень за охопленням одиниць досліджуваного об'єкта: суцільне обстеження всіх одиниць досліджуваної сукупності і часткове обстеження, коли спостереження піддається лише частина одиниць досліджуваної сукупності. У лісовому господарстві зазвичай обмежуються частковим (вибірковим) обстеженням.

Вибірку виробляють для отримання характеристики цілого, що підлягає вивченню. Та сукупність (як зазначалось вище), яка підлягає вивченню, називається загальною, або генеральною, а відібрані з неї одиниці спостереження представляють *часткову*, або *вибіркову*, сукупність. Відносна частина, яку становить число одиниць з даними значення ознаки від загального числа одиниць сукупності, називається в генеральної сукупності – *часткою*, а в вибіркової – *частковістю*.

Рослини підросту лісу, які були узяті з їх генеральної сукупності для досліджень, називаються *вибірковою* або *частковою* сукупністю, простіше *вибіркою*.

Окремі одиниці, що складають генеральну або часткову сукупність (x) (в нашому прикладі окремі рослини приросту або їх ознаки), є членами або варіантами відповідної вибіркової сукупності, загальна чисельність якої позначається (n).

Як суцільні, так і несуцільні спостереження за способом проведення можуть бути:

✓ По зв'язку з об'єктом – безпосереднє, експедиційне, кореспондентське і звітне;

✓ За часом проведення – безперервне, в міру виникнення явища (наприклад, фенологічне); періодичне (повторне) – через певні проміжки часу і одноразове або одноразове;

✓ За джерелами – власне спостереження, усне і письмове (анкетне) опитування та за документальними матеріалами – літературними, службовими і архівними даними.

✓ За особливостями відбору одиниць:

➢ відбір середніх типових одиниць, напр., застосовується в лісовій таксації;

➢ відбір випадкових одиниць, вироблений за спеціальним планом, який є в статистиці основним;

➢ по виду самого відбору:

➢ без повторне, коли відібрана одиниця після проведення спостережень над нею не повертається в генеральну сукупність;

➢ повторне, коли обстежена одиниця повертається в генеральну сукупність і може знову потрапити в відбір.

✓ За принципом відбору одиниць – випадковий, типовий (по групах, однорідних по будь-якій ознаці) і механічний. Відбір можна здійснювати або з усієї сукупності, а також з обмеженої сукупності. В останньому випадку одиниці з крайніми значеннями ознаки не беруться до уваги, відбір може вестися і з приватних сукупностей, на які на основі кількісних (наприклад, ступені товщини) або ж якісних (наприклад, класи росту дерев) ознак. розбита загальна сукупність Часткове спостереження методом випадкового відбору, при якому кожна одиниця досліджуваного об'єкту має однакову з іншими можливість потрапити до вибірки, забезпечує повну об'єктивність спостереження.

При роботах в лісі випадковість (об'єктивність) відбору здійснюється застосуванням механічного відбору за принципом без повторної вибірки. У цьому випадку будь-яка одиниця сукупності може потрапити в вибірку тільки один раз.

Принцип вибіркового методу теоретично обґрунтований математиком П. Л. Чебишевим (1821 – 1894), який довів, що при досить великій вибірці вибіркова середня може бути як завгодно близька до генеральної середньої. Звідси випливає, що частота ознаки в вибірковій сукупності може бути як завгодно близька до частки цієї ознаки в генеральній сукупності. Іншими словами – досить велика вибірка правильно відтворює особливості і властивості генеральної сукупності і тим краще, чим відносно більшою є вибірка.

Відбір одиниць можна здійснювати за методом випадкової вибірки (рандомно), коли з сукупності навмання, наосліп (наприклад, за жеребом) відбирають необхідну кількість одиниць. Даний метод частіше використовують в сільському господарстві, при біологічних та екологічних дослідженнях. Це зробити досить складно, а в умовах лісу практично не застосовується. Тому матеріал спостереження при дослідженнях в лісовій справі збирають шляхом механічного відбору.

Сутність цього способу полягає в тому, що всю сукупність механічно розбивають на число частин, однакових за розміром або кількістю одиниць, відповідних числу спостережень, і в кожній частині навмання вибирають одиницю для спостереження. Можна розбити і на число частин, в кілька разів менше кількості спостережень. В цьому випадку з кожної частини потрібно взяти навмання не по одній одиниці, а більше в стільки разів у скільки потрібно для спостереження.

Наприклад, при вивченні якості насіння (плодів) всю партію насіння можна механічно розбити на кілька частин або груп, однакових за кількістю або за вагою насіння, і з кожної такої частини або групи вибрати навмання партію насіння для спостереження. При відборі модельних дерев всю сукупність стовбурів на пробній площі (200-300 шт.) ділять за ступенями товщини (через 2-4 см) і відбір роблять всередині цих ступенів.

Є кілька широко застосовуваних способів відбору.

1. ***Спосіб стрічок.*** Поперек обстежуваної площі (наприклад, лісосіки), через однакову відстань закладають стрічки однієї і тієї ж ширини, наприклад, 1, 2 або 10 м, із суцільним обстеженням на кожній. Площа всіх смужок або стрічок, виражена у відсотках від всієї площі обстеження, дасть відсоток вибірки. Величина відсотка вибірки залежить від: величини обстежуваної площі (чим більше площа, тим менше може бути відсоток вибірки); ступеня коливання досліджуваної ознаки (чим більша ступінь мінливості ознаки, тим більша повинна бути вибірка). Тому число стрічок в різних випадках може бути різний. Тому, при виборі методики потрібно мати уявлення про розміри обстежуваних площ і особливості мінливості ознаки. Останнє визначається або за літературними даними, або за матеріалами господарства чи лісгоспу, або ж шляхом попередньої закладки досліду. Число одиниць спостереження N (в даному випадку – число стрічок, забезпечують результат з запланованою точністю) можна визначити за формулами:

$$\mathrm{N} = \frac{V^2}{\mathrm{p}^2}; \tag{2.1}$$

або

$$\mathrm{N} = \frac{\sigma^2}{m^2}; \tag{2.2}$$

де V – коефіцієнт мінливості;

σ – середньоквадратичне відхилення значень ознаки одиниць від їх середнього значення;

m – запланована точність в одиницях виміру ознаки;

p – точність досліду, %.

Методи знаходження V, σ, m, p будуть розглянуті нижче.

Знаючи необхідну кількість стрічок і довжину обстежуваної площі, відстань між центрами стрічок можна отримати поділом довжини обстежуваної площі на запроектоване число стрічок.

При неможливості попередньо підрахувати необхідне число стрічок треба брати не менше 5-10 стрічок і встановити відстань між ними розподілом довжини обстежуваної площі на число цих стрічок. Кожну стрічку доцільно розділити на три-п'ять частин, однакових по довжині, які і будуть складовими одиницями спостереження, що складаються з ряду одиниць сукупності (дерев).

При обстеженні лісових культур стрічками можуть служити ряди культур: один, два і більше в стрічці. Кількість відібраних рядів дерев в стрічках, виражена у відсотках від усього числа рядів на обстежуваній площі, представляє відсоток вибірки.

2. ***Спосіб майданчиків.*** Майданчики строго однакового розміру і форми закладають через одну і ту ж відстань одна від одної.

Ширина і довжина майданчиків в різних випадках може бути різною, тобто майданчики можуть бути квадратними (наприклад, 1 х 1, 2 х 2 м), прямокутними (наприклад, 1 х 2, 2 х 3, 2 х 4 м), круговими і т. д. Майданчики розміщують по площі або рядами, коли вони сформовані в один ряд, тобто на одній лінії як в поздовжньому, так і в поперечному напрямку, або ж в шаховому порядку. Це залежить від спостерігача. Відстань вимірюється від центру майданчиків.

Такий спосіб застосовують, наприклад, при обліку природного поновлення. При обліку культур майданчиками за облікові майданчики приймають посадкові або посівні місця. Останні для спостереження можна відбирати або цілими шпальтами, розташованими через однакові відстані один від одного, або ж окремими майданчиками, коли в межах смуги обстежують не всі майданчики, а лише через одну і ту їх кількість, наприклад – кожний п'ятий, кожний десятий і т. д., незалежно від того, який за своїм станом він виявиться.

Якщо число майданчиків N, необхідних для отримання результату з заданою точністю, вже відомо, то відстань *l* між центрами майданчиків в метрах визначається за формулою 2.3:

$$l = \sqrt{\frac{П}{N}}; \qquad (2.3)$$

де П – розмір обстежуваної площі, м^2.

Намітивши візири (ходові лінії обстеження), розташовані один від одного на розрахованій відстані *l,* кожен з них розбивають на відрізки довжиною теж в *l* метрів; в кінці цих відрізків (або тільки праворуч, або тільки зліва, або тільки на самому візирі) закладають майданчики, де б в натурі вони не припали, нехай навіть на дорозі або на прогалині. Пересувати їх не можна.

Так само можна закладати ями для визначення зараженості ґрунтів, наприклад травневим хрущем, майданчики на стовбурі для вивчення ступеня заселеності і враження шкідниками-комахами і т. п.

Обстежувану ділянку попередньо можна розбити на типові частини, однорідні за будь-якою ознакою, і кожну типову частину обстежити окремо (типова вибірка). Але розбивку цілого на типові частини для їх характеристики можна виробляти камерально – за матеріалами спостереження. Наприклад, обстежуючи велику неоднорідну площу лісових культур її слід розділити на відносно однорідні частини.

3. Спосіб візирів. На обстежуваній площі через одну і ту ж відстань один від іншого пробивають візири. На кожному шляху облікові одиниці (майданчики, екземпляри і т. п.) для спостереження можна відбирати двома способами.

Кожну енну (наприклад, десяту, двадцяту і т. д.) по ходу, але тільки завжди або зліва, або тільки праворуч, незалежно від того, яка за своїм станом виявиться відібрана одиниця спостереження.

Слід зауважити, що принцип відбору енної (п'ятої, десятої і т. д.) по порядку одиниці спостереження можна застосовувати, наприклад, і при відборі гілок дерев для вивчення плодоношення і в ряді інших випадків; при суцільних обліках – відбір для більш детального вивчення кожного енного дерева поспіль. Так беруть модельні і облікові дерева.

Облікові одиниці намічають через певне число метрів, наприклад – 10; 20; 50 м і т. д. – в залежності від величини обстежуваної площі. Тут обстежується та одиниця, яка виявиться якраз в кінці кожного такого відрізка візира або в найближчій до цієї точки. При цьому дотримується правило відбору, вказане в попередньому пункті.

Число відрізків, на яке будуть розбиті всі візири, дорівнює числу одиниць спостереження, що відбираються.

Отже, знаючи необхідну кількість одиниць спостережень, задавшись відстанню між візирами, можна підрахувати число і загальну довжину візирів і розділити останню на прийняте число одиниць спостереження. В результаті отримаємо необхідну відстань по візиру між одиницями, що відбираються, або інакше – довжини відрізків в кінці яких і здійснюється спостереження. Відібрані таким шляхом дерева звуться модельними, якщо їх зрубують, і обліковими – якщо їх не зрубують.

Описаним прийомом можна відбирати дерева для вивчення відсотку виходу ділової деревини, відсотків заражених дерев і т. д., піддаючи окомірним спостереженням (опису), скажімо, кожне десяте дерево. Для детального вивчення із зрубуванням дерева можна брати, наприклад, кожне двадцяте, сорокове і т. д., виходячи з того, яка кількість дерев потрібна для детального обстеження.

Завданням спостереження може бути виявлення впливу на об'єкт спостереження, наприклад на дерево, навколишнього середовища, скажімо, зімкнутої біогрупи, відкритого місця, намету лісу і т. д. Ці особливості середовища можуть також впливати на характер відновлення, на склад і стан живого надґрунтового покриву, на плодоношення різних рослин, на ріст і розвиток лісу і т. д. Тоді облікові площадки одного і того ж розміру закладають в цих типових умовах, вплив яких потрібно вивчити; вони нерідко піддаються повторним періодичним спостереженням, наприклад через кожні 5 або 10 років.

При зборі матеріалу слід створити такий фундамент з точних і безперечних фактів, на який можна було б спиратися, з яким можна було б співставляти результати інших дослідників. Тому, необхідно аналізувати не окремі факти, а всю сукупність фактів, що відносяться до розглянутого питання, без жодного винятку.

При плануванні спостережень важливим є розробка форми записів даних спостереження з чітким переліком всіх ознак, що підлягають обліку, і зазначенням одиниць і точності вимірювання. Від якості форми запису залежить і повнота одержуваних відомостей. Форма запису може бути облікова, коли в відомість заносяться дані спостережень по кожній одиниці спостереження в порядку обстеження, і карткова, коли всі дані про кожну одиницю спостереження заносяться в окрему картку або бланк. Характерним прикладом тут буде картка таксації модельного дерева. Карткова система запису зручніше в тому відношенні, що сильно полегшує камеральну обробку при зведенні і групуванні матеріалу за різними ознаками. В даний час все частіше для запису використовують носії інформації для комп'ютерів, які вставляють в спеціальні пристрої. У цьому випадку інформація відразу вводиться в ПК, виключаючи її набір, що багаторазово прискорює процес обробки інформації.

Перед початком досліджень розробляється методика проведення спостереження: уточняється одиниця спостереження, якщо вона складна, визначаються її розміри і форма (стрічки, пробні майданчики і площі); встановлюється точність кінцевого результату і необхідну кількість одиниць спостереження, спосіб проведення спостереження, відбору одиниць; оформлення одиниць спостереження в натурі; терміни спостережень і одиниці вимірювань. Складається календарний план робіт і кошторис витрат.

Узагальнюючи викладене про проведення спостережень, можемо зробити такі висновки.

- Спостереження є дослідною основою статистичного дослідження.
- Для того, щоб за даними вибірки можна було б з певним ступенем упевненості робити висновки про сукупність, вибіркове спостереження має бути правильно організовано. Тут вирішуються два основних питання:

- яке число спостережень є достатнім;
- які одиниці сукупності повинні бути обрані для спостереження, тобто що (хто) буде складати вибірку.

Перше питання може бути вирішене за допомогою таблиці досить великих чисел, яка наводиться в багатьох посібниках по статистиці, або із застосуванням спеціальних формул. Вони будуть описані нижче.

- ✓ Відбір одиниць для спостереження може бути спланований по-різному залежно від складу сукупності і відомостей про неї.
- ✓ Якщо сукупність варіює не в надто широких межах і якщо вибірка становить не менше 20% обсягу сукупності,

застосовують простий випадковий відбір одиниць або просте вибіркове спостереження.

Для цього зручно користуватися таблицею випадкових чисел.

✓ В лісовому господарстві скористатися таблицею випадкових чисел в натурі технічно важко і практично часто неможливо. Тому тут користуються систематичним вибірковим спостереженням. Наприклад, якщо належить узяти 10% -ну вибірку дерев з 800 штук, то випадковим порядком обирають перше, між інших 5-ти дерево, і після цього беруть кожне наступне через 10 номерів, тобто за номерами: 15, 25, 35 і т. д., закінчуючи номером 795.

✓ Якщо є відомості про те, що сукупність в своїх частинах неоднакова, наприклад з більш високим рівнем явища в одних частинах, ніж в інших, доцільно пошарове вибіркове спостереження.

Наприклад, відомо, що запас деревостанів менше варіює в межах класів віку. Тоді для отримання статистичних характеристик величини запасу всю сукупність деревостанів розчленовують на групи по віку. Отримаємо шари сукупності, для кожного з яких беруть незалежну вибірку і обчислюють її характеристики. Статистична обробка матеріалів досліду пошарової вибірки трохи складніша.

Але тільки спостереженнями і статистичним опрацюванням не обмежується збір інформації в лісівничих та інших дослідженнях.

Дуже часто на додаток до спостережень ставиться експеримент. У сукупності спостереження і експеримент практично вичерпні джерела первинної інформації в лісовому господарстві.

Спостереження зазвичай не вимагають втручання в нормальне функціонування об'єкта. У багатьох лісівничих дослідженнях вони є єдино можливими, наприклад, фенологічні спостереження, вивчення росту дерев і деревостанів, приживлюваності лісових культур і ін. Однак певна "пасивність" спостереження по відношенню до об'єкта дослідження не передбачає відсутності плану або системи: спостереження як метод наукового пізнання передбачає наявність суворого плану. Доброю ілюстрацією планованих спостережень є вибіркові методи інвентаризації лісових ресурсів на великих територіях, які проводяться в багатьох країнах.

Дослідження передбачає активний і цілеспрямований вплив на досліджуваний об'єкт або явище, певну керованість умов його проведення. Співвідношення ролі спостереження і експерименту досить складне. У науковому пізнанні завдання спостереження зазвичай більш скромне і зводиться частіше до опису та аналізу явищ, що спостерігаються і процесів. В експерименті сильніша теоретична сторона, рівень осмислення спостережуваних факторів; експеримент у своєму розпорядженні має засоби активного втручання в хід подій. Однак в лісовому господарстві, особливо при вивченні природних об'єктів, спостереження часто відіграє важливішу роль, ніж експеримент.

Для зручності класифікації можна виділити звичайний модельний експеримент і математично спланований або екстремальний. Звичайний модельний експеримент відрізняється виділенням досліджуваних зв'язків і ізоляцією їх від зовнішніх впливів; при цьому він може бути однофакторним і багатофакторним. Наприклад, беремо сіянець, поміщаємо його в штучне середовище і досліджуємо вплив на його ріст деякого добрива. Якщо ж цей сіянець спостерігати в природних умовах, то треба врахувати і опади, і температуру і інше, тобто багато чинників, а не одне добриво. Математичне планування експерименту (багатофакторного) дозволяє оптимізувати сам процес дослідження: заздалегідь вибрати найкращу (з точки зору мети роботи) математичну модель, застосувати послідовну стратегію і скорегувати напрямки досліджень після кожного етапу і т. д.

При будь-якому методі збір інформації, її обробку та використання будують за схемою: інформація – гіпотеза – модель – перевірка відповідності моделі вихідної інформації і об'єкту, для якого розроблена модель.

Запитання для самоконтролю:

1. Поясніть, що таке статистична сукупність. Наведіть приклади.
2. Що таке біологічна ознака, біометрична ознака, варіанта, об'єм вибірки. При поясненні наведіть приклади.
3. Дайте визначення поняттям «генеральна» і «вибіркова» сукупності з наведенням прикладів.
4. Назвіть основні положення для більш повного відображення генеральної сукупності при дослідження вибірки.
5. Які джерела інформації щодо біологічних об'єктів вам відомі?
6. Які можуть бути спостереження за способом проведення?
7. Способи відбору одиниць спостереження. Опишіть один із них.

РОЗДІЛ 3
ГРУПУВАННЯ ВИХІДНИХ ДАНИХ

3.1. Кількісний та якісний аналіз масових явищ

При розгляді масових явищ, коли маємо справу з великими масивами інформації (дерева в лісі, партії насіння, густота травостою, населення країни і т.д.) їх аналіз може бути якісний або (і) кількісний.

Якісний аналіз досліджуваного явища або процесу полягає у виділенні деяких його характерних властивостей, особливостей, ознак, що якісно відрізняються між собою всередині даної сукупності. Наприклад, при вивченні змішаного насадження, що складається з бука лісового і ялини європейської. Заклали пробну площу, на якій нарахували 220 дерев бука і 90 ялини. Для подальшого використання цього матеріалу необхідно зробити попередній якісний аналіз. Тобто розділити дерева по головній якості – по належності їх до ботанічних видів, тобто на бук і ялину. Наведений приклад простий. На практиці ж буває так, що зробити якісний аналіз важко, для цього застосовуються спеціальні методи.

Але якісного аналізу часто буває недостатньо, щоб зрозуміти деяке явище або процес, дати йому коректний математичний опис.

У цьому випадку необхідно використовувати кількісні методи дослідження.

Проведення якісного та кількісного дослідження починається з планування та постановки експерименту. У лісовому господарстві дослідження часто полягає в проведенні вимірювань на деяких виділених ділянках лісу (пробних площах) або в вимірі частини дерева.

При науковому або практичному дослідженні деякого явища або процесу потрібно з'ясувати його природу, закономірності зміни у часі або зв'язок з деякими параметрами. Для цього недостатньо провести спостереження або поставити експеримент. З'ясувати закономірності досліджуваного явища, отримати правильні висновки зі спостережень можна тільки в тому випадку, якщо буде зроблений коректний кількісний і якісний аналіз проведених спостережень, або, як їх ще називають, випадкових явищ.

3.2. Систематизація і групування вихідних даних

Будь-який аналіз проведених спостережень починається з систематизації спостережень. Першим її етапом є угруповання вихідних даних або варіантів. При постановці дослідження угруповання передбачають вже на етапі збору експериментального матеріалу, тобто на етапі спостережень.

Спостереження – це збір первинних даних про об'єкт, який підлягає вивченню. Кожне окреме спостереження називають варіантою або датою. *Групування* – це процес систематизації результатів спостережень для отримання закладеної в них інформації, а також виявлення закономірностей, які властиві досліджуваним ознакам.

Однією з форм групування є побудова статистичних рядів – рядів числових значень ознаки розміщених в певному порядку.

Для побудови ряду розподілу потрібно згрупувати всі варіанти у наперед встановлені інтервали (класи). Наприклад, визначаючи висоти в 6-8-річних культурах сосни, в робочий зошит записують окремо кожний вимір як наведено нижче.

Висота, м	0,8	1,5	3,1	2,2	0,4	1,1	1,6	1,4
	1,9	2,0	1,8	2,4	2,7	2,9	0,9	

Аналіз наведених величин, хоча їх відносно небагато, в представленому вигляді ускладнений. При великих масивах інформації аналіз окремих вимірювань переростає у велику проблему. Для її вирішення результати спостережень, як правило, систематизують. Систематизація полягає в угрупованні виміряних величин: товщини або висоти дерев, ваги тварин, розміру і ваги насіння і т. п. Для угруповання спостережень виділяють класи (при вимірах дерев їх називають ступені), за якими розносять виміряні величини. У наведеному прикладі доцільно виділити наступні класи висот:

0-0,5 0,51-1 1,01-1,5 1,51 - 2 2,01 - 2,5 2,51 - 3 3,01 - 3,5

Тоді запис вимірювань можна буде звести в таблицю (табл. 3.1).

У таблиці не обов'язково показувати сам інтервал, досить привести значення його середини.

Таблиця 3.1

Розподіл висоти сосни

Ступінь висоти, м	Кількість дерев, шт.
0,25	1
0,75	2
1,25	3
1,75	4
2,25	2
2,75	2
3,25	1
Всього	15

Дані, зведені в таблицю, дають більш наочне уявлення про висоту культур сосни на досліджуваній ділянці. До того ж обробку матеріалу легше проводити, коли маємо систематизовані дані. Тому саме така таблична форма найбільш часто використовується при проведенні досліджень в лісовому господарстві.

При упорядкуванні (систематизації) отриманих даних легко обробити їх математично і вивести статистичні показники, які вичерпно характеризуватимуть досліджувану сукупність. Проблема систематизації та групування займає велике місце в статистиці.

Помилкове угруповання даних може привести до неправильних висновків про суть досліджуваного явища.

Важливим показником статистичного ряду є його розмах. Це найбільш простий показник, який показує різницю між найбільшими і найменшими величинами (їх ще називають лімітами) в досліджуваному варіаційному ряду, тобто:

$$L = x_{max} - x_{min},$$

де L – розмах ряду,

x_{max}, x_{min} – максимальна і мінімальна величини (ліміти).

У багатьох дослідженнях в лісовому господарстві використовують даний показник. Так, в лісовій таксації при вивченні будови деревостану розмах ряду розподілу є обов'язковим.

У біометрії розмах ряду і ліміти визначаються при зведенні даних спостережень або вимірювань в статистичні сукупності, тобто в варіаційні ряди.

Однак для повного аналізу досліджуваного явища або для господарської оцінки деревостану цього недостатньо. Оскільки розмах ряду і ліміти схильні до значних коливань від однієї часткової сукупності до іншої. Тому використання цього показника обмежене. Отже, необхідно отримати ще й характеристики для сукупності, які були б виражені більш загальними цифровими показниками. З їх допомогою можна порівнювати різні ряди, що важко зробити за допомогою лімітів.

Наприклад, якщо відомо, що варіаційний ряд розподілених дерев в деревостані по товщині у одного насадження має розмах від 7 до 42 см, а в іншому від 12 до 53 см, то, здавалося б, можна зробити висновок про більш високу якість дерев другого насадження, тобто що у другому насадженні вони товщі. Однак ліміти не вказують на те, як розподіляються по досліджуваній ознаці, окремі члени сукупності. Ось чому для характеристики сукупності потрібні такі показники, які відображали б властивості всіх її членів.

Показник розмаху варіації (L) використовують при визначенні величини класового інтервалу (формули 3.1-3.3).

Існують два види рядів розподілу – безінтервальні та інтервальні.

Безінтервальні – це такі ряди, в яких значення ознаки (х) представлені лише конкретними цілими числами.

Наприклад. Розміри площ горизонтальних проекцій крон дерев (в м^2):

Площа, м^2 (х)	3	4	5	6	7	8	9	10	11
Кількість дерев, шт. (р)	1	0	2	3	5	2	1	0	1

Інтервальні ряди – це такі ряди, в яких значення ознаки, що досліджується, (х) представлено у відповідних межах (від ... і до ...), а в цих межах фіксуються відповідні варіанти (у). Наприклад, при вивченні значень діаметрів дерев слід враховувати:

Діаметри дерев, см (х)	6-8	9-11	12-14	15-17	18-20
Кількість дерев з такими діаметрами, шт. (р)	2	4	6	5	1

При складанні інтервальних рядів уся варіація ознаки від мінімальної (*min*) до максимальної (*max*) (*розмах розподілу*) ділиться на відповідні інтервали (у наведеному вище прикладі – 6-8, 9- 11 і т.д.). Ці інтервали мають ще назву – класи.

Значення ознаки, зведені в ряд, називають класовими варіантами, а число повторень їх у класах – численностями, або частотами класів. При вимірі дерев в лісі класи зазвичай називають ступенями товщини або ступенями висоти.

Статистичний ряд значень виміряної ознаки отримують шляхом визначення величини класу або інтервалу, розташування класів і розподілу в них всіх одиниць спостереження.

Кількість класів (і) визначається за формулами Г. А. Стерджеса (1926) (3.1) або за формулою К. Брукса і Н. Карузерса (1963) (3.2).

$$i = \frac{x_{\max} - x_{\min}}{1 + 3.32 \lg n}; \tag{3.1}$$

$$i = \frac{x_{\max} - x_{\min}}{5 \lg n}; \tag{3.2}$$

Пропонується ще одна формула для визначення величини класового інтервалу (k) (3.3):

$$k = \frac{x_{\max} - x_{\min}}{t}; \tag{3.3}$$

X max і *X min* – відповідно найбільше та найменше значення ознаки або варіант; t – число прийнятих класів.

Як k приймають кругле число, найближче до отриманого конкретного. При цьому дійсне число класів визначиться як частка від розмаху варіант (*X max – X min*) на округлене значення інтервалу. Як отриманого приватного приймають також кругле число.

Округлення робиться завжди в більшу сторону.

При проведенні біометричних досліджень оптимальною кількістю класів при наявності великої вибірки вважається 12.

Припустимо збільшувати або зменшувати цю кількість в залежності від обсягу спостережень.

Якщо ряд спостережень відносно невеликий (40-60 вимірів), а різниця між *Xmin і Xmax* невелика, то його доцільно звузити, тому що в іншому випадку ряд розподілу виявиться розмитим, в деяких класах може не виявитися виміряних величин.

Зазвичай рекомендоване число класів дорівнює 12 ± 3, тобто коливається від 9 до 15. У окремих випадках допустимо зменшити кількість класів до 8, як виняток – до 7. Менше і більшу кількість класів приймати не рекомендується, якщо це не пов'язано зі специфічними особливостями дослідження.

Найбільш просте угруповання при якісному аналізі. Так, якщо бук лісовий різниться за морфологічними формами, то розподіл дерев різних форм може бути виражено у відсотках від загальної кількості виміряних дерев, як це показано в таблиці 3.2.

Таблиця 3.2

Розподіл дерев бука лісового за морфологічними формами

Морфологічна форма	Кількість дерев, шт	Відсоток
Рожевооблямована ('Roseo-marginata')	36	18
Жовтолиста ('Zlatia')	39	19
Круглолиста ('Rotundifolia')	43	21
Плакуча ('Pendula')	45	22
Пурпуролиста ('Purpurea')	41	20
Всього	204	100

Окремим випадком якісної варіації є альтернативна, коли в сукупності можна виділити тільки дві групи. У членів однієї групи присутня певна якість (або ознака), у членів іншої групи її немає. Так, при дослідженні дубових культур, уражених кореневою губкою, ми ділимо дерева за альтернативною ознакою: здорові і хворі.

3.3. Складання рядів і таблиць розподілу

При проведенні спостережень у лісовому господарстві найчастіше мають справу з безперервним дискретним (перериvчастим) розподілом досліджуваної величини. Так, вимірюючи товщину дерева, ми міряємо кожне дерево. Їх сукупність являє собою деякий ряд розподілу, у якого є мінімальна і максимальна величина.

Для того, щоб надати дослідним матеріалам певну наочність і витягти з них необхідну статистичну інформацію про спостереження ознаки, матеріали спостереження піддають угрупованню.

Згруповані матеріали називають статистичними таблицями або статистичними рядами.

Ряди розподілу формуються шляхом *ранжирування* даних, тобто розміщення їх у зростаючому або спадаючому порядку величин їх значень.

Розглянемо наступний приклад. При вивченні особливостей будови крон дерев в робочий зошит занесені такі дані щодо розмірів площ їх горизонтальних проекцій (в м^2): 12, 8, 4, 4, 6, 8, 8, 7, 12, 10, 6, 6. В представленому вигляді – це ряд розподілу даних про розмір площ проекцій крон. Але якщо розмістимо наведені показники у зростаючому порядку, то будемо мати їх у вигляді ряду: 4, 4, 6, 6, 6, 7, 8, 8, 8, 10, 12, 12. Такий ряд зветься *ранжированим* рядом. Для більш зручного подальшого аналізу ранжирований ряд слід упорядкувати шляхом розміщення відповідно кожного розміру площі проекції крон (x) кількості дерев, які мали таку площу (p). Тоді цей ранжирований ряд буде мати вигляд:

Розмір варіанти	4	5	6	7	8	9	10	11	12
Кількість (повторюваність)	2	0	3	1	3	0	1	0	2

Такий упорядкований ранжирований ряд розподілу варіант зветься варіаційним рядом.

Варіаційний ряд – це подвійний ряд чисел, що складається з класових варіант і відповідних їм чисельностей, розмішених у порядку зростання або спадання.

Числа (2, 0, 3, 1, 3, 1, ... і т.д.), які показують скільки окремих варіант мають відповідний розмір (2, 3, 4, 5, ... і т.д.), звуться *вагою* або *частотами*, або *повторністю*. Частоти можуть бути представлені номінальними (іменними) значеннями або як відносна участь, тобто в % частоти від загальної кількості варіант, або в частках від одиниці.

Загальне число варіант, що входить в склад даного ряду розподілу або варіаційного ряду, зветься його *об'ємом*.

В процесі біометричних досліджень для позначення тих чи інших положень повинні застосовуватись чітко визначені символи. Ознака, яка вивчається, позначається однією з великих літер латинського алфавіту (X, Y, Z), кожна окрема варіанта в межах цієї ознаки позначається відповідною маленькою літерою (x, y, z).

Тобто, *наприклад*:

Довжина шишки ялини звичайної (X), а заміри довжин шишок окремих дерев ялини (x_1, $x_2 \ldots x_n$).

3.4. Прогнозування випадкової величини

Випадкові величини і їх прогнозування є основним об'єктом вивчення в біометрії. Прогнозування випадкових величин засноване на теорії ймовірності.

Теорія ймовірності – це одна з дисциплін математики. Детально її вивчають на математичних, фізичних і деяких технічних факультетах. Ми будемо розглядати лише деякі положення теорії ймовірності.

Одним з основних понять цієї теорії є імовірність. Імовірністю події А називають відношення числа випадків, що сприяють появі цієї події до числа всіх можливих випадків. Позначимо ймовірність буквою Р із зазначенням в дужках індексу події, в нашому випадку події А.

Її визначають за формулою (4):

$$P(A)=n/N; \tag{3.4}$$

де n – число випадків, що сприяють події А;

N – загальне число випадків.

Так, якщо в урні міститься 10 однакових перемішаних кульок, причому 4 з них чорні, а 6 – білі, то ймовірність вийняти навмання чорної кулі дорівнює Р (А) = 4/10 = 2/5 = 0,4, а ймовірність вийняти білу кулю Р (В) = 6/10 = 3/5 = 0,6.

Імовірність змінюється від нуля до одиниці. Імовірність, що дорівнює нулю, вказує, що подія є неможливою; імовірність, що дорівнює одиниці, означає, що подія є можливою або достовірною.

Якщо поява однієї події виключає появу іншої події, їх називають несумісними.

У зазначеному прикладі події А і В – несумісні. Сума ймовірностей несумісних подій дорівнює одиниці Р (А) + Р (В) = 1.

Події називають рівноможливими, якщо жодна з них не є більш можливою, ніж інша. Імовірності таких подій однакові.

У біологічних роботах ймовірність найчастіше встановити неможливо, так як вся досліджувана генеральна сукупність і її склад невідомі, наприклад, число дерев різної товщини на великій ділянці лісу. У таких випадках отримують аналог ймовірності на основі досвіду, тобто вибіркової сукупності. Для цього підраховують число випробувань, в яких подія практично з'явилася і відносять його до загального числа випробувань.

Це відношення називають відносною частотою події і виражають формулою (3.5):

$$W(A) = n / N \tag{3.5}$$

де n – число появ події;

N – загальне число випробувань.

Тривалі спостереження показали, що при однакових умовах випробувань і досить великому їх числі, відносна частота в різних дослідах змінюється мало, причому тим менше, чим більший об’єм вибірки. Вона коливається (варіює) близько деякого постійного числа. Це чудова властивість відносних частковостей називається стійкістю відносної частоти, або статистичною стійкістю.

Дуже характерний приклад пов'язаний з народженням людей різної статі.

Запитання для самоконтролю:

1. Що таке кількісний та якісний аналіз масових явищ? Поясніть їх принципи.
2. Що являє собою процес систематизації вихідних даних?
3. Що таке ранжирований і варіаційний ряди розподілу? Дайте визначення цим поняттям.
4. Дайте визначення поняттям «розмах варіації», «ліміти» варіації. Поясніть різницю між інтервальними і безінтервальними рядами.
5. Наведіть формули по визначенні кількості класів і їх величини.
6. Як складають ряди і таблиці розподілу?
7. В чому зміст прогнозування випадкової величини? Наведіть формули визначення імовірності події.

РОЗДІЛ 4
СЕРЕДНІ ВЕЛИЧИНИ В БІОМЕТРІЇ

4.1. Статистичні показники варіаційного ряду і їх класифікація

Варіаційні ряди можуть відрізнятися за значенням ознаки, навколо якого концентрується більшість варіант, тобто за величиною модального класу або ступеня. Значення цієї ознаки відображає центральну тенденцію, яка є типовою для даного ряду. Але частоти в ряду відрізняються за ступенем варіації, тобто за величиною відхилення від центральної тенденції ряду.

Відповідно до цього статистичні показники поділяються на дві групи: показники, які характеризують центральну тенденцію, або рівень ряду, і показники, що вимірюють ступінь варіації.

До першої групи належать різні середні величини: мода, медіана, середня арифметична, середня геометрична. До другої – варіаційний розмах (коефіцієнт варіації), середнє абсолютне відхилення, середнє квадратичне відхилення, або варіанса, дисперсія, коефіцієнти асиметрії та ексцесу.

4.2. Середні величини

Однією з найбільш важливих варіюючих ознак є середня величина їх значень. Вона характеризує не окремих представників біологічної сукупності, а всю сукупність в цілому, тобто вона характеризує групові властивості. Середня величина відображає внутрішній зв'язок, який існує між окремими варіантами і всієї їх сукупності. Середня величина є центром розподілу варіант, тобто окремих значень.

Середні значення, що застосовуються в біології, і зокрема в лісовому господарстві, діляться на ***степеневі***, або ***параметричні***, і середні ***порядкові,*** або ***непараметричні***.

Параметричні середні функціонально пов'язані із розподілом варіруючих ознак, тобто їх значення функціонально залежать від значень варіант, що представлені у ряді розподілу і визначаються з цих значень відповідними арифметичними діями.

Непараметричні (порядкові) середні функціонального зв'язку з розподілом варіант не мають. Вони лише характеризують структурні особливості цього розподілу (медіана, мода та інші).

В залежності від характеру біологічного об'єкту дослідження або його окремої ознаки для характеристики варіаційного ряду розподілу застосовуються різні види параметричних середніх – середня арифметична, середня квадратична, середня кубічна, середня гармонічна. Наведені види середніх відрізняються степенем, який застосовується в розрахунках по їх визначенню. Внаслідок цього вони ще мають назву ***степеневих***.

Загальна формула параметричних (або степеневих) середніх така:

$$\bar{x} = \sqrt[k]{\frac{\Sigma x^k}{n}}; \tag{4.1}$$

де: $\bar{x}$ − середня величина; x − значення окремої варіанти; Σ − знак підсумовування; n − об'єм сукупності, для якої вираховується середня; k − степінь, що визначає вид середнього значення.

Наприклад:

• при k = 1 одержуємо середню арифметичну (x);

• при k = –1 одержуємо середню гармонічну (x_h);

• при k = 2 одержуємо середню квадратичну (x_g);

• при k = 3 одержуємо середню кубічну (x_Q);

• при k = 4 і більше − одержуємо середню геометричну (x_q).

4.3. Середня арифметична

Середня арифметична (x) – це величина, яка є центром розподілу, навколо якої групуються всі варіанти статистичної сукупності.

Середня арифметична може бути **простою і виваженою**.

Середня арифметична (проста) ($\bar{x}$) – це частка від ділення суми всіх варіант сукупності на їх загальну кількість:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ...x_k}{n} = \frac{\Sigma x}{n} = \frac{1}{n}\Sigma x; \qquad (4.2)$$

Це і є загальна формула середньої арифметичної, де x_1 x_2, x_3, ..., x_k позначають варіанти, що входять до складу даної сукупності; *Σ* – знак підсумовування; *n* – загальне число варіант, або обсяг вибіркової сукупності.

Наприклад, для варіант (сходів сіянців бука, см) 6, 6, 7, 9, 9, 10 середня арифметична буде наступна $\bar{x}$ = (6+6+7+9+9+10)/6 = 47/6 = 7,8

Для ряду, розділеного на класи, тобто для варіаційного ряду, середню арифметичну обчислюють як зважену величину:

$$\bar{x}_в = (n_1 x_1 + n_2 x_2 + \ldots + n_n x_n)/(n_1 + n_2 + \ldots n_n) = (\Sigma n x_i)/N; \qquad (4.3)$$

де x_1, x_2, ..., x_n – класові варіанти (серединні значення класів); n_1, n_2, ..., n_n – частоти відповідних класів; N – загальне число варіант (обсяг ряду) або загальне число спостережень.

Тобто, якщо варіанти згрупувати у варіаційний ряд (приклад з сіянцями бука) то будемо мати:

x, довжина сіянців, см	6	7	9	10
частота, р	2	1	2	1

І відповідно:

$\bar{x}_в = (6 \cdot 2 + 7 \cdot 1 + 9 \cdot 2 + 10 \cdot 1)/6 = 7{,}8$ см

Середня арифметична, як і деякі інші середні, відома давно. Вона широко використовується при дослідженні сукупностей в науці, техніці, біології та лісовому господарстві.

Середня арифметична є узагальнюючою величиною, яка вбирає в себе всі особливості досліджуваної сукупності або ряду розподілу. Вона відображає рівень всієї сукупності в цілому, дає вільну, узагальнену характеристику досліджуваного ознаки.

Цифрове значення середньої арифметичної як таке може не зустрітися ні в одному конкретному випадку в сукупності. Може виявитися, що жодна варіанта не буде їй рівною. Наприклад, ми вимірюємо товщину дерев з округленням до 1 см: 8, 15, 16, 17 ... 30 см. Середня арифметична виміряної сукупності може скласти дробове число, наприклад, 24,8. У виміряної сукупності жодного такого виміру немає.

Або ж такий приклад. Якщо при наведенні статистики середнє значення прояву симптомів хвороби у людей було 28,4. Такого по відношенні до людей, тварин і навіть рослин не може бути.

В цьому сенсі середня арифметична є абстрактною величиною. але водночас вона і конкретна.

Середня арифметична виражається в тих же одиницях виміру, що і варіанти ряду. При її визначенні відхилення зі знаком (+) і (-) взаємно погашаються, відкидаються випадкові коливання, відхилення від центральної тенденції, від рівня варіаційного ряду і виступає загальний закон явища. Розкривається то типове, що характерне для всієї сукупності в цілому.

У той же час потрібно застерегти від можливих помилок в розумінні середньої арифметичної. Середня арифметична характеризує всю сукупність в цілому, а не окремі члени сукупності.

Слід пам'ятати, що середня арифметична має сенс тільки по відношенню до якісно однорідної сукупності. Так, не можна обчислювати середню вагу тварин різного віку або середній обсяг дерев різних деревних видів.

4.4. Інші види середніх величин

4.4.1. Середня геометрична

При вивченні середнього темпу зростання досліджуваного ознаки середня арифметична не придатна. Замість неї обчислюють середню геометричну M_q (або) X_q. Її визначаємо за формулою:

$$Xq = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\Pi x}; \qquad (4.4)$$

де $x_1, x_2, \ldots, x_n$ – темпи зростання (величини, що показують, у скільки разів збільшувалася ознака від періоду до періоду); n – число періодів.

При n > 2 формулу зручніше застосовувати в логарифмічному вигляді:

$$log\, x_q = \frac{\log x_1 + log\, x_2 + \cdots . + \log x_n}{n} = \frac{\Sigma\, log x_n}{n}; \qquad (4.5)$$

Якщо дані, для яких обчислюють середню геометричну, представлені різними численностями (n_i) в межах виділених класів (X_i), то застосовується формула:

$$lg\, x_q = (n_1\, lg\, x_1 + n_2\, lg\, x_2 + \ldots + n_n\, lg\, x_n)/ N; \qquad (4.6)$$

Застосування середньої геометричної можна пояснити наступним прикладом.

Нехай виміряли об'єм саджанця дуба в см^3 в момент посадки (1 рік), а також в 5, 10, 15 і 20 років. Нехай ці об'єми складуть відповідно 8, 560, 2800 і 11200 і 22400 см^3. Таке відношення об'ємів сіянців через рівні проміжки часу (5 років) будуть наступні:

$$x_1 = \frac{560}{8} = 70; \quad x_2 = \frac{2800}{560} = 5; \quad x_3 = \frac{11200}{2800} = 4; \quad x_4 = \frac{22400}{11200} = 2;$$

Число розглянутих періодів дорівнює 4, тобто N = 4.

За формулою (4.4) обчислимо середню геометричну:

$$Xq = \sqrt[4]{70 \cdot 5 \cdot 4 \cdot 2} = \sqrt[4]{2800} \approx 7{,}274;$$

Це означає, що об'єм дуба від 1-го до 20 років збільшується в середньому за кожен період в 7,274 рази.

4.4.2. Середня квадратична

Основна мета вимірювання діаметрів дерев в деревостані – це визначення запасу деревини. Для обчислення запасу треба знати суму площ поперечних перерізів виміряних дерев (g_i), а потім скористатися відповідною формулою, куди Σg_i входить як співмножник (більш детально вивчення таких питань розглядається при вивченні дисципліни «лісова таксація» на 3 курсі). Суму площ перетинів дерев в деревостані можна визначити, помноживши число дерев на площу перетину середнього дерева. Відомо, що площа перетину дерева на висоті 1,3 м прирівнюється до площі круга, і її знаходять за формулою $g_i = \pi d_i^2/ 4$, де $\pi = 3{,}14159$, d_i – діаметр на висоті 1,3 метра. Нам необхідно з'ясувати, яку величину має представляти середній діаметр, щоб він був репрезентативним показником для досягнення зазначеної мети.

Застосування середньої арифметичної величини в цьому випадку неприйнятно, а необхідно обчислити середню квадратичну. Її знаходять за формулою:

$$x_{кв} = \sqrt{\frac{1}{n}\Sigma x_i^2 n_i}; \tag{4.7}$$

4.4.3. Середня гармонійна

Ця величина застосовується для обчислення середньої характеристики ознак, які представляють собою відношення двох інших варіюючих величин. Середню гармонійну визначають за формулою:

$$\bar{x}_h = \frac{N}{\sum\frac{1}{x}} \tag{4.8}$$

або

$$\bar{x}_h = \frac{N}{\sum\frac{n}{x}}; \tag{4.9}$$

де n – вага окремих значень.

У практиці лісового господарства ця величина застосовується дуже рідко.

4.5. Структурні (непараметричні) середні величини – медіана і мода

До непараметричних середніх величин відносять медіану і моду.

Медіана (Ме) – варіанта, яка ділить ряд розподілу наполовину; в обидві сторони від медіани розташовується однакова кількість варіант.

$$Me = x_{\min} + i\frac{0{,}5N - \sum n}{n_e} \tag{4.10}$$

де x_{min} – нижня межа інтервалу, який містить медіану; i – величина інтервалу; N – чисельність вибіркової сукупності; Σn – сумарна чисельність до інтервалу, в якому знаходиться серединне значення; n_e – чисельність інтервалу, де знаходиться серединне значення.

Мода (Мо) – варіанта, яка найчастіше зустрічається у даній сукупності. Визначити моду можна за наступною формулою:

$$Mo = X_{min} + i\ (n_2 - n_1/2n_2 - n_1 - n_3); \qquad (4.11)$$

де X_{min} – нижня межа модального інтервалу; i – величина інтервалу; n_1 – чисельність інтервалу, що знаходиться перед модальним; n_2 – чисельність модального інтервалу; n_3 – чисельність інтервалу, що знаходиться після модального.

Медіану зазвичай застосовують в тих випадках, коли потрібно забезпечити мінімум абсолютної величини відхилень між вихідними даними і використовуваним показником. Наприклад, логічно вимагати дотримання такої умови для показника, що відображає середню тривалість життя дерев у різновіковому лісі. Мода представляє інтерес, коли встановлюють деяку типову властивість явища чи об'єкта: колір жолудів, врожайність насіння на плюсових деревах і т. д.

Розглянемо як саме визначати моду і медіану на наступному прикладі.

x	4	5	6	7	8	9	10	11	12	13	14	15	16
p	15	38	55	90	115	153	131	135	114	82	42	15	15

Довжина листків дуба (см).

Приклад визначення медіани:

1. Визначаємо, що медіана знаходиться в класі 10, варіанта, яка ділить ряд розподілу наполовину; тобто зліва знаходиться від цього класу 466 варіант, а справа 403 варіанти.

2. За характером класів (x) бачимо, що інтервал класу дорівнює 1, нижня границя класу, де знаходиться медіана – 9,5.

3. Ряд накопичення частот до класу, де знаходиться Ме, закінчується числом: 15 + 38 + 55 + 90 + 115 + 153 = 466.

4. Частота класу, де знаходиться медіана – 131.

5. За формулою 4.10 знаходимо значення медіани.

$$Me = x_{\min} + i\frac{0{,}5N - \sum n}{n_e} = 9{,}5 + 1\ \frac{0{,}5\times1000-466}{131} = 9{,}76;$$

Приклад визначення моди:

1. Визначаємо, що мода знаходиться в класі 9, так як ця варіанта найчастіше зустрічається у даній сукупності.

2. За характером класів (х) бачимо, що інтервал класу дорівнює 1, нижня границя класу, де знаходиться мода – 8,5.

3. Чисельність інтервалу, що знаходиться перед модальним – 115.

4. Чисельність модального інтервалу – 153.

5. Чисельність інтервалу, що знаходиться після модального – 131.

6. За формулою 4.11 знаходимо значення моди.

$Mo = X_{min} + i\ (n_2 - n_1/2n_2 - n_1 - n_3)$ = 8,5+1((153-115)/2·153 – 115-131) = 9,1

Зазвичай мода, так як і медіана є величиною, близькою до середньої арифметичної і навіть співпадати з нею при повному симетричному розподілі. Тому для симетричних або ж слабо скошених розподілів мода визначається за формулою 4.12 або 4.13.

$$Mo = 3\bar{x} - 2Me \qquad (4.12)$$

$$Mo = \bar{x} - 3\ (\bar{x} - Me) \qquad (4.13)$$

Мода і медіана – є допоміжними характеристиками варіаційного ряду. В дослідженнях, для аналізу використовуються рідко.

4.6. Способи визначення середньої арифметичної

Існують спрощені способи, що дозволяють швидко і точно визначати середню величину. В даній роботі будуть розглянуті наступні: спосіб умовної середньої і спосіб підсумування, визначення середнього значення з декількох середніх та їх помилок, визначення добутків, часток і різниці між середніми значеннями з врахуванням їх помилок.

Спосіб умовної середньої. Цей спосіб полягає в тому, що значення однієї з варіант (будь-якої) приймається умовно за середню величину варіант даного ряду розподілу. Від цієї варіанти (A) вліво і вправо від неї (або догори і вниз) по ряду розподілу обчислюється відхилення (різниця) значення кожної варіанти від значення обраної варіанти. Якщо значення варіант будуть розміщені у наростаючому порядку, то результати, що розмістяться ліворуч від умовного середнього будуть мати знак "мінус", а праворуч − "плюс". Між значеннями різниці (a) і частотою (p) кожної варіанти (x) визначаються добутки (pa), які також будуть зі знаками "мінус" і "плюс". Ці добутки підсумовуються (Σpa).

Після цього знаходиться середнє арифметичне значення (x) як:

$$\bar{x} = A + \frac{\Sigma pa}{n}; \tag{4.14}$$

де A – умовна середня, a − різниця між сумами першого і другого неповних рядів накопичених частот одержаних кумуляцією частот з протилежних кінців варіаційного ряду до умовної середньої (A), n − загальне число варіант в даній сукупності.

Спосіб підсумування. Цей спосіб полягає в наступному. З протилежних кінців ряду розподілу до умовної середньої ($A_{ум}$) здійснюється підсумування частот (p). Ці дві частоти варіаційного ряду звуться неповними рядами накопичення частот.

$$\bar{x} = A_{ум} + \frac{d}{n}; \tag{4.15}$$

де d – сума неповних рядів накопичення частот;

Визначення середнього значення з декількох середніх та їх помилок. В окремих видах дослідів виникає необхідність визначити середнє значення з двох або більшої кількості середніх, які мають свої помилки репрезентативності. В таких випадках середнє значення визначається звичайним шляхом, тобто сумування значень всіх середніх і діленням одержаної суми на їх кількість. Узагальнена (середня) помилка знаходиться за формулою:

$$m_x = \frac{1}{n_i}\sqrt{m_1{}^2 + m_2{}^2 + m_3{}^2 + m_n{}^2}; \qquad (4.16)$$

Визначення добутків, часток і різниці між середніми значеннями з врахуванням їх помилок. В окремих випадках є необхідність визначити *добуток середніх значень* з їх помилками (10). Для цього застосовується звичайне множення значень середніх.

$$\bar{x}_1 \pm m_1 \times \bar{x}_2 \pm m_2 = \bar{x}_1 \times \bar{x}_2 \pm \bar{x}_1 \times \bar{x}_2 \sqrt{\left(\frac{m_1}{x_1}\right)^2 + \left(\frac{m_2}{x_2}\right)^2} \qquad (4.17)$$

Помилка одержаного добутку визначається як:

$$m_n = \bar{x}_1 \times \bar{x}_2 \sqrt{\left(\frac{m_1}{x_1}\right)^2 + \left(\frac{m_2}{x_2}\right)^2}; \qquad (4.18)$$

Для визначення частки від ділення середніх значень з врахуванням їх помилок застосовується формула:

$$\frac{\bar{x}_1 \pm m_1}{\bar{x}_2 \pm m_2} = \frac{\bar{x}_1}{\bar{x}_2} \pm \frac{\bar{x}_1}{\bar{x}_2}\sqrt{\left(\frac{m_1}{x_1}\right)^2 + \left(\frac{m_2}{x_2}\right)^2}; \qquad (4.19)$$

Різниця між середніми величинами (x_1–x_2) з урахуванням їх помилок визначається так:

$$x_1 \pm m_1 - x_2 \pm m_2 = [\![x_1 - x_2]\!] \pm \sqrt{m_1^2 + m_2^2}; \qquad (4.20)$$

Узагальнюючи викладене, зазначимо, що в лісовій справі найбільше застосування знаходять середня арифметична і середня квадратична. Решта середніх застосовують зрідка.

Запитання для самоконтролю:

1. Які середні величини ви знаєте?

2. Наведіть формули середніх: арифметичної простої і виваженої, геометричної, квадратичної. Коли вони використовуються?

3. Що таке мода і медіана?

4. Які способи визначення середньої арифметичної ви знаєте?

5. В чому суть способу умовної середньої?

РОЗДІЛ 5
ПОКАЗНИКИ ВАРІАЦІЇ

5.1. Варіація як явище і її джерела

Ряд розподілу характеризується кількома параметрами. Одним з них є розглянуті нами середні значення. Вони характеризують, то значення варіаційного ряду, щодо якого в певному сенсі розташовуються всі інші чисельні значення, властиві даному ряду розподілу випадкової величини.

У той же час існують інші важливі характеристики рядів випадкових величин. У їх числі однією з основних є розмах (діапазон) випадкових величин від мінімальної до максимальної. Ця величина характеризує варіабельність, або мінливість, ряду розподілу.

Щоб дати більш повну характеристику рядів розподілу випадкової величини, введена характеристика їх варіації, або мінливості, яка характеризує ступінь розсіювання випадкової величини щодо середнього значення. Для вираження варіації використовують спеціальні величини: коефіцієнт варіації, середнє квадратичне відхилення, дисперсію і інші.

Основна причина варіації полягає в тому, що за своєю природою будь-які біологічні об'єкти, навіть якщо належать до однорідної сукупності, відрізняються один від одного. Дерева одного виду, одного віку, що ростуть в однакових умовах зростання мають різну висоту і товщину, що викликано як генетичними властивостями кожної особини, так і деякими особливостями їх територіального розміщення. Подібні приклади можна привести для будь-яких біологічних об'єктів: люди навіть однієї раси і національності відрізняються по росту, тварини одного виду мають різну вагу і т. д. Основна причина цього, як вже було зазначено, – наявність біологічної (в першу чергу генетичної) різноманітності.

Мінливість відбувається і через помилки вимірювань. Домогтися абсолютної точності вимірювань дуже важко (і дорого), а часто і не потрібно.

Практику влаштовує деякий рівень точності, його-то і витримують.

Наприклад, при обліку дерев їх вимірюють за ступенями товщини. Точність вимірювань може підвищуватися, якщо це буде потрібно в окремих випадках. Так, в умовах ринку підвищуються вимоги до точності оцінки об’єму деревини, особливо для її найціннішої частини. При проведенні вимірювань крім допустимих відхилень в вимірах можливі і помилки, особливо при масових вимірах. Найбільш розповсюджені з них це:

• помилки інструментальні або систематичні;

• помилки випадкові.

Групу помилок *інструментальних* складають помилки, які мають місце внаслідок несправності інструмента (ваги, штангенциркуль, анемометр, висотомір та ін.). Усунення таких помилок полягає в ретельній перевірці приладів за правилами стандартизації і внесенням відповідних поправок до показників приладів.

Випадкові помилки мають місце в епізодичних проявах, коли, наприклад, випадково вимірювальний прилад встановлений невірно, або коли одержаний невірний результат внаслідок якоїсь випадкової події, що порушила методику спостереження.

Випадкові помилки можуть бути *поодинокими* і *систематичними*.

Поодинокі випадкові помилки трапляються під впливом причини, яка швидко зникає. Одержаний в такому випадку результат різко відрізняється від інших результатів, його легко виявити і усунути від подальшого опрацювання. Систематична випадкова помилка має місце тоді, коли причина, що її викликає, діє постійно на всьому періоді досліджень. Виявити таку помилку по результатах одного дослідження дуже важко, оскільки всі результати спостереження будуть мати помилку одного розміру і знаку. Тому, щоб уникнути таких випадків спостереження роблять в декількох повторностях.

Слід мати на увазі те, що жоден показник об'єкту, що вимірюється, не можна за допомогою будь-якого приладу визначити з абсолютною точністю. Так, яким би точним не був штангенциркуль, він не дасть абсолютно точний розмір діаметра, що вимірюється, якби не була відрегульована вага, вона не дасть абсолютно точне значення маси наважки і таке інше. В зв'язку з цим існує поняття точність виміру.

Точність виміру − це ступінь відповідності одержаного результату справжньому значенню параметру, що вимірюється. Вона визначає надійність дослідження, його досконалість. Кожне нове вимірювання того ж самого параметру і тим самим приладом дає свій результат з відповідним відхиленням одержаного значення від справжнього. Це відхилення зветься абсолютною похибкою (помилкою) (Δ). Отже, *абсолютна* *помилка* − це різниця між справжнім значенням величини, що вимірюється (a_c), і одержаним результатом виміру (а): $\Delta = \mathbf{a_c - a}$.

Відрізняють також відносну помилку (Δ_i) як:

$$\Delta_i = \pm\frac{\Delta}{a_c}\,100\%; \qquad (5.1)$$

Наведена особливість, яка полягає в тому, що жоден прилад не може дати результат виміру, який абсолютно точно відповідає справжньому значенню ознаки, обумовила встановлення для всіх вимірювальних приладів так званої граничної помилки. Гранична ***помилка*** − це межі, в яких точність функціонального приладу вважається допустимою для його використання в дослідженнях. В межах стандартних граничних помилок ($\Delta_г$) знаходиться об'єктивно існуюче значення показника, що вимірюється, тобто

$$x = a \pm \Delta_г; \qquad (5.2)$$

де x − об'єктивно існуюча величина ознаки, що вимірюється; а − показник величини ознаки, який дав прилад; $\Delta_г$ − гранична помилка приладу.

Зменшенню діапазону граничної помилки також сприяє проведення вимірів в декількох повторностях.

Хоча помилки вимірювань і є однією з причин варіювання, але ця причина не така велика як природна біологічна мінливість.

5.2. Типи варіювання

Отримані в результаті спостережень значення досліджуваної ознаки називають **варіантами**. Варіанти в біологічних об'єктів вказують на різноманітність (або варіювання) досліджуваної властивості. Наприклад, дерева відрізняються один від одного по діаметру, висоті, об’єму, санітарному стану.

Залежно від характеру досліджуваної ознаки розрізняють варіювання *безперервне, переривчасте (дискретне) і атрибутне*. Безперервне і дискретне варіювання притаманне кількісним ознакам, а атрибутне – якісним.

При ***безперервному*** варіюванні окремі значення ознаки виражають мірою протяжності, об'єму і т. д. Окремі варіанти можуть мати будь-яке значення міри, яке змінюється в певних межах.

Товщина дерев в деревостані, наприклад, від самого тонкого до самого товстого може приймати самі різні значення міри протяжності. Тільки в залежності від мети дослідження (вимірювання) виражають її в класах товщини: в кілька сантиметрів, в цілих сантиметрах, в десятих або сотих частках сантиметра.

При ***дискретному*** варіюванні окремі значення ознаки виражають абстрактними числами, найчастіше цілими. Наприклад, число сходів сосни на обліковій ділянці має дискретне варіювання, тому що вони, так само як і число насіння в наважці, виражаються цілими числами.

При ***атрибутному*** варіюванні значення ознаки виражають в якісних показниках. Це може бути ступінь забарвлення, консистенції, пошкодженість або стійкість, а також форма, вид і т.д. Кількісно ці ознаки виражають в абсолютних числах, частках одиниці, відсотках, балах і т. д. Наприклад, розрізняють колір кори на деревах, форму крони дерев (куляста, пірамідальна і т.д.), густота розчину, ступінь пошкодження дерев шкідниками: сильна, слабка та ін.

Окремим випадком атрибутивного варіювання є альтернативне, при якому значення ознаки розглядають в альтернативній формі, тобто протиставляючи здорові хворим, сильні – слабким, пофарбовані – незабарвленим, присутні – відсутнім і т. д. В альтернативній формі можна уявити і кількісні ознаки, протиставляючи, наприклад, високі дерева в деревостані низьким, панівні дерева – пригніченим, здорові – сухим або всихаючим.

5.3. Характеристики варіаційних рядів і їх обчислення. Межі і розмах варіації

Основними показниками варіаційного ряду крім середнього значення є абсолютне і відносне значення його меж, яке виражається величиною дисперсії і коефіцієнта варіації.

Одним з показників амплітуди варіації служать так звані ліміти (від лат. *Limes* – межа, кордон), тобто значення мінімальної і максимальної варіант вибіркової сукупності. Цей показник (*Lim*) вказує фактичні межі варіабельності ознаки. Тому його зазвичай вказують поряд з іншими біометричними показниками в зведених статистичних таблицях. Значення лімітів полягає в їх конкретності.

Величина варіації може бути оцінена і по різниці між максимальною і мінімальною варіантами сукупності. Цей показник отримав назву ***розмаху варіації***. Наприклад, межі першого розподілу (n_1), рівні: min = 8 і max = 60 одиницям, звідки розмах цього ряду дорівнює 60 - 8 = 52. Другий ряд (n_2) має межі варіації від 16 до 52 одиниць, його розмах дорівнює 52-16 = 36 одиниць. Найнижчим цей показник виявляється в третьому ряду (n_3), розмах якого дорівнює 24 одиницям: 44-20 = 24.

Розглянуті показники варіації цілком об'єктивні і прості.

Але в силу притаманних їм недоліків вони мало придатні для вимірювання варіабельності ознак. Справа в тому, що ці показники нестійкі: вони залежать від багатьох випадкових причин і при повторних вибірках можуть різко змінювати своє значення. Головний же недолік зазначених показників полягає в тому, що вони не відображають істотні риси варіювання.

Зі сказаного випливає, що ліміти і межі варіації, хоча і дають певне, конкретне уявлення про величину мінливості ознак, не можуть служити основним мірилом варіабельності біологічних величин.

5.4. Середнє лінійне відхилення

Для вимірювання варіації можна використовувати центральний момент першого порядку як одну з характеристик варіаційного ряду, що представляє суму відхилень варіант від середньої арифметичної, віднесену до загальної кількості варіант даної сукупності.

$$\Delta = \frac{\Sigma\, \mathrm{I}(x_i - x)\mathrm{I}}{n}; \qquad (5.3)$$

Цей показник, що називають середнім лінійним відхиленням, може мати значення тільки за умови, що відхилення варіант від середньої арифметичної беруться без урахування знаків, так як в противному випадку $\sum(x - x_{сер.}) = 0$.

5.5. Середнє квадратичне відхилення

Щоб подолати недоліки лінійного відхилення, прийнято відхилення варіант від середньої арифметичної зводити в квадрат і суму квадратів відхилень відносити до загальної кількості спостережень, тобто до обсягу вибірки. Отриманий таким чином показник служить *центральним моментом* другого порядку, він характеризує ***дисперсію або розсіювання варіант біля середньої арифметичної***. Цей показник, називається дисперсією або варіансою і позначається грецькою буквою σ^2.

Дисперсія (σ^2) – середній квадрат відхилень варіант від середньої величини, який вказує на характер розсіювання числових значень ознак розподілу чисельностей, а також описує мінливість варіант відносно середнього значення.

Стандартне відхилення (σ) – показник, який характеризує ступінь розсіювання варіант навколо середнього значення, описує криву нормального розподілу і дає уявлення про найбільш ймовірну середню помилку окремого спостереження даної сукупності.

Стандартне (середньоквадратичне) відхилення виражається наступною формулою:

$$S_x(\sigma_x) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}\ ; \tag{5.4}$$

де Σ – знак підсумовування добутків відхилень варіант хі від їх середньої х;

n – загальне число спостережень, або обсяг вибірки.

(n -1) – число ступенів свободи.

Величина (N-1) називається числом ступенів свободи. Вона показує, що в обмеженої сукупності (а будь-яка вибірка завжди має обмежений об’єм) всі варіанти вільні приймати будь-які значення, крім однієї, значення якої визначається різницею між сумою всіх інших варіант і об’ємом вибірки.

Середнє квадратичне відхилення, як і середня арифметична виражається в тих же величинах, що і ознака. Вибірка, в якій розсіювання варіант біля середньої арифметичної більше, характеризується і більшою величиною середнього квадратичного відхилення і, навпаки, при меншій варіабельності ознаки середнє відхилення виявляється меншим.

У лісовому господарстві найчастіше використовують σ замість σ^2. Перевага середнього квадратичного відхилення проти дисперсії пояснюється практичною зручністю: у разі використання σ ми маємо міру розсіювання, виражену в тих же величинах, що і середнє значення.

У порівнянні із середнім лінійним відхиленням середньоквадратичне відхилення більш точно характеризує варіабельність ознак.

5.6. Способи обчислення середньоквадратичного відхилення

Способів обчислення середньоквадратичного відхилення варіаційного ряду як і інших його показників (статистик) є кілька. В даний час для обчислення статистик розроблені комп'ютерні програми.

Вони відразу дають шуканий результат, але дослідник, який працює з варіаційним рядом, повинен розуміти суть досліджуваного явища. Тому необхідно знати алгоритм проведених обчислень, що ми зараз і розглянемо.

Одним з найбільш простих способів є, так званий, прямий або довгий. Його раціонально використовувати для невеликого числа спостережень, які не згруповані в варіаційний ряд. Робота виконується в такий спосіб. Після того як обчислена середня арифметична, потрібно визначити відхилення від неї кожної варіанти, тобто знайти значення, $x_1 - \tilde{x}$, $x_2 - \tilde{x}$, $x_3 - \tilde{x}$ і т. д. Потім кожне відхилення зводиться в квадрат і, якщо варіанти повторюються, квадрати відхилень множаться на відповідні частоти, і результати підсумовуються. Отримана сума ділиться на загальне число спостережень без одиниці (n-1), і з отриманого числа знаходять квадратний корінь.

5.7. Моменти розподілу варіаційного ряду

Момент розподілу варіаційного ряду є середні величини відхилень варіант ряду від будь-якого числа. Для визначення моменту розподілу можуть бути застосовані: значення середньої арифметичної ряду розподілу (х), усяке довільне число (А) або 0 (нуль). В залежності від того, яке значення обрано для порівняння, моменти розподілу розрізняють на: 1. Центральний момент розподілу (М) − коли для порівняння значень варіант застосовується значення середнього арифметичного цього ряду розподілу. Цей момент розподілу позначається літерою М і визначається за формулою:

$$\text{М} = \frac{\sum p\ (\text{x} - \tilde{x})}{\text{n}}; \tag{5.5}$$

2. Умовний момент розподілу, коли для його визначення застосовується будь-яке число. Цей момент розподілу позначається літерою b і визначається за формулою:

$$\text{b} = \frac{\sum p\ (\text{x} - \text{A})}{\text{n}} = \frac{\sum p\text{a}}{n}; \tag{5.6}$$

3. Початковий момент розподілу, коли для його визначення застосовується 0 (нуль). Цей момент позначається літерою m і визначається за формулою:

$$\text{М} = \frac{\sum p\ (\text{x} - 0)}{\text{n}} = \frac{\sum p\text{x}}{n}; \tag{5.7}$$

Як бачимо, початковий момент розподілу є одним з варіантів умовного моменту. Для характеристики окремих особливостей рядів розподілу використовуються моменти розподілу, піднесені до різних степенів. Тоді відповідно вони звуться: момент першого порядку, момент другого порядку, момент третього порядку і т. д.

Коефіцієнт варіації (v_x) – відносна величина, яка вказує на ступінь мінливості ознаки варіаційного ряду.

$$v_x = \frac{\sigma_x}{\bar{x}} * 100\%; \tag{5.8}$$

при v_x < 10% – варіація незначна;

при v_x - 10 - 20% – варіація середня;

при v_x > 20% – варіація значна.

Запитання для самоконтролю:

1. Які показники варіації ви знаєте?
2. Які причини варіації біометричної ознаки?
3. При проведенні вимірювань які можливі помилки? Охарактеризуйте їх.

4. Охарактеризуйте типи варіювання.
5. Дайте визначення поняттям «середнє лінійне відхилення», «середнє квадратичне відхилення». Наведіть формули.
6. Що таке моменти розподілу варіаційного ряду? Які моменти розподілу ви знаєте?
7. Що таке асиметрія? Наведіть формули.
8. Що таке ексцес? Наведіть формули.

РОЗДІЛ 6
ФУНКЦІЇ РОЗПОДІЛУ. НОРМАЛЬНИЙ РОЗПОДІЛ. АСИМЕТРИЧНІ РЯДИ

6.1. Поняття про види розподілу

Масиви випадкових величин зазвичай розподіляються не хаотично, а по деякому закону. Ці закони розподілу призводять до різних видів розподілів. Всі розподіли можна виразити графічно, що в основному і робили в лісовому господарстві до 50-60 років XX століття. У той же час найбільш коректно висловлювати розподіл через деякі математичні функції. В даний час це стало основним методом опису розподілів, тому що застосування комп'ютерів зробило таку роботу простою, швидкою і доступною. У той же час ми повинні ясно розуміти суть досліджуваного явища. Цього не досягти, якщо дослідник використовує комп'ютер тільки як користувач.

Розподіли можуть бути **дискретними і безперервними.** Так, якщо ми розглядаємо деякий розподіл, наприклад, дерева в деревостані, який виражається конкретними числами: 4, 8, 12, ..., то він буде дискретним. При описі розподілу безперервною функцією розподіл стає безперервним. Часто між ними важко провести межу: все залежить від мети дослідження, величини класового проміжку і т. д.

Все, що може бути виміряне або перелічено в живій природі, називають величиною постійною або змінною. В залежності від обставин ці величини можуть приймати різні значення. Змінну величину вважають визначеною, якщо заздалегідь, до досліду можна вказати її значення. Якщо ж в одних і тих же умовах змінна величина може приймати різні значення, які заздалегідь вказати не можна, вона називається випадковою величиною. Поняття випадкової величини, як і поняття випадкової події, відноситься до фундаментальних в теорії ймовірностей.

Випадкові величини бувають залежними і незалежними. Випадкові величини називають **незалежними**, якщо ймовірність будь-якого значення однієї величини (X) не залежить від того, які значення приймає інша величина (У). В іншому випадку ці величини знаходяться в залежності одна від іншої і називаються **взаємозалежними** випадковими величинами. Наприклад, ми вивчаємо деяку ділянку лісу. Там є певний грунт, на якому ростуть дерева. На одному і тому ж ґрунті можуть рости різні дерева: сосна, береза, ялина та інші. Механічний склад ґрунту не залежить від того, який деревний вид сьогодні тут росте. Таким чином при розгляді системи ґрунт-дерево характеристики ґрунту, скажімо, відсоток фізичної глини, буде величиною **незалежною**. У той же час висота дерева, та й сам породний склад лісової ділянки визначається ґрунтовими умовами, тобто показники росту дерев залежать від ґрунтових характеристик і є величиною **залежною**. Якщо ми будемо розглядати діаметр і висоту дерева, то виявимо, що зі зміною одного показника змінюється і другий. Це величини **взаємозалежні.**

Ще один приклад. Такі показники, як висота, продуктивність рослин залежить від кількості опадів. Але кількість опадів не залежить від висоти чи продуктивності рослин.

Існують різні типи випадкових величин. Для лісівника і біолога найбільш істотне значення мають дискретні і безперервні випадкові величини. *Дискретна випадкова величина* приймає лише окремі рахункові значення, для яких можна вказати ймовірності, або частоти. *Безперервна випадкова величина* може приймати будь-які значення в деякому заданому інтервалі. Вказати ймовірності або частоти її окремих значень неможливо, тому вони відносяться до тих значень, які ця величина приймає в інтервалі (від – до), причому цей інтервал може бути яким завгодно – і великим і малим.

6.2. Емпіричні функції розподілу

У лісовій біометрії ми, як правило, маємо справу з емпіричними функціями розподілу. Це означає, що, виконуючи деякі вимірювання випадкової величини, наприклад, діаметри дерев в лісі, отримуємо розподіл і визначаємо вид цього розподілу, керуючись графіком деякої функції розподілу.

Дуже важливо знати розподіли ознак, що найбільш часто зустрічаються в лісових дослідженнях. Це дозволяє застосовувати різні способи обробки вибіркових даних не сліпо, а продумано, а також правильно оцінювати результати дослідів і спостережень, об'єктивно і точно порівнювати їх між собою.

В емпіричних розподілах, тобто тих, які спостерігаємо, проводячи вимірювання в лісі, кидається в очі одна важлива особливість – переважаюче накопичення варіант в центральних класах і поступове спадання їх числа в міру віддалення від середньої арифметичної варіаційного ряду. Ця особливість, яка становить одну з характерних рис варіювання біологічних ознак взагалі і лісогосподарських зокрема, факт фундаментального значення, що має широке поширення в природі.

Таку картину отримуємо, якщо проаналізуємо розподіл людей по росту, диких тварин, наприклад зайців, по вазі, величину ступні у дорослих людей однієї статі і т. д.

Таку закономірність, тобто концентрацію найбільших численностей в середині ряду розподілу, вперше описав бельгійський статистик А. Кетле в 1835 році, який досліджував розподіл кількох тисяч солдатів американської армії по росту.

Дерева в лісі розподіляються по товщині, висоті та іншими ознаками аналогічно. Якщо побудуємо графік за даними замірів стовбурів дерев, де по горизонталі відкласти ступені товщини, а по вертикалі – кількість дерев, то можна відзначити, що отримана крива піднімається від тонких ступенів до середніх, досягає тут максимуму, а потім знову знижується. Такий розподіл дерев за товщиною характеризується симетричною, одновершинною лінією, яку називають **кривою нормального розподілу.** Ця загальна закономірність свідчить про те, що найбільша кількість дерев припадає на середні ступені товщини, а менша – на крайні, тобто найтонші і найтовстіші.

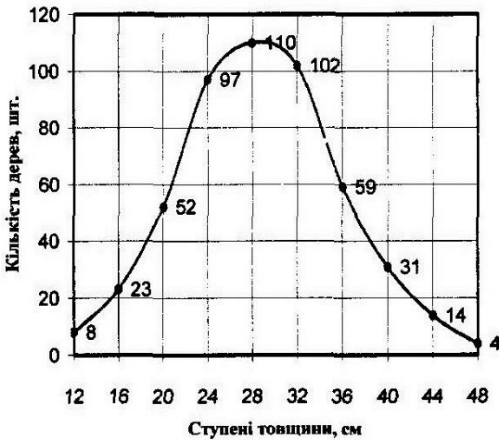


Рис. 6.1. Крива нормального розподілу

Такий розподіл дерев характерний для нормальних насаджень, які складаються з одного елементу лісу, створених висіванням або садінням і мали однаковий догляд до змикання крон.

6.3. Функція нормального розподілу і її параметри

Вище показано, що розподіл випадкових величин в біологічних, в тому числі і лісових, сукупностях носить закономірний характер. Є багато функцій, якими описують названі розподіли. Найбільш універсальним і використовуваним частіше інших є рівняння кривої нормального розподілу або функція Гауса-Лапласа. Її суть полягає в тому, що частота відхилення окремих варіант від середньої арифметичної даної сукупності є функція їх величини. Імовірність частоти тієї чи іншої варіанти в генеральній сукупності і визначається цією функцією.

Графічно функція нормального розподілу схожа з графіками на рисунку 1. У той же час графік, який строго відповідає рівнянню кривої нормального розподілу, виглядає наступним чином (рисунок 6.2).

Рівняння кривої нормального розподілу вказує залежність теоретичних чисельностей f(x) або y від значень x, і є випадковою величиною, що безперервно розподіляється.

Є різні форми вираження цієї кривої. Основна форма написання цього виразу:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad (6.1)$$

де **f(x)** – ордината або висота кривої на будь-якій відстані від $\bar{X}$, тобто центру розподілу. Вправо від цього центру випадкова величина x має позитивні, а вліво – негативні значення.

σ або середньоквадратичне відхилення характеризує амплітуду коливання окремих значень випадкової величини біля середньої арифметичної;

(x - a) – відхилення варіанти від середньої арифметичної, що є серединою ряду;

Вираз $\frac{\mathbf{1}}{\boldsymbol{\sigma\sqrt{2\pi}}}$ – максимальна ордината, що відповідає точці $\bar{X}$. У міру віддалення від цієї точки, тобто центру розподілу, щільність значень випадкової величини падає і крива асимптотично наближається до осі абсцис;

Так як $\boldsymbol{\pi}$ **– 3,14593** і **е** – основа натуральних логарифмів, рівне **2,7183**, є постійними величинами, отже величина $\frac{\text{x}-\text{a}}{\boldsymbol{\sigma}} = \boldsymbol{t}$ є не що інше як нормоване відхилення. Ця величина має велике значення для дослідження властивостей нормального розподілу.

Знайдені для різних значень t величини f(x) дають ординати нормальної кривої. Неодмінною умовою нормування служить вимога, щоб вся площа, яка знаходиться під кривою ймовірності (нормальної кривої), дорівнювала одиниці.

Якщо прийняти σ = 1, то рівняння (6.1) буде мати наступний вигляд:

$$f(x) = \frac{1}{2\pi}. e^{-\frac{1}{2}(x-a)^2}; \qquad (6.2)$$

Крива, що описується цим рівнянням, отримала назву стандартизованої кривої розподілу, або кривої Гауса-Лапласа. Вона відображає закон нормального розподілу з площею під кривою, що дорівнює одиниці.

Щоб ордината f(x) відображала не імовірні, а абсолютні чисельності, тобто частоти варіант, потрібно в чисельник правої частини рівняння (6.1) ввести в якості додаткових множників **N** – загальне число варіант даної сукупності і **і** – величину класового інтервалу (якщо варіація розбита на класи).

Тоді рівняння (6.1) приймає наступний вигляд:

$$f(x) = \frac{N \cdot i}{\sigma\sqrt{2\pi}}. e^{-\frac{1}{2}(t)^2}; \qquad (6.3)$$

Тут f(x) позначає теоретичну частоту варіаційного ряду, що відповідає нормованому відхиленні t.

Крива нормального розподілу має такі властивості:

- Однозначно визначається двома параметрами: $\bar{X}$ – середнім значенням і σ – середньоквадратичним відхиленням.
- Крива симетрична відносно середнього значення ($\bar{X}$) і має колоколоподібну форму, яка залежить від величини **σ**, що є параметром масштабу, а положення визначається $\bar{X}$.
- Крива має один максимум, рівний $\boldsymbol{\frac{1}{\sigma\sqrt{2\pi}}}$ і дві точки перегину на відстані ± σ від $\bar{X}$.
- Відгалуження кривої асимптотично наближаються до осі абсцис на відстань ± ∞.

Отже, нормальний закон розподілу відображає функціональну залежність між величиною ознаки і її частотою в генеральній сукупності. Чим більше відхилення варіанти від середньої величини, тим менше її частота, і навпаки, чим менше варіанта відхиляється від середньої арифметичної, тим більше її частота в даній сукупності. Отже, ймовірність відхилення кожної варіанти від середньої арифметичної є функція нормованого відхилення. Ця функція виражається за допомогою асимптотичних, тобто наближених формул (6.1), а також графічно (рис. 6.2) і в формі таблиць.

Користуючись таблицею значень ймовірності, що відповідають різним значенням нормованого відхилення t, можна за двома параметрами – x і σ – побудувати теоретичний варіаційний ряд, що має значення при порівняльній оцінці емпіричних розподілів.

Вище було показано, що нормальний розподіл є широко поширеною закономірністю в живій природі, в тому числі і в лісових насадженнях. Такому явищу повинна бути деяка переконлива причина. Її привів математик і механік А.М. Ляпунов (1857 – 1918), довівши в 1901 році граничну теорему Ляпунова, яка відноситься до області теорії ймовірностей.

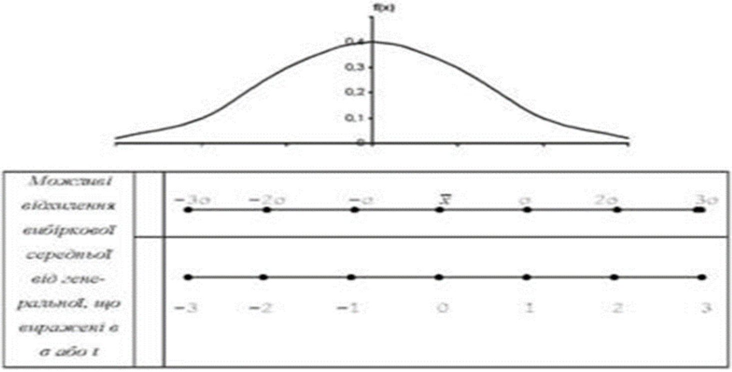


Рис. 6.2. Крива нормального розподілу

З огляду на фундаментальність нормального закону розподілу і його велику практичну значимість, наведемо короткий виклад теореми Ляпунова.

Теорема Ляпунова стверджує, що якщо значення незалежних випадкових величин будуть малі в порівнянні з їх сумою, то при необмеженому зростанні числа цих величин розподіл їх суми стає наближено нормальним.

Умови теореми Ляпунова є настільки широкими, що в багатьох випадках їх можна припускати такими, що виконуються. Тому, якщо є підстава розглядати досліджувану величину як суму багатьох незалежних випадкових величин, вплив кожної з яких на цю суму є практично мізерним, то, якщо навіть розподіл складових величин нам невідомий, можна часто заздалегідь бути упевненим, що величина, яка вивчається має нормальний розподіл.

Завдяки цьому стає зрозумілим, що, наприклад, розподіл випадкових помилок при вимірах буде нормальним. Точно так же нормальним буде розподіл фізичних ознак людей, розподіл механічних властивостей матеріалів і т. д. Цей висновок повністю підтверджується численними дослідженнями.

Таким чином, теорема Ляпунова дає пояснення тому важливому положенні, що в багатьох випадках величини мають нормальний розподіл.

Причину такої відповідності нормальної кривої отриманим при спостереженні рядам розподілу можна бачити у виконанні тих самих умов, на підставі яких теоретично з'являється нормальна крива.

Можна припустити, що окремі досліджувані значення є результатом безлічі вельми незначних незалежних між собою причин, кожна з яких може призвести до дуже малих позитивних або негативних відхилень від середнього значення досліджуваної величини.

Прикладів, коли застосовують закон, розкритий теоремою Ляпунова, безліч. Так, наприклад, на деревостани, які ростуть без втручання людини, діє безліч абіотичних і біотичних факторів. Ресурси росту дерева визначаються родовими ознаками конкретного насіння, місця, де воно зростає (умови зростання мають певну неоднорідність навіть в межах однієї таксаційної ділянки), сусідніми деревами, випадковими пошкодженнями (снігом, тваринами, блискавкою тощо) і багатьма іншими причинами. Кожна з причин носить випадковий характер і її вплив на їх сукупний результат призводить до нормального розподілу. Тут ми виключаємо антропогенний фактор, особливо рубки догляду, вплив яких великий і цілеспрямований, що має наслідком інші розподіли діаметрів дерев.

6.4. Асиметрія, ексцес

Поряд з практично симетричними розподілами зустрічаються і скошені, асиметричні ряди. Аналітично вони характеризуються порушенням рівності між модою, медіаною і середньою арифметичною розподілу. Графічно вони виражаються асиметричними кривими розподілу. Прийнято розрізняти правосторонню, або негативну, асиметрію і позитивну або лівобічну. У випадках правобічної, або негативної, асиметрії варіанти накопичуються переважно в правій частині ряду; вершина такого ряду зміщена вправо. У разі лівосторонньої асиметрії права гілка кривої, починаючи від вершини, більша за ліву.

Асиметрія (*А*) – міра відхилення розподілу чисельностей від нормальної кривої симетричного розподілу відносно максимальної ординати. При А = 0 – крива симетрична, А < 0 – крива має правосторонню асиметрію; А > 0 – крива має лівосторонню асиметрію.

Пірсон запропонував оцінювати ступінь асиметрії по різниці між середньою арифметичною і модою, віднесеної до величини середнього квадратичного відхилення:

$$A_s = \frac{\bar{x} - M_o}{\sigma}; \tag{6.4}$$

де *As* – міра скошеності рядів розподілу, або коефіцієнт асиметрії.

Як показник асиметрії також може служити потроєна різниця між середньою арифметичною і медіаною, віднесена до величини середнього квадратичного відхилення, тобто

$$A_s = \frac{3(\bar{x} - M_e)}{\sigma}; \tag{6.5}$$

Величина цього показника звичайно не виходить за межі -3 і +3, що вказує на негативну чи позитивну асиметрію. При симетричному розподілі коефіцієнт асиметрії дорівнює нулю. Міра кососсті менше 0,5 вважається малою, від 0,5 до 1 – середньою, вище 1 – великою.

Найбільш досконалим показником асиметрії служить центральний момент третього порядку, віднесений до кубу середнього квадратичного відхилення:

$$A_s = \sum pa^3 \Big/ n\sigma^3 \text{ або } A_s = \frac{\Sigma p(\bar{x} - x_i)^3}{n\sigma^3}; \tag{6.6}$$

Поряд з симетричними і скошеними розподілами варіаційні ряди можуть бути гостро- і тупо вершинними. Ця властивість розподілу називається *ексцес*.

Ексцес (*Ex*) – міра відхилення фактичної кривої від нормальної, яка характеризує ступінь концентрації варіант навколо середнього значення.

Величина ексцесу вимірюється *центральним* моментом четвертого порядку, віднесеним до середнього квадратичного відхилення четвертого ступеня.

$$E_x = \frac{\Sigma pa^4}{n\sigma^4} - 3; \tag{6.7}$$

Для строго симетричних розподілів ексцес дорівнює нулю. Крива є гостро вершинною при $Ex > 0$, а тупо вершинною – при $Ex < 0$.

При $Ex \leq 0{,}2$ ексцес практично відсутній. Якщо ж $0{,}5 \leq Ex \leq 1$, ексцес вважається помітним, але невеликим. Крайній ступінь позитивного ексцесу теоретично безмежний. Граничне значення від'ємного ексцесу дорівнює -2, що вказує на наявність двох варіаційних рядів, тобто рядів із самостійними центрами розподілу, об'єднаних в одній загальній сукупності.

Запитання для самоконтролю:

1. В чому суть функції Гауса-Лапласа?
2. Що таке рівняння нормального розподілу? Наведіть основну форму його вираження.
3. Якими можуть бути розподіли? Наведіть приклади.
4. Наведіть приклади нормального розподілу в лісовій галузі і в інших видах діяльності.

РОЗДІЛ 7
КОРЕЛЯЦІЙНИЙ АНАЛІЗ

7.1. Сутність кореляційного аналізу в біології

В природі гармонійно поєднуються різні ознаки. От, наприклад, у лісі існують закони і закономірності, за якими лісовий біогеоценоз розвивається і на основі яких існує. Подібні закономірності притаманні і біологічним об'єктам, і людській спільноті.

Між морфологічними елементами живих організмів існує певний взаємозв’язок, один з яких полягає в тому, що із збільшенням або зменшенням розміру одного елементу (х) відповідно збільшуються або зменшуються розміри іншого елементу (у). При цьому можливі такі варіанти: 1) коли даному значенню однієї залежної ознаки відповідає чітко визначене одне значення другої ознаки; 2) коли даному значенню одного залежного фактору відповідає декілька значень другого фактору за умов, що ця група значень розміщена у певних межах щодо середнього їх значення.

Залежності та зв'язки в природі та суспільстві, що мають спільні методи їх статистичного виміру, а також взаємна залежність двох величин (або явищ), коли зміна однієї з них веде до закономірної зміни іншої називається **кореляцією,** зв'язком або залежністю. Останні два слова – це синоніми терміну “кореляція”. Це поняття широко використовується в науці, особливо в біології, фізиці, хімії. Наприклад, у лісовому господарстві відомі тісні зв'язки між діаметром та висотою дерева. Для однієї породи, наприклад у бука лісового, при збільшенні діаметра, як правило, спостерігається і збільшення висоти. У свою чергу висота дерева у певному віці залежить від родючості ґрунту. З висотою, отже, і з родючістю ґрунтів, тісно пов'язана продуктивність деревостанів. Форма стовбура залежить від густоти насадження і т. д. Про зв’язки і залежності між окремими об’єктами, елементами, ознаками чи факторами в любій галузі можна наводити безліч прикладів.

Відомі функціональні та кореляційні зв'язки. До функціональних відносять закони математики та фізики. Тут зміна однієї з величин завжди веде до обов'язкової, чітко позначеної і завжди певної зміни іншої величини.

Залежність, коли певному значенню одного елемента або його ознаки (х) відповідає лише одне значення другого елемента або ознаки (у), зветься функційною і визначається як $y = f(x)$.

Наприклад, ріст дерева в висоту – біологічний закон росту рослин. Параметри росту (швидкість збільшення висоти дерева з віком) для конкретного деревного виду, скажімо, дуба звичайного, – це закономірність.

У природі під впливом різних факторів довкілля зв'язок між ознаками проявляється в вигляді кореляційного зв'язку, чи кореляції. Зв’язок між окремими ознаками, коли одному значенню даного елемента (х) відповідає декілька значень умовно залежного елементу (у), одержав назву кореляційного зв’язку (залежності).

У цьому випадку кожному значенню однієї ознаки тут відповідає не одне, а кілька значень іншої ознаки, тобто їх розподіл.

Одна з ознак (зазвичай легша або та, яку можна точніше виміряти) приймають за факторіальну, а іншу – результативну. Іноді в умовному значенні одну називають незалежною, а іншу – залежною від першої. Тобто, незалежна ознака (х) зветься аргументом, а залежна (у) − його функцією.

Статистичне дослідження кореляції зводиться до встановлення факту зв'язку, визначення її форми, спрямованості та тісноти. Висвітлення параметрів цієї залежності, тобто питання, які значення (у) відповідають даному значенню (х) і як змінюються значення (у) із зміною значення (х), складає сутність кореляційного аналізу. На основі загального аналізу явища інколи можна судити про наявність зв'язку між ознаками. Але це не часто так буває. Наприклад, про наявність кореляції між такими таксаційними показниками дерева: товщиною та висотою можна сказати ще до їхнього виміру. Наприклад, як вже зазначалося, що зі збільшенням діаметра дерева стає більшою і його висота. Так, досліджували діаметр і висоту дерев ялиці білої. Дерева з діаметрами 20, 23, 27, 32 см мали відповідно висоти цих стовбурів 22, 26, 27, 30 м. Тобто висота дерева ялиці з діаметром стовбура 20 см становила 22 м, з діаметром 23 см – 26 м і т. д.

В інших випадках наявність кореляції між ознаками, що вивчаються, не можна передбачити настільки точно. Наприклад, без вимірювань та подальшого аналізу важко оцінити зв'язок форми стовбура з проекцією крони. В таких випадках висновок про наявність/відсутність кореляції між ознаками встановлюють на основі вимірювання та статистичного аналізу його результатів.

В сукупності біологічних організмів або явищ завжди знайдеться певна їх кількість, у якої при однаковому значенні даного елементу (х) значення елементів, що порівнюються (у) будуть різними. Наприклад, якщо у деревостані відібрати 15 дерев з однаковим діаметром 24 см. Висоти у всіх цих 15 дерев обов'язково будуть різними − від 20 до 28 м. Візьмемо ще 15 дерев діаметром 28 см. Їх висоти також будуть різними, але вже у межах 26 − 32 м. Отже, діаметру (х) 24 см буде відповідати середня висота дерев, припустимо 24 м, а діаметру 28 см − 29 м. Серед двох взаємозалежних елементів (ознак) за бажанням дослідника як аргумент і функція може бути обрано будь-яка з них.

Тобто, як бачимо з наведеного прикладу, із зміною ознаки одного умовно незалежного елемента (х) (діаметр дерева, см) обов'язково зміниться ознака іншого умовно залежного елементу (у) (висота дерева, м), але ця зміна математично не є суворо функційною, тобто при кожному значенню (х) значення (у) будуть різними.

Тому із збільшенням або зменшенням ознаки (х) морфологічного органу у сукупності організмів будуть збільшуватись або зменшуватись значення (у) другої ознаки, але їх конкретні номінальні значення будуть різні. Це означає, що одному конкретному зміненому значенню (х) буде відповідати певна кількість збільшених або зменшених різних значень (у).

Термін "кореляція" (лат. correlatio − співвідношення, зв'язок) вперше застосував Ж. Кюв'є ("Лекції по порівняльній анатомії", 1806). Згодом явище кореляції було використане і розвинене в генетиці, біології і впроваджене в біометрію Пірсоном і Гальтоном (1886).

У біологічних об'єктах, процесах і явищах біологічної природи кореляційний зв'язок може мати прояв:

- між окремими морфологічними елементами живих організмів;
- між морфологічними елементами організмів і їх якісними характеристиками;

- між елементами оточуючого середовища і морфологічними, а також якісними характеристиками біологічних об’єктів;
- між окремими фізико-біологічними процесами.

Кореляційний зв’язок зветься ***простим***, коли він описує характер залежності лише по двох ознаках, і ***множинним***, якщо зміну результативної ознаки вивчають у зв'язку з впливом або зміною кількох факторіальних ознак.

Наприклад, коли ми розглядаємо зв'язок діаметр дерева – висота, це проста кореляція. Зв'язок, коли висоту дерева вивчають залежно від ґрунтової родючості, повноти деревостану, складу насадження буде множинною.

За формою розрізняють кореляцію ***лінійну***, коли залежність між (х) і (у) із зміною (х) відображається прямою лінією, або ***криволінійну***, коли така залежність відображається лінією із відповідним скривленням. Наприклад, між довжиною і товщиною коренів молодих дерев у насадженні можна очікувати лінійну кореляцію. Але не можна очікувати такої ж форми кореляції у дерев, що ростуть у старих насадженнях. Для встановлення форми зв’язку необхідно провести статистичний аналіз. Такий аналіз дає відповідь про форму зв'язку і в тих випадках, коли на основі біологічного аналізу її встановити важко або взагалі неможливо.

Крім того, розрізняють ***пряму*** (не прямолінійну, а пряму), коли із збільшенням однієї ознаки (х) в середньому збільшується значення і іншої (у), і ***зворотну***, коли із збільшенням значень однієї ознаки (х) значення іншої (у) зменшується і навпаки.

Прикладом прямого зв'язку є дослідження діаметра та висоти дерева. Зворотній зв'язок ми можемо спостерігати, вивчаючи вплив шкідників – хвоє листогризів (для прикладу, звичайного соснового пильщика, самшитової вогнівки і т. п.) на приріст дерев. Із збільшенням кількості особин на одному дереві приріст зменшується.

Слід зазначити, що кореляційному аналізу в дослідженнях завжди повинен передувати біологічний аналіз факторів, що досліджуються.

Відобразити кореляційну залежність двох ознак, тобто провести таке співставлення можливо такими способами:

• скласти кореляційний ряд з рядів значень двох аналізованих ознак;

• за допомогою кореляційної грати, в якої з одного боку (по вертикалі) розміщується значення однієї ознаки, а з другого боку (по горизонталі) − значення другої ознаки;

• за допомогою так званої лінії регресії, абсциси якої пропорційні значенням першої ознаки, а ординати − значенням другої ознаки.

Для визначення особливостей кореляційних зв'язків застосовують такі показники, що розраховуються за допомогою методів варіаційної статистики:

• коефіцієнт кореляції;

• кореляційне відношення;

• міра криволінійності зв'язку.

7.2. Показники кореляції і їх визначення

7.2.1. Коефіцієнт кореляції як міра лінійного зв'язку

Коефіцієнт кореляції – числовий показник простої лінійної кореляції, який описує напрям і тісноту зв'язку між досліджуваними величинами, вимірює зв'язок лише при лінійній формі залежності, а його абсолютне значення знаходиться в межах від –1 до +1. При значенні $r = 0$ – зв'язок відсутній; при +1 – пряма кореляційна залежність; а при –1 – зворотна. Він широко використовується в науці для вимірювання ступеня лінійної залежності між двома змінними.

Показник був розроблений Карлом Пірсоном зі схожої ідеї, представленої Френсісом Гальтоном в 1880-х рр.

Коефіцієнт кореляції Пірсона (r) розраховують за формулами 7.1. та 7.2:

$$r = \frac{\sum a_x a_y}{n\sigma_x \sigma_y}; \tag{7.1}$$

$$r = \frac{\Sigma a_x a_y}{n\sigma_x\sigma_y} = \frac{\Sigma a_x a_y}{\sqrt{n\sigma_x^2 \times n\sigma_y^2}} = \frac{\Sigma a_x a_y}{\sqrt{\Sigma a_x^2 \times \Sigma a_y^2}}; \tag{7.2}$$

В формулі 7.1. x_i та y_i – випадкові (досліджувані) величини, $\bar{x}$ та $\tilde{y}$ – середні значення для x і y.

У чисельнику формули 7.2. стоїть сума добутків відхилень варіант від середньої арифметичної по одному ряду (X) на відповідні відхилення варіант від середньої арифметичної по іншому ряду (У) тобто: $a_x = x_x - \bar{x}_x$ і $a_y = x_y - \bar{x}_y$; $\sum a_x a_y = \sum\left[(x_x - \bar{x}_x)(x_y - \bar{x}_y)\right]$. У знаменнику σ_x і σ_y – середньоквадратичні відхилення розподілів X та Y; n – число зіставлюваних пар або число спостережень.

Макет допоміжної таблиці для розрахунків коефіцієнта кореляції:

x	y	a_x $a_x = x_x - \bar{x}_x$	a_y $a_y = x_y - \bar{x}_y$	a_x a_y	a_x^2	a_y^2

Тіснота кореляційного зв’язку є показник, що визначає наближення (або віддалення) ступеню кореляції до показника функційного зв’язку. Якщо функційному зв’язку дати оцінку 1, то віддалення від 1 (0,9; 0,8; 0,7, ..., 0,3; 0,2; 0,1; 0) буде характеризувати тісноту відповідного кореляційного зв’язку. Показник r > 0,9 свідчить про дуже тісний зв'язок, показник r = 0,7 − 0,9 − про тісний зв’язок, зв'язок, в якому r = 0,5 − 0,7 характеризує значний, r = 0,3 − 0,4 − помірний, а r < 0,3 − слабкий зв'язок. Тобто:

Кореляція	Зворотна (негативна)	Пряма (позитивна)
Відсутня	−0.09 до 0.0	0.0 до 0.09
Низька	−0.3 до −0.1	0.1 до 0.3
Середня	−0.5 до −0.3	0.3 до 0.5
Висока	−1.0 до −0.5	0.5 до 1.0

Оцінка коефіцієнта кореляції при малих і великих вибірках

Оскільки окремі варіанти, що досліджуються є величинами випадковими, то і коефіцієнт кореляції, який визначається для відповідної вибірки варіант, є величина випадкова. Звідси виникає необхідність визначати ступінь його наближення до показника генеральної сукупності значень r. Цей показник позначається літерою (ρ). Для вирішення наведеного завдання також застосовується нульова гіпотеза. Вона полягає в припущенні, що $\rho = 0$, тобто, що між випадковими величинами X і Y кореляція відсутня.

Для перевірки цієї нульової гіпотези використовується порівняння показників з критерієм t − Стьюдента (t_s). Значення t при досить значній кількості спостережень (n ≥100) є відношенням коефіцієнта кореляції до своєї помилки (m_r), яка визначається за формулою 7.3:

$$t = r/m_r\ ; \quad \text{або} \quad t = \frac{r\sqrt{(N-1)}}{\sqrt{1-r^2}}\ ; \qquad (7.3)$$

Помилка коефіцієнта кореляції при досить значній кількості спостережень (n ≥100) визначається за формулою 7.4:

$$m_r = \pm\sqrt{\frac{1-r^2}{N-1}}\ ; \qquad (7.4)$$

де N – об’єм вибірки.

Якщо в досліді кількість спостережень менше 100 ($n < 100$), за критерій для перевірки нульової гіпотези приймається показник достовірності коефіцієнта кореляції t (формула 7.5), для визначення якого вираховується помилка коефіцієнта кореляції за формулою 7.6:

$$t = r/m_r \text{ або } t = \frac{r\sqrt{(N-2)}}{\sqrt{1-r^2}}; \tag{7.5}$$

$$m_r = \pm\sqrt{\frac{1-r^2}{N-2}} \tag{7.6}$$

Критичне (стандартне) значення критерія $t_{5\%}$ знаходиться з таблиць (додаток 1) за числом ступенів вільності ν = 'N-2. Порівнюючи фактичне і критичне значення критеріїв, роблять відповідний висновок:

якщо $t_ф > t_{5\%}$ – коефіцієнт кореляції достовірний; встановлену тісноту зв’язку доведено.

якщо $t_ф < t_{5\%}$ – коефіцієнт кореляції недостовірний; причиною може бути недостатній об'єм вибірки або відсутність зв’язку.

В біології є багато прикладів, де необхідно вміти правильно використовувати показники кореляції і аналізувати досліджуваний матеріал.

Приклад визначення коефіцієнта кореляції на малих вибірках

Визначення тісноти зв’язку і оцінки його достовірності для малої вибірки розглянемо на наступному прикладі. Провели заміри надземної частини і довжини коріння у самосіву бука лісового, їх дані занесені в таблицю 7.1.

Таблиця 7.1

Обчислення коефіцієнта кореляції між висотою рослин і довжиною коріння однорічного самосіву бука лісового

Надземна частина, см	Довжина кореня, см	a_x	a_y	$a_x\ a_y$	a_x^2	a_y^2
x	y					
11	12,5	-1,4	-1,1	1,54	1,96	1,21
14	14,8	1,6	1,2	1,92	2,56	1,44
12	12,8	-0,4	-0,8	0,32	0,16	0,64
13	13,6	0,6	0	0	0,36	0
11	12,4	-1,4	-1,2	1,68	1,96	1,44
12	13,0	-0,4	-0,6	0,24	0,16	0,36
13	14,6	0,6	1	0,6	0,36	1
12	14,2	-0,4	0,6	-0,24	0,16	0,36
13	14,0	0,6	0,4	0,24	0,36	0,16
13	14,1	0,6	0,5	0,3	0,36	0,25
124 (Σ)	136 (Σ)	0	0	**6,6** ($\Sigma\ a_x\ a_y$)	**8,4** ($\Sigma\ a_x^2$)	**6,86** ($\Sigma\ a_y^2$)
12,4 ($\bar{x}$)	13,6 ($\bar{y}$)					

Алгоритм визначення коефіцієнта кореляції:

1. Визначаємо середнє арифметичне значення довжини надземної маси самосіву бука лісового

$\tilde{x} = \Sigma x/n = 124/10 = 12{,}4$ см

2. Визначаємо середнє арифметичне значення довжини кореневої системи самосіву $\bar{y} = \Sigma y/n = 136/10 = 13{,}6$ см

3. Знаходимо відхилення кожної окремої варіанти від середньої арифметичної рядів Х ($a_x = x_x - \bar{x}_x$) і У ($a_y = x_y - \bar{x}_y$) і заносимо в стовпці 3 і 4 таблиці 7.1.

4. Знаходимо добутки відхилень варіант від середньої арифметичної (a_x a_y) по обох рядах і їх суму і заносимо в стовпчик 5.

5. Знаходимо квадрати відхилень a_x^2 і a_y^2 та заносимо в стовпчики 6 і 7 таблиці 7.1.

6. Отримані дані підставляємо у формулу:

$r = \Sigma\, a_x\, a_y / \sqrt{\sum a_x^2 \Sigma a_y^2} = 6{,}6 / \sqrt{8{,}4 \cdot 6{,}86} = 0{,}86$

7. Оскільки в досліді кількість спостережень менше 100 (n = 10), то для визначення помилки коефіцієнта кореляції використовуємо формулу 7.6.

Для даних таблиці 7.1. при r = 0,9 маємо таку оцінку:

$$m_r = \pm\sqrt{\frac{1-0{,}86^2}{10-2}} = 0{,}18$$

Вираховуємо t-критерій і порівнюємо його з табличним значенням:

$t_r = r/m_r = 0{,}86/0{,}18 = 4{,}8$;

Отримане значення t_r більше як $t_{0,05} = 2{,}3$.

Отже, коефіцієнт кореляції достовірний; доведено, що між ознаками висотою рослин і довжиною коріння однорічного самосіву бука лісового є тісний зв'язок.

Цю форму оцінки не можна застосовувати за інших гіпотез, крім $\rho = 0$. Так само критерій t не можна використовувати для побудови довірчого інтервалу. Р.А. Фішер запропонував для цих цілей z – перетворення величини r.

Величина z:

$$Z = \frac{1}{2} ln \frac{1+r}{1-r}; \qquad (7.7)$$

Розподіл цієї величини мало залежить від чисельності вибірки і від значення коефіцієнта кореляції в генеральній сукупності. Перевага z (зет) перед r в тому, що z, міняючи своє значення від -∞ до +∞, швидко наближається до нормального розподілу при збільшенні числа спостереження (n). Тому показник зет дає досить надійні результати на малих вибірках, так як саме в таких випадках коефіцієнт кореляції може сильно відрізнятися від свого значення в генеральній сукупності. Навіть при значенні, близьких до одиниці, при обчисленні на малих вибірках, показник z має нормальний розподіл з дисперсією:

$$\sigma_z^2 = \frac{1}{N-3}; \quad (7.8)$$

Перетворення коефіцієнта кореляції в величину зет проводять за допомогою спеціальної таблиці, розробленої Фішером (додаток 3).

Для оцінки достовірності і встановлення довірчого інтервалу, по якому з достатньою імовірністю можна судити про величину коефіцієнта кореляції в генеральній сукупності необхідно зробити наступне. По емпіричному значенню коефіцієнта кореляції з таблиці в додатку 3 визначають значення z. Так для вибірки висоти рослин і довжини коріння однорічного самосіву бука лісового (таблиця 7.1) при r = 0,93; z = 1,658. Помилка $\sigma_z = \sqrt{\frac{1}{10-3} = 0{,}378}$, звідси t_z = 1,658 x $\sqrt{7}$ = 4,39. По таблиці Стьюдента (таблиця в додатку 1) для k = 10-2 = 8 і P = 0,05 знаходимо t_{st} = 2,31. Оскільки $t_ф$ = 4,39 > t_{st} = 2,31, нульова гіпотеза не зберігається.

За величиною показника $\Delta_z = t\,\sigma_z$ = 2,3x0,378 = 0,8694 знаходимо межі довірчого інтервалу для показника зет. Нижня межа – 0,7886 (1,658 – 0,8694), верхня межа – 2,5274 (1,658 + 0,8694). Переводимо значення зет в величини коефіцієнта кореляції і знаходимо його довірчі межі: нижня межа – 0,66, а верхня – 0,99 (при цьому використовується допоміжна таблиця значень зет, що відповідають величинам коефіцієнта кореляції (додаток 3)).

Для генеральної сукупності ρ величина коефіцієнта кореляції знаходиться в межах $0,66<r<0,99$, що свідчить про те, що значення коефіцієнта для вибірки $r = 0,9$ визначено достатньо точно.

Про наявність чи відсутність зв’язку між різними явищами, характеристиками чи ознаками можна підтвердити і іншими прикладами. Так, наприклад, немає зв’язку між частотою скошування трави і площею газону, або ж висотою середньовікового деревостану і величиною популяції зайців. Однак, між частотою скошування газону і густотою травостою є зв'язок. Як і є середній зв’язок між величиною приросту в $м^3$/га і повнотою насадження. Від родючості ґрунту залежить бонітет насадження, від висоти дерева – його діаметр, від густоти, складу травостою – урожайність зеленої маси і т. д.

Визначення коефіцієнта кореляції на великих вибірках

При наявності великої вибірки визначити коефіцієнт кореляції описаним способом важко. В таких випадках вибірку розподіляють у варіаційні ряди таким чином, щоб чисельності у них були загальними, відповідали парним значенням класів чи класових варіант ознак Х і У, що корелюють. Для цього будують кореляційну таблицю (таблиці 7.2 і 7.3 в даному розділі), яка виглядає як сітка, утворюється пересіканням рядків і стовпчиків, число яких дорівнює числу класів одного і другого варіаційних рядів.

В верхньому рядку і в першому стовпчику таблиці розміщують класи варіаційних рядів, а також їх середні значення (класові варіанти) або тільки одні класові варіанти. Частота варіант з’єднаних рядів позначається символом p_{xy}, займає в таблиці своє місце, клітинку, так, що по характеру розміщення частот в клітках кореляційної таблиці можна бачити наявність і напрямок, а частково і тісноту зв’язку між ознаками.

Якщо частоти розміщуються переважно по діагоналі з лівого верхнього кута сітки до нижнього правого кута (за умови, що ряди розміщуються в напрямку від мінімальних до максимальних значень ознак зліва направо і зверху вниз), як в таблиці 7.2. Це вказує на позитивний зв'язок між ознаками.

Таблиця 7.2

Y \ X	x_1	x_2	x_3	x_4	. . .	x_k	P_y
y_1	p_{xy}				. . .		Σy_1
y_2		p_{xy}	p_{xy}		. . .		Σy_2
y_3		p_{xy}	p_{xy}	p_{xy}	. . .		Σy_3
y_4				p_{xy}	. . .	p_{xy}	Σy_4
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
y_k					. . .	p_{xy}	Σy_k
Px	Σx_1	Σx_2	Σx_3	Σx_4	. . .	Σx_k	$\Sigma p_{xy} = n$

Якщо ж частоти розміщуються переважно в напрямку з нижнього лівого кута до правого верхнього кута кореляційної решітки (таблиця 7.3), це свідчить про наявність від’ємного зв’язку між ознаками.

Таблиця 7.3

Y \ X	x_1	x_2	x_3	x_4	. . .	x_k	P_y
y_1					. . .	p_{xy}	Σy_1
y_2				p_{xy}	. . .		Σy_2
y_3		p_{xy}	p_{xy}	p_{xy}	. . .		Σy_3
y_4		p_{xy}	p_{xy}		. . .		Σy_4
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
y_k	p_{xy}	p_{xy}			. . .		Σy_k
Px	Σx_1	Σx_2	Σx_3	Σx_4	. . .	Σx_k	$\Sigma p_{xy} = n$

Якщо ж кореляція між ознаками відсутня, частоти розподіляються по клітинках кореляційної таблиці більш менш рівномірно.

Методи визначення коефіцієнта кореляції

При формуванні даних досліджень у вигляді кореляційних таблиць вирахувати коефіцієнт кореляції різними способами.

Спосіб умовних середніх є одним із найбільш поширених методів визначення коефіцієнта кореляції. Називають його ще способом умовного початку. Від умовних середніх Ах і Ау беруть відхилення класів (або класових варіант). Коефіцієнт кореляції визначають по такій формулі:

$$r = \frac{\Sigma p a_x a_y - n b_x b_y}{n \sigma_x \sigma_y}; \tag{7.9}$$

де а – відношення відхилень класів чи класових варіант від умовних середніх Ах і Ау до величин класових інтервалів, тобто $a_x = (x_i - A)$: i_x і $a_y = (y_i - A)$: i_y; р – частоти класових варіант; $b = \frac{\Sigma pa}{n}$ – умовний момент першого порядку; n – загальне число спостережень (об'єм вибірки); σ_x і σ_y – середні квадратичні відхилення рядів, що вираховуються без множення на величину класового інтервалу.

Спосіб підсумування дозволяє досить легко визначити коефіцієнт кореляції. Згідно даного способу замість відхилень класів, класових варіант від умовних середніх ставляться порядкові номери класів в напрямку від менших значень варіант до більших, які перемножують на відповідні частоти, що в результаті дає величину $pa_x a_y$. Замість центральних чи умовних моментів першого порядку в даному способі використовується перша (S') і друга (S') суми повних накопичених рядів, які отримують кумуляцією частот кожного варіаційного ряду в напрямку від великих значень ознак Х і У до їх менших значень.

Коефіцієнт кореляції в такому випадку визначають за такою формулою:

$$r = \left(\frac{\Sigma p a_x a_y}{n} - \frac{S'_x}{n} - \frac{S'_y}{n}\right) : \sigma_x \sigma_y; \qquad (7.10)$$

де $\Sigma p a_x a_y$ – сума похідних частот кореляційної решітки (p_{xy}) на відповідні порядкові номери класів (a_x і a_y), які розміщені в напрямку від менших до більших значень класових варіант; S' – сума першого повного ряду накопичення частот, який отримують кумуляцією частот кожного ряду в зворотному напрямку до порядкової нумерації класів, тобто від більших значень класових варіант до їх менших значень; σ_x і σ_y – середні квадратичні відхилення рядів; n – загальна кількість парних спостережень (об'єм вибірки).

7.2.1.2. Множинна кореляція

При визначенні коефіцієнта кореляції між двома ознаками X і У не враховується їх залежність від варіювання інших ознак. Є випадки, коли дві величини, начебто взаємозалежні, але більш докладний аналіз показує, що ця «взаємозалежність» є відображенням того факту, що вони обидві корелюють з деякою третьою величиною або з сукупністю величин. У природі ми часто зустрічаємося з такими явищами, коли зміна однієї величини (функції) визначається зміною одного, а кількох аргументів. Наприклад, висота дерев чи їх діаметр залежать від багатьох чинників, зокрема від родючості ґрунту, його вологості, віку дерев, повноти насадження і т.д. Все це дає можливість говорити про множинну кореляцію.

Коефіцієнт множинної кореляції – це показник тісноти зв'язку (лінійного) між однією залежною величиною та сукупністю незалежних. Визначається між трьома досліджуваними ознаками X,Y і Z за наступною формулою:

$$r_{xyz} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2r_{xy}r_{xz}r_{yz}}{1 - r_{xy}^2}}; \qquad (7.11)$$

де r_{xy} r_{xz} і r_{yz} – парні коефіцієнти лінійної кореляції між ознаками X і У, X і Z , Y і Z.

Коефіцієнт множинної кореляції має тільки позитивний знак.

В практиці загальний коефіцієнт множинної кореляції використовується не часто. Частіше використовуються часткові (парціальні) коефіцієнти, які вимірюють силу зв'язку між двома варіюючими ознаками при постійному значенні третьої досліджуваної ознаки, яка знаходиться чи може знаходитися в кореляційній залежності з іншими двома.

Частковий коефіцієнт кореляції між ознаками X і Y при виключенні впливу на цей зв'язок третьої ознаки Z визначається наступним чином:

$$r_{xy(z)} = \frac{r_{xy} - r_{xz} \times r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}; \qquad (7.12)$$

Відповідно коефіцієнт кореляції між ознаками X і Z при виключенні впливу на цей зв'язок третьої ознаки Y визначається наступним чином:

$$r_{xz(y)} = \frac{r_{xz} - r_{xy} \times r_{yz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{yz}^2)}}; \qquad (7.12.1)$$

Частковий коефіцієнт кореляції між ознаками Y і Z при виключенні впливу на цей зв'язок третьої ознаки X визначається наступним чином:

$$r_{yz(x)} = \frac{r_{yz} - r_{xy} \times r_{xz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{xz}^2)}}; \qquad (7.12.2)$$

Як визначити коефіцієнт множинної кореляції та часткові коефіцієнти через розрахунок парних коефіцієнтів кореляції розглянемо на прикладі.

На облікових ділянках визначили висоту травостою, см (X), кількість пагонів на м^2 (Y) і масу рослин, кг (Z) конюшини лучної. Отримані дані і окремі розрахунки, необхідні для визначення множинної кореляції наведені в таблиці 7.4

Таблиця 7.4

Допоміжна таблиця для розрахунків множинної кореляції

X	Y	Z	X^2	Y^2	Z^2	X Y	Y Z	X Z
94	459	5,7	8836	210681	32,49	43146	2616,3	535,8
79	484	3,7	6241	234256	13,69	38236	1790,8	292,3
88	455	4,3	7744	207025	18,49	40040	1956,5	378,4
78	532	3,4	6084	283024	11,56	41496	1808,8	265,2
75	370	3,2	5625	136900	10,24	27750	1184	240
68	384	3,0	4624	147456	9	26112	1152	204
72	402	3,9	5184	161604	15,21	28944	1567,8	280,8
83	420	4,1	6889	176400	16	34860	1680	332
78	468	3,2	6084	219024	10,24	36504	1497,6	249,6
85	524	4,6	7225	274576	21,16	44540	2410,4	391
800	**4498**	**39**	**64536**	**2050946**	**158,08**	**361628**	**17664,2**	**3169,1**

Щоб вирахувати коефіцієнти множинної кореляції, потрібно спочатку знайти значення парних коефіцієнтів зв'язку між цими ознаками. Для цього розраховуємо суми квадратів відхилень варіант від їх середніх арифметичних:

$\Sigma(x_i - \overline{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 64536 - \frac{800^2}{10} = 536;$

$\Sigma(y_i - \overline{y})^2 = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 2050946 - \frac{4498^2}{10} = 2050946 - 2023200{,}4$
$= 27745{,}6;$

$\Sigma(z_i - \overline{z})^2 = \Sigma z^2 - \frac{(\Sigma z)^2}{n} = 158{,}08 - \frac{39^2}{10} = 158{,}08 - 152{,}1 = 5{,}98.$

Знаходимо середнє квадратичне відхилення:

$\sigma_x = \sqrt{\frac{536}{10}} = 7{,}321;\ \sigma_y = \sqrt{\frac{27745{,}6}{10}} = 52{,}67;\ \sigma_z = \sqrt{\frac{5{,}98}{10}} = 0{,}77$

Далі розраховуємо:

$$\Sigma(y_i - \overline{y})(x_i - \overline{x}) = \Sigma yx - \frac{\Sigma y \times \Sigma x}{n} = 361628 - \frac{4498 \times 800}{10} = 1788;$$

$$\Sigma(y_i - \overline{y})(z_i - \overline{z}) = \Sigma yz - \frac{\Sigma y \times \Sigma z}{n} = 17664{,}2 - \frac{4498 \times 39}{10} = 122;$$

$$\Sigma(x_i - \overline{x})(z_i - \overline{z}) = \Sigma xz - \frac{\Sigma x \times \Sigma z}{n} = 3169{,}1 - \frac{800 \times 39}{10} = 49{,}1;$$

Звідси:

$$r_{xy} = \frac{\Sigma(y_i - \overline{y})(x_i - \overline{x})}{n\sigma_x\sigma_y} = \frac{1788}{10 \times 7.321 \times 52.67} = 0{,}46;$$

$$r_{yz} = \frac{\Sigma(y_i - \overline{y})(z_i - \overline{z})}{n\sigma_z\sigma_y} = \frac{122}{10 \times 0.77 \times 52.67} = 0{,}3;$$

$$r_{xz} = \frac{\Sigma(x_i - \overline{x})(z_i - \overline{z})}{n\sigma_z\sigma_x} = \frac{49.1}{10 \times 0.77 \times 7.321} = 0{,}87;$$

Знайдені значення парних коефіцієнтів дозволяють розрахувати величини парціальних коефіцієнтів кореляції:

$$r_{xy(z)} = \frac{r_{xy} - r_{xz} \times r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} = \frac{0{,}46 - 0{,}87 \times 0{,}3}{\sqrt{(1 - 0.87^2)(1 - 0.3^2)}} = 0{,}4;$$

$$r_{xz(y)} = \frac{r_{xz} - r_{xy} \times r_{yz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{yz}^2)}} = \frac{0{,}87 - 0{,}46 \times 0{,}3}{\sqrt{(1 - 0.46^2)(1 - 0.3^2)}} = 0{,}86;$$

$$r_{yz(x)} = \frac{r_{yz} - r_{xy} \times r_{xz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{xz}^2)}} = \frac{0{,}3 - 0{,}46 \times 0{,}87}{\sqrt{(1 - 0.46^2)(1 - 0.87^2)}} = -0{,}2;$$

Тісний зв'язок виявився між висотою травостою і масою рослин конюшини лучної, незалежно від впливу на цю залежність кількості пагонів на 1м^2.

Критерієм для нульової гіпотези є:

$$t = \frac{r_{\text{парц}}\sqrt{n - v}}{\sqrt{1 - r_{\text{парц}}^2}}; \qquad (7.13)$$

де n – об’єм вибірки, а v – кількість досліджуваних ознак. Нульова гіпотеза, тобто припущення про незалежне варіювання двох ознак при виключенні впливу третього (чи інших досліджуваних ознак) відкидається, якщо $t_ф \geq t_{st}$ для $k = n - 3$ і прийнятого рівня імовірності Р. Так, в даному прикладі:

$$t_{xz(y)} = \frac{0.86\sqrt{10-3}}{\sqrt{1-0.86^2}} = 4{,}45;$$

$$t_{xy(z)} = \frac{0.4\sqrt{10-3}}{\sqrt{1-0.4^2}} = 0{,}42.$$

Для Р = 0,05 і k = 7 за таблицею Стьюдента $t_{st} = 2{,}365$. Між висотою травостою і масою рослин конюшини лучної є тісний зв'язок, так як $t_ф = 4{,}45 \geq t_{st}$ і нульова гіпотеза відкидається,. Між іншими досліджуваними ознаками зв'язок відсутній або ж для формування висновків недостатній об’єм вибірки.

На основі отриманих значень парціальних коефіцієнтів використовуючи формулу 7.11 визначаємо загальний коефіцієнт множинної кореляції.

7.2.2. Кореляційне відношення

Зв'язок між випадковими величинами ***x*** та ***y*** не завжди виражається у вигляді прямої лінії. Тому, К. Пірсон запропонував показник кореляційного відношення, який описує тісноту зв’язку при будь-якій формі залежності. Попереднє визначити криволінійність зв’язку, можна, побудувавши графік функції, на якому буде видно форму залежності.

Важливо вміти вираховувати кореляційне відношення. Це стосується, насамперед, досліджень, при визначенні коефіцієнта кореляції було з’ясовано, що між ознаками існує зв'язок, який є слабким або ж середнім, тобто r < 0,7 – 0,8. І навіть тоді, коли зв'язок між ознаками є сильним, але при цьому має криволінійний характер.

На відміну від коефіцієнта кореляції кореляційне відношення описує двосторонній зв'язок. Пояснити таку властивість можна на такому прикладі. Візьмемо кілька парних значень (результати замірів діаметрів і висот молодих дерев сосни).

Діаметр, см (x)	8	6	5	4	4	6	4	8
Висота, м (y)	10	7	7	6	7	7	5	8

Ранжируємо (розміщуємо в порядку зростання) цю сукупність по X:

x	4	4	4	5	6	6	8	8
y	5	6	7	7	7	7	8	10

→

x	4	5	6	8
$\bar{y}_i$	6	7	7	9

Оскільки деякі значення X повторяються, то розподіляємо сукупність таким чином:

Тут $\bar{y}_i$ – групові середні значень Y, які відповідають значенням величини X, тобто, наприклад, (5+6+7):3 = 6 і т. д.

Якщо ранжирувати по Y, то отримуємо наступний розподіл:

y	5	6	7	7	7	7	8	10
x	4	4	4	5	6	6	8	8

→

y	5	6	7	8	10
$\bar{x}_i$	4	4	5,25	8	8

Ці ряди складається з різної кількості груп, яким відповідають неоднакові групові середні. Видно, що залежність між змінними X і Y виражається по різному в залежності від того, по значеннях якої з них ранжирується сукупність.

Кореляційне відношення дорівнює кореню квадратному із співвідношень сум квадратів відхилень групових або часткових середніх ($\bar{x}_y$ і $\bar{y}_x$) від загальних середніх ($\bar{x}_c$ і $\bar{y}_c$) до суми квадратів відхилень окремих варіант (x_i і y_i) від загальних середніх даної сукупності. Кореляційне відношення виражають грецькою літерою η.

Кореляційне відношення характеризує зв'язок між варіюючими ознаками двосторонньо – X по Y і Y по X, тому виражається не одним, а двома показниками $\eta_{y/x}$ та $\eta_{x/y}$ (формула 7.14).

Для визначення кореляційного відношення (формули 7.14 a; 7.14 b) необхідно вирахувати дисперсії групових середніх (формули 7.15 a; 7.15 b) та загальні дисперсії (формули 7.16 a; 7.16 b).

$$\eta_{y/x} = \sqrt{\frac{\sigma_{yx}^2}{\sigma_y^2}}\ ; \qquad \eta_{x/y} = \sqrt{\frac{\sigma_{xy}^2}{\sigma_x^2}} \qquad (7.14\ a;\ 7.14\ b)$$

$$\sigma_{yx}{}^2 = \frac{\sum(y_x - y_c)^2 n_i}{N-1}\ ; \qquad \sigma_{xy}{}^2 = \frac{\sum(x_y - x_c)^2 n_i}{N-1}\ ; \quad (7.15\ a;\ 7.15\ b)$$

$$\sigma_y^2 = \frac{\sum(y_i - y_c)^2}{N-1}\ ; \qquad \sigma_x^2 = \frac{\sum(x_i - x_c)^2}{N-1}\ ; \quad (7.16\ a;\ 7.16\ b)$$

де y_x і x_y – групові (часткові) середні; y_i і x_i – окремі варіанти; y_c і x_c – загальні середні; n_i – число спостережень в окремих групах; N – загальне число спостережень.

Для обчислення кореляційного відношення створюють допоміжну таблицю, макет якої поданий нижче.

x	y	Число спосте-режень (n_i)	Групові середні, $\bar{y}_x$; $\bar{x}_y$	$\bar{y}_x - \bar{y}_c$ $\bar{x}_y - \bar{x}_c$	$(\bar{y}_x - y_c)^2$ $(\bar{x}_y - \bar{x}_c)^2$	$n_i(\bar{y}_x - y_c)^2$ $n_i(\bar{x}_y - \bar{x}_c)^2$	$y_i - y_c$ $x_i - \bar{x}_c$	$(y_i - y_c)^2$ $(x_i - \bar{x}_c)^2$

Оцінюють достовірність кореляційного відношення за визначенням помилки репрезентативності (формула 7.17) і критерію Стьюдента (формула 7.18), або ж критерієм для оцінки нульової гіпотези є зведена формула 7.18.1.

$$m_{ή} = \pm\sqrt{\frac{(1-ή)^2}{N-2}}; \qquad (7.17)$$

$$t = \frac{\eta_{y/x}}{m_\eta}; \qquad (7.18)$$

$$t = \eta\sqrt{\frac{n-2}{1-\eta^2}} \qquad (7.18.1)$$

Значення $t_{5\%}$ беруть з таблиць (додаток 1) за числом ступенів свободи $\nu = N$-2

при $t_ф < t_{5\%}$ – кореляційне відношення достовірне;

при $t_ф > t_{5\%}$ – кореляційне відношення не достовірне.

Визначення кореляційного відношення (таблиця 7.5) розглянемо на прикладі розподілу 10-ти сіянців бука лісового по висоті рослин і довжині коріння, для якого ми визначали коефіцієнт кореляції (див. таблицю 7.1).

Таблиця 7.5

Обчислення кореляційного відношення для зв'язку висоти рослин (X) по довжині коріння (Y) однорічного самосіву бука лісового

Надземна частина, см	Довжина кореня, см	Число спосте-режень (n_i)	Групові середні $\bar{y}_x$	$\bar{y}_x - y_c$	$(\bar{y}_x - \bar{y}_c)^2$	$n_i(\bar{y}_x - \bar{y}_c)^2$	$y_i - \bar{y}_c$	$(y_i - \bar{y}_c)^2$
x	y_i							
11	12,5	2	12,45	-1,15	1,3225	2,645	-1,1	1,21
	12,4						-1,2	1,44
12	12,8	3	13,3	-0,3	0,09	0,27	-0,8	0,64
	13,0						-0,6	0,36
	14,2						0,6	0,36
13	13,6	4	14,075	0,475	0,2256	0,9024	0	0
	14,6						1	1
	14,0						0,4	0,16
	14,1						0,5	0,25
14	14,8	1	14,8	1,2	1,44	1,44	1,2	1,44
	$\bar{y}_c$ = 13,6					$\sum$ = **5,257**		$\sum$ = **6,86**
						0,5841 (σ_{yx}^2)		**0,7622** (σ_y^2)

Алгоритм визначення кореляційного відношення:

1. Визначаємо середнє арифметичне значення довжини кореневої системи самосіву

$y_c = \Sigma y / n$ = 136/10 = 13,6 см.

2. Визначаємо число групових спостережень n_i, наприклад у сіянців висотою 11 см довжина коріння становила 12,5 см і

12,4 см, тобто n_i = 2 і т.д. Заносимо результати в стовпчик 3 таблиці 7.5.

3. Знаходимо групові середні (y_x = (12,5 + 12,4)/2 = 12,45) і т. д.

4. Знаходимо відхилення групових середніх (y_x) від загальної середньої (y_c) і заносимо в стовпчик 5 таблиці 7.5.

5. Знаходимо квадрати відхилень групових середніх від загальної середньої – $(y_x - y_c)^2$ (стовпчик 6) і з врахуванням числа спостережень в окремих групах $n_i (y_x - y_c)^2$ (стовпчик 7).

6. Знаходимо відхилення окремих варіант (y_i) від загальної середньої (y_c) та квадрати цих відхилень $(y_i - y_c)^2$ і заносимо відповідно в стовпчики 8 і 9 таблиці 7.5 .

7. Визначаємо дисперсію групових середніх за формулою 7.15 а, використовуючи результати обчислення таблиці 7.5 (стовпчик 7) – $\sum n_i(y_x - y_c)^2$.

$${\sigma_{yx}}^2 = \frac{5,2574}{9} = 0,5841;$$

8. Визначаємо загальну дисперсію за формулою 7.16 а, використовуючи результати обчислення з таблиці 7.5 (стовпчик 9) – $\sum (y_i - y_c)^2$.

$$\sigma_y^2 = \frac{6,86}{9} = 0,7622;$$

9. Визначаємо кореляційне відношення за формулою 7.14 а:

$$\eta_{y/x} = \sqrt{\frac{0,5841}{0,7622}} = 0,87;$$

10. Знаходимо помилку кореляційного відношення за формулою 7.17 і критерій Стьюдента за формулою 7.18:

$$m_{\acute{\eta}} = \pm\sqrt{\frac{(1-\acute{\eta})^{2}}{N-2}} = \pm\sqrt{\frac{(1-0,87)^{2}}{8}} = \pm 0,045;$$

$$t = \frac{\eta_{y/x}}{m_{\eta}} = \frac{0,87}{0,045} = 19;$$

Для ν = N - 2 = 8 критичне значення t (додаток 1) навіть при достовірності 99,9% дорівнює 5,041 і є меншим, ніж фактичне значення t (5.041 < 19), що свідчить про те, що кореляційне відношення достовірне.

Таким же методом знаходимо кореляційне відношення довжини коренів (Y) по надземній частині (X), ранжируючи сукупність по Y і допоміжної таблиці по визначенні дисперсій. В даному прикладі $\eta_{x/y} = 1$.

Визначення кореляційного відношення для великих вибірок

При наявності великої кількості спостережень визначення кореляційного відношення описаним способом є досить трудомістким. Тому рекомендується вибірку групувати в кореляційну таблицю, таку ж як і при визначенні коефіцієнта кореляції для великих вибірок (таблиці 7.2 і 7.3). Для цього необхідно виконати наступне:

1) вибірка розбивається на класи окремо по кожній ознаці;

2) побудова кореляційної решітки;

3) по кожному класу X (і по класу Y) визначається часткові (групові) середні ($\bar{y}_x$) і загальна середня $\bar{y}_c$;

4) знаходять відхилення групових середніх від загальної середньої ряду ($\bar{y}_x - \bar{y}_c$);

5) знайдені відхилення підносять до квадрату і множаться на частоти кожного класу (кількість ознак в класі) $(\bar{y}_x - \bar{y}_c)^2$;

6) знаходять суму квадратів відхилень $\Sigma p(\bar{y}_x - \bar{y}_c)^2$;

7) знаходять суму квадратів відхилень загальної варіації ознаки $\Sigma p\ (y_i - \bar{y}_c)^2$, тобто визначають відхилення окремих варіант від загальної середньої, які підносять до квадрату, множать на частоти, а отримані величини додають;

8) визначають кореляційне відношення за формулою 7.14 і оцінюють за критерієм Стьюдента (формула 7.18) для числа

вільності k = n – 2 і принятого рівня значимості р. Помилка ($m_{ή}$) кореляційного відношення для великих вибірок визначається за формулою:

$$m_{ή} = \frac{1 - ή^2}{\sqrt{n}};$$

7.2.3. Міра лінійності зв’язку (К)

Для встановлення форми зв’язку використовують показник – міру лінійності, яку обчислюють за наступною формулою:

$$К = \eta_{ух}^2 - r_{ух}^2\,; \tag{7.19}$$

При цьому якщо К ≥ 0,1 – зв'язок криволінійний;

К< 0,1 – зв'язок прямолінійний;

К = 0 – зв'язок строго прямолінійний.

Основну помилку і достовірність міри криволінійності визначають за наступними формулами:

$$m_k = \pm\frac{2}{\sqrt{N}}\sqrt{K - K^2(2-K)} \tag{7.20}$$

$$t_ф = \frac{K}{m_K} \tag{7.21}$$

При $t_ф < 3$ – значення достовірне;

$t_ф > 3$ – значення не достовірне.

В даній лекції було розглянуто приклади визначення коефіцієнта кореляції та кореляційного відношення (підрозділ 2.1. та 2.2). Залишилося визначити характер зв’язку. Для вихідних даних: висоти рослин і довжини коріння однорічного самосіву бука лісового коефіцієнт кореляції та кореляційне відношення мали однакове значення (r = ή = 0,9). Такий результат визначення цих показників свідчить про прямолінійний характер зв’язку.

Для встановлення форми зв’язку необхідно використати показник міри криволінійності (К) (формула 7.19).

К = $\eta_{ух}^2 - r_{ух}^2$ = 0,87^2 – 0,86^2 = 0,02;

Оскільки К< 0,1, то зв'язок прямолінійний.

Основна помилка m_k дорівнює 0,017 і достовірність міри криволінійності зв'язку дорівнює 1,1. Оскільки $t_ф$ < 3, то значення достовірне.

Отримані величини показників кореляції (r = 0,86; η = 0,87; К < 0,1) вказують на прямолінійний зв'язок між досліджуваними ознаками висотою рослин і довжиною коріння однорічного самосіву бука лісового. Однак такий однозначний, близький до функціонального зв'язок зустрічається далеко не завжди, особливо в біології, де числовому значенню однієї ознаки відповідає не одне значення, а цілий ряд варіюючих значень іншої ознаки.

Показники кореляції є величинами випадковими, можуть і не дорівнювати нулю навіть і при незалежному варіюванні ознак. Більш детально про такі випадки розглянемо на наступному прикладі. Виміряли висоту рослин і довжину коріння, як і в попередньому прикладі, тільки не в однорічних, а у трьох-, чотирьох- і п'ятирічних сіянців бука лісового. Необхідно з'ясувати, чи є взаємозв'язок між досліджуваними ознаками, його тісноту (табл. 7.6). і характер (таблиці 7.7 і 7.8).

Коефіцієнт кореляції визначаємо за алгоритмом, описаним вище і з використанням допоміжної таблиці 7.6.

r = Σ a_x a_y / $\sqrt{\sum a_x^2 \Sigma a_y^2}$ = 37 /$\sqrt{141{,}2 \cdot 56{,}4}$ = 0,4

Помилка коефіцієнта кореляції:

$$m_r = \pm\sqrt{\frac{1-0{,}4^2}{20-2}} = 0{,}22$$

$t_r = r/m_r$ = 0,4/0,2 = 2

Показник достовірності коефіцієнта кореляції t_r менше $t_{0,05}$ = 2,1

Таблиця 7.6

Визначення коефіцієнту кореляції між висотою рослин і довжиною коріння 3-5 річних сіянців бука лісового

Надземна частина, см	Довжина кореня, см	a_x	a_y	$a_x\ a_y$	a_x^2	a_y^2
x	y					
31	25,3	3,2	1,1	3,52	11,9025	1,21
32	26,4	4,2	2,2	9,24	19,8025	4,84
30	24,4	2,2	0,2	0,44	6,0025	0,04
28	25,9	0,2	1,7	0,34	0,2025	2,89
24	24,5	-3,8	0,3	-1,14	57,0025	0,09
26	21,0	-1,8	-3,2	5,76	2,4025	10,24
26	25,3	-1,8	1,1	-1,98	2,4025	1,21
30	25,1	2,2	0,9	1,98	6,0025	0,81
32	26,2	4,2	2,0	8,4	19,8025	4,0
26	24,7	-1,8	0,5	-0,9	2,4025	0,25
24	22,6	-3,8	-1,6	6,08	20,7025	2,56
24	26,0	-3,8	1,8	-6,84	12,6025	3,24
29	23,5	1,2	-0,7	-0,84	2,1025	0,49
27	22,8	-0,8	-1,4	1,12	0,3025	1,96
28	26,2	0,2	2,0	0,4	0,2025	4,0
30	23,7	2,2	-0,5	-1,1	6,0025	0,25
28	22,4	0,2	-1,8	-0,36	0,2025	3,24
31	24,2	3,2	0	0	11,9025	0
25	20,4	-2,8	-3,8	10,64	6,5025	14,44
25	23,4	-2,8	-0,8	2,24	6,5025	0,64
556 (Σ)	484 (Σ)	0	0	**37,0** (Σ $a_x\ a_y$**)**	**141,2** (Σa_x^2)	**56,4** (Σ a_y^2)
27,8 ($\bar{x}$)	24,2 ($\bar{y}$)					

В даному прикладі при визначенні коефіцієнта кореляції (r) між висотою рослин і довжиною коріння 3-5 річних сіянців бука лісового з'ясували, що зв'язок позитивний, середній (помірний). Однак оскільки значення t_r менше, хоч не значно за $t_{0,05}$ = 2,1, тому помірний зв'язок між досліджуваними ознаками може мати випадковий характер. В таких випадках необхідно провести додаткові дослідження.

В даному прикладі, при визначенні коефіцієнта кореляції було з'ясовано, що між ознаками є імовірність існування помірного зв'язку, тобто r < 0,7 – 0,8, тому необхідно правильно розрахувати кореляційне відношення, щоб з'ясувати характер зв'язку для даних, поданих нижче, в таблиці 7.7.

Таблиця 7.7

Визначення кореляційного відношення між висотою рослин і довжиною коріння 3-5 річних сіянців бука лісового

Надземна частина, см x	Довжина кореня, см y_i	Число спостережень (n_i)	y_x	$y_x - y_c$	$(y_x - y_c)^2$	$n_i(y_x - y_c)^2$	$y_i - y_c$	$(y_i - y_c)^2$
9,5	22,6	3	24,36	0,16	0,0256	0,0768	-1,6	2,56
	24,5						0,3	0,09
	26,0						1,8	3,24
25	20,4	2	21,9	-2,3	5,29	10,58	-3,8	14,44
	23,4						-0,8	0,64
26	21,0	3	23,66	-0,54	0,2916	0,8748	-3,2	10,24
	25,3						1,1	1,21
	24,7						0,5	0,25
27	22,8	1	22,8	-1,4	1,96	1,96	-1,4	1,96
28	22,4	3	24,8	0,6	0,36	1,08	-1,8	3,24
	25,9						1,7	2,89
	26,2						2	4
29	23,5	1	23,5	-0,7	0,49	0,49	-0,7	0,49
30	24,4	3	24,4	0,2	0,04	0,12	0,2	0,04
	23,7						-0,5	0,25
	25,1						0,9	0,81
31	24,2	2	24,75	0,55	0,3025	0,605	0	0
	25,3						1,1	1,21
32	26,2	2	26,3	2,1	4,41	8,82	2	4
	26,4						2,2	4,84
	y_c = 24,2					$\sum =$ **24,6**		$\sum =$ **56,4**
						1,2947 (σ_{yx}^2)		**2,9684** (σ_y^2)

Визначаємо кореляційне відношення за формулою 7.14 та допоміжних розрахунків, поданих в таблицях 7.7 і 7.8.

Таблиця 7.8

Визначення кореляційного відношення між висотою рослин і довжиною коріння 3-5 річних сіянців бука лісового

Надземна частина, см	Довжина кореня, см	Число спостережень	Групові середні, $\bar{x}_y$	$\bar{x}_y - x_c$	$(\bar{x}_y - \bar{x}_c)^2$	$n_i(\bar{x}_y - \bar{x}_c)^2$	$x_i - \bar{x}_c$	$(x_i - \bar{x}_c)^2$
y	x_i	(n_i)						
20,4	25	1	25	-2,8	7,84	7,84	-2,8	7,84
21,0	26	1	26	-1,8	3,24	3,24	-1,8	3,24
22,4	28	1	28	0,2	0,04	0,04	0,2	0,04
22,6	24	1	24	-3,8	14,44	14,44	-3,8	14,44
22,8	27	1	27	-0,8	0,64	0,64	-0,8	0,64
23,4	25	1	25	-2,8	7,84	7,84	-2,8	7,84
23,5	29	1	29	1,2	1,44	1,44	1,2	1,44
23,7	30	1	30	2,2	4,84	4,84	2,2	4,84
24,2	31	1	31	3,2	10,24	10,24	3,2	10,24
24,4	30	1	30	2,2	4,84	4,84	2,2	4,84
24,5	24	1	24	-3,8	14,44	14,44	-3,8	14,44
24,7	26	1	26	-1,8	3,24	3,24	-1,8	3,24
25,1	30	1	30	2,2	4,84	4,84	2,2	4,84
25,3	26	2	28,5	0,7	0,49	0,98	-1,8	3,24
	31						3,2	10,24
25,9	28	1	28	0,2	0,04	0,04	0,2	0,04
26,0	24	1	24	-3,8	14,44	14,44	-3,8	14,44
26,2	28	2	30	2,2	4,84	9,68	0,2	0,04
	32						4,2	17,64
26,4	32	1	32	4,2	17,64	17,64	4,2	17,64
	x_c = 27,8					$\sum$ = **120,7**		$\sum$ = **141,2**
						6,3526 (σ_{yx}^2)		**7,4315** (σ_y^2)

$$\eta_{y/x} = \sqrt{\frac{1,2947}{2,9684}} = 0,66; \qquad \eta_{x/y} = \sqrt{\frac{6,3526}{7,4315}} = 0,92;$$

В результаті є два значення кореляційного відношення $\eta_{y/x}$ = 0,66 та $\eta_{x/y}$= 0,92, які вказують на двосторонній зв'язок між досліджуваними ознаками.

Знаходимо помилку кореляційного відношення за формулою 7.17 і критерій Стьюдента за формулою 7.18:

$$m_{ή} = \pm\sqrt{\frac{(1-ή)^{2}}{N-2}} = \pm\sqrt{\frac{(1-0{,}66)^{2}}{18}} = \pm 0{,}08;$$

$$t_{y/x} = \frac{\eta_{y/x}}{m_{\eta}} = \frac{0{,}66}{0{,}08} = 8; \qquad t_{x/y} = 0{,}92\sqrt{\frac{20-2}{1-0{,}92^{2}}} = 9{,}95$$

Для γ = N - 2 = 18 критичне значення t (додаток 1) навіть при достовірності 99,9% дорівнює 3,922 і є меншим, ніж фактичні значення t (3,9 < 8 і 10), що свідчить про те, що кореляційне відношення достовірне.

Для встановлення форми зв'язку використовуємо показник міри криволінійності (К) (формула 7.19).

К = $\eta_{ух}^{2} - r_{ух}^{2}$ = $0{,}66^2 - 0{,}4^2 = 0$;

Оскільки К = 0,27, то зв'язок криволінійний.

$m_k = \pm\frac{2}{\sqrt{20}}\sqrt{0{,}27 - 0{,}27^{2}(2-0{,}27)} = 0{,}16$

$t_ф = \frac{K}{m_K} = t_ф = \frac{0{,}27}{0{,}16} = 1{,}68$

Основна помилка і достовірність міри криволінійності зв'язку дорівнюють нулю. Оскільки $t_ф < 3$, то значення достовірне.

7.2.4. Кореляція між якісними ознаками

Кореляція встановлюється між ознаками, що не піддаються вимірам і не розподіляються в варіаційний ряд, по наявності одного чи кількох ознак в залежності від наявності інших альтернативних ознак, що досліджуються.

Коефіцієнт асоціації

Визначення взаємозв'язку між ознаками, які групуються в чотирьох клітинну кореляційну решітку, здійснюють за допомогою коефіцієнта асоціації Дж. Юла, позначається літерою r_a і розраховується за формулою:

$$r_a = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}; \qquad (7.22)$$

де a, b, c, d – чисельності альтернативних ознак, розміщених в клітинках кореляційної таблиці.

Наприклад, сосна з плоским апофізом шишок і чорним насінням схрещена з сосною, яка має гачкуватий апофіз і біле насіння. Отримали рослини:

З плоским апофізом шишок і чорним насінням………75

З плоским апофізом шишок і білим насінням………...16

З гачкуватим апофізом шишок і чорним насінням..….14

З гачкуватим апофізом шишок і білим насінням……..68

Необхідно з'ясувати чи є зв'язок між гачкуватим апофізом шишок і білим насінням. Отримані дані групуємо в вигляді таблиці:

X / У	Чорне насіння	Біле насіння	Σn_x
Плоский апофіз шишок	a = 75	b = 16	a + b = 91
Гачкуватий апофіз шишок	c = 14	d = 68	c + d = 82
Σn_y	a + c = 89	b + d = 84	n = 173

З таблиці відповідні дані заносимо в формулу 7.22 і знаходимо значення коефіцієнта асоціації.

$$r_a = \frac{75 \times 68 - 14 \times 16}{\sqrt{89 \times 84 \times 91 \times 82}} = 0{,}65;$$

Коефіцієнт асоціації виражається в частках від одиниці, його величина змінюється в межах від нуля до одиниці. Чим сильніший зв'язок між ознаками, тим вищий коефіцієнт асоціації.

Ранговий коефіцієнт кореляції

Одним з відомих непараметричних показників зв'язку є ранговий коефіцієнт кореляції Спірмена, який визначається по формулі:

$$r_{\text{р}} = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}; \qquad (7.23)$$

тут Σ – знак суми; d – різниця між рангами сполучених значень ознак X та Y, тобто $d = x_{\text{р}} - y_{\text{р}}$; n – об'єм вибірки чи загальне число спостережень.

Згідно даного способу, щоб з'ясувати наявність зв'язку між ознаками, необхідно ранжирувати їх значення і спостерігати, як вони розміщуються по відношенні один до одного. Якщо значенням ознаки, що зростає X відповідає зростаюче значення іншої ознаки Y, то між ними позитивний зв'язок. Якщо при зростанні значення однієї ознаки значення іншої послідовно спадають, це свідчить про наявність негативного зв'язку. Коли зв'язок між ознаками відсутній, ранжированим значенням однієї з них будуть відповідати невпорядковані значення іншої ознаки.

Позначивши ранжировані значення ознак порядковими числами натурального ряду, можна розраховувати ранги цих значень і по їх різниці визначати степінь взаємозв'язку між ознаками. У випадку, якщо наявний фунціональний зв'язок ранги ранжированих значень ознак повністю співпадають між собою і різниця між ними буде дорівнювати нулю. В таких випадках $r_{\text{р}}$=1. Якщо обидві ознаки варіюють незалежно одна від одної, то ранговий коефіцієнт дорівнює нулю. Отож, ранговий коефіцієнт може змінюватися в межах від нуля до одиниці, тобто виражається в частках від одиниці з позитивним і негативним значенням, тобто -1$< r_{\text{р}} < 1$.

Як і всі попередньо розглянуті показники, коефіцієнт рангової кореляції є величиною випадковою, і тому потребує оцінки достовірності його значень. Нульова гіпотеза в припущенні, що в генеральній сукупності $r_{\text{р}}$= 0, тобто кореляція між ознаками відсутня. Критерієм оцінки цієї гіпотези є:

$$t_p = r_p \sqrt{\frac{n-2}{1-r_p^2}} \tag{7.24}$$

Нульова гіпотеза відкидається, якщо фактичне t_p перевищить критичне значення або буде дорівнювати t_{st} (додаток 1) для вибраного рівня значимості Р і числа ступеня вільності k = n – 2.

Переваги рангового коефіцієнта в тому, що він дозволяє вимірювати степінь взаємозв’язку між варіюючими ознаками незалежно від закону їх розподілу і форми зв’язку. Даний коефіцієнт кореляції дозволяє вимірювати зв'язок між ознаками, які не піддаються точному кількісному вимірюванню, але можуть бути ранжировані по місцезнаходженні, що займають в загальній сукупності спостережень. Разом з цим ранговий коефіцієнт не є настільки точним, як параметричні показники – коефіцієнт кореляції чи кореляційне відношення. Тому ранговий коефіцієнт кореляції слід використовувати лише в тих випадках, коли по певних причинах не можна застосувати параметричні показники. Наприклад, коли досліджувані ознаки, що варіюють оцінюють балами, відсотками або іншими умовними одиницями.

Ранговий коефіцієнт в окремих випадках знаходить широке використання в практиці, але ні в якому разі не може замінити параметричні показники кореляційного зв’язку.

Запитання для самоконтролю:

1. Поясніть суть сумарного показника зв'язку.
2. Що таке функціональна залежність і кореляція.
3. Дайте визначення коефіцієнта кореляції, поясніть його суть.
4. Що таке кореляційне відношення та визначення його достовірності.
5. Міра криволінійності зв’язку. Наведіть формулу.
6. Які є непараметричні показники для визначення кореляційного зв’язку?
7. Що таке множинна кореляція?

РОЗДІЛ 8
РЕГРЕСІЙНИЙ АНАЛІЗ

8.1. Поняття про регресію

Показники коефіцієнт кореляції, кореляційне відношення і показник міри криволінійності, розглянуті в попередньому розділі, являють собою споріднену групу характеристик, які визначають тісноту і характер зв'язку між досліджуваними ознаками, напрямок зв'язку (прямий або зворотній) і його форму (прямолінійний або криволінійний). Але великий інтерес, що займає вагоме місце в галузі статистичного аналізу масових явищ має уявлення про те, наскільки в середньому може змінитися ознака, що варіює, при зміні на одиницю вимірювання іншої, пов'язаної з нею ознаки, чого наведені показники не визначають.

Функція, з допомогою якої можна за величиною однієї ознаки (X) знаходити середні (очікувані) значення іншої ознаки ($\bar{y}_x$), пов'язаної з X кореляційно, називається ***регресією***. Це слово латинської мови − *regressio* від *redior* − повертаюсь. В математиці воно означає імовірну залежність середнього значення будь-якої величини від іншої величини. Статистичний аналіз регресії отримав назву ***регресійного аналізу***.

Регресійний аналіз тісно пов’язаний з кореляційним. Але є відмінності між показниками кореляції та регресії. Насамперед, показники регресії вимірюють взаємозв’язок між ознаками X і Y, враховуючи зміни X залежно від змін Y і, навпаки, зміни У і X. Винятком служать так звані ряди динаміки, або часові ряди, що показують зміну ознак у часі. Регресія таких рядів є односторонньою. Крім того, показники регресії характеризують залежність між змінними X і У за їх абсолютним значенням, тоді як показники кореляції – величини відносні, що вимірюють тісноту зв'язку між ознаками в частках одиниці.

Регресія – це ряд групових середніх y_x (або x_y), що показує динаміку мінливості ознаки У (або X) залежно від зміни значень ознаки X (або У).

Термін "регресія" в біологію ввійшов завдяки Ф. Гальтону, який вивчав зв'язок між ростом батьків та їх дітей. Він встановив "закон регресивної спадковості", коли діти дуже високих і дуже низьких батьків мають тенденцію "регресувати" в своєму розвитку в бік середнього, для даної популяції людей росту.

Аналіз регресійної залежності може бути представлений у вигляді рядів регресії, у вигляді кривої регресії, а також у вигляді крапкового графіку.

Емпіричні ряди регресії − це подвійні ряди значень ознак аргументу і значень відповідних ознак функції, які формуються з даних, одержаних в експериментальних дослідженнях. При їх графічному відтворенні критерії значень аргументу відкладаються на осі абсцис, а критерії значень функцій − на осі ординат. Якщо для кожного значення (х) вирахувати по декількох вимірах середнє значення (у), то одержані результати при наявності кореляційної залежності, нанесені на систему координат дають ***емпіричну лінію регресії*** (лінію, яка з'єднує середні значення залежної ознаки (у) відносно значення незалежної ознаки (х)), тобто її графічне зображення. Відповідно до характеру кореляційного взаємозв'язку лінія регресії може бути прямою або кривою. Для емпіричних ліній характерно те, що в межах одних інтервалів аргументу функція може мати або підвищене (крива іде догори) або понижене (крива іде вниз), або від'ємне прирощення (крива іде в зворотному напрямку). Крапковий графік – це, коли на системі координат відкладаються крапки, які фіксують відповідні значення функції щодо аргументу.

Вибір математичного рівняння, яке в найбільшій мірі відповідає характеру залежності між ознаками, що досліджуються, знаходження конкретних значень коефіцієнтів цього рівняння, визначення стандартної помилки обчисленого значення; оцінка точності визначення рівняння; знаходження теоретичних (найбільш ймовірних) значень залежної ознаки; вирівнювання емпіричних рядів; оцінка точності вирівнювання; визначення ефекту регресії при вимірюванні варіації залежної ознаки є завданнями регресійного аналізу після проведення кореляційного аналізу.

Регресивним аналізом визначається також ступінь варіювання залежної ознаки по відношенню до незалежної ознаки.

8.2. Види і особливості регресії

Як вже відзначалось, зв'язок між взаємозалежними змінними величинами X і Y має математичний вираз у вигляді: y = f (x). Наведене рівняння характеризує кореляційний характер зв'язків між X і Y і, водночас, відображає регресійну особливість, тобто є регресивним рівнянням. Це рівняння дозволяє визначити середні значення однієї ознаки (Y) за відповідними значеннями іншої ознаки (X). В правій частині даного рівняння коефіцієнт (f) є числовим значенням, яке фіксує, наскільки змінюється значення (Y) відносно значення (X). Це є *коефіцієнт регресії або міра регресії.* Він показує, на яке число збільшиться значення (Y) при збільшенні (X) на відповідну одиницю міри. При графічному зображенні від величини (f) залежить крутість розміщення лінії, що з'єднує одержані значення (Y).

Для всіх значень (x), що аналізуються, цей коефіцієнт в разі прямолінійної залежності завжди лишається постійним коефіцієнтом, що визначає міру збільшення або зменшення (y) щодо значень (x) або (x) щодо (y). Фактичне положення лінії (Y) відрізняється від лінії (X) на відповідну сталу величину (a), від якої залежить положення лінії f (x) в системі координат.

Отже, лінія регресії в системі координат в разі залежності y = f (x) буде визначатись формулою прямої лінії: y = a + bx. Тут (a) – вільний член (сталий коефіцієнт), який означає рівень регресії, тобто її положення над віссю абсцис, який відраховується на початку координат від осі (x); а (b) − коефіцієнт пропорційності, який зветься коефіцієнтом регресії або мірою регресії, і вказує наскільки збільшиться (або зменшиться) значення змінної ознаки (y), якщо значення (x) зміниться на одиницю виміру.

Криволінійна регресія має місце тоді, коли із зміною (*X*) на відповідну одиницю виміру (*Y*) з кожним наступним збільшенням (*X*) змінюється не на сталу величину, а на якусь частинку більше або менше. Розмір цієї частинки регламентується ще одним коефіцієнтом "*C*", перемноженим на (x^2), тобто Cx^2. Тоді формула лінійної регресії приймає вигляд $y = a + bx + cx^2$, а крива на графіку буде відповідно згинатись від ординати (*X*) вверх (якщо значення Cx^2 буде додатнім) або вниз (значення cx^2 – від'ємне).

Криволінійність регресій може бути більш складною в порівнянні з наведеною. Тоді у математичну формулу регресії будуть включатись коефіцієнти, перемножені на наступні степені (x) – x^3; x^4 і т. д.

Відображення характеру залежності ознак (*Y*) і (*X*) відповідним математичним рівнянням зветься її *апроксимацією*

(апроксимація – від лат. approximatio – зближення, тобто наближене зображення одних математичних об'єктів іншими, наприклад: ламаних ліній – кривими.)

Лінія регресії і відповідно до неї формула апроксимації обираються згідно, характером залежності одержаних експериментальних даних, який апробується за даними ранжированих рядів розподілу, а краще – за характером їх графічного виразу.

Звичайно за лінію вирівнювання приймають геометрично найбільш прості лінії: пряму, параболічну, логарифмічну, для яких за експериментальними даними складають рівняння зв'язку.

При цьому можуть бути такі випадки.

Якщо із збільшенням однієї ознаки спостерігається пропорційне збільшення або зменшення другої ознаки, за лінією вирівнювання обирається пряма лінія, рівняння якої $y = a + bx$; (рис 8.1 а).

Якщо зміна залежного показника виражається плавною кривою з одним вигином, для вирівнювання беруть параболу другого порядку: $y = a + bx + cx^2$ (рис. 8.1 b).

В більш складних випадках – для кривих ~ подібної форми, що мають два вигини, для вирівнювання експериментальних даних використовують параболу третього порядку: $y = a + bx + cx^2 + dx^3$ (рис. 8.1 с).

Якщо із збільшенням незалежної ознаки спостерігається уповільнене збільшення другої ознаки, застосовують логарифмічну криву: $y = \frac{x^2}{a+bx+cx^2}$ (рис.8.2 а).

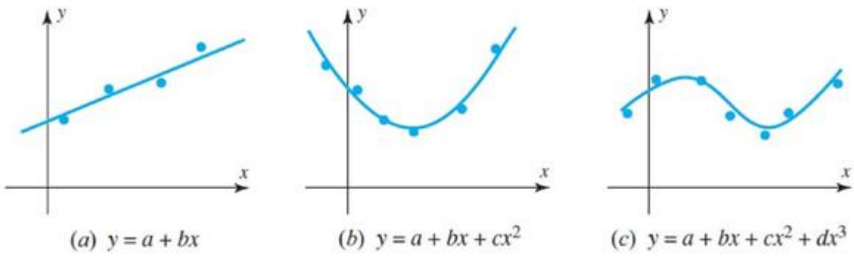
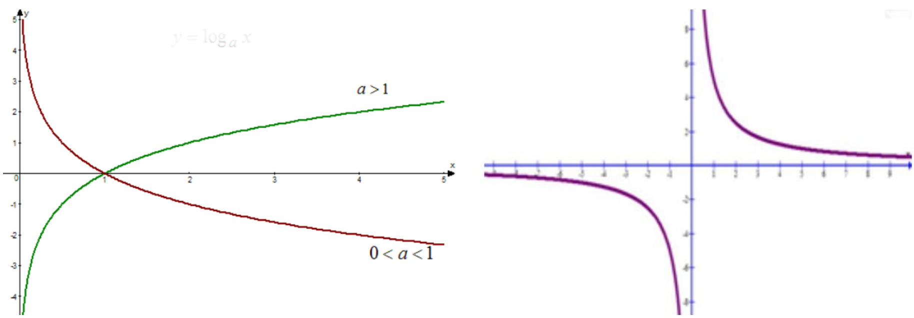


Рис. 8.1. Графічне відображення лінійної (a)
і параболічної (b і c) залежності.

Коли залежна ознака при збільшенні незалежної поступово зростає і це зростання переходить в пропорційне збільшення, для вирівнювання береться крива типу $y = ax^b + c$.



(a) логарифмічні криві (b) криві гіперболи

Рис. 8.2. Графічне відображення логарифмічної гіперболічної залежності

Дуже часто ознаки біологічних об'єктів знаходяться в зворотній залежності одна від одної. Таку залежність обраховують рівняннями гіперболи: $y = \frac{a+b}{x}$; або $y = \frac{a}{x} + b$ (рис.8.2.b).

В наведених формулах x − незалежна змінна; y − залежна змінна; a, b, c – постійні коефіцієнти, які підлягають визначенню.

Доцільно перед проведенням розрахунків рівнянь залежності графічно встановити відповідність характеру розподілу ознак тієї або іншої кривої розподілу.

8.3. Вирівнювання кривих регресії

Лінія регресії описується (апроксимується) відповідною математичною формулою.

Процес розрахунку (визначення) середнього значення (y) для кожного відповідного значення (x) за відповідною математичною формулою має назву ***вирівнювання експериментальних даних.***

Якщо по осі абсцис відкласти класи (інтервали) значення незалежної ознаки, а по осі ординат − класи залежної ознаки і досліджувані дані спостережень нанести на графік, то, з'єднавши крапки на графіку прямими лініями, одержимо деяку ламану лінію. Але якщо спостережень (y) було в достатньо великій кількості, ця лінія повинна бути плавною кривою. Тобто ця лінія має вигляд ламаної внаслідок недостатньої кількості спостережень.

Після одержання ламаної лінії для позначення вибіркової сукупності варіант (тобто експериментальну або дослідну лінію) її необхідно вирівняти, тобто замінити на плавну згладжену лінію, що означає знайти закон її зміни у вигляді математичної формули.

Для цього застосовуються наступні методи.

Графічній спосіб

Графічний спосіб вирівнювання полягає в тому, що після нанесення на графік емпіричних даних у вигляді ламаної лінії або у вигляді окремих крапок на око визначаються серединні крапки лінії регресії, які після цього з'єднуються суцільною лінією, яка є вирівняною лінією регресії.

Це найбільш простий спосіб, який не потребує розрахунків. Недоліком є те, що він не виключає вплив індивідуальних властивостей дослідника на результати вирівнювання. Тому, у випадках, коли потребується більша точність вирівнювання рядів надають перевагу іншим.

Вирівнювання за способом ковзної середньої

Більш точні результати вирівнювання маємо, коли спочатку за даними двох або трьох послідовних експериментальних значень ряду регресії знаходимо середнє значення і потім з'єднуємо їх плавною кривою лінією. Наприклад, є емпіричні дані про довжину коренів висоту рослин сіянців сосни звичайної:

Довжина коріння, см	4	4	5	5	5	6	6	6	7	7
Довжина стовбурців, см	3,0	3,1	3,5	3,5	4,1	3,5	4,0	5,0	5,0	5,3

Знаходимо суму перших двох значень ряду: 3,0 + 3,1 + 3,5 = 9,6; Потім визначаємо суму наступних трьох значень, що знаходяться за першим: 3,1 + 3,5 + 3,5 = 10,1; потім 3,5 + 3,5 + 4,1 = 11,1 і т. д. до кінця ряду. Отримані суми ділимо на число доданків, тобто в даному прикладі 3 і отримуємо середні значення ряду: 3,2; 3,4; 3,7; 3,7; 3,9; 4,1; 4,7; і 5,1.

Спосіб ковзної середньої простий і особливо зручний в тих випадках, коли емпіричний ряд включає велику кількість значень і втрата двох із них (крайніх) помітно не впливає на його структуру. Крім того, середні величини можна отримувати із двох, трьох і більше значень емпіричного ряду.

Наведені вище два способи дають лише орієнтовні результати, які можна використовувати лише для загальних уявлень про характер регресії.

Вирівнювання ліній регресії за способом найменших квадратів

В 1806 році Гаусом був запропонований спосіб, в основу якого покладена вимога, щоб сума квадратів відхилень варіант від середньої арифметичної була найменшою, тобто $\Sigma(y_i - y)^2 = min$. Звідси і назва способу.

Для рішення вирівнювання за цим способом необхідно:

1. Провести групування фактичного матеріалу в емпіричні ряди. Скласти графічне зображення емпіричних рядів.
2. Скласти початкове рівняння регресії.
3. В початкове рівняння підставити відповідні емпіричні ряди, утворюючи систему нормальних рівнянь.
4. Рішення сумісно одержаних рівнянь визначають їх параметри.
5. Підставивши значення параметрів в загальне початкове рівняння, одержують емпіричне рівняння регресії, яке виражає функціональну залежність між змінними X і Y.

8.4. Лінійна регресія

Зв'язок між X і Y можна відобразити за допомогою математичного рівняння загального виду ***y = f (X)***, де символом f (X) позначена форма рівняння, яка більш або менш повно відображає функціональну залежність середньої величини однієї змінної ($\bar{y}_x$) від значення другої змінної (x). Такі математичні рівняння мають назву ***кореляційних*** або ***регресійних.***

Як вже зазначалося, залежність між біологічними ознаками може бути різною, але в більшості випадків мають вираз простого рівняння лінійної залежності:

$$y = a + bx, \tag{8.1}$$

де a і b − параметри рівняння − "a" − вільний член рівняння, а "b" є показником пропорційності, він зветься ***коефіцієнтом***

регресії. Тому насамперед розглянемо визначення коефіцієнтів лінійної залежності.

При обчисленні коефіцієнтів рівнянь регресії можна використовувати наступні методи: графічний спосіб, спосіб вибраних точок, спосіб найменшої помилки, за центральним відхиленням, за способом Маркова, за коефіцієнтом кореляції, спосіб найменших квадратів, за числовими коефіцієнтами (метод Труля). Деякі з цих методів визначення коефіцієнтів рівнянь регресії розглянемо детальніше.

Окомірно будують лінію регресії, яка проходить через центр сукупності точок (графічний спосіб).

Коефіцієнт **a** визначають, замірявши відрізок від початку координат до точки перетину прямої з віссю OY; заміряючи кут нахилу **a** і через нього визначають коефіцієнт **b** за формулою:

$$b = \mathrm{tg}\, a \qquad (8.2)$$

Для визначення параметрів a і b застосовується система нормальних рівнянь:

$$\begin{cases} \Sigma y = an + b\Sigma x \\ \Sigma xy = a\Sigma x + b\Sigma x^2 \end{cases}; \qquad (8.3)$$

де n – об'єм парних спостережень або число членів емпіричного ряду регресії.

Щоб скласти систему нормальних рівнянь по об'єму парних спостережень або числу членів емпіричного ряду регресії n, необхідно попередньо визначити значення таких величин: Σy; Σx; Σxy; Σx^2, для цього складається допоміжна таблиця.

Визначення коефіцієнтів рівнянь за способом найменших квадратів для рівняння прямої $y = a + bx$ розглянемо на прикладі залежності довжини коріння (y) від висоти рослин (x) однорічного самосіву бука лісового (таблиця 8.1).

Таблиця 8.1

Розрахунок допоміжних величин для визначення коефіцієнтів рівнянь за способом найменших квадратів для рівняння прямої y = a + bx

x	*y*	*xy*	x^2
11	12,5	137,5	121
14	14,8	207,2	196
12	12,8	153,6	144
13	13,6	176,8	169
11	12,4	136,4	121
12	13,0	156	144
13	14,6	189,8	169
12	14,2	170,4	144
13	14,0	182	169
13	14,1	183,3	169
(Σ) 124	(Σ) 136	1693	1546

Отож, маємо Σx = 124; Σy = 136; Σxy = 1693; Σx^2 = 1546; n = 10 (таблиця 8.1). Графічне зображення подано на рисунку 8.1 (а).

Відповідно, нормальні рівняння в даному випадку будуть такі:

$$\begin{cases} 10a + 124b = 136 - 1 \text{ рівняння} \\ 124a + 1546b = 1693 - 2 \text{ рівняння} \end{cases}$$

Поділивши всі члени на коефіцієнти при b, які дорівнюють 124 та 1546, маємо:

$$\begin{cases} 0{,}0802a + b = 1{,}0950 \\ 0{,}0806a + b = 1{,}0967 \end{cases}$$

Розв’язуємо систему рівнянь:

В першому рівнянні системи 0,0802a + b = 1,0950 знаходимо: b =1,0950 – 0,0802a, підставляємо в друге рівняння: 0,0806a+1,0950 – 0,0802 a = 1,0967. Отримуємо 0,0004a = 0,0017, звідси a =

0,0017/0,0004 = 4,25. Потім знаходимо b =1,0950 – 0,0802×4,25 = 0,7541.

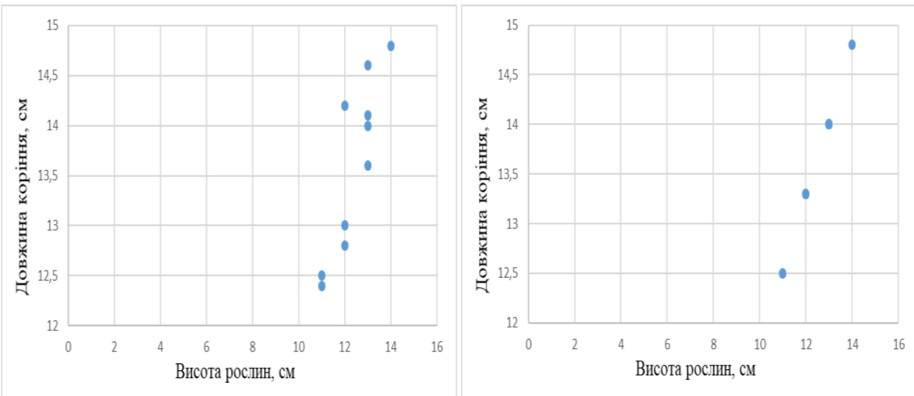
Отож, рівняння регресії: y = 4,25 + 0,7541x;

Підставляємо в це рівняння замість x конкретні значення висоти рослин, можна визначити середню довжину коріння. А саме:

Висота рослин, x	11	14	12	13	11	12	13	12	13	13
Довжина коріння, $\bar{y}$	12,5	14,8	13,3	14	12,5	13,3	14	13,3	14	14

Графічне відображення регресії довжини коріння на висоту однорічного самосіву бука лісового після вирівнювання ліній регресії за способом найменших квадратів подано на рисунку 8.3 б.

З спільного рішення системи нормальних рівнянь, можна отримати формули (8.4 а; b – 8.7 а; b), які значно полегшують роботу при визначенні параметрів *a* і *b*, особливо для великих вибірок і при наявності багаточисельних числових значень взаємозв’язаних ознак.



а б

Рис. 8.3. Регресія довжини коріння на висоту однорічного самосіву бука лісового (а – до вирівнювання; б – після вирівнювання)

Вільний член рівняння лінійної функції – сталий коефіцієнт **а** визначається за формулами (8.4 а і b):

$$a_{y/x} = y_c - b_{y/x}\, x_c; \qquad a_{x/y} = x_c - b_{x/y}\, y_c; \qquad (8.4\ a;\ b)$$

де коефіцієнти регресії $b_{y/x}$ і відповідно $b_{x/y}$ визначають за формулами:

$$b_{y/x} = \frac{\sum xy - nx_c y_c}{\sum x^2 - nx_c^2}; \qquad b_{x/y} = \frac{\sum xy - nx_c y_c}{\sum y^2 - ny_c^2} \qquad (8.5\ a;\ b)$$

або $$b_{y/x} = \frac{n\sum xy - \Sigma x \cdot \Sigma y}{n\sum x^2 - (\Sigma x)^2}; \qquad b_{x/y} = \frac{n\sum xy - \Sigma x \cdot \Sigma y}{n\sum y^2 - (\Sigma y)^2}; \qquad (8.6\ a;\ b)$$

або $$b_{y/x} = \frac{\Sigma xy - \frac{1}{n}(\Sigma x \cdot \Sigma y)}{\sum x^2 - \frac{1}{n}(\Sigma x)^2}; \quad b_{y/x} = \frac{\Sigma xy - \frac{1}{n}(\Sigma x \cdot \Sigma y)}{\sum x^2 - \frac{1}{n}(\Sigma x)^2}; \qquad (8.7\ a;\ b)$$

Визначення параметрів а і b за наведеними формулами розглянемо, наприклад, для даних таблиці 8.1. Використовуючи формули 8.4 та 8.5, знайдемо емпіричне рівняння регресії довжини коріння на висоту однорічного самосіву бука лісового. Попередньо розрахуємо допоміжні величини Σy, Σx, Σxy, Σx^2, Σy^2 (таблиця 8.2) і підставляємо у формули 8.5 а і 8.4 а:

$$b_{y/x} = \frac{1693 - 10 \cdot 12{,}4 \cdot 13{,}6}{1546 - 10 \cdot 12{,}4^2} = \frac{6{,}6}{8{,}4} = 0{,}7857;$$

а = 13,6 - 0,7857·12,4=3,86;

Звідки, у = 3,86 + 0,7857х;

Аналогічно визначаються параметри а і *b* для знаходження емпіричного рівняння регресії висоти однорічного самосіву бука лісового на довжину коріння, при цьому використовуються формули 8.4 *b* та 8.5 *b*.

Таблиця 8.2

Розрахунок допоміжних величин для визначення коефіцієнтів рівнянь

x	*y*	*xy*	x^2	y^2
11	12,5	137,5	121	156,25
14	14,8	207,2	196	219,04
12	12,8	153,6	144	163,84
13	13,6	176,8	169	184,96

Продовження таблиці 8.2

11	12,4	136,4	121	153,76
12	13,0	156	144	169
13	14,6	189,8	169	213,16
12	14,2	170,4	144	201,64
13	14,0	182	169	196
13	14,1	183,3	169	198,81
(Σ) 12,4	(Σ) 136	1693	1546	1856,46

Параметр b (коефіцієнт регресії) можна визначати *за коефіцієнтом кореляції* за такими формулами:

$$b_{x|y} = r\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{\Sigma(y_i-\bar{y})^2}}; \qquad b_{y|x} = r\sqrt{\frac{\Sigma(y_i-\bar{y})^2}{\Sigma(x_i-\bar{x})^2}}; \qquad (8.8 \text{ a; b})$$

або з врахуванням середніх квадратичних відхилень (σ_y і σ_x) взаємозв'язаних розподілів та класового інтервалу (i_y; i_x):

$$b_{x|y} = r\cdot\frac{\sigma_y\cdot i_y}{\sigma_x\cdot i_x}; \qquad b_{y|x} = r\cdot\frac{\sigma_x\cdot i_x}{\sigma_y\cdot i_y}; \qquad (8.9 \text{ a; b})$$

По значеннях коефіцієнта регресії $b_{x|y}$ і $b_{y|x}$ визначається коефіцієнт кореляції:

$$r = \sqrt{b_{y|x}\cdot b_{x|y}}; \qquad (8.10)$$

8.5. Оцінка достовірності показників регресії

Показники регресії, як і всі інші вибіркові параметри, є величинами випадковими: їх значення може не співпадати з значеннями генеральної сукупності. Для визначення можливої похибки, з якої визначаються вибіркові показники відносно своїх генеральних параметрів, є помилки репрезентативності, які дозволяють з тією чи іншою імовірністю встановлювати довірчі межі для генеральних параметрів по даних вибіркових спостережень, оцінювати статистичну достовірність показників регресії.

Помилку рівняння регресії визначають за формулою:

$$m_{yx} = \sqrt{\frac{\Sigma(y_i - y_t)^2 n_i}{n-f}}; \qquad (8.11)$$

де $y_i - y_t$ – відхилення фактичних значень результуючої ознаки від теоретичних;

n – кількість пар спостережень;

f – число коефіцієнтів рівняння.

Достовірність коефіцієнтів лінійної регресії визначають за наступними формулами:

коефіцієнта а:

$$t_{ф} = \frac{a}{m_{yx}\sqrt{\frac{1}{n}+\frac{1}{n}\left(\frac{x_C}{\sigma_x}\right)^2}}; \qquad (8.12)$$

коефіцієнта b:

$$t_{ф} = \frac{b\sigma_x\sqrt{n-1}}{m_{yx}}; \qquad (8.13)$$

де m_{yx} – помилка рівняння регресії;

а – вільний член рівняння;

X_c – середнє значення незалежної ознаки;

σ_x – стандартне відхилення ряду X;

b – коефіцієнт лінійної регресії;

Оцінку достовірності показників регресії розглянемо на прикладі, наведеному в таблиці 8.1. Визначили ряд допоміжних величин, розрахунок яких поданий в таблиці 8.3.

Визначаємо: помилку рівняння регресії – $m_{yx} = \sqrt{\frac{1{,}69}{8}} = 0{,}45$ см

достовірність коефіцієнта **а** лінійної регресії – $t_{ф} = \frac{3{,}86}{0{,}45\sqrt{\frac{1}{10}+\frac{1}{10}\left(\frac{12{,}4}{0{,}966}\right)^2}} = 2{,}10$

достовірність коефіцієнта **b** лінійної регресії – $t_{ф} = \frac{0{,}7857\cdot 0.966\sqrt{10-1}}{0{,}45} = 5{,}06$;

При v = n – f = 10-2 = 8, $t_{5\%}$= 2,31

Отже, коефіцієнт **a** є достовірним (2,10<2,31), а коефіцієнт регресії **b** не достовірним (5,06 >2,31) на 5% рівні значимості.

Таблиця 8.3

Допоміжна таблиця для визначення помилки рівняння регресії

x	y_i	y_t	$y_i - y_t$	$(y_i - y_t)^2 n_i$
11	12,5	12,5	0	0
14	14,8	14,8	0	0
12	12,8	13,3	-0,5	0,25
13	13,6	14	-0,4	0,16
11	12,4	12,5	-0,1	0,01
12	13,0	13,3	-0,3	0,09
13	14,6	14	0,6	0,36
12	14,2	13,3	0,9	0,81
13	14,0	14	0	0
13	14,1	14	0,1	0,01
(Σ) 124	(Σ) 136			$\Sigma(y_i - y_t)^2 n_i = \mathbf{1{,}69}$

Коли відомі *середні квадратичні відхилення* і *коефіцієнт кореляції* вибіркових розподілів, *помилку коефіцієнта* регресії визначають за формулами:

$$m_{byx} = \frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-r^2}{n-2}}; \qquad m_{bxy} = \frac{\sigma_x}{\sigma_y}\sqrt{\frac{1-r^2}{n-2}}; \qquad (8.14\ a;\ b)$$

Помилку коефіцієнта регресії можна вирахувати без визначення середніх квадратичних відхилень за такими формулами:

$$m_{byx} = \frac{\Sigma a_y^2 - \frac{(\Sigma a_x a_y)^2}{\Sigma a_x^2}}{(n-2)\cdot\Sigma a_x^2}; \qquad m_{bxy} = \frac{\Sigma a_x^2 - \frac{(\Sigma a_x a_y)^2}{\Sigma a_y^2}}{(n-2)\cdot\Sigma a_y^2}; \qquad (8.15\ a;\ b)$$

де a_y і a_x – відхилення варіант від середніх арифметичних, тобто $a_y = y_i - \bar{y}$ і $a_x = x_i - \bar{x}$.

8.6. Довірча зона лінійної регресії

Помилка регресії дозволяє визначити імовірну зону, або область, тих випадкових відхилень вибіркової лінії, в межах яких знаходиться справжня лінія регресії, тобто лінійна регресія генеральної сукупності. Максимальна помилка репрезентативності чи похибка (Δ) вибіркової лінії регресії визначається формулами:

$$\Delta_{y/x} = t\sigma_{y|x}; \qquad \Delta_{y/x} = t\sigma_{y|x}; \qquad (8.16\ a;\ b)$$

Звідси лінійна регресія, з врахуванням можливої максимальної похибки може бути виражена наступними рівняннями:

$$\bar{y}_x = (a + b_{y|x}x) \pm t\sigma_{y|x}; \qquad (8.17)$$

$$\overline{x_y} = (a + b_{x|y}y) \pm t\sigma_{x|y}; \qquad (8.18)$$

За допомогою цих рівнянь можна визначити довірчі межі лінії регресії для будь якого рівня імовірності (Р). Так, маючи на увазі, що в межах від $\bar{y} - 2{,}31\sigma$ до $\bar{y} + 2{,}31\sigma$ ($t = 2{,}31$, при $v = n - f$ з таблиці додатку 1 для прикладу, розглянутому в даному розділі) знаходяться майже всі 100% варіант сукупності нормального розподілу, за помилкою регресії $\sigma_{y|x} = 0{,}45$ визначаємо величину максимальної похибки $\Delta_{y/x} = 2{,}31 \cdot 0{,}45 = 1$ см. Відкладаємо цю величину в обидві сторони від теоретично визначеної за способом найменших квадратів лінії регресії, отримаємо межі довірчого інтервалу для даного рівня достовірності. Тобто, наприклад, значення першої варіанти $x = 11$ (таблиця 8.1), підставляємо в рівняння регресії $y = (3{,}86 + 0{,}7857 \cdot 11) \pm 1$. А це означає, що:

нижня межа – 12,5 – 1 = 11,5;

верхня межа – 12,5 + 1 = 13,5.

Якщо для всіх варіант знайти межі довірчого інтервалу і провести через ці крапки прямі (вздовж лінії прогресії), отримаємо довірчу зону (область) можливих (випадкових) відхилень вибіркової лінії регресії від її положення в генеральній сукупності. На рисунку 8.4 пунктиром позначено межі довірчої зони.

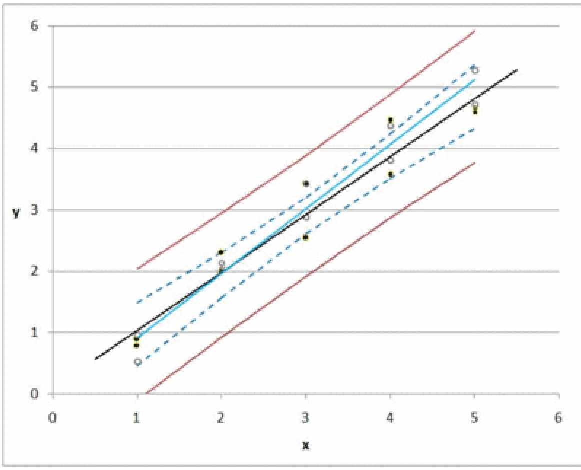


Рис. 8.4. Лінія регресії та межі її довірчої зони

Аналіз емпіричних регресій має практичне значення. За рівнянням регресії, можна визначити, наприклад, відхилення від норми, яка прийнята для популяції рослин, тварин і т. п.

8.7. Параболічна залежність

Параболічна регресійна залежність найбільш часто зустрічається при біометричних дослідженнях. Загальне рівняння параболи: $y = a + bx + cx^2 + dx^3 + ... + ex^n$, тобто рівняння параболи є розвитком лінійної функції, де X має першу ступінь (a+bx), з подальшим зростанням ступеня при X.

В цій формулі y і x – змінні величини (залежна і незалежна); a, b, c, d, e – параметри (коефіцієнти) рівняння.

В залежності від того, на якому члені полінома зупиняться розрахунки, формується парабола першого, другого, третього і т. д. порядку, і застосовується для розрахунків параметрів відповідна система рівнянь.

Система нормальних рівнянь для визначення параметрів параболи другого порядку $y = a + bx + cx^2$:

$$\begin{cases} \Sigma y = an + b\Sigma x + c\Sigma x^2; \\ \Sigma yx = a\Sigma x + b\Sigma x^2 + c\Sigma x^3; \\ \Sigma yx^2 = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4; \end{cases} \quad (8.19)$$

Для складання системи нормальних рівнянь за емпіричними даними необхідно визначити значення: Σy; Σx; Σxy; Σx^2; Σx^3, Σx^4; Σyx^2. Це робиться із застосуванням наступного макету допоміжної таблиці:

x	y	xy	x^2	yx^2	x^3	x^4

Визначити параметри рівняння регресії можна, розв’язавши систему нормальних рівнянь.

8.8. Рівняння показової, степеневої функцій, логарифмічної і гіперболічної залежностей

Особливість кореляційних зв'язків між багатьма ознаками біологічних об'єктів часто вимагають застосування для їх апроксимації більш складних функціональних залежностей, виявлення яких доцільно досліджувати застосуванням напівлогарифмічних і логарифмічних видів парабол. Застосування логарифмів дозволяє складні криволінійні залежності замінити на прямолінійні.

Важливе значення має простота моделі та кількість коефіцієнтів, що підлягають визначенню. Тому коло виразів, з яких вибирають рівняння регресії, має бути обмежений по можливості простими функціями. З них особливе місце посідають ті, які шляхом алгебраїчних перетворень можуть бути приведені до лінійного виду. Це дозволяє застосувати відносно прості, але глибоко розроблені методи статистичної оцінки. До них в першу чергу належать показові, степеневі, логарифмічні та гіперболічні функції. У таблиці 8.4 наведено типи рівнянь, які найчастіше використовують у біологічних дослідженнях. Ці моделі з однією незалежною змінною і з двома коефіцієнтами, що приводяться до лінійного вигляду. У таблиці показаний вид перетворення та критерії, що дозволяють оцінити прийнятність даного виразу.

До напівлогарифмічних відносяться параболи виду

$y = a + b\,lg\,x$ і $y = a + b\,lg\,x + c(lg\,x)^2$.

Логарифмічні параболи описуються рівняннями

$\lg y = a + b \lg x$ і $\lg y = a + b \lg x + c(\lg x)^2$.

Система нормальних рівнянь логарифмічно перетворених показових функцій типу $y=ae^{bx}$ для визначення параметрів a і b :

$$\begin{cases} \Sigma \lg y = nlg\,a + lg\,b\Sigma x; \\ \Sigma x \lg y = \lg a\,\Sigma x + \lg b\Sigma x^2\,; \end{cases} \qquad (8.26)$$

Рішення рівнянь напівлогарифмічних і логарифмічних парабол аналогічно рішенню простих парабол, з тією різницею, що тут значення варіант ознак, що досліджуються, представлені у логарифмічному вигляді.

Гіперболічна залежність описується рівнянням: $y = \frac{a+b}{x}$. Для визначення параметрів цього рівняння використовується система нормальних рівнянь:

$$\begin{cases} \Sigma y = a\Sigma \frac{1}{x} + bn \\ \Sigma \frac{y}{x} = a\Sigma \frac{1}{x^2} + b\Sigma \frac{1}{x} \end{cases}; \qquad (8.27)$$

Для того, щоб скласти систему рівнянь, необхідно визначити значення Σy; $\Sigma\frac{y}{x}$; $\Sigma\frac{1}{x}$ і $\Sigma\frac{1}{x^2}$, і застосовується така форма допоміжної таблиці:

x	*y*	x^2	*y/x*	*1/x*	$1/x^2$

Таблиця 8.4

Деякі елементарні функції, що найчастіше використовуються в біології, однією змінною та перетворення їх до лінійного вигляду

Рівняння	Тип перетворення	Рівняння прямої	
$y=ax^b$	логарифмічне	lgy = lga+blgx	(8.20)
$y=ae^{bx}$	логарифмічне	lgy=lga+(blge)x	(8.21)
$y=a+bx^n$ (n – відоме)	заміна $x'=x^n$ $y'=y$	$y'=a+bx'$	(8.22)
$y=a+\frac{b}{x}$	заміна $x'=1/x$, $y'=y$	$y'=a+bx'$	(8.23)
$\frac{1}{x}=a+bx$	заміна $x'=x$, $y'=1/y$	$y'=a+bx'$	(8.24)
$\frac{x}{y}=a+bx$ $y=\frac{x}{a+bx}$	заміна $x'=x$, $y'=x/y$	$y'=a+bx'$	(8.25a) (8.25b)

Залежність між Х і У, яка відповідає в загальному гіперболічній кривій, може бути виражена рівнянням степеневої функції $y=ax^b$, яке логарифмічно перетворено в рівняння прямої lgy=lga+blgx. Для визначення параметрів a і b цього рівняння служить система нормальних рівнянь:

$$\begin{cases} \Sigma lg\,y = nlga + b\Sigma lgx \\ \Sigma(lgx \cdot lgy) = lga \cdot \Sigma lgx + b\Sigma(lgx)^2 \end{cases}; \quad (8.28)$$

Цю систему можна розв’язувати різними способами, але зручніше знаходити параметри регресії за формулами:

$$a = \frac{\Sigma lgy \cdot \Sigma(lgx)^2 - \Sigma(lgx \cdot lgy)\Sigma lg\,x}{n\Sigma(lgx)^2 - \Sigma lg\,x \cdot \Sigma lg\,x}; \quad (8.29)$$

$$b = \frac{n\Sigma(lgx \cdot lgy) - \cdot \Sigma lg\,x \cdot lgy}{n\Sigma(lgx)^2 - \Sigma lg\,x \cdot \Sigma lg\,x}; \quad (8.30)$$

Із яких видно, що спочатку необхідно визначити Σlgy; $\Sigma lg\,x$; $\Sigma(lgx \cdot lgy)$; $\Sigma(lgx)^2$.

Запитання для самоконтролю:

1. Що таке регресія та емпіричний ряд регресії?
2. Які завдання регресійного аналізу?
3. Які види регресії ви знаєте?
4. Які способи вирівнювання кривих регресії вам відомі?
5. Як визначати коефіцієнти (параметри) рівняння регресії лінійної залежності?

РОЗДІЛ 9
ДИСПЕРСІЙНИЙ АНАЛІЗ

9.1. Поняття про дисперсійний аналіз, його суть та задачі

9.2. Найпростіша схема варіювання при аналізі за одним фактором

9.3. Принцип проведення дисперсійного аналізу за алгоритмом на прикладах

9.4. Дисперсійний аналіз однофакторних комплексів великих груп

9.5. Оцінка сили впливу факторів на результативну ознаку

9.6. Двофакторний дисперсійний аналіз

9.1. Поняття про дисперсійний аналіз, його суть та задачі

Зв’язки між ознаками і явищами складні. Біологічні об'єкти завжди пов'язані з оточуючим середовищем, складною системою стосунків, вертикальних та горизонтальних зв'язків.

Сумарний вплив багатьох факторів як біотичних, так і абіотичних на біологічний об’єкт зумовлюють прояв ознак, формування його індивідуальних особливостей. В свою чергу сума індивідуальних ознак біологічних об'єктів формує особливості їх угрупувань.

При проведенні еколого-біологічних та лісівничих досліджень часто необхідно з'ясувати: яким чином той чи інший фактор вплинув на підсилення або послаблення тієї чи іншої характеристики. У зв'язку з величезним антропогенним навантаженням на природне середовище, це питання набуває особливо важливого значення з необхідністю встановлювати силу окремого антропогенного впливу.

Відомо, що живий організм реагує у своєму розвитку одночасно на зміну декількох факторів, внаслідок чого визначення характеру впливу одного з них є досить важким, а іноді і невирішуваним. Вивчаючи вплив різних доз або різного складу добрив, термінів і способів внесення їх у ґрунт на підвищення продуктивності сої, конюшини або іншої сільськогосподарської культури чи їх суміші (напр. бобово-злакової травосуміші), також необхідно враховувати вплив відповідного комплексу факторів, частину з яких дослідник може регулювати за власним бажанням, а прояв іншої частини від нього не залежить. Так, наприклад, вивчаючи вплив технології доглядів за молодими рослинами деревних порід, дослідник повинен враховувати вплив факторів – зміну аерації корененасиченого шару ґрунту, зміну температури, зміну вологості і т. д. Або ж, вивчаючи питання впливу підбору трав на якість корму для великої рогатої худоби, також необхідно враховувати одночасну дію інших умов оточуючого середовища – вміст поживних елементів в ґрунті, його вологість, середньодобову температуру і т. п.

У генетиці, лісовій селекції, лісовідновленні, ентомології та в інших галузях біології, при проведенні досліджень часто треба визначити, наскільки істотний вплив одного або кількох факторів на кінцевий результат. Вирішення поставлених завдань можливо лише порівнянням результатів досліду, в яких даний фактор виступає у відповідній градації. А оскільки всі фактори, що впливають на результат, виступають в комплексі, то застосування знайомого нам методичного прийому парного порівняння середньоарифметичних показників рядів розподілу стає занадто трудомістким, що потребує багато розрахункових робіт. Не завжди тут можна застосувати кореляційний чи регресійний аналіз із отриманням різного виду моделей.

У цьому випадку англійський вчений Р. А. Фішер (1925) запропонував використовувати порівняння вибіркових дисперсій і проводити статистичний аналіз результатів спостережень, що залежать від різних одночасно діючих факторів, робити вибір цих факторів та оцінку "сили" їх впливу.

Дисперсія – (від лат. dispersio – розсіювання) є мірою відхилення показників варіаційного ряду від середнього арифметичного значення цього ряду ($\bar{x}$). Цей спосіб порівняно з іншими біометричними методами є більш ефективним і отримав назву "дисперсійний аналіз". Суть його полягає у встановленні ролі окремих факторів у варіюванні тієї чи іншої ознаки.

Основні поняття і терміни. Характеристики (ознаки) біологічних об'єктів, які змінюються під впливом тих чи інших чинників називаються ***результативними ознаками***, а причини зміни характеристик біологічних об'єктів називаються ***факторами***. Наприклад, висота, щільність, кущистість травостою є ознаками, а такі засоби впливу на ці ознаки, як дози добрив, терміни їх внесення, полив відносяться до категорії факторів.

В дисперсійному аналізі фактори є ***регульовані*** і ***нерегульовані.*** Регульовані, це ті чинники, які можна контролювати. Наприклад, внесення добрив в ґрунт, норми і строки поливу культур і т. п. Нерегульовані фактори – це ті, що не можна контролювати, але вони можуть впливати чи не впливати на досліджувані ознаки. Наприклад, вміст хімічних елементів в ґрунті, кількість опадів в регіоні, рівень ґрунтових вод і т. д.

Регульовані фактори прийнято позначати великими буквами латинського алфавіту – А, В, С і т. д., а результативні ознаки – Х, У і т. д. На стадії планування дослідження передбачається: скільки градацій фактору слід ввести, регулюючи його силу (кількість доглядів за деревостанами, дозу добрив і т. ін.). Зазвичай кожен фактор А, що вивчається, має не одне, а кілька значень, які називають ***градаціями*** або рівнями фактору А. У межах же кожного рівня окремі змінні (варіанти) приймають різні значення, тобто. спостерігається випадкова варіація. Те саме стосується і більш складних випадків, коли у спільному варіюванні бере участь кілька факторів, кожен з яких може мати свої рівні.

В залежності від числа досліджуваних факторів, по яких проводиться дисперсійний аналіз, розрізняють *однофакторні, двофакторні* та *багатофакторні* дисперсійні комплекси. А по тому, як результати спостережень (варіанти) групуються по градаціях комплексу – порівну, пропорційно чи нерівномірно, розрізняють *ортогональні* (сюди відносять рівномірні і пропорціональні) і *неортогональні* (або нерівномірні) комплекси. Цей поділ має значення при визначенні факторіальної суми квадратів відхилень, що буде розглянуто нижче.

Розглянемо це питання більш детально. Нам відомо, що біологічні ознаки варіюють під впливом різних чинників, і, включаючи ті, що вивчаються в дослідженні.

Нульова гіпотеза. Як і в інших випадках статистичного аналізу, при дисперсійному аналізі слід виходити з прийняття нульової гіпотези, а саме: що даний фактор А (або В або С і т. д.) не впливає на ознаку. Якщо правильна нульова гіпотеза, то $\sigma_{\mathrm{A}}^2 = 0$ (те ж стосується і σ_{B}^2; σ_{C}^2 і т.д.), тобто вся варіація є випадковою.

Щоб відкинути нульову гіпотезу, треба довести, що σ_{A}^2 є достовірною, тобто з ймовірністю не меншою, ніж 0,95, або з рівнем значимості 0,05 відрізняється від нуля. Достовірність значення σ_{A}^2 може бути встановлена шляхом ділення його на помилку, тобто:

$$\sigma_{\mathrm{A}}=\frac{\mathrm{A}}{m_A}; \qquad (9.1)$$

9.2. Найпростіша схема варіювання при аналізі за одним фактором

Для того, щоб розуміти сенс розрахунків при дисперсійному аналізі, дуже важливо з самого початку ясно представляти можливу варіацію в тих групах, на які розбивається фактичний матеріал.

Розберемо найпростішу схему, коли аналізується вплив лише одного фактору, що може приймати різні градації, або кількісні рівні:

$$1, 2, \ldots, i, \ldots, a \qquad (9.2)$$

Окремі спостереження (варіанти) розбиваються на групи згідно градацій фактору, що досліджується або по спостереженнях у природі. Важливо, що фактор, що вивчається, тільки один, наприклад: вид добрива, що вноситься в розсаднику, або відношення до різних деревних видів, або вплив отрутохімікатів, або роль способів обробки ґрунту і т. д. За наявності двох або декількох факторів знадобляться більш складні схеми.

Число спостережень (варіант) у кожній групі n, але рівне число в групах не обов'язково. При нерівному числі можна виходити із середнього числа $n_{i.}$

$$N = an \ (= an_i) \qquad (9.3)$$

Зазвичай різні рівні прийнято позначати літерою i, а окремі варіанти (спостереження) – літерою j. Тому кожну варіанту, незалежно від того, де вона знаходиться, можна позначати як xij. В межах кожного рівня (групи) окремі варіанти приймають випадкові значення:

$$x.i\ , x_{.2}, x.з\ , \ldots, x.j\ , \ldots, x.n$$

Суми варіант по кожній групі (у графі «суми по групах») позначені літерами T_1, T_2,... , T_i,….T_a. Загалом їх можна позначити T_i.

Таблиця 9.1

Схема варіювання при відмінностях груп по одному фактору

Групи по одному фактору	Окремі варіанти (спостереження) x_{ij}							Суми по групах	Середні по групах x_i
	1	2	3	….	j	….	n		
1	x_{11}	x_{12}	x_{13}		x_{1j}		x_{1n}	$\sum x_1=T_1$	$\bar{x}_1$
2	x_{21}	x_{22}	x_{23}		x_{2j}		x_{2n}	$\sum x_2=T_2$	$\bar{x}_2$
:									
:									
i	x_{i1}	xi_2	xi_3		xij		x_{in}	$\sum x_i=T_i$	$\bar{x}_i$
:									
:									
a	x_{a1}	x_{a2}	x_{a3}		xaj		x_{an}	$\sum x_a=T_a$	$\bar{x}_a$
								$\sum xij=T$	$\bar{x}$

Типи варіювання варіант та їх характеристика

Після введення всіх цих позначень можна приступити до аналізу варіювання даних, представлених у таблиці 9.1.

Можна виділити 3 типи, або напрямки, варіювання:

а) загальне варіювання всіх варіант (x_{ij}), незалежно від того, в якій групі вони знаходяться навколо загальної середньої $\bar{x}$;

б) варіювання групових середніх $\bar{x_i}$, або, інакше, середніх кожного рівня даного досліджуваного фактору, навколо загальної середньої $\bar{x}$;

в) варіювання варіант x_{ij} всередині кожної групи навколо кожної групи повної середньої x_i.

Для характеристики цих варіювань при проведенні дисперсійного аналізу використовуються вже відомі з колишніх розділів цього навчального посібника величини:

а) суми квадратів відхилень від середньої арифметичної;

б) середні квадрати відхилень, тобто суми квадратів, поділені на кількість ступенів свободи. Це дисперсії σ^2.

Суми квадратів

Для всіх трьох типів варіювання можна визначити суми квадратів відхилень.

Величина *загальної варіації ознаки* (D_y) – це сума квадратів відхилень окремих варіант від середньої арифметичної всього комплексу спостережень. Цю величину можна розділити на частини, одна з яких відображає вплив на ознаку досліджуваного, тобто контрольованого фактору (D_x), інша частина визначається впливом на ознаку інших неконтрольованих чинників, що не досліджуються, тобто «випадкових» (D_z). Таке варіювання можна зобразити рівнянням:

$$\Sigma_{ij}\left(\mathrm{x}_{\mathrm{i}j}-\bar{\mathrm{x}}\right)^2=\Sigma n_i(\bar{\mathrm{x}}_i-\bar{\mathrm{x}})^2+\Sigma\Sigma(x_i-\bar{\mathrm{x}}_i)^2, \qquad (9.4)$$

де $\Sigma_{ij}\left(\mathrm{x}_{\mathrm{i}j}-\bar{\mathrm{x}}\right)^2$ характеризує величину загальної варіації ознаки, тобто варіювання її значень навколо середньої арифметичної всієї сукупності спостережень (D_y). А позначення **ij** вказує на те, що підсумування проводиться по варіантах всіх груп.

Сума квадратів для групових середніх:

$$\Sigma n_i(\bar{\mathrm{x}}_i-\bar{\mathrm{x}})^2 \qquad (9.5)$$

відображує частку загальної варіації, що відноситься до варіювання групових середніх ($\bar{\mathrm{x}}_i$) навколо загальної середньої ($\bar{\mathrm{x}}$), позначується символом (D_x). Щоб ця величина була того ж порядку, що і перша, введено множник n_i, тобто середнє число варіант у кожній групі. Якщо кількість варіант у всіх групах однакова, то просто n.

Сума квадратів відхилень варіант від групових середніх всередині кожної групи, іншими словами випадкової варіації всередині окремої групи:

$$\Sigma_{\text{i}}\left[\Sigma_j\left(x_{ij}-\bar{\text{x}}_i\right)^2\right] \qquad (9.6)$$

позначається символом D_z, – це об'єднана характеристика варіювання значень ознаки всередині окремих груп. Два символи суми вказує на те, що підсумування відбувається двічі: всередині кожної групи, тобто по окремих j (від 1 до n), а потім по всіх рівнях i – від 1 до a.

Тобто, якщо підсумувати вище наведене: $D_y = D_x + D_z$.

Степені свободи (вільності)

Щоб розрахувати середні квадрати (дисперсії), необхідно розділити кожну суму квадратів на відповідні їм числа ступенів свободи, які будуть наступними:

для загальної дисперсії:

$$K_y = N - 1 \qquad (N = an) \qquad (9.7)$$

де N – загальна кількість варіант дисперсійного комплексу.

для дисперсії групових середніх:

$$K_x = a - 1 \qquad (9.8)$$

де a – кількість градацій фокусного фактору.

для випадкової варіації всередині групи:

$$K_z = a\,(n - 1) = n\,a - a = N\text{-}a, \qquad (9.9)$$

де a – кількість градацій фокусного фактору, а n – кількість показників в окремій градації.

Відповідно, сума чисел ступенів свободи для групових середніх і для варіації всередині груп має дорівнювати числу ступенів свободи для загальної дисперсії:

$$(N - a) + (a - 1) = N - 1 \qquad (9.10)$$

Загальна схема дисперсійного аналізу при одному факторі показана в таблиці 9.2.

Дисперсійний аналіз дозволяє *оцінювати* значущість впливу окремих факторів, а також їх відносну роль у загальній мінливості. Оцінками впливу різних факторів на ознаку є дисперсії – відношення сум квадратів відхилень до ступенів вільностей.

Види дисперсій, які фігурують в дисперсійному аналізі, отримують наступним чином.

Загальна дисперсія ($\sigma_{заг}^2$), викликана впливом на складові дисперсійного комплексу всіх факторів, вже відома нам дисперсія, основана на сумі квадратів відхилень окремих варіант (x_{ij}) від загальної середньої ($\bar{x}$)(формула 9.11 в таблиці 9.2).

Таблиця 9.2

Схема дисперсійного аналізу (аналізу дисперсії) при одному факторі

Джерело варіювання	Сума квадратів відхилень (D)	Число степенів свободи (K)	Середній квадрат (дисперсія, σ^2)	Номер формули для σ^2
Загальне (всі варіанти)	$\Sigma_{ij}\left(x_{ij}-\bar{x}\right)^2$	N – 1	$\frac{1}{N-1}\Sigma_{ij}\left(x_{ij}-\bar{x}\right)^2$	(9.11)
Групове середнє (фактор А)	$\Sigma n_i(\bar{x}_i-\bar{x})^2$	a - 1	$\frac{1}{a-1}\Sigma n_i(\bar{x}_i-\bar{x})^2$	(9.12)
Варіанти всередині груп (випадкове відхилення)	$\Sigma_i\left[\Sigma_j\left(x_{ij}-\bar{x}_i\right)^2\right]$	N-a	$\frac{1}{N-a}\Sigma_i\left[\Sigma_j\left(x_{ij}-\bar{x}_i\right)^2\right]$	(9.13)

З таблиці 9.2 видно, що загальна варіація розкладається на 2 компоненти: один з них – це варіація групових середніх (за градаціями фактору А) навколо загальної середньої $\bar{x}$; інший – варіація окремих варіант усередині груп. Останню варіацію можна розглядати як випадкову в тому сенсі, що вона створюється багатьма неконтрольованими факторами крім досліджуваного фактору А. Тобто, під час дослідження одержуємо декілька вибіркових сукупностей, в яких фактор (-и), який нас цікавить діє з різною силою. Обрахувавши біометричні показники вибіркових сукупностей, ми знаємо, що дія досліджуваного фактору, як і дія інших факторів, вплинула на отримані показники, але поки що не можемо сказати в якій мірі.

Якщо фактор, який досліджується значимо впливає на результативну ознаку, то групи показників будуть відрізнятись одна від одної тим сильніше, чим сильніше вплив досліджуваного фактору. Це знайде свій прояв в різниці групових середніх ($\bar{x}_i$) вибіркових сукупностей і наприкінці – в степені розсіяння цих групових середніх навколо загальної середньої арифметичної.

Відношення факторіальної суми квадратів відхилень групових середніх від загальної середньої до відповідної кількості ступенів вільності і є **факторною (міжгруповою) дисперсією** ($\sigma_{факт}^2$). Число ступенів вільності для факторної дисперсії дорівнює а – 1, де а – кількість градацій фокусного фактору (формула 9.12 в таблиці 9.2). Те, що в середині градацій досліджуваного фактору А має місце відхилення окремих групових варіант (x_i) від групових середніх ($\bar{x}_i$), свідчить про вплив на результативну ознаку і інших факторів, окрім досліджуваного.

Сума квадратів цих відхилень по всіх градаціях комплексу поділена на відповідну кількість ступенів вільності **є залишковою (випадковою) дисперсією** ($\sigma_{зал.}^2$) (формула 9.13).

Отож, суть дисперсійного аналізу є розкладання загальної дисперсії комплексу спостережень на його складові частини: факторну дисперсію, що виникає під впливом досліджуваного фактору, і залишкову, – під впливом всіх інших факторів, з чого і з'ясовується вплив окремих факторів на досліджувану ознаку. Про достовірність оцінок судять по критерію F-Фішера.

$$F = \left(\frac{\sigma^2_{факт}}{\sigma^2_{зал}}\right); \qquad (9.14)$$

Зіставляючи факторну дисперсію з залишковою і, порівнюючи результат із теоретичним значенням критерію F-Фішера, для прийнятого рівня значимості (частіше – 5 %, рідше 1 % або 0,1 %) та ступенів вільності факторної та залишкової дисперсії, яке знаходять по спеціальній таблиці, розробленій Фішером (додаток 2), ми дізнаємось – чи є статистично значимою різниця між цими видами дисперсій, тобто чи достовірна різниця між впливом досліджуваного фактору і всіх інших.

Нульова гіпотеза полягає в припущенні відсутності впливу організованого фактора на результативну ознаку. Вона відкидається, якщо $F_{факт} \geq F_{табл}$. В даному випадку приймається, що вплив організованого (досліджуваного) фактора на результативну ознаку не є випадковим, тобто стверджується наявність достовірної різниці генеральних параметрів за градаціями дисперсійного комплексу.

Якщо значення $F_{факт} \leq F_{табл}$, то це означає, що нульова гіпотеза підтверджується, тобто та різниця, яка зафіксована між груповими середніми, є статистично недостовірною. По встановленні достовірності статистично значимого впливу досліджуваного фактору на результативну ознаку, якщо необхідно встановити, яка з градацій фактору впливає на неї найбільш значимо, слід попарно порівняти середні.

9.3. Принцип проведення дисперсійного аналізу за алгоритмом на прикладах

Алгоритм дисперсійного аналізу однофакторних комплексів

№	Зміст організаційної або математичної дії
1	Групування вибіркових матеріалів у комбінаційну таблицю дисперсійного комплексу
2	Визначення значень: середнього арифметичного всього комплексу ($\bar{x}$) і групових середніх за градаціями організованого фактора (x_i)
3	Визначення загальної суми квадратів відхилень (D_y), тобто суми квадратів відхилень варіант від загальної середньої: $D_y = \Sigma x^2 - \frac{(\Sigma x)^2}{N}$
4	Визначення міжгрупової суми квадратів відхилень, яка дорівнює сумі квадратів відхилень групових середніх від загальної середньої з урахуванням статистичної ваги (n_i) групових середніх: • у випадку рівних чисел варіант в градаціях комплексу – $D_x = n_i \Sigma \bar{x}_i^2 - \frac{(\Sigma x)^2}{N}$; • у випадку різної кількості варіант в градаціях комплексу – $D_x = \Sigma[n_i(x_i - x)^2]$;
5	Визначення внутрішньогрупової суми квадратів, тобто суми квадратів відхилень групових варіант від групових середніх: $D_z = \Sigma x^2 - n\Sigma \bar{x}_i^2$
6	Визначення дисперсій (середніх квадратів відхилень): • загальна: $\sigma_{заг}^2 = D_y / (N-1)$; • факторна: $\sigma_{факт}^2 = D_x / (a-1)$, де a – кількість груп; • остаточна: $\sigma_{зал}^2 = D_z / (N-a)$
7	Визначення фактичного значення критерію $F_{факт.} = \frac{\sigma_{факт}^2}{\sigma_{ост}^2}$;
8	Порівняння фактичного значення критерію F з його табличним (стандартним) значенням для відповідного рівня значимості (p) і даних чисел

Принцип проведення дисперсійного аналізу за наведеним алгоритмом розглянемо на прикладі.

Визначали вплив удобрення на приріст травостою першого укосу в висоту. Облік проводили в середині червня. Результати вимірювань подані в таблиці 9.1.

1. Групуємо результати замірів у комбінаційну таблицю дисперсійного комплексу (таблиця 9.3).

2. Визначаємо значення: середнього арифметичного всього комплексу ($\bar{x}$) і групових середніх за градаціями організованого фактора ($\bar{x}_i$) (останній стовпчик таблиці 9.1).

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots x_k}{n} = 111{,}39$$

3. Визначаємо загальну суму квадратів відхилень (D_y), тобто суму квадратів відхилень варіант від загальної середньої: $D_y = \Sigma x^2 - \frac{(\Sigma x)^2}{N}$. Для цього визначаємо:

Σx = 94+85+92+90+92+86+90+91+120+105+112…+…123+129 =5347.

$(\Sigma x)^2 = (5347)^2 = 28590409$

$\Sigma x^2 = 94^2 + 85^2 + 92^2 + 90^2 + 92^2 + 86^2 + 90^2 + 91^2 + 120^2 + 105^2 + 112^2 \dots + \dots 123^2 + 129^2 = 602095$

$D_y = 602095 - \frac{28590409}{48} = 6461{,}48$

Таблиця 9.3

Вплив удобрення на приріст травостою в висоту

Варіанти досліду	Приріст, см								$\bar{x}_i$
	1	2	3	4	5	6	7	8	
контроль – без добрив	94	85	92	90	92	86	90	91	90
$P_{30}K_{30}$ – щорічно навесні	120	105	112	108	100	110	118	114	110,875

Продовження таблиці 9.3

$P_{60}K_{60}$ – щорічно навесні	119	120	121	110	110	116	105	118	114,875
$N_{30}P_{60}K_{60}$ – щорічно навесні	118	120	115	116	118	121	109	117	116,75
$N_{60}P_{60}K_{60}$ – щорічно навесні	117	111	100	118	112	117	110	105	111,25
гній 30 т/га восени + $N_{30}P_{105}$ навесні	128	127	124	119	121	126	123	129	124,625

4. Визначаємо міжгрупову суму квадратів відхилень, яка дорівнює сумі квадратів відхилень групових середніх від загальної середньої з урахуванням статистичної ваги (n_i) групових середніх. Оскільки в прикладі рівне число варіант в градаціях комплексу, то використовуємо формулу: $D_x = n_i\Sigma\bar{x}_i^2 - \frac{(\Sigma x)^2}{N}$.

$\Sigma\bar{x}_i^2 = 90^2+110{,}875^2+114{,}875^2+116{,}75^2+111{,}25^2+124{,}625^2 = 75128$

$D_x = n_i\Sigma\bar{x}_i^2 - \frac{(\Sigma x)^2}{N} = 8 * 75128 - \frac{28590409}{48} =$ 601024-595633,5=5390,5

5. Визначимо внутрішньо групову суму квадратів, тобто суму квадратів відхилень групових варіант від групових середніх: $D_z = \Sigma x^2 - n\Sigma\bar{x}_i^2$.

$D_z = 602095 - 601024$=1071

6. Розраховуємо дисперсії (середні квадрати відхилень):

• загальну: $\sigma_{заг}^2 = D_y / (N\text{-}1)$;

$\sigma_{заг}^2 = 6461{,}48 / 47 = 137{,}47$;

• факторну: $\sigma_{факт}^2 = D_x / (a\text{-}1)$, де a – кількість груп:

$\sigma_{факт}^2 = 5390{,}5 / 5 = 1078{,}1$

• остаточну: $\sigma_{ост}^2 = D_z / (N\text{-}a)$:

$\sigma_{ост}^2 = 1071 / 42 = 25{,}5$.

7. Визначаємо фактичне значення критерію $F_{факт.} = \frac{\sigma^2_{факт}}{\sigma^2_{ост}}$:

$F_{факт.} = \frac{1078{,}1}{25{,}5} = 42{,}2$

8. Порівнюємо фактичне значення критерію F з його табличним (стандартним) значенням для рівня значимості (p = 0,95) і даних чисел: значення $F_{факт.} = 42{,}2$ більше, ніж $F_{5\%.} = 2{,}4$.

Висновок: оскільки $F_{факт.} > F_{5\%}$, то вплив удобрення на приріст травостою у висоту є статистично достовірним.

9.4. Дисперсійний аналіз однофакторних комплексів великих груп

При обробці складніших однофакторних комплексів, що групуються у вигляді кореляційної таблиці, суми квадратів відхилень розраховуються по наступних формулах.

Загальна дисперсія:

$$D_y = \Sigma(x^2 n_x) - \frac{(\Sigma x n_x)^2}{N}; \qquad (9.15)$$

Міжгрупова (факторіальна):

$$D_x = \Sigma \frac{(\mathrm{px})^2}{\mathrm{n_A}} - \frac{(\Sigma x n_x)^2}{N}; \qquad (9.16)$$

Внутрішньогрупова або ж випадкова:

$$D_z = \Sigma(x^2 n_x) - \Sigma \frac{(\mathrm{px})^2}{\mathrm{n_A}}; \qquad (9.17)$$

де x – окреме значення або варіанти результативної ознаки (градації X); р – частоти окремих значень результативної ознаки по градаціях фактору А; n_x – сума частот по градаціях ознаки X;

n_A – сума частот по градаціях фактору А; N – загальне число спостережень, тобто $\Sigma n_x = \Sigma n_A = N$.

Розглянемо на прикладі. Досліджували вплив органо-мінерального удобрення на урожайність сіна багаторічних трав. Облікували 50 ділянок. Варіанти удобрення: 1. Без добрив – контроль; 2. $P_{30}K_{30}$ – щорічно; 3. $P_{60}K_{60}$ – щорічно; 4. $N_{30}P_{60}K_{60}$ – щорічно; 5. $N_{60}P_{60}K_{60}$ – щорічно; 6. 30 т/га гною + $N_{30}P_{105}$ під покривну культуру; 7. 30 т/га гною восени + $N_{30}P_{105}$ весною наступного року – раз в три роки; 8. щорічно 10 т/га гною восени + $N_{10}P_{35}$ навесні (в таблиці 9.3 варіанти удобрення відповідно позначені як фактор А від 1 до 8). Урожай на ділянках значно варіює. Необхідно з'ясувати чи достовірний вплив різної кількості і складу органо-мінерального удобрення на урожай багаторічних трав.

В таблиці 9.4 вказані результати дослідження, а також допоміжні величини, які необхідні для розрахунку суми квадратів відхилень. Значення ***px*** отримані в результаті множення частот ***(p)*** на числа ***(x)*** по градаціях результативної ознаки. Величини $\mathbf{xn_x}$ отримані в результаті множення значень результативної ознаки **(x)** на суми частот $(\mathbf{n_x})$ по рядках таблиці.

Таблиця 9.4

Приклад однофакторного дисперсійного аналізу з допомогою комбінативної кореляційної таблиці

Урожай сіна, т/га, Х	Фактор А: удобрення								$\mathbf{n_x}$	$\mathbf{xn_x}$	$\mathbf{x^2 n_x}$
	1	2	3	4	5	6	7	8			
14						1	4	1	6	84	1176
13			2	2	3	3	2	2	14	182	2366
12		2	1	2	2	3	1	2	13	156	1872
11		3	2	1	1			1	8	88	968
10		2		1					3	30	300
9	1								1	9	81

Продовження таблиці 9.4

8	3								3	24	192
7	2								2	14	98
n_A	6	7	5	6	6	7	7	6	50	587	7053
px	47	77	60	71	74	89	94	75	Σpx = 587		
$\frac{(px)^2}{n_A}$	368,16	847	720	840,16	912,66	1131,57	1262,28	937,5	$\Sigma\frac{(px)^2}{n_A} =$ 7019,33		

Підставляємо знайдені допоміжні величини в формули 9.15 – 9.17 і знаходимо суми квадратів відхилень:

$D_y = \Sigma(x^2 n_x) - \frac{(\Sigma x n_x)^2}{N}$ = 7053 - $\frac{587^2}{50}$ = 7053-6891,38 = 161,62

$D_x = \Sigma\frac{(px)^2}{n_A}$ - $\frac{(\Sigma x n_x)^2}{N}$ = 7019,33 – 6891,38 = 127,95

$D_z = \Sigma(x^2 n_x) - \Sigma\frac{(px)^2}{n_A}$ = 7053 – 7019,33 = 33,67

Степені свободи для даного прикладу: $К_y = N$-1 = 50-1 = 49; $К_x$ = a – 1 = 8-1 = 7; $К_z = N$- a = 50 – 8 = 42

Визначаємо дисперсії комплексу:

$\sigma_y^2 = \frac{161,62}{49}$ = 3,29

$\sigma_x^2 = \frac{127,95}{7}$ = 18,27

$\sigma_z^2 = \frac{33,67}{42}$ = 0,8

Критерій достовірності:

$F_{факт.} = \frac{18,27}{0,8} = 22,8$ > F_{st} = 2,2 при $К_x$ = 7 і $К_z$ = 42 і рівні достовірності Р = 0,01.

Висновок: вплив органо-мінерального удобрення на врожай багаторічних трав є достовірним.

9.5. Оцінка сили впливу факторів на результативну ознаку

Дисперсійний аналіз дозволяє встановити не тільки достовірність, але і силу впливу досліджуваних факторів на результативну ознаку. Сила впливу фактору визначається як відсоток факторної варіації в загальній варіації ознаки і розраховується за формулою, запропонованою Плохинським М.О.:

$$\acute{\eta}_{x}^{2} = \frac{D_x}{D_y} \quad \text{або} \quad \acute{\eta}_{x}^{2} = 1 - \frac{D_x}{D_y} \qquad (9.18)$$

Помилка репрезентативності даного показника визначають за формулою:

$$m_{\acute{\eta}x}^{2} = (1 - \eta_{x}^{2})\frac{K_x}{K_Z}; \qquad (9.19)$$

де $K_x = a - 1$ (а – число градацій фактору А); $K_Z = N - a$ (N – об’єм дисперсійного комплексу).

Нульова гіпотеза відкидається, якщо $F_ф = \frac{\eta_x^2}{m_{\eta x}^2} \geq F_{st}$ для $K_1 = a - 1$ і $K_2 = N - a$.

Так, для прикладу, наведеного вище, знаходимо силу впливу удобрення на урожайність сіна багаторічних трав за методом Плохинського (формула 9.18).

$$\eta_x^2 = \frac{D_x}{D_y} = \frac{127{,}95}{161{,}62} = 0{,}79 \text{ або } 79\%.$$

Помилка репрезентативності даного показника $m_{\eta x}^2 = (1 - 0{,}79)\frac{7}{42} = 0{,}035.$

Критерій $F_ф = \frac{0.79}{0{,}035} = 22{,}6$ більше F_{st} (При $K_1 = 7$ і $K_2 = 42$) (додаток 4).

Оскільки $F_ф > F_{st}$, нульова гіпотеза відкидається, ступінь впливу удобрення на урожайність сіна багаторічних трав є високою і достовірною.

9.6. Двофакторний дисперсійний аналіз

Дисперсійний аналіз є одним з найдосконаліших біометричних методів. За його допомогою можна аналізувати вплив не лише одного фактору, а двох, трьох і таке інше. В такому випадку факторна дисперсія розглядається як сума дисперсій, викликаних дією досліджуваних факторів А, В, С.

За участі у загальній варіації двох факторів А та В аналіз ускладняється наявністю взаємодії між цими факторами. Тому загальна сума квадратів при двофакторній схемі розкладається на 4 компоненти: а) варіація під впливом фактору А; б) варіація під впливом фактору В; в) варіація під спільним впливом А та В, тобто взаємодії А та В; і г) випадкові відхилення.

Крім того, треба пам'ятати, що за двофакторної схеми кожен рівень одного фактору повинен поєднуватися з будь-яким рівнем другого фактору А.

Дисперсійний аналіз двофакторних рівномірних комплексів, що складаються із порівняно невеликих груп, проводиться за наступним алгоритмом.

1. Всі варіанти дисперсійного комплексу сумують. Знайдену суму підносять до квадрату і відносять до загального числа спостережень.

2. Потім кожну варіанту комплексу підносять до квадрату і знаходять суму квадратів (Σx^2).

3. Необхідно розрахувати суму квадратів відхилень. Для цього можна використати наступні робочі формули:

Загальна для всього комплексу:

$$D_y = \Sigma x^2 - \frac{(\Sigma x)^2}{N}; \tag{9.20}$$

Міжгрупова (по узгодженню градацій):

$$D_x = \Sigma \frac{(\Sigma x_i)^2}{n_i} - \frac{(\Sigma x)^2}{N}; \tag{9.21}$$

Залишкова або випадкова:

$$D_z = D_y - D_x; \tag{9.22}$$

По фактору А:

$$D_A = \Sigma \frac{(\Sigma x_A)^2}{n_A} - \frac{(\Sigma x)^2}{N}; \tag{9.23}$$

По фактору В:

$$D_B = \Sigma \frac{(\Sigma x_B)^2}{n_B} - \frac{(\Sigma x)^2}{N}; \qquad (9.24)$$

По взаємодії факторів АВ:

$$D_{AB} = D_x - D_A - D_B; \qquad (9.25)$$

Тут х – окремі варіанти дисперсійного комплексу; Σx_i – сума варіант по градаціях комплексу; Σx_A – сума варіант по фактору А; Σx_B – сума варіант по фактору В; n_A – число варіант по фактору А; n_B – число варіант по фактору В; n_i – число варіант в окремих клітинках таблиці; N – загальне число варіант, що складають дисперсійний комплекс.

4. Визначивши суми квадратів, встановлюють степені свободи, які дорівнюють: для загальної дисперсії $K_y = N - 1$, дисперсії по фактору А - K_A = a-1, дисперсії по фактору В - K_B = b-1, по взаємодії двох факторів АВ - K_{AB}=(a-1)(b-1) = $K_A \cdot K_B$, залишкової чи випадкової дисперсії – $K_z = N - ab$.

5. Відношенням суми квадратів відхилень до чисел степенів свободи визначають дисперсії: загальну – $\sigma_y^2 = \frac{D_y}{K_y}$; по фактору А - $\sigma_A^2 = \frac{D_A}{K_A}$; по фактору В - $\sigma_B^2 = \frac{D_B}{K_B}$; по взаємодії факторів АВ - $\sigma_{AB}^2 = \frac{D_{AB}}{K_{AB}}$; і залишкову, яка показує вплив на результативну ознаку випадкових нерегульованих факторів – $\sigma_z^2 = \frac{D_z}{K_z}$.

6. Визначають достовірність відношення дисперсій за критерієм F Фішера.

$$F_A = \frac{\sigma_A^2}{\sigma_z^2};\ F_B = \frac{\sigma_B^2}{\sigma_z^2};\ F_{AB} = \frac{\sigma_{AB}^2}{\sigma_z^2};$$

Розглянемо на прикладі. Досліджували вплив удобрення та обробка інсектицидами від колорадського жука на урожайність картоплі. Результати досліду вказані в таблиці 9.5. Необхідно з’ясувати чи випадковими чи достовірними є вплив добрив і обробка інсектицидами на урожай картоплі.

Таблиця 9.5

Результати двохфакторного досліду з агротехніки картоплі

Варіанти добрив (B)	Інсектициди для захисту від шкідників												$\bar{x}_i$
	Препарат A_1				Препарат A_2				Препарат A_3				
	1	2	3	4	1	2	3	4	1	2	3	4	
B_1	21,3	21,8	20,5	21	19,4	22	21,4	19,3	20,8	19,4	20	18,9	20,5
B_2	18,2	18	17	19,8	20,7	18,5	18,3	17,7	19,2	18	17,3	17	18,3
B_3	18,9	19,6	19,7	19,5	17	17,2	16,3	16,7	16,9	17,9	17,2	18	17,9

Позначаємо як фактор А – інсектициди для захисту картоплі, фактор В – органо-мінеральне удобрення картоплі і проводимо дисперсійний аналіз результатів дослідження. Відмітимо, що обидва фактори мають однакову кількість градацій – по 3 варіанти, тобто a=b=3. Дисперсійний комплекс рівномірний, оскільки по групах однакова кількість варіант (n_i = 4) (таблиця 9.5). В комплексі 36 варіант ($a \cdot b \cdot n_i = N$). Підраховуємо:

1. $\Sigma x = 21{,}3+21{,}8+20{,}5\ldots..+18 = 680{,}4$

2. $\Sigma x^2 = 21{,}3^2+21{,}8^2+20{,}5^2\ldots..+18^2 = 12948{,}3$

3. $(\Sigma x)^2 = (680{,}4)^2 = 462944{,}16$

4. Знаходимо суму квадратів відхилень для всього комплексу:

$$D_y = \Sigma x^2 - \frac{(\Sigma x)^2}{N} = 12948{,}3 - \frac{462944{,}16}{36} = 12948{,}3 - 12859{,}56 = 88{,}74$$

5. Щоб знайти факторіальну суму квадратів відхилень, підсумовуємо варіанти в окремих клітинках таблиці, потім результати підсумовуємо по рядках і стовпчиках (по градаціях факторів А і В) (таблиця 9.6).

Таблиця 9.6

Визначення допоміжних величин для розрахунку факторіальної суми квадратів відхилень

A / B	A_1	A_2	A_3	Σx_B	$(\Sigma x_B)^2$	$\frac{(\Sigma x_B)^2}{n_B}$
B_1	84,6	82,1	79,1	245,8	60417,64	5034,8
B_2	73	75,2	71,5	219,7	48268,09	4022,3
B_3	77,7	67,2	70	214,9	46182,01	3848,5
Σx_A	235,3	224,5	220,6	680,4	154867,74	**12905,6**
$(\Sigma x_A)^2$	55366,09	50400,25	48664,36	154430,7	$n_A = b \cdot n_i = 3 \cdot 4 = 12$	
$\frac{(\Sigma x_A)^2}{n_A}$	4613,8	4200,02	4055,36	**12869,2**	$n_B = a \cdot n_i = 3 \cdot 4 = 12$	

Залишається визначити ще одну допоміжну величину $\Sigma \frac{(\Sigma x_i)^2}{n_i}$. Для цього кожне значення таблиці 9.6, за винятком підсумкових чисел, підносимо до квадрату і розділяємо на $n_i = 4$ (таблиця 9.7).

Таблиця 9.7

Визначення допоміжної величини $\boldsymbol{\Sigma \frac{(\Sigma x_i)^2}{n_i}}$

A / B	A_1	A_2	A_3	Ділення на $n_i = 4$			
				A_1	A_2	A_3	Сума
B_1	7157,16	6740,41	6256,81	1789,29	1685,1	1564,2	5038,59
B_2	5329	5655,04	5112,25	1332,25	1413,76	1278,06	4024,04
B_3	6037,29	4515,84	4900	1509,32	1128,96	1225	3863,28
							12925,94

Розрахунок суми квадратів відхилень.

Факторіальної суми квадратів відхилень за формулою 9.21:

$D_x = \Sigma \frac{(\Sigma x_i)^2}{n_i} - \frac{(\Sigma x)^2}{N}$ = 12925,94 - 12859,56 = 66,38

Залишкової суми квадратів відхилень за формулою 9.22:

$D_z = 88{,}74 - 66{,}38 = 22{,}36$

По фактору А за формулою 9.23:

$$D_A = \Sigma\frac{(\Sigma x_A)^2}{n_A} - \frac{(\Sigma x)^2}{N} = 12869{,}2 - 12859{,}56 = 9{,}64$$

По фактору В за формулою 9.24:

$$D_B = \Sigma\frac{(\Sigma x_B)^2}{n_B} - \frac{(\Sigma x)^2}{N} = 12905{,}6 - 12859{,}56 = 46{,}04$$

По взаємодії факторів АВ (формула 9.25):

$D_{AB} = 66{,}38 – 9{,}64 – 46{,}04 = 10{,}7$

З'ясовуємо степені свободи. Для загальної дисперсії $K_y = N - 1 = 36 – 1 = 35$, факторіальної А – $K_A = a\text{-}1 = 3\text{-}1 = 2$, факторіальної В – $K_B = b\text{-}1 = 3\text{-}1 = 2$, по взаємовпливу АВ – $K_{AB} = K_A \cdot K_B = 2 \cdot 2 = 4$, залишкової або випадкової – $K_z = N – ab = 36 - 3 \cdot 3 = 27$.

Знаходимо дисперсії, результати вказані в підсумковій таблиці дисперсійного аналізу (таблиця 9.8).

Таблиця 9.8

Підсумкова таблиця дисперсійного аналізу

Варіації	Степені свободи	Сума квадратів відхилень	Середній квадрат (σ^2)	$F_{факт}$	F_{st}	
					P = 0,05	P = 0,01
По фактору А	2	9,64	4,82	5,8	3,4	5,5
По фактору В	2	46,04	23,02	27,7	3,4	5,5
АВ	4	10,7	2,675	3,22	2,7	4,1
Залишкова	27	22,36	0,83	-	-	-
Загальна	35	88,74	-	-	-	-

З таблиці 9.7 видно, що на урожайність картоплі впливають обидва фактори. Вплив органо-мінерального удобрення (фактор В) на результативну ознаку є найбільш суттєвим. Відмічається незначний вплив взаємодії двох факторів А і В.

Визначаємо силу впливу фактору В на результативну ознаку:

$$\eta_B^2 = \frac{D_B}{D_y} = \frac{46,04}{88,74} = 0,52 \text{ або } 52\%.$$

Помилка репрезентативності даного показника:

$$m_{\eta B}^2 = (1 - \eta_B^2)\frac{K_B}{K_Z} = (1 - 0,52)\frac{2}{27} = 0,036$$

$F_ф = \frac{0,52}{0,036} = 14,4$, що більше $F_{st} = 5,5$ для $K_1 = K_B = 2$, $K_2 = Kz = 27$ і $p = 0,01$

Отож, сила впливу удобрення на урожайність картоплі, що дорівнює 52% всіх впливів є достовірною.

Визначаємо силу впливу фактору А на результативну ознаку:

$$\eta_A^2 = \frac{D_A}{D_y} = \frac{9,64}{88,74} = 0,11 \text{ або } 11\%.$$

Помилка репрезентативності даного показника:

$$m_{\eta A}^2 = (1 - \eta_A^2)\frac{K_A}{K_Z} = (1 - 0,11)\frac{2}{27} = 0,066$$

$F_ф = \frac{0,11}{0,066} = 1,7$, що менше $F_{st} = 5,5$ для $K_1 = K_A = 2$, $K_2 = Kz = 27$ і $p = 0,01$

Отож, вплив захисту картоплі від колорадського жука на її урожайність, що дорівнює 11% є не достовірним.

Визначаємо силу впливу взаємодії двох факторів А і В на результативну ознаку:

$$\acute{\eta}_{AB}^2 = \frac{D_{AB}}{D_y} = \frac{10,7}{88,74} = 0,12 \text{ або } 12\%.$$

Помилка репрезентативності даного показника:

$$m_{\acute{\eta} AB}^2 = (1 - \acute{\eta}_{AB}^2)\frac{K_A}{K_Z} = (1 - 0,12)\frac{4}{27} = 0,13$$

$F_ф = \frac{0,12}{0,13} = 0,92$, що менше $F_{st} = 2,7$ для $K_1 = K_{AB} = 4$, $K_2 = Kz = 27$ і $p = 0,05$.

Запитання для самоконтролю:

1. Поняття про дисперсійний аналіз.
2. Передумови та постановка задачі однофакторного дисперсійного аналізу.
3. Загальна, факторна та залишкова суми квадратів відхилень та зв’язок між ними.
4. Загальна, факторна та залишкова дисперсії.
5. Алгоритм однофакторного дисперсійного аналізу за Фішером.
6. Назвіть основні принципи проведення дисперсійного двофакторного аналізу.

РОЗДІЛ 10
ОСНОВИ КОМП'ЮТЕРНОЇ ГРАМОТНОСТІ В ДОДАТКУ ДО БІОМЕТРІЇ

10.1. Рішення задач описової статистики засобами MS EXCEL

10.2. Розрахунок параметрів кореляції з використанням засобів Вставка функцій і Майстер діаграм MS Excel

10.3. Однофакторний дисперсійний аналіз з використанням засобів Аналіз даних MS Excel.

10.1. Рішення задач описової статистики засобами MS EXCEL

З огляду на те, що обрахування статистичних характеристик в біометрії – це в більшості випадків трудомісткий процес, на який доводиться витрачати велику кількість часу. Крім того необхідно виконувати значну кількість взаємопов'язаних розрахункових операцій з кожною варіантою ряду спостережень, а це монотонність однотипних операцій, що повторюються, і тому збільшується вірогідність випадкових помилок. Звичайно, масштаби цієї роботи прямо залежать від кількості варіант та виду аналізу, що проводиться.

Особлива складність обчислень із значною витратою часу притаманна кореляційному, регресивному та дисперсійному аналізам багатовимірних масивів спостережень, який інколи складається з декількох рядів по 100–200 варіант. Крім того, часто приходиться повторяти процес заради уникнення випадкових помилок в обчисленнях, що ще більше займає часу в дослідника.

Для спрощення процесу розрахунків статистичних характеристик вже давно були розроблені розрахунки у вигляді алгоритмів, спеціальних таблиць та матриць. Це лише систематизує процес розрахунків, робить його більш організованим.

Для усунення основних недоліків неавтоматизованих обчислень: трудомісткості та високої вірогідності помилки було запропоновано обробляти біологічні дані за допомогою комп'ютерної техніки.

Для освоєння застосування стандартних функцій MS Excel для вирішення задач описової статистики розглянемо приклад. Заміряли висоту лучного газону, результати подані в таблиці 10.1. Необхідно виконати наступні дії:

1. Запустити MS Excel: **Пуск / Програми / MS Excel** і перейменувати ярличок робочого аркуша **Лист 1**: подвійне клацання по ярличку і надрукувати поверх виділення **Статистика 1**. Ввести вихідні дані та заголовки статистичної таблиці за **зразком 1**: виділити комірку А1 клацанням миші / ввести текст заголовка і зафіксувати клацанням по інструменту Enter 7 / розташувати заголовок по центру стовпців А-Е – виділити комірки А1: Е1 і натиснути інструмент **Об'єднати і помістити в центрі** / аналогічно виділяючи послідовно комірки А2 – Е11 ввести числа вихідних даних таблиці 10.1:

Таблиця 10.1

Заміри висоти травостою лучного газону

18	21	20	17	24
17	18	23	24	18
22	22	27	23	17
16	17	22	24	25
22	21	19	20	16
21	23	26	17	22
18	24	15	27	22
24	22	21	24	21
18	19	22	27	24
25	22	17	21	21

В комірках G2 – G14 заголовки рядків статистичної таблиці і число вибірок:

середнє	СРЗНАЧ
Середньоквадратичне відхилення	СТАНДОТКЛОН
Дисперсія	ДИСПА
Медіана	МЕДИАНА
Мода	МОДА
Асиметрія	СКОС
Ексцес	ЕКСЦЕС
Найменше значення	МИН
Найбільше значення	МАКС
Кількість вибірок	50

Виконати розрахунки зазначених в заголовках рядків статистичної таблиці параметрів, вставляючи при допомозі засобів **Вставка функцій** розрахункові формули. Наприклад, для розрахунку середнього виділити клацанням миші комірку H2, клацнути інструмент **Вставка функцій** / у вікні **Майстер функцій** в полі **Категорії** клацнути **Статистичні**, в полі **Функція** за допомогою смуги прокрутки перегорнути список назви функцій, знайти і клацнути **СРЗНАЧ** і **ОК** / у вікні вставки функції праворуч від поля **Число1** клацнути кнопку згортання / виділити мишею діапазон комірок A2: E11 (утримуючи ліву кнопку миші) / в згорнутому вікні вставки функції клацнути кнопку розгортання / ОК. Аналогічно вставити інші формули.

2. Сформувати таблицю частот досліджуваної величини, виконавши угруповання даних і розрахунки безпосереднім введенням формул і за допомогою засобу **Вставка функцій**:

- вставити формулу для обчислення мінімального числа інтервалів групування за допомогою засобу **Вставка функцій**: виділити комірку A14 і ввести «мін. кількість інтервалів», виділити комірку B14 / інструмент **Вставка функцій** / в полі **Категорії** клацнути **Математичні**/ в полі **Функція** знайти і вибрати **ОКРУГЛ** і **ОК**/ у вікні вставки функції встановити курсор в полі **Число розрядів** і ввести 0 (округлення до цілого числа), встановити курсор в поле **Число** і ввести 5* (множник) / в інструменті вибору функції (лівий верхній кут робочої книги) клацнути кнопку списку і вибрати позицію **Інші функції**... / у вікні **Майстер** функцій вибрати категорію **LOG10** з категорії **Математичні** і **ОК** / у вікні **вставки функції** в поле **Число 1** ввести посилання з числом вибірок **Н14** і **ОК**.

- вставити формулу для розрахунку ширини інтервалу за допомогою введення з клавіатури: виділити комірку А15 і ввести «ширина інтервалу», виділити комірку В15 / ввести знак = (дорівнює) і знак ((дужка) / клацнути комірку з максимальним значенням Н10 і натиснути клавішу F4 для переходу до абсолютного посилання / ввести знак - / клацнути клітинку з мінімальним значенням Н9 і натиснути клавішу **F4** / ввести знак) (дужка) і знак / (похила риска) і клацнути комірку В14 з числом інтервалів / **Enter**.

- аналогічно в комірці А20-А27 вставити формули для обчислення правих меж інтервалів: клацнути комірку А20, ввести знак = (дорівнює) / клацнути клітинку з мінімальним значенням Н9 і натиснути клавішу **F4** для переходу до абсолютного посилання / ввести знак + (плюс) і клацнути комірку з значенням ширини інтервалу В15 / **Enter**; в комірку А21 ввести формулу = А20 + $ В $ 15; в комірці А22 і нижче розтягнути формулу з комірки А20 за допомогою автозаповнення: після введення формули в А21 вказати на нижній правий кут комірки А21 до появи маркера автозаповнення у формі +, натиснути ліву кнопку миші і, утримуючи її, протягнути виділення комірки до А27 і відпустити кнопку миші.

- вставити формулу для розрахунку частот із застосуванням функції масивів: виділити діапазон комірок В20-В27 / інструмент **Вставка функцій** / знайти і вибрати функцію **ЧАСТОТА** з категорії **Статистичні** та **ОК** / у вікні вставки функції праворуч від поля **Масив даних** клацнути кнопку згортання / виділити мишею діапазон комірок A2: E11 / клацнути кнопку розгортання / справа від поля **Масив інтервалів** клацнути кнопку згортання / виділити мишею діапазон комірок A20: A27 / клацнути кнопку розгортання / одночасно натиснути клавіші **Ctrl, Shift, Enter** для фіксації функції масиву.

3. Побудувати гістограму для досліджуваної величини із застосуванням майстра діаграм: виділити діапазон комірок з таблицею частот A20: B27 / інструмент **Майстер діаграм** / на вкладці **Нестандартні** в полі **Тип** вибрати **Графік / Гістограма 2** і кнопка **Далі** / в вікні **Вихідні дані** на вкладці **Діапазон даних** включити перемикач в стовпцях / на вкладці **Ряд** клацнути кнопку згортання праворуч від поля **Підписи по осі X** / виділити діапазон комірок A20: A27 і клацнути кнопку розгортання / в полі **Підписи другої осі X** внести діапазон комірок B20: B27 і кнопка **Далі** / в вікні розміщення діаграми включити перемикач **на наявному листі** і **ОК**.

Розглянемо ще один приклад**.** Необхідно виконати процедуру генерації випадкових чисел і проаналізувати їх за допомогою засобів **Аналіз даних** і **Майстер діаграм** MS Excel.

1. Для цього необхідно перейти на вільний робочий лист книги і перейменувати його в **Генерація даних.**

2. Підключити надбудову **Пакет аналізу** MS Excel: **Сервіс / Надбудови** / в вікні Надбудови встановити прапорець **Пакет аналізу** та **ОК**.

3. Виконати генерацію 30 випадкових чисел, розподілених відповідно до нормального закону з нульовим середнім і дисперсією 1: клацнути комірку A1 і **Сервіс / Аналіз даних** / в полі зі списком **Інструмент аналізу** клацнути позицію **Генерація випадкових чисел** і **ОК** / в полі **Число змінних** ввести 1, в поле **Число випадкових чисел** ввести 30, розкрити список поля **Розподіл** і вибрати позицію **Нормальне**, ввести в полях **Середнє** – 0, **Стандартне відхилення** – 1, в розділі **Параметри виведення**

включити перемикач **вихідний інтервал**, клацнути кнопку згортання / клацнути комірку А1 і кнопку розгортання, **ОК**.

4. Змінити розрядність даних, для цього необхідно зменшити число знаків після коми до двох: виділити діапазон комірок А1: А30 / клацнути інструмент **Зменшити розрядність** чотири рази.

5. Виконати процедуру описової статистики по згенерованих даних: **Сервіс / Аналіз даних / Описова статистика** і **ОК** / в вікні **Описова статистика** в полі **Вхідний інтервал** ввести посилання на діапазон комірок А1: А30 / в розділі **Групування** включити перемикач по стовпцях і забрати прапорець **Мітки в першому рядку** / в розділі **Параметри виведення** включити перемикач **Вихідний інтервал** і клацнути комірку С1 / встановити прапорець **Підсумкова статистика** і **ОК**.

6. Побудувати гістограму за даними стовпчика А: **Сервіс / Аналіз даних / Гістограма** і **ОК** / в вікні **Гістограмма** в розділі **Вхідні дані** в поле **Вхідний інтервал** ввести посилання на діапазон комірок А1: А30 і встановити прапорець Мітки / в розділі параметри виводу включити перемикач **Вихідний інтервал** і вказати будь-яку вільну комірку робочого листа / встановити прапорець **Інтегральний відсоток** і **Висновок графіка** і **ОК**. Додати на побудовану гістограму 2 лінії тренду: поліноміальної зі ступенем 4 і змінного середнього на 2 точки: клацнути правою клавішею миші по рядах значень, на контекстному меню вибрати позицію **Додати лінію тренда** / на вкладці тип вибрати Поліноміальна, в поле **ступінь** ввести 4 і т.д.

10.2. Розрахунок параметрів кореляції з використанням засобів Вставка функцій і Майстер діаграм MS Excel

Розглянемо приклад розрахунку коефіцієнта кореляції (перевірки наявності зв'язку між двома змінними) з використанням засобів Вставка функцій і Майстер діаграм MS Excel. Спочатку необхідно провести візуальний аналіз даних і вже потім провести розрахунок коефіцієнта кореляції в рамках завдання перевірки наявності зв'язку між двома змінними з використанням засобів Вставка функцій і Майстер діаграм MS Excel. Для цього необхідно виконати наступні дії.

1. Запустити MS Excel і перейменувати ярличок робочого листа в **Діаграма розсіювання.**

2. Сформувати масив вихідних. Пропонуємо розглянути на даних результатів вимірювань діаметра і висоти у 10 дерев сосни звичайної за зразком:

x_i 19,617,7 13,5 21,9 15,1 11,4 21,1 16,4 21,5 19,3
y_i 19,018,9 17,4 19,2 19,8 10,7 20,2 19,1 22,1 20,5

3. Використовуючи засіб Майстер діаграм, побудувати діаграму розсіювання: виділити діапазон комірок А1: В11 / інструмент **Майстер діаграм** / тип діаграми **Точкова** виду 1 і **Далі** / на вкладці **Діапазон даних** встановити прапорець **Ряди в стовпцях** і **Далі** / ввести заголовки діаграми **Діаграма розсіювання**, осі Х – діаметр , осі У – висота дерев / зняти відображення легенди / встановити відображення основних і проміжних ліній сітки по обох осях / розташувати діаграму на наявному листі/ додати лінійний тренд з включенням відображення рівняння (вкладка **Параметри** вікна **Лінія тренду**).

4. Проаналізувати отримані дані і зробити попередній висновок про наявність лінійного зв'язку між розглянутими ознаками.

5. Розрахувати коефіцієнт кореляції для досліджуваних вибірок. Виділити клітинку **А12** і ввести текст **Коефіцієнт кореляції** / виділити комірку **В12** / клацнути інструмент **Вставка функцій** / у вікні **Майстер функцій** в полі **Категорії** клацнути **Статистичні**, в полі **Функція** за допомогою смуги прокрутки перегорнути список назви функцій, знайти і клацнути **КОРЕЛЛ** і **ОК** / в полях **Массив1** і **Массив2** ввести послідовно посилання на діапазони **А1: А11** і **В1: В11** відповідно і **ОК**.

6. Перевірити значимість отриманого значення коефіцієнта кореляції за критерієм Стьюдента. Виділити клітинку **A13** і ввести текст **t-статистика** / виділити комірку **B13** і ввести формулу за зразком **= B12 * КОРЕНЬ (СЧЕТ (A2: A11) -2) / КОРЕНЬ (1-B12 * B12)**. Виділити клітинку **A14** і ввести текст **Критичне значення**. Виділити клітинку B14 і ввести формулу за зразком = **СТЪЮДРАСПОБР (0,05;СЧЕТ(A2:A11)-2).** Якщо значення t більше табличного (критичного), то приймається наявність значного лінійного зв'язку (відкидається припущення про відсутність зв'язку).

Для того, щоб провести візуальний і кореляційний аналіз даних в рамках завдання перевірки наявності зв'язку між двома змінними з використанням засобів **Аналіз даних** і **Майстер діаграм** MS Excel пропонуємо:

1. Перейти на **Лист 2** книги і перейменувати його ярличок в **Корел. аналіз.** Скопіювати попередній приклад на новий лист: виділити діапазон комірок A1: B11 листа **Діаграма розсіювання /** клацнути правою кнопкою миші для виклику контекстного меню / **Копіювати** / перейти на новий робочий лист / виділити комірку **A1 /** клацнути правою кнопкою миші і **Вставити.**

2. Розрахувати кореляцію для досліджуваних даних, використовуючи групу **Кореляція** засоби **Аналіз даних**: **Сервіс / Аналіз даних** / в полі зі списком **Інструмент аналізу** клацнути позицію **Кореляція** і **ОК** / в полі **Вхідний інтервал** ввести посилання на діапазон A1: B11 / встановити прапорці в **Групування за стовпцями** і **Мітки в першому рядку** / в полі **вихідний інтервал** вказати посилання на клітинку D1 і **ОК**.

10.3. Однофакторний дисперсійний аналіз з використанням засобів Аналіз даних MS Excel

Щоб провести однофакторний дисперсійний аналіз впливу одного фактора на характеристики декількох експериментальних груп з використанням засобів **Аналіз даних** MS Excel необхідно виконати наступні дії:

1. Перейти на **Лист 3** книги і перейменувати його ярличок в **Однофакт. аналіз.**

2. Сформувати масив вихідних даних впливу на висоту тополевих саджанців заготовки живців за зразком:

Фактор (живці різної довжини)	Висота по повторностях					
1	14	22	18	27	6	45
2	26	41	47	32	35	27
3	25	43	28	21	13	26
4	15	16	12	14	20	16

3. Запустити процедуру однофакторного дисперсійного аналізу: **Сервіс / Аналіз даних / Однофакторний дисперсійний аналіз** / вказати діапазон вхідних значень В2: G16, групування за **стовпцями**, прапорець **Мітки** зняти, вказати клітинку вихідного діапазону комірок **В5**.

4. Проаналізувати отримані результати: порівняти дисперсію всередині груп (характеризує вплив випадкової складової) і між групами (характеризує вплив досліджуваного фактора – живці різної довжини). Якщо вони значимо відрізняються (рівень значимості Р = 0,05), то вважається, що фактор впливає на досліджувану змінну. Чи є цей вплив достовірним визначають на порівнянні F фактичного і теоретичного значення статистики Фішера. Відмінність вважається значимою, якщо розрахункове значення більше теоретичного.

Запитання для самоконтролю:

1. Які основні переваги розрахунку біометричних показників за допомогою комп’ютерної техніки? Чи є недоліки?

2. Опишіть дії, які необхідно виконати для рішення задач описової статистики засобами MS EXCEL.

3. Опишіть дії, які необхідно виконати для розрахунку коефіцієнта кореляції засобами MS EXCEL.

4. Опишіть особливості дисперсійного аналізу засобами MS EXCEL.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Барановський Д.І. Біометрія в програмному середовищі MS Excel. / Д.І. Барановський, О.М. Гетманець, А.М. Хохлов. Х.:СПД ФО Бровін О.В., 2017. 90 с.
2. Багинский В.Ф. Биометрия в лесном хозяйстве: учебное пособие для студентов высших учебных заведений, обучающихся по специальностям «Лесное хозяйство», «Лесоинженерное дело», «Садово-парковое строительство» / В.Ф. Багинский, О.В. Лапицкая. Гомель: ГГУ им. Ф. Скорины, 2010. 374 с.
3. Горбунов Л.В. Біометрія: навчальний посібник / Л.В. Горбунов, М.Ф. Клещев. Х.: НТУ «ХПИ», 2014. 160 с.
4. Горошко М.П. Біометрія: Навчальний посібник. / Горошко М.П., Миклуш С.І., Хомюк П.Г. Львів: Камула, 2004. 236 с.
5. Горошко М.П. Практикум з лісової біометрії. / Горошко М.П., Миклуш С.І., Хомюк П.Г. Львів: УкрДЛТУ, 1999. 108 с.
6. Лакин Г.Ф. Биометрия: учебное пособие для биологических специальностей вузов. / Лакин Г.Ф.: Высшая школа, 1980. 294 с.
7. 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.
8. Чепур С.С. Біометрія: методичний посібник. Ужгород: Видавництво УжНУ «Говерла», 2020. 40 с.

ДОДАТКИ

Додаток 1

Стандартні значення критерію Стъюдента

Число ступенів волі $u=n_1+n_2-2$	Критерій Стъюдента t_{st} при імовірності безпомилкового заключення p			
	0.1	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.952
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.684	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.732	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.723	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.714	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
∞	1.645	1.960	2.326	2.576

Додаток 2

Значення критерію Фішера F

U_2	Степінь волі для більшої дисперсії U_1											
	3	4	5	6	8	10	12	16	24	30	50	∞
5%-ний рівень значущості $F_{0,05}$												
3	9,3	9,1	9,0	8,9	8,8	8,8	8,7	8,7	8,6	8,6	8,6	8,5
4	6,6	6,4	6,3	6,2	6,0	5,9	5,8	5,8	5,8	5,7	5,7	5,6
5	5,4	5,2	5,1	5,0	4,8	4,7	4,7	4,6	4,5	4,5	4,4	4,4
6	4,8	4,5	4,4	4,3	4,2	4,1	4,0	3,9	3,8	3,8	3,8	3,7
7	4,4	4,1	4,0	3,9	3,7	3,6	3,6	3,5	3,4	3,4	3,4	3,2
8	4,1	3,8	3,7	3,6	3,4	3,4	3,4	3,2	3,1	3,1	3,0	2,9
9	3,9	3,6	3,5	3,4	3,2	3,2	3,1	3,0	2,9	2,9	2,8	2,7
10	3,7	3,5	3,3	3,2	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,5
12	3,5	3,3	3,1	3,0	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,3
16	3,2	3,1	2,9	2,8	2,7	2,5	2,4	2,3	2,2	2,2	2,1	2,0
18	3,2	2,9	2,8	2,7	2,5	2,4	2,3	2,3	2,2	2,1	2,0	1,9
24	3,0	2,8	2,6	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,9	1,7
40	2,8	2,6	2,5	2,3	2,2	2,1	2,0	1,9	1,8	1,7	1,7	1,5
120	2,7	2,4	2,3	2,2	2,0	1,9	1,8	1,7	1,6	1,6	1,5	1,3
∞	2,6	2,4	2,2	2,1	1,9	1,8	1,8	1,6	1,5	1,5	1,4	1,0
1%-ний рівень значущості $F_{0,01}$												
3	29,5	28,7	28,2	27,9	27,5	27,0	26,8	26,6	26,5	26,5	26,4	26,1
4	16,7	16,0	15,5	15,2	14,8	14,5	14,4	14,2	13,9	13,8	13,7	13,5
5	12,1	11,4	11,0	10,7	10,3	10,1	9,9	9,7	9,5	9,4	9,2	9,0
6	9,8	9,2	8,8	8,5	8,1	7,9	7,7	7,5	7,3	7,2	7,1	6,9
7	8,4	7,8	7,5	7,2	6,8	6,6	6,5	6,3	6,1	6,0	5,8	5,6
8	7,6	7,0	6,6	6,4	6,0	5,8	5,7	5,5	5,3	5,2	5,1	4,9
9	7,0	6,4	6,1	5,8	5,5	5,3	5,1	4,9	4,7	4,6	4,5	4,3
10	6,6	6,0	5,6	5,4	5,1	4,9	4,7	4,5	4,3	4,3	4,1	3,9
12	6,0	5,4	5,1	4,8	4,5	4,3	4,2	4,0	3,8	3,7	3,6	3,4
16	5,3	4,8	4,4	4,2	3,9	3,7	3,6	3,4	3,2	3,1	3,0	2,8
20	4,9	4,4	4,1	3,9	3,6	3,4	3,2	3,0	2,9	2,8	2,6	2,4
30	4,5	4,0	3,7	3,5	3,2	3,0	2,8	2,7	2,5	2,4	2,2	2,0
60	4,1	3,6	3,3	3,1	2,8	2,6	2,5	2,3	2,1	2,0	1,9	1,6
120	3,9	3,5	3,2	3,0	2,7	2,4	2,3	2,2	1,9	1,9	1,7	1,4
∞	3,8	3,3	3,0	2,8	2,5	2,3	2,2	2,0	1,8	1,7	1,5	1,0

Додаток 3

Значення Z, що відповідає величинам коефіцієнта кореляції r

r	Соті долі коефіцієнта кореляції r									
	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	0,111	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	0,498	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,663	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,776	0,793	0,811	0,829	0,848
0,7	0,867	0,887	0,908	0,929	0,951	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
0,9	1,472	1,528	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

Додаток 4

Значення F-критерію Фішера
(верхній ряд – 99,9%, середній – 99%, нижній – 95%)

	K_1																								
K_2	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	K_2
3	167,5	148,5	141,1	137,1	134,6	132,9	131,8	130,6	130,0	129,5	128,9	128,3	127,7	127,1	126,5	125,9	125,6	125,3	125,0	124,7	124,4	124,1	123,8	123,5	3
	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,7	27,4	27,5	27,1	27,1	26,9	26,8	26,7	26,6	26,5	26,4	26,4	26,3	26,2	26,2	26,1	26,1	
	10,1	9,6	9,3	9,1	9,0	8,9	8,8	8,8	8,8	8,8	8,8	8,7	8,7	8,7	8,7	8,6	8,6	8,6	8,6	8,6	8,6	8,5	8,5	8,5	
4	74,1	61,2	66,1	53,4	51,7	50,5	49,8	49,0	48,6	48,2	47,8	47,4	47,0	46,6	46,2	45,8	45,6	45,4	45,2	45,0	44,7	44,5	44,3	44,1	4
	21,2	18,8	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,7	14,5	14,4	14,2	14,1	14,0	13,9	13,9	13,8	13,7	13,7	13,6	13,5	13,5	13,5	
	7,7	6,9	6,6	6,4	6,3	6,2	6,1	6,0	6,0	6,0	5,9	5,9	5,9	5,8	5,8	5,8	5,8	5,7	5,7	5,7	5,7	5,7	5,6	5,6	
5	47,0	36,6	33,2	31,1	29,8	28,8	28,2	27,6	27,3	27,0	26,7	26,4	26,1	25,8	25,4	25,1	24,9	24,8	24,6	24,5	24,3	24,1	24,0	23,8	5
	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	10,0	9,9	9,7	9,6	9,5	9,4	9,3	9,2	9,1	9,1	9,1	9,1	9,0	9,0	
	6,6	5,8	5,4	5,2	5,1	5,0	4,9	4,8	4,8	4,7	4,7	4,7	4,6	4,6	4,6	4,5	4,5	4,5	4,4	4,4	4,4	4,4	4,4	4,4	
6	35,5	27,0	23,7	21,9	20,8	20,0	19,5	19,0	18,8	18,5	18,3	18,0	17,7	17,5	17,2	16,9	16,8	16,6	16,5	16,4	16,2	16,1	15,9	15,8	6
	13,4	10,9	9,8	9,2	8,8	8,5	8,3	8,1	8,0	7,9	7,8	7,7	7,6	7,5	7,4	7,3	7,2	7,1	7,1	7,0	7,0	6,9	6,9	6,9	
	6,0	5,1	4,8	4,5	4,4	4,3	4,2	4,1	4,1	4,1	4,0	4,0	4,0	3,9	3,9	3,8	3,8	3,8	3,8	3,7	3,7	3,7	3,7	3,7	
7	29,2	21,7	18,8	17,2	16,2	15,5	15,1	14,6	14,4	14,2	13,9	13,7	13,5	13,2	13,0	12,7	12,6	12,5	12,3	12,2	12,1	12,0	11,8	11,7	7
	12,3	9,6	8,5	7,9	7,5	7,2	7,0	6,8	6,7	6,6	6,5	6,5	6,4	6,2	6,2	6,1	6,0	5,9	5,9	5,8	5,8	5,7	5,7	5,7	
	5,6	4,7	4,4	4,1	4,0	3,9	3,8	3,7	3,7	3,6	3,6	3,6	3,5	3,5	3,4	3,4	3,4	3,3	3,3	3,3	3,3	3,3	3,2	3,2	
8	25,4	18,5	15,8	14,4	13,5	12,9	12,5	12,0	11,8	11,6	11,4	11,2	11,0	10,8	10,5	10,3	10,2	10,1	10,0	9,9	9,7	9,6	9,5	9,4	8
	11,3	8,7	7,6	7,0	6,6	6,4	6,2	6,0	5,9	5,8	5,7	5,7	5,6	5,5	5,4	5,3	5,2	5,1	5,1	5,0	5,0	4,9	4,9	4,9	
	5,3	4,6	4,1	3,8	3,7	3,6	3,5	3,4	3,4	3,3	3,3	3,3	3,2	3,2	3,2	3,1	3,1	3,1	3,0	3,0	3,0	3,0	2,9	2,9	
9	22,9	16,4	13,9	12,6	11,7	11,1	10,8	10,4	10,2	10,0	9,8	9,6	9,4	9,2	8,9	8,7	8,6	8,5	8,4	8,3	8,1	8,0	7,9	7,8	9
	10,6	8,0	7,0	6,4	6,1	5,8	5,6	5,5	5,4	5,3	5,2	5,1	5,0	4,9	4,8	4,7	4,6	4,6	4,5	4,5	4,4	4,4	4,3	4,3	
	5,1	4,3	3,6	3,6	3,5	3,4	3,3	3,2	3,2	3,1	3,1	3,1	3,0	3,0	2,9	2,9	2,8	2,8	2,8	2,8	2,8	2,7	2,7	2,7	

Продовження таблиці

K_2	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	K_2
	21,0	14,9	12,3	11,3	10,5	9,9	9,6	9,2	9,0	8,9	8,7	8,5	8,3	8,1	7,8	7,6	7,5	7,4	7,3	7,2	7,1	7,0	6,9	6,8	
10	10,0	7,9	6,6	6,0	5,6	5,4	5,2	5,1	5,0	4,9	4,8	4,7	4,6	4,5	4,4	4,3	4,3	4,2	4,1	4,1	4,0	4,0	3,9	3,9	10
	5,0	4,1	3,7	3,5	3,3	3,2	3,1	3,1	3,0	3,0	2,9	2,9	2,9	2,8	2,8	2,7	2,7	2,7	2,6	2,6	2,6	2,6	2,6	2,5	
	19,7	13,8	11,6	10,4	9,6	9,1	8,8	8,4	8,2	8,0	7,8	7,6	7,4	7,3	7,1	6,9	6,8	6,7	6,6	6,5	6,3	6,2	6,1	6,0	
11	9,7	7,2	6,2	5,7	5,3	5,1	4,9	4,7	4,6	4,5	4,5	4,4	4,3	4,2	4,1	4,0	3,9	3,9	3,8	3,7	3,7	3,7	3,6	3,6	11
	4,8	4,0	3,6	3,4	3,2	3,1	3,0	3,0	2,9	2,9	2,8	2,8	2,7	2,7	2,7	2,6	2,6	2,5	2,5	2,5	2,4	2,4	2,4	2,4	
	18,6	12,3	10,8	9,6	8,9	8,4	8,1	7,7	7,5	7,4	7,2	7,0	6,8	6,7	6,5	6,3	6,2	6,1	6,0	5,9	5,7	5,6	5,5	5,4	
12	9,3	6,9	6,0	5,4	5,1	4,8	4,7	4,5	4,4	4,3	4,2	4,2	4,1	4,0	3,9	3,8	3,7	3,6	3,6	3,5	3,5	3,4	3,4	3,4	12
	4,8	3,9	3,5	3,3	3,1	3,0	2,9	2,9	2,8	2,8	2,7	2,7	2,6	2,6	2,5	2,5	2,5	2,4	2,4	2,4	2,4	2,3	2,3	2,3	
	17,8	12,3	10,2	9,1	8,4	7,9	7,6	7,2	7,0	6,9	6,7	6,5	6,3	6,2	6,0	5,8	5,7	5,6	5,5	5,4	5,3	5,2	5,1	5,0	
13	9,1	6,7	5,7	5,2	4,9	4,6	4,4	4,3	4,2	4,1	4,0	4,0	3,9	3,8	3,7	3,6	3,5	3,4	3,4	3,3	3,3	3,2	3,2	3,2	13
	4,7	3,8	3,4	3,2	3,0	2,9	2,8	2,8	2,7	2,7	2,6	2,6	2,6	2,5	2,5	2,5	2,4	2,4	2,3	2,3	2,3	2,3	2,2	2,2	
	17,1	11,8	9,7	8,6	7,9	7,7	7,1	6,8	6,6	6,5	6,3	6,1	5,9	5,8	5,6	5,4	5,3	5,2	5,1	5,0	4,9	4,8	4,7	4,6	
14	8,9	6,5	5,6	5,0	4,7	4,5	4,3	4,1	4,0	3,9	3,9	3,8	3,7	3,6	3,5	3,4	3,3	3,3	3,2	3,1	3,1	3,1	3,0	3,0	14
	4,6	3,7	3,3	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,6	2,5	2,5	2,4	2,4	2,4	2,3	2,3	2,2	2,2	2,2	2,2	2,1	2,1	
	16,6	11,3	9,3	8,3	7,6	7,1	6,8	6,5	6,3	6,2	6,0	5,8	5,6	5,5	5,3	5,1	5,0	4,9	4,8	4,7	4,6	4,5	4,4	4,3	
15	8,7	6,4	5,4	4,9	4,6	4,3	4,1	4,0	3,9	3,8	3,7	3,7	3,6	3,5	3,4	3,3	3,2	3,1	3,1	3,0	3,0	2,9	2,9	2,9	15
	4,5	3,7	3,3	3,1	2,9	2,8	2,7	2,6	2,6	2,6	2,5	2,5	2,4	2,4	2,3	2,3	2,3	2,2	2,2	2,2	2,1	2,1	2,1	2,1	
	16,1	11,0	9,0	7,9	7,3	6,8	6,5	6,2	6,1	5,9	5,8	5,6	5,4	5,3	5,1	4,9	4,8	4,7	4,6	4,5	4,4	4,3	4,2	4,1	
16	8,5	6,2	5,3	4,8	4,4	4,2	4,0	3,9	3,8	3,7	3,6	3,5	3,5	3,4	3,3	3,2	3,1	3,0	3,0	2,9	2,9	2,8	2,8	2,8	16
	4,5	3,6	3,2	3,0	2,9	2,7	2,7	2,6	2,5	2,5	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,2	2,1	2,1	2,1	2,0	2,0	2,0	

Продовження таблиці

K_2	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	K_2
	15,7	10,7	8,7	7,7	7,0	6,6	6,3	6,0	5,8	5,7	5,5	5,3	5,1	5,0	4,8	4,6	4,5	4,4	4,3	4,3	4,2	4,1	4,0	3,9	
17	8,4	6,1	5,2	4,7	4,3	4,1	3,9	3,8	3,7	3,6	3,5	3,5	3,4	3,3	3,2	3,1	3,0	2,9	2,9	2,8	2,8	2,7	2,7	2,7	17
	4,5	3,6	3,2	3,0	2,8	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,2	2,1	2,1	2,1	2,0	2,0	2,0	2,0	2,0	
	15,4	10,4	8,5	7,5	6,8	6,4	6,1	5,8	5,6	5,5	5,3	5,1	5,0	4,8	4,7	4,5	4,4	4,3	4,2	4,1	4,0	3,9	3,8	3,7	
18	8,3	6,0	5,1	4,6	4,2	4,0	3,8	3,7	3,6	3,5	3,4	3,4	3,3	3,2	3,1	3,0	2,9	2,8	2,8	2,7	2,7	2,6	2,6	2,6	18
	4,4	3,5	3,2	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,9	
	15,2	10,2	8,3	7,3	6,6	6,2	5,9	5,6	5,5	5,3	5,2	5,0	4,8	4,7	4,5	4,4	4,2	4,1	4,0	3,9	3,8	3,7	3,6	3,5	
19	8,2	5,9	5,0	4,5	4,2	3,9	3,8	3,6	3,5	3,4	3,4	3,3	3,2	3,1	3,0	2,9	2,9	2,8	2,7	2,6	2,6	2,5	2,5	2,5	19
	4,4	3,5	3,1	2,9	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,3	2,3	2,2	2,1	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,9	1,9	
	14,8	10,0	8,1	7,1	6,5	6,0	5,7	5,4	5,3	5,1	5,0	4,8	4,7	4,5	4,4	4,2	4,1	4,0	3,9	3,8	3,7	3,6	3,5	3,4	
20	8,1	5,8	4,9	4,4	4,1	3,9	3,7	3,6	3,4	3,4	3,3	3,2	3,1	3,0	2,9	2,9	2,8	2,7	2,6	2,6	2,5	2,5	2,4	2,4	20
	4,3	3,5	3,1	2,9	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,9	1,8	1,8	
	14,6	9,8	7,9	7,0	6,3	5,9	5,6	5,3	5,2	5,0	4,9	4,7	4,5	4,4	4,2	4,0	3,9	3,8	3,7	3,6	3,5	3,5	3,4	3,3	
21	8,0	5,8	4,9	4,4	4,0	3,8	3,6	3,5	3,4	3,3	3,2	3,2	3,1	3,0	2,9	2,8	2,7	2,6	2,6	2,5	2,5	2,4	2,4	2,4	21
	4,3	3,5	3,1	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,9	1,8	1,8	1,8	
	14,4	9,6	7,8	6,8	6,2	5,8	5,5	5,2	5,1	4,9	4,8	4,6	4,4	4,3	4,1	3,9	3,8	3,7	3,6	3,5	3,5	3,4	3,3	3,2	
22	7,9	5,7	4,8	4,3	4,0	3,8	3,6	3,4	3,3	3,3	3,2	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,3	22
	4,3	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,8	1,8	
	14,2	9,5	7,7	6,7	6,1	5,6	5,4	5,1	5,0	4,8	4,7	4,5	4,3	4,2	4,0	3,8	3,7	3,6	3,5	3,5	3,4	3,3	3,2	3,1	
23	7,9	5,7	4,8	4,3	4,0	3,7	3,5	3,4	3,3	3,2	3,1	3,1	3,0	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,3	2,3	23
	4,3	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,8	1,8	

K_2	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500
	14,0	9,3	7,6	6,6	6,0	5,6	5,3	5,0	4,9	4,7	4,6	4,4	4,2	4,1	3,9	3,7	3,6	3,5	3,4	3,4	3,3	3,2	3,1	3,0
24	7,8	5,6	4,7	4,2	3,9	3,7	3,5	3,4	3,2	3,2	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2
	4,3	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,9	1,8	1,8	1,8	1,7	1,7
	13,9	9,2	7,5	6,5	5,9	5,5	5,2	4,9	4,8	4,6	4,5	4,3	4,2	4,0	3,9	3,7	3,6	3,5	3,4	3,3	3,2	3,1	3,0	2,9
25	7,8	5,6	4,7	4,2	3,9	3,6	3,5	3,3	3,2	3,1	3,0	3,0	2,9	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,2
	4,2	3,4	3,0	2,8	2,6	2,5	2,4	2,3	2,3	2,2	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7
	13,7	9,1	7,4	6,4	5,8	5,4	5,1	4,8	4,7	4,5	4,4	4,2	4,1	3,9	3,8	3,6	3,5	3,4	3,3	3,2	3,1	3,0	2,9	2,8
26	7,7	5,5	4,6	4,1	3,8	3,6	3,4	3,3	3,2	3,1	3,0	3,0	2,9	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,2	2,2	2,1	2,1
	4,2	3,3	3,1	2,7	2,6	2,5	2,4	2,3	2,2	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7	1,7
	13,6	9,0	7,3	6,3	5,7	5,3	5,1	4,8	4,7	4,5	4,4	4,2	4,0	3,9	3,7	3,5	3,4	3,3	3,2	3,2	3,1	3,0	2,9	2,8
27	7,7	5,5	4,6	4,1	3,8	3,6	3,4	3,3	3,1	3,1	3,0	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,3	2,2	2,2	2,2	2,1	2,1
	4,2	3,3	3,1	2,7	2,6	2,5	2,4	2,3	2,2	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7	1,7
	13,5	8,9	7,2	6,3	5,7	5,2	5,0	4,7	4,6	4,4	4,3	4,1	4,0	3,8	3,7	3,5	3,4	3,3	3,2	3,1	3,0	2,9	2,8	2,7
28	7,6	5,4	4,6	4,1	3,8	3,5	3,4	3,2	3,1	3,0	2,9	2,9	2,8	2,7	2,6	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,1
	4,2	3,3	2,9	2,7	2,6	2,4	2,4	2,3	2,2	2,2	2,1	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,7	1,7	1,6
	13,4	8,9	7,1	6,2	5,6	5,2	5,0	4,7	4,6	4,4	4,3	4,1	4,0	3,8	3,6	3,4	3,3	3,2	3,1	3,0	2,9	2,8	2,7	2,6
29	7,6	5,4	4,5	4,0	3,7	3,5	3,3	3,2	3,1	3,0	2,9	2,9	2,8	2,7	2,6	2,5	2,4	2,3	2,3	2,2	2,1	2,1	2,1	2,0
	4,2	3,3	2,9	2,7	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7	1,6	1,6
	13,3	8,8	7,1	6,1	5,5	5,1	4,9	4,6	4,5	4,3	4,2	4,0	3,9	3,7	3,6	3,4	3,3	3,2	3,1	3,0	2,9	2,8	2,7	2,6
30	7,6	5,4	4,5	4,0	3,7	3,5	3,3	3,2	3,1	3,0	2,9	2,8	2,7	2,7	2,5	2,5	2,4	2,3	2,2	2,2	2,1	2,1	2,0	2,0
	4,2	3,3	2,9	2,7	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7	1,6	1,6

Продовження таблиці

K_2	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	K_2
32	13,2	8,7	7,0	6,0	5,4	5,0	4,8	4,5	4,4	4,2	4,1	3,9	3,8	3,7	3,5	3,4	3,2	3,2	3,1	3,0	2,8	2,7	2,6	2,5	32
	7,5	5,3	4,5	4,0	3,7	3,4	3,2	3,1	3,0	2,9	2,9	2,8	2,7	2,6	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	2,0	
	4,1	3,3	2,9	2,7	2,5	2,4	2,3	2,2	2,2	2,1	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,7	1,6	1,6	1,6	
34	13,1	8,6	7,0	6,0	5,4	5,0	4,8	4,5	4,4	4,2	4,1	3,9	3,8	3,6	3,5	3,3	3,2	3,1	3,0	2,9	2,8	2,6	2,6	2,5	34
	7,4	5,3	4,4	3,9	3,6	3,4	3,2	3,1	3,0	2,9	2,2	2,8	2,7	2,6	2,5	2,4	2,3	2,2	2,1	2,1	2,0	2,0	1,9	1,9	
	4,1	3,3	2,9	2,6	2,5	2,4	2,3	2,2	2,1	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,6	1,6	1,5	
36	13,0	8,6	6,9	5,9	5,3	4,9	4,7	4,4	4,3	4,1	4,0	3,8	3,7	3,6	3,4	3,3	3,1	3,1	3,0	2,9	2,7	2,6	2,5	2,4	36
	7,4	5,2	4,4	3,9	3,6	3,3	3,2	3,0	2,9	2,9	2,8	2,7	2,6	2,5	2,4	2,3	2,3	2,2	2,1	2,0	2,0	1,9	1,9	1,9	
	4,1	3,3	2,9	2,6	2,5	2,4	2,3	2,2	2,1	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,6	1,6	1,5	
38	12,9	8,5	6,8	5,8	5,3	4,9	4,7	4,4	4,3	4,1	4,0	3,8	3,7	3,5	3,4	3,2	3,1	3,0	2,9	2,8	2,7	2,6	2,5	2,4	38
	7,3	5,2	4,3	3,9	3,5	3,3	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,5	2,4	2,3	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	
	4,1	3,2	2,8	2,6	2,5	2,3	2,3	2,2	2,1	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,6	1,5	1,5	
40	12,8	8,4	6,7	5,8	5,2	4,8	4,6	4,3	4,2	4,0	3,9	3,7	3,6	3,5	3,3	3,2	3,0	3,0	2,9	2,8	2,6	2,5	2,4	2,3	40
	7,3	5,2	4,3	3,8	3,5	3,3	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,5	2,4	2,3	2,2	2,1	2,0	2,0	1,9	1,9	1,8	1,8	
	4,1	3,2	2,8	2,6	2,4	2,3	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,7	1,6	1,6	1,5	1,5	1,5	
42	12,7	8,3	6,7	5,7	5,2	4,8	4,6	4,3	4,2	4,0	3,9	3,7	3,6	3,4	3,3	3,1	3,0	2,9	2,8	2,7	2,6	2,4	2,4	2,3	42
	7,3	5,1	4,3	3,8	3,5	3,3	3,1	3,0	2,9	2,8	2,7	2,6	2,5	2,5	2,3	2,3	2,2	2,1	2,0	1,9	1,9	1,8	1,8	1,8	
	4,1	3,2	2,8	2,6	2,4	2,3	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,6	1,5	1,5	1,5	
44	12,5	8,3	6,6	5,6	5,1	4,7	4,5	4,2	4,1	3,9	3,8	3,6	3,5	3,4	3,2	3,1	2,9	2,9	2,8	2,7	2,5	2,4	2,3	2,2	44
	7,2	5,1	4,3	3,8	3,5	3,2	3,1	2,9	2,8	2,7	2,7	2,6	2,5	2,4	2,3	2,2	2,1	2,1	2,0	1,9	1,9	1,8	1,8	1,7	
	4,1	3,2	2,8	2,6	2,4	2,3	2,2	2,2	2,1	2,0	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,6	1,5	1,5	1,5	

K_2	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	K_2
46	12,4	8,1	6,5	5,6	5,0	4,6	4,4	4,1	4,0	3,8	3,7	3,5	3,4	3,3	3,1	3,0	2,8	2,8	2,7	2,6	2,4	2,3	2,2	2,1	46
	7,2	5,1	4,2	3,8	3,4	3,2	3,0	2,9	2,8	2,7	2,7	2,6	2,5	2,4	2,3	2,2	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	
	4,0	3,2	2,8	2,6	2,4	2,3	2,2	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,8	1,7	1,7	1,5	1,6	1,6	1,5	1,5	1,5	1,5	
48	12,3	8,1	6,5	5,6	5,0	4,6	4,4	4,1	4,0	3,8	3,7	3,5	3,4	3,3	3,1	3,0	2,8	2,8	2,7	2,6	2,4	2,3	2,2	2,1	48
	7,2	5,1	4,2	3,7	3,4	3,2	2,9	2,8	2,8	2,7	2,6	2,6	2,5	2,4	2,3	2,2	2,1	2,0	2,0	1,9	1,8	1,8	1,7	1,7	
	4,0	3,2	2,8	2,6	2,4	2,3	2,2	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,8	1,7	1,7	1,6	1,6	1,6	1,5	1,5	1,5	1,4	
50	12,2	8,0	6,4	5,4	4,9	4,5	4,3	4,0	3,9	3,7	3,6	3,4	3,3	3,2	3,0	2,9	2,7	2,7	2,6	2,5	2,3	2,2	2,1	2,0	50
	7,2	5,1	4,2	3,7	3,4	3,2	3,0	2,9	2,8	2,7	2,6	2,6	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,9	1,8	1,8	1,7	1,7	
	4,0	3,2	2,8	2,6	2,4	2,3	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,5	1,5	1,5	1,5	1,4	
55	12,1	7,9	6,3	5,4	4,9	4,5	4,3	4,0	3,9	3,7	3,6	3,4	3,3	3,2	3,0	2,9	2,7	2,7	2,6	2,5	2,3	2,2	2,1	2,0	55
	7,1	5,0	4,1	3,7	3,4	3,1	3,0	2,8	2,7	2,7	2,6	2,5	2,4	2,3	2,2	2,1	2,1	2,0	1,9	1,8	1,8	1,7	1,7	1,6	
	4,0	3,2	2,8	2,5	2,4	2,3	2,2	2,1	2,0	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,5	1,5	1,5	1,4	1,4	
60	12,0	7,8	6,2	5,3	4,8	4,4	4,2	3,9	3,8	3,6	3,5	3,3	3,2	3,1	2,9	2,8	2,6	2,6	2,5	2,4	2,2	2,1	2,0	1,9	60
	7,1	5,0	4,1	3,6	3,3	3,1	2,9	2,8	2,7	2,6	2,6	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,9	1,8	1,7	1,7	1,6	1,6	
	4,0	3,1	2,8	2,5	2,4	2,2	2,2	2,1	2,0	2,0	1,9	1,9	1,9	1,8	1,7	1,7	1,6	1,6	1,6	1,5	1,5	1,4	1,4	1,4	
65	11,9	7,7	6,1	5,2	4,7	4,3	4,1	3,8	3,7	3,5	3,4	3,2	3,1	3,0	2,8	2,7	2,5	2,5	2,4	2,3	2,1	2,0	1,9	1,8	65
	7,0	5,0	4,1	3,6	3,3	3,1	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,8	1,8	1,7	1,6	1,6	1,6	
	4,0	3,1	2,4	2,5	2,4	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,5	1,5	1,5	1,4	1,4	1,4	
70	11,6	7,6	6,0	5,2	4,7	4,3	4,1	3,8	3,7	3,5	3,4	3,2	3,1	3,0	2,8	2,7	2,5	2,5	2,3	2,2	2,1	2,0	1,8	1,7	70
	7,0	4,9	4,1	3,6	3,3	3,1	2,9	2,8	2,7	2,6	2,5	2,4	2,3	2,3	2,1	2,1	2,0	1,9	1,8	1,7	1,7	1,6	1,6	1,5	
	4,0	3,1	2,7	2,5	2,3	2,2	2,1	2,0	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,5	1,5	1,4	1,4	1,4	1,3	

Продовження таблиці

K_2	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	K_2
	11,6	7,5	6,0	5,1	46,0	4,2	4,0	3,7	3,6	3,4	3,3	3,1	3,0	2,9	2,7	2,6	2,4	2,4	2,3	2,2	2,0	1,9	1,8	1,7	
80	7,0	4,9	4,0	3,6	3,2	3,0	2,9	2,7	2,6	2,5	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,8	1,8	1,7	1,6	1,6	1,5	1,5	80
	4,0	3,1	2,7	2,5	2,3	2,2	2,1	2,0	2,0	1,9	1,9	1,9	1,8	1,8	1,7	1,6	1,6	1,5	1,5	1,4	1,4	1,4	1,3	1,3	
	11,5	7,4	5,9	5,0	4,1	4,1	3,9	3,7	3,6	3,4	3,3	3,1	3,0	2,8	2,7	2,5	2,4	2,3	2,2	2,1	1,9	1,8	1,7	1,6	
100	6,9	4,8	4,0	3,5	3,2	3,0	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,2	2,1	2,0	1,9	1,8	1,7	1,6	1,6	1,5	1,5	1,4	100
	3,9	3,1	2,7	2,5	2,3	2,2	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,5	15,0	1,4	1,4	1,3	1,3	1,3	
	11,4	7,4	5,8	5,0	4,5	4,1	3,9	3,6	3,5	3,3	3,2	3,0	2,9	2,8	2,6	2,5	2,3	2,3	2,1	2,0	1,9	1,8	1,6	1,5	
125	6,8	4,8	3,9	3,5	3,2	2,9	2,8	2,6	2,6	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,8	1,7	1,7	1,6	1,5	1,5	1,4	1,4	125
	3,9	3,1	2,7	2,4	2,3	2,2	2,1	2,0	1,9	1,9	1,9	1,8	1,8	1,7	1,6	1,6	1,5	1,5	1,4	1,4	1,4	1,3	1,3	1,2	
	11,3	7,3	5,7	4,9	4,4	4,0	3,8	3,5	3,4	3,2	3,1	2,9	2,8	2,7	2,5	2,4	2,2	2,2	2,0	1,9	1,8	1,7	1,5	1,4	
150	6,8	4,7	3,9	3,4	3,1	2,9	2,8	2,6	2,5	2,4	2,4	2,3	2,2	2,1	2,0	1,9	1,8	1,7	1,7	1,6	1,5	1,4	1,4	1,3	150
	3,9	3,1	2,7	2,4	2,3	2,2	2,1	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,6	1,6	1,5	1,5	1,4	1,4	1,3	1,3	1,2	1,2	
	11,2	7,2	5,6	4,8	4,3	3,9	3,7	3,5	3,4	3,2	3,1	2,9	2,8	2,6	2,5	2,3	2,2	2,1	1,9	1,8	1,7	1,6	1,4	1,3	
200	6,8	4,7	3,9	3,4	3,2	2,9	2,7	2,6	2,5	2,4	2,3	2,3	2,2	2,1	2,0	1,9	1,8	1,7	1,6	1,5	1,5	1,4	1,3	1,3	200
	3,9	3,0	2,6	2,4	2,3	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,5	1,4	1,4	1,3	1,3	1,3	1,2	1,2	
	11,0	7,1	5,6	4,7	4,2	3,8	3,6	3,4	3,3	3,1	3,0	2,8	2,7	2,5	2,4	2,2	2,1	2,0	1,9	1,8	1,6	1,5	1,4	1,3	
400	6,7	4,7	3,8	3,4	3,1	2,8	2,7	2,5	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,8	1,7	1,6	1,6	1,5	1,4	1,3	1,2	1,2	400
	3,9	3,0	2,6	2,4	2,2	2,1	2,0	2,0	1,9	1,8	1,8	1,8	1,7	1,7	1,6	1,5	1,5	1,4	1,4	1,3	1,3	1,2	1,2	1,1	
	10,9	7,0	5,5	4,7	4,2	3,8	3,6	3,4	3,3	3,1	3,0	2,8	2,7	2,5	2,4	2,2	2,1	2,0	1,8	1,7	1,6	1,5	1,3	1,2	
1000	6,7	4,6	3,8	3,4	3,3	3,0	2,8	2,7	2,5	2,4	2,3	2,3	2,2	2,1	2,0	1,9	1,8	1,7	1,6	1,5	1,4	1,4	1,2	1,1	1000
	3,8	3,0	2,6	2,4	2,2	2,1	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,6	1,6	1,5	1,5	1,4	1,4	1,3	1,3	1,2	1,1	1,1	

K_2	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	K_2
	10,8	6,9	5,4	4,6	4,1	3,7	3,5	3,3	3,2	3,0	2,9	2,7	2,6	2,4	2,3	2,1	2,0	1,9	1,7	1,6	1,5	1,4	1,2	1,1	
>1000	6,6	4,6	3,8	3,3	3,0	2,8	2,6	2,5	2,4	2,3	2,2	2,2	2,1	2,0	1,9	1,8	1,7	1,6	1,5	1,4	1,4	1,2	1,1	1,0	>1000
	3,8	3,0	2,6	2,4	2,2	2,1	2,0	1,9	1,9	1,8	1,8	1,7	1,7	1,6	1,6	1,5	1,5	1,4	1,3	1,3	1,2	1,2	1,1	1,0	

Навчальне видання

БІОМЕТРІЯ

Навчальний посібник

Укладач
Чепур Світлана Степанівна,
кандидат сільськогосподарських наук,
доцент кафедри лісівництва
географічного факультету УжНУ

Гарнітура Times New Roman.
Формат 60х84/16.
Ум.друк.арк. 11,39. Обл.вид.арк. 6,38.
Зам. № 39. Наклад 100 прим.

Оригінал-макет виготовлено
у редакційно-видавничому відділі ДВНЗ «УжНУ».
88000, м.Ужгород, вул. Законовецької, 89.
E-mail: dep-editors@uzhnu.edu.ua

Видавництво Ужгородського національного університету «Говерла».
88000, м. Ужгород, вул. Капітульна, 18.
Свідоцтво про внесення до державного реєстру
видавців, виготівників і розповсюджувачів видавничої продукції
Серія 3т № 32 від 31 травня 2006 року

Б 63

Біометрія: навчальний посібник / Укладач: С.С. Чепур. Ужгород: Вид-во УжНУ «Говерла», 2023. 196 с.
ISBN 978-617-7825-90-5

Посібник включає основи планування дослідження і методи статистичної обробки результатів спостереження.

Для студентів лісогосподарських, біологічних, екологічних факультетів вищих навчальних закладів. Може бути корисним для аспірантів, науковців, практиків, спеціалістів сільської, лісової та суміжних галузей.

УДК 57.087.1(075.8)