В. Б. Левченко
І. В. Шульга
П. І. Трофименко
А. А. Романюк
Г. М. Мачульський

БІОМЕТРІЯ В ЛІСОВОМУ ГОСПОДАРСТВІ З ОСНОВАМИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

В. Б. Левченко, І. В. Шульга, П. І. Трофименко,
А. А. Романюк, Г. М. Мачульський

БІОМЕТРІЯ В ЛІСОВОМУ ГОСПОДАРСТВІ З ОСНОВАМИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

НАВЧАЛЬНИЙ ПОСІБНИК

Рекомендовано для здобувачів закладів вищої, фахової пердвищої освіти, науково-педагогічних працівників спеціальностей: лісове, садово-паркове господарство, екологія.

Житомир
Вид-во ЖДУ ім. І. Франка
2026

УДК 630:57.081:519.22/.25] (075.8)
Л 38

Навчальний посібник друкується за рішенням:

вченої ради ННІ професійної освіти та технологій Національного університету «Чернігівський колегіум» імені Т. Г. Шевченка (м.Чернігів) протокол № 12 від 30.06.2025 р.,

методичної ради Житомирського агротехнічного фахового коледжу (м. Житомир), протокол № 1 від 28.08.2025 р.

Рецензенти: Назаренко Віталій Васильович, кандидат с.-г. наук, доцент, доцент кафедри лісових культур, меліорацій та садово-паркового господарства Державного біотехнологічного університету (м. Харків).
Сидоренко Сергій Григорович, канд. с. - г. наук, старший дослідник, завідувач сектору УкрНДІЛГА ім. Г. М. Висоцького.

Автори: Левченко Валерій Борисович, кандидат с.-г. наук, доцент;
Шульга Ігор Володимирович, кандидат с.-г. наук, доцент;
Трофименко Петро Іванович, доктор с.-г. наук, доцент;
Романюк Алла Андріївна, спеціаліст вищої категорії, викладач-методист, Заслужений працівник освіти України;
Мачульський Григорій Миколайович, кандидат с.-г. наук, доцент.

Біометрія в лісовому господарстві з основами математичної статистики.

Б 142. Навчальний посібник для студентів спеціальності 205 «Лісове господарство» освітньо-професійного ступеня «фаховий молодший бакалавр», початкового рівня (короткого циклу) вищої освіти «молодший бакалавр», першого рівня вищої освіти «бакалавр», другого рівня вищої освіти «магістр». За ред. кандидата с.-г. наук, доцента В. Б. Левченко: - Житомир: Вид-во ЖДУ ім. І. Франка, 2026 р. – 288 с., іл.

Навчальний посібник містить виклад понять про біометрію та варіаційну статистики. Розкрито методи збору, обробки та аналізу біометричної інформації в лісовому господарстві Викладено методологію планування експерименту. Наведено приклади з практики лісівничих досліджень та денрохроноїндикції. Посібник орієнтований на студентів, магістрантів, аспірантів та викладачів лісогосподарських закладів вищої освіти, наукових та інженерно-технічних працівників лісового господарства, лісовпорядкування та садово-паркового господарства, екологів та викладачів біології.

ВСТУП

В лісовому господарстві широко використовуються методи лісової біометрії. Без математико-статистичної обробки сьогодні немислиме проведення лісівничих досліджень. Переважна більшість нормативів, що застосовуються у лісовому господарстві, розроблено з використанням біометричних методів. Застосування комп'ютерів, наявність різноманітних програм для статистичної обробки данних, а в перспективі і використання штучного інтелекту в лісовпорядкуванні та лісогосподарській діяльності робить біометричні методи доступними широкому колу працівників лісового господарства. В той же час лісівник не може бути лише користувачем комп'ютерних результатів біометричних вимірів та обчислень. Він повинен розуміти суть явища, або процесу, що вивчається, розбиратися в алгоритмі, механізмах та особливостях обчислень, які виконав комп'ютер за заданою програмою.

Саме з цією метою у закладах вищої освіти, що готують фахівців лісового господарства для роботи в лісогосподарських філіях ДП «Ліси України», в лісовпорядних організаціях, в садово-парковому господарстві, а також для лісопатологів, мисливствознавців за відповідними освітньо-професійними програмами викладається курс лісової біометрії.

Для того, щоб курс біометрії в лісовому, садово-парковому, мисливському господарстві, який певною мірою вимагає знання математики, був засвоєний, студентам у сучасних закладах вищої освіти викладають досить фундаментальний курс математики. Цей курс за обсягом годинного забезпечення дещо менший, ніж на спеціалізованих математичних та технічних факультетах, але теж достатньо широкий: математика за кількістю відпущених на її вивчення годин зіставна з такими основними лісогосподарськими компонентами, як: лісівництво, лісовпорядкування, лісова таксація, лісознавство та ін., що достатньо для розуміння та засвоєння курсу біометрії в лісовому господарстві. Проте для організації навчального процесу з курсу біометрії, а також для її використання у практиці лісового господарства потрібен сучасний навчальний посібник «Математичні методи у лісовому господарстві», який включав би в себе в т.ч. математичну та варіаційну статистику, які застосовуються в галузі досить тривалий час - з другої половини XIX століття.

Згадаймо відому школу «форстматематиків», яскравим представником якої був академік А. К. Турський. Як самостійна дисципліна, математична статистика почала викладатися в лісівничих вищих навчальних закладах з 30-х років минулого століття. Звичайно, з'явилися відповідні навчальні посібники. Їхніми авторами з 30-х і до початку 60-х років були відомі вчені-статистики О. В. Тюрін, М. Л. Дворецький, М. Г. Здорік, К. Є. Нікітін та інші. З другої половини 60-х та у 70-ті роки видано посібники, написані за

авторством О. А. Труллем, І. І. Гусєвим, Е. М. Фалалєєвим, П. К. Верхуновим, Л. П.Зайченко, Г. Л. Кравченко та Є. С. Мурахтановим та іншими. В 1977 вийшла остання книга з цієї серії – підручник «Варіаційна статистика» за авторством доктора фізико-математичних наук, професора, академіка Н. М. Свалову. Крім вище перерахованих посібників в лісівничій практиці користуються підручниками, де викладається біометрія у загальнобіологічному та загальнотехнічному аспектах. Їх авторами є доктори наук, професори: А. К. Митропольський, Г. Ф. Лакін, Н. А. Плохінський, П. В. Рокіцький, Г. М. Зайцев, В. Л. Вознесенський, Д. Дюче, М. Г. Кенуй, Є. І. Гурський, Б. Л. Ван дер Ванден та ін. Всі ці книги, випущені в основному 50 і більше років тому давно стали бібліографічною рідкістю. До того ж підручники та навчальні посібники, видані у 60-70 роки минулого століття і раніше, вже морально застаріли. Багато з них зорієнтовані на досить обмежене на той час знання математики лісівниками. У той же час надмірно математизований виклад є перешкодою для засвоєння курсу студентами-лісівниками і сьогодні. Тому потрібно, зберігши все те раціональне, що міститься у фундаментальних наукових працях, що були раніше напрацьовані, підготувати новий, більш комунікативний курс біометрії в лісовому господарстві.

В останнє десятиліття було опубліковано низку навчальних посібників з біометрії в лісововому господарстві за авторством М. В. Устінова (Харків) та львівських вчених М. П. Горошко, С. І. Міклуша та П. Г. Хомюка (НЛТУ України), але вони досить насичені складними формулами та біометричними розрахунками, що є складним для осмислення та засвоєння сучасних знань студентами лісівничих закладів вищої освіти. Наприкінці 70-х років 20-го століття, професори К. Є. Нікітін та А. З. Швиденко випустили чудову книгу «Методи та техніка обробки лісівничої інформації». Ця книга призначена в основному для аспірантів та наукових працівників. В даний час вона теж вже бібліографічна рідкість; крім того, для студентів окремі її розділи дещо складні. В силу всього вище описаного, видання сучасного посібника із застосування біометрії в лісовому господарстві є досить актуальним. Швидше за все, варто було б видати кілька посібників різного рівня складності, щоб у студентів та практичних лісівників був вибір доступного їм посібника за рівнем підготовки і вимогам, як це було і раніше.

Під час підготовки цього посібника автори, природньо, спиралися на загально-методичні розробки, що були напрацьовані в галузі математичної статистики та біометрії і в свій час викладені у підручниках відомого математика та статистика професора О. К. Митропільського, математичних теоретиків: Н. А. Плохінського, П. Ф. Рокицького, Г. Ф. Лакіна, Н. М. Свалова, Я. С. Перельмана, Б. Р. Шельмана та інших. Особливе місце під час написання посібника зайняла монографія К. Е. Нікітіна та А. З. Швиденка, де в

сучасній інтерпретації подано багато питань біометрії, що не були висвітлені в інших виданнях, наприклад, система кривих розподілів Джонсона. Тому назване видання широко використане у посиланнях в цьому посібнику. Враховано і нові моменти щодо оцінки вибіркових показників, наведених М. П. Горошко із співавторами. Статистичні таблиці (величини різних критеріїв, функції розподілу і т.д.) зазвичай складені давно та ідентичні для всіх навчальних посібників. Тому вони зазначені в даному посібнику без змін з бібліографічними посиланнями на книги-першоджерела.

Не претендуючи на авторську оригінальність викладу самих теоретичних аспектів формул, їх висновків та тому подібне, що є прерогативою фахівців-математиків, автори посібника зосередили свою увагу на доступності викладу для розуміння матеріалу лісівниками, студентами, а також на практичні додатки до сфери застосування біометричних методів у лісовому господарстві.

В даний час біометричні методи широко застосовуються фаховими молодшими бакалаврами, молодшими бакалаврами, бакалаврами, магістрантами, здобувачами наукового ступеня доктор філософії, науковими та науково-технічними працівниками під час проведення досліджень у лісовому господарстві. Тому справжній курс біометрії, крім того мінімуму знань, який необхідний здобувачам вищої освіти (він визначається освітньо-професійною програмою та регулюється викладачем курсу), повинен задовольняти і потреби вище перерахованої категорії користувачів.

Наведені приклади та ілюстрації взяті в основному з практичних завдань, які свого часу потрібно було вирішувати авторам у своїй багаторічній науково-дослідницькій роботі лісівничого, екологічного та біологічного спрямування. В окремих випадках використані приклади інших авторів, на які в тексті зроблено посилання.

Під час підготовки посібника авторами було враховано власний багаторічний досвід наукової математико-статистичної обробки та аналізу вихідних матеріалів при проведенні наукових досліджень.

Представлений навчальний посібник буде корисним здобувачам освіти за перерахованим вище спеціальностями, а також магістрантам, аспірантам, викладачам, науковцям та співробітникам проектних установ лісогосподарського напряму. Цим посібником зможуть користуватися екологи, біологи та фахівці з лісо-інженерної справи.

Автори приносять глибоку подяку рецензентам посібника: кандидату с.-г. наук, доценту, доценту кафедри лісових культур, меліорацій та садово-паркового господарства Державного біотехнологічного університету (м. Харків) Назаренку Віталію Васильовичу, а також кандидату с.-г. наук, старшому досліднику, завідувачу сектору УкрНДІЛГА ім. Г. М. Висоцького Сидоренку Сергію Григоровичу за цінні поради та зауваження, що дозволили покращити книгу.

РОЗДІЛ 1. БІОМЕТРІЯ ЯК НАУКА

1.1 Визначення біометрії як спеціальної дисципліни під час підготовки інженерів лісового господарства, її цілі та завдання.

1.2 Особливості біометрії як науки та її місце у ряді інших наук.

1.3 Історія виникнення та розвитку математичної статистики та біометрії.

1.4 Лісова біометрія як частина загальної біометрії, її значення для розвитку лісового господарства.

1.1 Визначення біометрії як наукової дисципліни, її мета та завдання.

Слово «біометрія» походить від грецьких слів «bios» - життя і «metreo» - вимірюю, тобто це наука про планування та проведення вимірювань у живій природі, їх узагальненні та аналізі результатів цих вимірів.

Життя, як і всі явища матеріального світу, мають дві нерозривно пов'язані сторони: якісну, сприйняту безпосередньо органами чуття, та кількісну, що виражається числами за допомогою розрахунків та заходів.

При дослідженні різних явищ природи застосовують одночасно і якісні, і кількісні показники. Безсумнівно, що тільки в єдності якісної та кількісної сторін найбільш повно розкривається сутність явищ, що вивчаються. Однак насправді доводиться користуватися, дивлячись за обставинами, або якісними, або переважно кількісними показниками, пам'ятаючи про те, що якість та кількість матерії знаходяться в діалектичній єдності, взаємно переходять одна в одну. Безсумнівно, що кількісні методи як більш об'єктивні та точні мають перевагу перед якісною характеристикою предметів. Недарма ще в давнину достовірність пізнання природи пов'язувалась з математикою - наукою точною, що вивчає кількісні відносини та просторові форми реальної дійсності. Спираючись на кількісні показники можна отримати більш достовірну інформацію про предмети, що дозволяє глибше осягнути їх якісну своєрідність.

Кількісні методи не обмежуються одними лише вимірами або підрахуванням живих істот та продуктів їхньої життєдіяльності. Самі собою результати вимірювань, хоч і мають відоме значення, ще недостатні для того, щоб зробити з них необхідні, повні та достовірні висновки.

Цифрові дані, зібрані у процесі масових випробувань, тобто вимірювань або обліку досліджуваного об'єкта, - це лише сирий фактичний матеріал, який потребує відповідної математичної обробки. Без обробки, впорядкування та систематизації цифрових даних не вдається всебічно висвітлити вкладену в них інформацію, оцінити надійність окремих сумарних показників, переконатися у достовірності або недостовірності виявлених між ними відмінностей.

Ця робота вимагає від фахівців певних знань, вмінь правильно узагальнювати та аналізувати зібрані в досліді дані.

Система цих знань і складає ***зміст біометрії - науки, що розглядає питання статистичного аналізу результатів досліджень як в галузі теоретичної, так і практичної біології, у тому числі й в лісовому господарстві.***

1.2 Особливості біометрії як науки та її місце у ряді інших наук.

Одна з характерних особливостей біометрії як науки полягає в тому, що вона не має прямого та безпосереднього відношення до питань техніки вимірювань чи обліку живих істот. Це справа спеціалізованих наук – таких, як ботаніка, зоологія, лісознавство, дендрологія, лісова таксація, лісовпорядкування, ентомологія, агрономія та ін. Вони мають свої об'єкти дослідження і стосовно них розробляють методику вимірювань та кількісного обліку об'єктів, що вивчаються. В даний час широко застосовуються біометричні способи планування біологічних експериментів, дослідів, польових досліджень. *Але головним завданням біометрії як і раніше залишається обробка результатів вимірювань та обліку результатів дослідів для того, щоб за небагатьма числовими показниками робити висновки про сутність явищ, що вивчаються.* Тому з чисто формальної сторони, біометрія представляє сукупність математичних методів, застосовуваних для обробки результатів біологічних досліджень. Ці методи вона запозичує переважно з галузі математичної статистики та теорії ймовірностей, що становлять технічну основу біометрії. Отже, біометрія - це математична статистика в додаток до явищ живої природи.

Порівнюючи названі науки, треба мати на увазі, що математична статистика та теорія ймовірностей є науками суто теоретичними, абстрактними. Вони вивчають статистичні узагальнення без відношення до специфіки елементів, що входять до їх складу. Методи математичної статистики та теорії ймовірностей, що лежить в її основі можуть бути прикладені до різних галузей знання, включаючи і гуманітарні науки, біометрія ж - наука емпірична, конкретна: вона досліджує виключно біологічні угрупування, переслідуючи не математичні, а біологічні цілі. На цій підставі не можна ставити знак рівності між біометрією та математичною статистикою, повністю їх ототожнювати. Біометрія має свій об'єкт дослідження, своє місце в системі біологічних наук: теорія ймовірностей та математична статистика – розділи сучасної математики, а біометрія належить до біологічних наук. Відношення біометрії до математики аналогічне такому, яке існує між ботанікою та методикою її викладання. Не маючи власних методів дослідження, які вона запозичує з математики, біометрія займає особливе положення, будучи відносно самостійним розділом

біології, що виник на стику математичних та біологічних наук. Біометрія є результатом розвитку теорії ймовірностей та математичної статистики. Її часто називають варіаційною статистикою. Теорія ймовірності має справу із випадковими явищами. *У наукових дослідженнях, техніці та масовому виробництві часто доводиться зустрічатися з явищами, що при неодноразовому відтворенні одного й того ж досвіду в незмінних умовах протікають щоразу дещо по-іншому. Такі явища називають випадковими.* Наприклад, при стрільбі результат кожного окремого пострілу буде випадковим.

Проводячи експериментальне вивчення будь-якого явища та систематизуючи результати дослідження (ті ж результати стрільби) у вигляді графічної залежності, ми переконуємося в тому, що за досить великої кількості експериментальних точок виходить не крива, а деяка смуга, тобто має місце випадкове розміщення експериментальних точок. Вимірюючи діаметри і висоти в лісі і задаючи отримані значення на графік, ми теж побачимо набір точок, але за досить великою їх кількістю серед спостережень, вимальовується деяка лінія, близька до параболи або до іншої функції, що має схожий графік.

При вирішенні багатьох практичних завдань випадковими відхиленнями точок від кривої можна знехтувати, припускаючи, що в цих умовах досліджень явище протікає цілком закономірно. В цьому і виявляється основна закономірність, що властива цьому явищу. За цією закономірністю, застосовуючи той чи інший математичний апарат, можна передбачити результат досліду щодо його заданих умов. В міру розвитку різних галузей науки стає необхідним вивчати випадкові явища, щоб навчитися передбачати дії випадкових факторів та враховувати їх у практичному вирішенні завдань. У лісовому господарстві така потреба з'явилася наприкінці XIX століття, що дало поштовх до використання методів математичної статистики та започаткувало розвитку лісової біометрії.

Математична наука, що вивчає загальні закономірності випадкових явищ незалежно від їх конкретної природи та дає методи кількісної оцінки впливу випадкових факторів на різні явища, називається теорією ймовірностей.

Основою наукового дослідження в теорії ймовірностей є дослід та спостереження. На теорії ймовірностей базується вся статистика - математична, економічна, соціальна. Ми живемо в епоху статистики. Чи не в кожному своєму аспекті всі явища природи, а також людська та інша діяльність піддаються певною мірою за допомогою статистичних показників аналізу та прогнозу. Існують дві діаметрально протилежні точки зору на статистику, широко доступну нині населенню. Згідно з однією з них, опубліковані статистичні дані містять у собі деякі смислові якісті, щось подібне до

того, що приписували числам Піфагора, і мають такий ступінь відсутності похибки, що їх можна приймати на віру беззастережно. Це, звичайно, так само абсурдно, як і інша, ще більш поширена думка про те, що можна сфабрикувати статистичні дані, які доведуть все, що завгодно, а тому, отже, вони насправді нічого не доводять.

Останнє твердження знайшло відображення у гумористичному вигляді. Відомий злий жарт. Є три роду брехні: брехня вимушена або просто брехня, їй є виправдання; брехня нахабна, на яку немає виправдання, і статистика. Але нічого зі сказаного не має відношення до математичної статистики. Останній немає необхідності брехати взагалі. Обидві точки зору помилкові тому, що вони засновані на незнанні чи нерозумінні мети, меж та вимог справжньої статистичної теорії та практики. Математична статистика - це сувора наука, а тому, як і будь-яка наука, вона об'єктивна й безпристрасна.

Математична наука про методи кількісного аналізу масових явищ, що враховує одночасно і якісну своєрідність цих явищ, називається статистикою.

Найчастіше виявлення загальної закономірності необхідно більше для спостережень. Але мало здійснити спостереження. Необхідно застосувати спеціальні способи їхньої обробки. Більше того, спостереження мають бути сплановані та організовані методично, інакше їх цінність різко знизиться. Методи та правила, необхідні формули для організації спостережень та обробки отриманого матеріалу дає математична статистика.

Загалом і теорія ймовірностей, і математична статистика – це розділи математики. Зв'язки сучасної біології та лісового господарства з математикою багатосторонні, вони дедалі більше розширюються. Але не всі кількісні методи, що використовуються в біології, складають зміст біометрії. Не можна, наприклад, ототожнювати біометрію з кібернетикою і з так званою «математичною біологією», у яких є свої, особливі завдання, що не збігаються із завданнями біометричного аналізу угрупувань. Іншою характерною особливістю біометрії як науки є те, що її методи можуть бути використані не тільки до одиничних об'єктів або до окремих результатів спостережень, їх сукупностей, явищ масового характеру, якщо ми розглядаємо наприклад окремо взяту особину і порівнюємо її з населенням, до якої вона належить, а і до масових явищ або групи об'єктів. Інакше кажучи, загальні та поодинокі явища не просто співіснують, вони взаємно обумовили один одного. Не можна уявити спільноту без її членів, як неможливий ліс без дерев певного ботанічного виду чи певної дендрологічної форми. Деревними породами в дендрології та лісовому господарстві прийнято називати деревні види. Кожен лісівник знає, що певна сума дерев – це ще не ліс. Ліс відрізняється від сукупності дерев не тільки

кількісно, а й якісно. Ось ці якісні відмінність і повинні відображати закони біометрії.

З першого погляду здається, що між загальним та окремим, цілим і частиною немає жодної різниці, і що закони які діють у сфері одиничних і масових явищ одні й ті самі. Але це не так. Множинність, угрупування або загальне не є проста арифметична сума, що входять до її складу одиниць. У сфері масових явищ, тобто у сукупностях, діють свої, властиві їм статистичні закони, які лише загалом характеризують поодинокі явища. Також і закони, властиві поодиноким явищам, що не відображають повною мірою загальних закономірностей, що виявляються лише у сфері статистичних сукупностей. У цій суперечливій єдності і полягає внутрішній зв'язок між частиною та цілим, між одиничними явищами та їх сукупністю. Біометрія допомагає виявляти цей зв'язок та оцінювати значення окремих факторів у світлі загальних закономірностей, властивих угрупуванню загалом. Зрештою, не можна не відзначити ще одну характерну рису біометрії - її своєрідна мова знаків, символів, рівнянь, графіків та формул. Усі ці умовні позначення, призначені для економного висловлювання думки, біометрія запозичує з математики. Відомо, що математична логіка відрізняється лаконічністю та точністю доказових формулювань, переконливістю висновків. Завдяки символіці вдається порівняно простими та точними засобами виражати зміст складних та різноманітних явищ природи, що значно полегшує розуміння властивих їм закономірностей. Графіки, рівняння та формули, оскільки вони містять у собі найбільш суттєве та типове у явищах, служать свого роду математичними моделями цих явищ. Математичне моделювання в даному випадку аналогічно схематичним побудовам у вигляді малюнків, графіків та інших зображень, що широко використовуються в педагогічній та науково-дослідній роботі. У наш час суцільної комп'ютеризації саме математичне моделювання є основним методом описи виявлених законів та закономірностей у живій природі, в тому числі і в лісових науках: лісознавстві, лісівництві, лісовій таксації, захисті ліси, лісовій пірології, лісової генетики та селекції і т.д. Зрозуміло, будь-які схеми та моделі дають лише деяке зображення реальної дійсності. Але саме в цьому і полягають їх великі методичні здібності. Достатньо вдягнути думки у форму символів і знаків, геометричних фігур і рівнянь, як це дає широкі можливості для глибокого та всебічного пізнання явищ, швидкого руху на шляху до істини. Але символи і взагалі біометричні показники набувають певного сенсу лише тоді, коли вони відповідають змісту висловлюваного ними процесу, перебувають у тісному зв'язку з конкретними завданнями біологічного дослідження. Слід зважити на те, що символіка не тільки не виправдовує себе, але може навести дослідника до помилок та помилок. Справа в тому, що в стислості і точності

числових характеристик, зручності висловлювати біологічні явища мовою математичних формул і рівнянь укладено як великі методичні можливості,так і небезпека відриву від конкретних явищ, а це веде до помилок, створює видимість істини там, де її насправді немає. Тому слід висловлювати математичними формулами чи графіками те, що очевидно саме собою. У багатьох випадках біометричні дані, зведені в статистичні таблиці, виявляються настільки переконливими, що не потребують жодної додаткової обробки. Подивимося на дві таблиці, що показують хід росту соснових деревостанів в лісорослинних умовах A_{1-2}, B_{2-3} Перганського та Копищанського природоохоронних науково-дослідних відділень Поліського природного заповідника.

Таблиця 1.1

Динаміка середніх висот соснового деревостану в лісорослинних умовах A_{1-2} Перганського природоохоронного науково-дослідного відділення Поліського природного заповідника (середнє за результатами дендрохронологічного моніторингу за 2021-2025 роки)

Вік, років	Середня висота в різних типах лісу, м	
	брусничний I^a клас бонітету	сфагновий V клас бонітету
20	10,2	3,6
40	19,1	7,1
60	25,4	10,2
80	29,9	12,9
100	33,0	15,0

З таблиці 1.1 відразу видно, що зростання сосняку брусничного типу набагато інтенсивніше, ніж сосняку сфагнового. Тут можна навести порівняння за методиками біометрії, але результат заздалегідь зрозумілий. Подивимося тепер інші матеріали. Розглянемо динаміку двох однорідних деревостанів, що ростуть у однакових умовах, але відрізняються тим, що в одному з них відбувся вітровал, тобто природна стихія прорідила деревостан. В обох ділянках визначили середню висоту. Результати вимірювань наведено у таблиці 1.2.

Таблиця 1. 2

Зміна середньої висоти в не зрідженому та зрідженому сосновому деревостані в лісорослинних умовах B_{2-3} Копищанського природоохоронного науково-дослідного відділення Поліського природного заповідника (середнє за результатами дендрохронологічного моніторингу за 2021-2025 роки)

Вік, років	Середня висота, м	
	не зріджений деревостан	зріджений деревостан після вітровалу
20	8,3	8,3
40	14,4	14,6
60	19,5	19,4
80	23,2	23,7
100	25,7	26,1

З таблиці 1.2 випливає, що зріджений деревостан має де в чому середню більшу висоту, але відмінності невеликі. Тому без додаткового аналізу робити висновки про те, що зріджений деревостан після вітровалу має більшу середню висоту ніж, де вітровалу не було, не можна. Щоб дати відповідь на це питання, треба звернутися до відповідних біометричних методів та оцінити ступінь достовірності спостережуваних у досліді відмінностей. Методику такої оцінки буде показано нами далі. З усього вище сказаного аж ніяк не виходить, що треба якось обмежувати застосування математичних методів у лісівництві. Йдеться не про обмеження, а про правильне використання цих методів у лісівничих дослідженнях. Сама по собі біометрія не веде до помилок та похибок. Вони виникають при невмілому, механічному використанні біометричних показників без врахування їх конструктивних особливостей та теоретичного обґрунтування. У роботі лісівника-дослідника однаково неприйнятні як біометричне жанглювання, що перетворює роботу в безплідну і шкідливу гру в цифри, і примітивізм в оцінці числових показників, що веде насправді до відмови від застосування математичних методів у лісовому господарстві. Істина, як це завжди буває, полягає не в крайнощах, а в розумному підході до виконання наукових досліджень.

1.3 Історія виникнення та розвитку математичної статистики та біометрії.

Біометрія сама по собі наука відносно молода. Перші досліди її застосування відносяться до робіт Бореллі, що на рубежі XVII та XVIII століть. Він робив математичні розрахунки руху тварин. На початку XVIII століття французький вчений Реомюр (1683-1757) (згадаємо, що є температурна шкала Реомюра), шукав математичні

закони будови бджолиних стільників. Але як повноцінна наука біометрія з'явилася лише до кінця XIX століття. Термін «біометрія» був уведений у науку англійським вченим антропологом
Фр. Гальтоном (1822-1911) у 1889 році для позначення кількісних методів, що застосовуються в галузі біологічних досліджень. Надалі Дункер (1899) запропонував іншу назву – «варіаційна статистика», яка теж стала на той час досить популярною, адже виражала більш точний зміст цього предмета. В даний час використовуються обидва ці терміни, хоча буквальний зміст їх неоднаковий. Слово «біометрія» (від латинського bios - життя, та «metron» - міра) означає проведення біологічних вимірів, а термін «варіаційна статистика» (від латинського «variatio» - зміна, коливання і «status» - стан, становище речей) розуміється як опис спостережень, їх математична обробка. Набагато тривалішу історію має як загальна статистика, так і математична чи варіаційна статистика. Розвиток статистики почалося в епоху античності, тобто ця наука має стародавнє коріння. Наприклад, використання середнього значення було добре відомо ще за життя Піфагора (VI ст. до н.е.), а згадки про статистичні обстеження зустрічаються і в біблійні часи. Статистика поступово розвивалася там, де у ній виникала потреба. Насамперед статистичні методи почали застосовувати для аналізу економіки та явищ суспільного життя, і лише пізніше вони проникли у біологію. Так, Ісаак Ньютон (1642-1727), чий внесок у відкриття диференціального та інтегрального обчислення став видатною подією в математиці, а його теорія всесвітнього тяжіння та «ньютонове яблуко» відомі будь-якому старшокласнику, є можливо найбільш помітною фігурою в галузі розвитку сучасної статистики, хоча сам Ньютон навряд чи чув колись про існування цієї науки. Інші математики, чиї імена відомі насамперед завдяки роботам у галузі теоретичної математики, опосередковано зробили для розвитку статистики більше, ніж багато хто з вчених, які безпосередньо спеціалізувалися на цій науці. Двома найвидатнішими представниками таких вчених-математиків є Абрахам де Муавр (1667-1754) та Карл Гаусс (1777-1855). Щодо самих вчених-статистиків, то варто згадати бельгійця Адольфа Кетле (1796-1874), який першим застосував сучасні методи збору даних. Можливо, здасться дещо несподіваним, що й відома англійська медична сестра середини XIX століття (кримська війна 1854-55 рр.) Флоренс Найтінгейл (1820-1910) все своє життя була палкою прихильницею застосування статистики. Ф. Найтінгейл, яка згодом працювала на керівних посадах, доводила, що адміністратор може мати успіх тільки в тому випадку, якщо він у своїй діяльності керуватиметься даними, одержуваними за допомогою статистики, та що законодавці та політики часто зазнавали невдач через те, що їх статистичні пізнання були недостатні.

Двома іншими вченими, які зробили значний внесок у розвиток статистики, є два англійці - Френсіс Гальтон (1822-1911) і Карл

Пірсон (1857–1936). Гальтон, родич Чарльза Дарвіна (1809-1882), серйозно зацікавився проблемою спадковості, до аналізу якою він незабаром застосував статистичні методи. Гальтон та Пірсон зробили значний внесок у розвиток теорії кореляції, яку ми детально розглянемо нижче. Найбільш відомим вченим у галузі статистики у XX столітті був Рональд Фішер (1890-1962). Фішер продуктивно працював з 1912 по 1962 р., і багато його досліджень надали суттєвий вплив на сучасну статистику. У XX столітті статистичні методи офіційно введені в Сполучених Штатах Америки для навчання в усіх коледжах. Там читалися невеликі курси статистики. Протягом перших тридцяти років XX століття поступово зростало значення статистики у дослідженні проблем психології. При цьому зазначимо, що в даний період психологія як наука не була самостійною і часто розглядалася лише як одна із розділів філософії.

Застосування математичної статистики не оминуло і лісову галузь. У лісовому господарстві нашої України математична статистика почала використовуватися з кінця XIX століття в основному завдяки працям відомих лісівників, таксаторів професорів В. Є. Свириденка, А. К. Турського та П. С. Погребняка. Вже у 20-30 роки минулого століття математична статистика стала невід'ємною частиною досліджень в лісовому господарстві. Однією з причин широкого застосування статистичних показників в останні роки є все зростаюча легкість обробки великих масивів чисел. Сучасні комп'ютери дозволяють у короткий час проаналізувати величезну кількість статистичних даних, що раніше було неможливо.

Для впровадження математики в біологію та лісове господарство наприкінці XIX і на початку XX століття були серйозні підстави. Одним із них був перехід від описового методу вивчення явищ життя до експериментального. Хоча і при описовому підході можливе встановлення математичних закономірностей (прикладом можуть бути закони руху небесних тіл), проте переважає в цьому випадку якісна оцінка. Експеримент неминуче вимагає кількісного вираження явищ та процесів. Створення фізіології, генетики, радіобіології та інших експериментальних напрямів біології спричинило розробку численних математичних прийомів та методів дослідження. Велику роль відіграли і суто практичні причини. Розроблені методи стали широко застосовуватися в зоології, ботаніці, лісівництві, лісовій таксації та інших біологічних науках. Нарешті, найважливішою обставиною, яка визначила використання математичних та математико-статистичних методів, з'явилося встановлення того факту, що багатьом біологічним явищам властиві статистичні закономірності які виявляються при вивченні сукупностей, але непридатні до окремих одиниць цих сукупностей.

Коли фізики перейшли від вивчення поведінки окремих фізичних тіл до вивчення поведінки множин молекул, електронів, вони вступили у сферу дії статистичних законів. На цій основі

створилася особливий напрям фізики – статистична фізика, що вивчає властивості та поведінка систем, що складаються з величезної кількості окремих частинок. В основі багатьох фізичних явищ, таких, як радіоактивний розпад, термодинамічні явища та деякі інші, лежать статистичні закономірності. З їх відкриттям закономірності, встановлені емпірично, наприклад, закони термодинаміки, отримали більш глибоке обґрунтування і були виведені зі статистичних імовірнісних законів. Фізики на початку XIX століття довго не могли примиритися, що в мікросвіті діють статистичні закони. Раніше здавалося, що у фізиці все детерміновано, тобто на однозначну дію настає однозначний результат. Квантова механіка це відкинула. Навіть великому Альберту Ейнштейну (1879-1955) дивно було сприйняти таке, і він неодноразово повторював: «Невже Бог грає з нами в кеглі». Приблизно таке ж становище спостерігається і зараз у ряді галузей біології. Коли зоологи та ботаніки перейшли від вивчення окремих «типових» представників виду до вивчення багатьох особин одного виду, вони виявили масові явища статистичної природи. Риби, рачки, молюски, коловратки, водорості, інфузорії та інші тварини, рослини характеризуються мінливістю, варіацією за самими різноманітним ознакам. Такою ж варіацією мають і організми, культивовані людиною: колосся пшениці різняться кількістю зерен у колосі, вагою окремих зерен; звірі одного виду мають різну масу, у них варіює екстер'єр та забарвлення. Дуже мінливими об'єктами є лісові насадження. Навіть в однорідному деревостані ми бачимо велику різноманітність дерев: по висоті, діаметру, формі крони і т. п. При вивченні біологічних сукупностей, що є типово статистичними, виявилося доцільним застосувати методи математичної статистики, яку в додатку до біології стали називати біологічною статистикою. Ще її називають варіаційною статистикою. Поле для застосування статистичних методів в біології дуже значне, так як багато екологічних, генетичних, цитологічних, мікробіологічних, радіобіологічних явищ - масові по своєїй природі. У них беруть участь не одна особина або клітина, не одна частка, не одна бактерія чи вірусна частка, не одне дерево, а множини, тобто сукупності клітин, часток, бактерій, особин виду, сімей, дерев та ін. Здійснення подій у таких сукупностях може бути оцінено ймовірностями, а їх аналіз вимагає застосування статистичних методів. Статистичні методи суттєво необхідні і під час постановки експериментів, оскільки тільки з їх допомогою можна встановити, залежить чи спостерігається відмінність між дослідними та контрольними ділянками лісу від впливу фактора, що вивчається, або ж воно суто випадково, тобто визначається багатьма іншими, не контрольованими факторами, які не піддаються обліку. Розуміння та облік статистичних закономірностей допомагають експериментатору скласти методично обґрунтований план дослідів, правильно їх провести і, нарешті, зробити з них об'єктивними висновки. При

цьому треба пам'ятати, що жодна математична і статистична обробка не допоможе, якщо досліди були проведені методично неправильно або дані зібрані недбало. Роль математики та математичної статистики в біології особливо зросла у зв'язку з розвитком теорії інформації та кібернетики в цілому та багатьох пов'язаних з ними галузей математики, серед яких чинне місце займають теорія ймовірності, математична статистика та математична логіка. Застосування комп'ютерів на порядок прискорило і розширило застосування статистичних методів у біології взагалі та у лісівництві зокрема. Використання математики у сучасній біології не обмежується лише статистичними методами. Тому біометрія (або біоматематика, як її іноді називають) ширше, ніж біологічна статистика. Вона використовує також прийоми та методи з інших галузей математики: диференціального та інтегрального обчислень, теорії чисел, матричної алгебри і т.д. Впровадження математики в біологію спочатку виражалося у використанні окремих математичних та математико-статистичних методів для вивчення тих чи інших біологічних питань та обробки даних, одержаних із природи або в лабораторії. Такі питання, як мінливість морфологічних, фізіологічних та екологічних ознак тварин і рослин та встановлення впливу на них зовнішніх та внутрішніх факторів, кількісний облік та процеси, що відбуваються в популяціях, подібність і різницю між видами, підвидами та інші систематичними категоріями, зростання індивідуальне та зростання популяцій, можуть вивчатися лише за допомогою математичних та математико-статистичних методів. Більше того, у різних галузях біології (генетика, еволюційне вчення, селекція, фізіологія), а також практично у всіх лісових науках: лісівництві, лісовій таксації та ін. відповідні біологічні процеси або явища тепер виражаються в математичній формі. Таким чином, пройшовши тривалий шлях розвитку, математична статистика та біометрія сьогодні постають перед нами як стрункі та високорозвинені науки, що мають велике практичне значення.

1.4 Лісова біометрія як частина загальної біометрії і її значення для розвитку лісового господарства.

Нас, як лісівників, щодо біометрії найбільше цікавить той її розділ, який називається «лісова біометрія».

Лісова біометрія - це розділ біометрії, змістом якого є планування та організація кількісних експериментів у лісознавстві, лісівництві, лісовій таксації та 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

математичних підходів, що застосовуються у загальній біометрії. Методи лісової біометрії широко використовуються під час аналізу процесів у природних явищах, властивих деревам та деревостанам, а також окремим ділянкам лісу. Як приклад можна навести аналіз успадкованості ознак дерев та деревостанів, оцінка ефективності різних лісогосподарських заходів. Лісова біометрія – необхідний методичний інструмент для проведення наукових досліджень у лісі, при розробці різних нормативних та довідкових матеріалів. Наприклад, всі лісівники користуються сортиментними та об'ємними таблицями для обліку лісу на корені, застосовують спеціальні таблиці для обліку готової лісопродукції, вимірюючи довжину сортименту та його діаметр у верхньому відрізі тощо. У цих таблицях наведені середні величини, отримані на основі обробки методами біометрії великого експериментального (його ще називають первинним чи вихідним) матеріалу. Далеко не кожна колода певної довжини і діаметра має об'єм який точно збігається з тим, що наведено у довіднику для параметрів довжини та діаметра. Але, використовуючи методи лісової біометрії, вчені отримали обсяги з достатнім ступенем наближення до реальності; крім того, точність визначення збільшується у разі зростання числа вимірів. На цьому невеликому прикладі, а їх дуже багато, ми бачимо, що ні лісова наука, ні лісогосподарська практика не можуть обійтися без використання у своїй роботі лісової біометрії. Методи лісової біометрії широко застосовуються також при обробці лісовпорядного матеріалу в геоінформаційній системи «Лісові ресурси» за допомогою комп'ютерних програм, а також при аналізі даних аеро- та космічної зйомки. У лісовій біометрії широко застосовуються методи математичного аналізу, математичне моделювання. В даний час найбільш раціональним є застосування біометричних методів із використанням комп'ютерів. Майже всі обчислення сьогодні проводяться лише на комп'ютерах.

У той же час, щоб правильно використовувати методи лісової біометрії, необхідно добре знати лісництво, лісову таксацію, лісові культури та інші освітні компоненти. де застосовують ці методи. Без доброго знання тієї компоненти, в якій ми працюємо, біометричні методи одержання коректного результату не гарантують. Механічне, бездумне жонглювання цифрами, зведеними в моделі, нехай навіть обробленими статистичними методами, неприпустимо, так як може призвести до хибних висновків. Перш ніж виводити модель, треба уявити загальну біологічну закономірність. Наприклад, відомо, що дерево росте вгору; діаметр дерева не може зменшуватись і т. д. Це найпростіші приклади, де все очевидно. Але є багато випадків, коли бездумне застосування статистичних моделей наводить навіть при аналізі росту дерев до негативних величин. Щодо наведених прикладів помилку виявити легко, але далеко ще не всі закономірності настільки очевидні. Саме тому, починаючи роботу з

матеріалом, який буде оброблений біометричними методами, треба дуже добре знати свій предмет: таксацію, лісову селекцію тощо. Дуже часто ставлять рівність між лісовою біометрією та математичною чи варіаційною статистикою. Слід пам'ятати, що математична статистика ширша за біометрію, а тим більше, лісову біометрію, так як охоплює все коло природних явищ живої та неживої природи, додаючи сюди техніку, суспільство та економіку. Коли ми надалі говоритимемо про співвідношення математичної статистики та лісової біометрії, то повинні мати на увазі, що об'єктом застосування статистичних методів у нас будуть дерева, деревостани, лісові біогеоценози і т. д. Варіаційна статистика широко використовується в дослідженнях не тільки для оцінки та аналізу результатів вимірювань, але й для планування експерименту. Біометрія як наука не стоїть на місті. Тут розробляють нові підходи та вдосконалюються старі способи. Це розширює коло розв'язуваних завдань у лісовому господарстві. Узагальнюючи сказане, можна зробити такі висновки:

- Лісова біометрія широко використовується у лісовому господарстві.

- Без знання лісової біометрії неможливо зрозуміти суть та значення багатьох нормативів і величин, які широко застосовуються в лісовому господарстві.

- Лісова біометрія - складова частина методики наукових досліджень в лісовій справі. Будь-якому досліднику її обов'язково треба знати та знати добре та всебічно. На додаток до всього відзначимо, що це цікава та захоплююча наука. Вона має свою строгу логіку, свої закони. Водночас час лісова біометрія непроста наука, яка вимагає для свого освоєння серйозної та кропіткої роботи, гарної загально-біологічної, лісівничої та математичної підготовки.

РОЗДІЛ 2. СТАТИСТИЧНІ СУКУПНОСТІ

2.1 Статистичні сукупності та статистичні спостереження. Статистичні вибірки.
2.2 Генеральна й вибіркова сукупність та їх обсяг.
2.3 Методи збору та обробки інформації в лісовій біометрії.
2.4 Дедуктивний та індуктивний методи у лісовій біометрії.

2.1 Статистичні сукупності та статистичні спостереження. Статистичні вибірки.

Вивчення біологічних явищ проводиться не за окремими спостереженнями, які можуть виявитися випадковими, нетиповими, такими, що неповно виражають сутність цього явища, а за значною групою однорідних спостережень, які дають більш повну інформацію про досліджуваний об'єкт. Наприклад, вивчаючи лісові екосистеми, скажімо, соснові деревостани в певних лісорослинних умовах ми не обмежуємося вимірами 1-2 або 5-10 дерев, а проводимо виміри на всій пробній площі, де росте не менше 200 стовбурів сосни, тобто вивчаємо ту їхню сукупність, яка вже може репрезентувати ліс.

Деяка кількість однорідних предметів чи об'єктів, що об'єднуються за тією чи іншою ознакою для спільного вивчення, називають статистичною сукупністю.

При цьому зовсім не обов'язково, щоб сукупність складалася з безлічі особин одного виду та віку, наприклад з однорідних дерев сосни. Вона може бути утворена в результаті численних випробувань, тобто проб, спостережень і т. п., що проводяться на тому самому індивідуумі. Наприклад, статистичною сукупністю будуть дані спостережень за генерацією умовних рефлексів у одного собаки чи кішки, фенологічні спостереження за одним або декількома деревами дуба і т. д. Таким чином, сукупність поєднує якесь число однорідних спостережень чи реєстрацій. Сукупностями є дерева, популяції мурах, заготовлені під час полювання шкірки, рослини на дослідних ділянках тощо. Поняття сукупності застосовується не тільки до тварин чи рослин. Такими самими сукупностями є молекули газу в певному об’ємі, населення міста і т. п. До складу сукупності входять різні члени, або одиниці: популяції тварин - кожна окрема тварина; для стада корів одиницею є кожна корова; для сукупності шкір, добутих в результаті полювання - кожна шкура тварини, для сукупності насіння сосни звичайної - кожна насінина; для деревостану – дерево; щодо дерева - його клітини тощо.

Загальна властивість предмета, що вивчається, називається його ознакою.

Так, щодо лісового насадження, ознаками будуть: деревний вид, вік дерев та деревостанів, розмір, форма або колір насіння,

особливості вегетації (у дуба є форми, що рано чи пізно розпускаються) і т. д. Число одиниць сукупності називають ***об'ємом сукупності,*** й позначають літерою ***N***. Одиниця сукупності може характеризуватись певними ознаками, наприклад: корови - надоями за лактацію, вагою, мастю; молекули газу – швидкостями їх руху; насіння сосни звичайної – формою, кольором, вагою; дерева – товщиною (діаметром стовбура), висотою і т.д. Кожна досліджувана ознака набуває різних значень у різних одиниць сукупності, вона змінюється у своєму значенні від однієї одиниці сукупності до іншої.

Ця різниця між одиницями сукупності називається варіацією чи дисперсією, тобто - розсіянням.

Ми говоримо – «ознака варіює». Це означає, що вона набуває різних значень у різних членів сукупності, наприклад у корів цієї породи, насіння одного виду, дерев одного виду тощо.

Елементи, що входять до складу однієї сукупності, називаються її членами, чи варіантами.

Останній термін походить від латинського ***varians***- змінюється.

Варіанти - це окремі спостереження чи деякі числові значення ознаки. Так, якщо позначити певну ознаку через Х (велике), то його значення або варіанти позначатимуться через х (мале), тобто. як x_1 x_2, x_3, x_4, ... $x_к$.

Загальна кількість варіантів, що входять до складу цієї сукупності, називається її об'ємом і позначається літерою ***N*** або ***Σn***. Саму ж варіюючу величину, тобто величину, що змінюється під впливом багатьох випадкових причин, і може приймати різні значення, називають випадкової змінною n_i. Варіанти є її числовими значеннями. У цьому випадку значок (індекс) ***i*** - порядковий номер варіанту. В той же час, незважаючи на різницю між варіантами, що входять в сукупність, остання має внутрішню однорідність. Члени сукупності подібні до ряду важливих ознак. Для прикладу – заячі шкури, що були добуті під час полювання не однакові за розмірами, якістю хутра, забарвленням, але всі вони – шкури особин одного й того ж виду – зайця русака або зайця біляка. Зерна пшениці озимої відрізняються одине від одного за вагою та іншим біохімічним складом, фізіологічними ознаками, але всі вони – зерна пшениці озимої, а не ячменю, хоча обидві ці сільськогосподарські культури могли б бути гіпотетично вирощені на одному полі.

Дерева варіюють за розміром, але всі вони одного виду, наприклад, сосни звичайної. Жолуді дуба звичайного відрізняються по розмірам, вазі, кольору, але вони - насіння одного деревного виду - дуба черешатого і т.д.

Найчастіше до складу сукупностей входять окремі особини. Так, наприклад, при характеристиці дерев сосни звичайної на лісонасінневій плантації, за одиницю сукупності можна взяти кожне окреме дерево. Однак одиницею сукупностей може бути саме дерево, або окрема його характеристика. В даному випадку

допустимо взяти врожай шишок чи насіння за певні роки. Тоді за загальною кількістю дерев на лісонасінневій плантації, скажімо 200 штук, кількість варіантів, що отримуються за кілька років збору насіння складатиме 600, 800, 1000 чи іншу величину. Можна вивчати варіацію тієї чи іншої ознаки в часі навіть на одному дереві. Як відомо, розмір насіння та його вагова величина змінюється залежно від ботанічних та абіотичних факторів. Вивчаючи цю зміну на одному дереві за кілька років, теж отримуємо статистичну сукупність, яка вивчається методами біометрії. Такою ж сукупністю є час розпускання бруньок на тому самому дереві, але в різні роки. Сукупністю буде також довжина та вага шпильок сосни звичайної на одній гілці цього виду і т.д. Отже, сума спостережень чи вимірів є також сукупність.

Кожне окреме спостереження, в якому встановлюється значення випадкової змінної називається одиницею цієї сукупності.

Сукупність може складатися з інших, окремо взятих сукупностей. Так, сукупність із всіх диких тварин одного виду розкладається на окремі частки сукупності – окремі популяції. У межах однієї популяції можна виділити ще більш окремі сукупності, наприклад, потомство певних самців чи самок. Вивчаючи деревостани сосни, їх можна розділити по галузях, лісогосподарських філіях в межах господарських користувань або лісорослинних районах. В усіх випадках, ми стикаємось з постійними відмінностями як в середині окремих частин сукупностей, так і між самими ними.

2.2 Генеральна й вибіркова сукупність та їх обсяг.

Найбільшу загальну сукупність називають генеральною. Це теоретично нескінченно велика або, принаймні наближується до нескінченності сукупність всіх одиниць чи складових, які можуть бути до неї віднесено. Так, якби можна було описати всі особини цього виду, наприклад всі дерева сосни звичайної в лісорослинних умовах Поліського природного заповідника, 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, мисловствознавець, який вивчає в лісі той чи інший вид мисливських тварин добуває кілька екземплярів, і по них прагне зробити висновок про всі особини цього виду. Лісівник закладаючи пробну площу, де 300 дерев, робить висновки про всю цю їх сукупність. Питання про те, якою мірою за вибірковою сукупності можна робити висновки про генеральну сукупність, належить до найважливіших теоретичних і практичних питань у біологічній статистиці. Це питання буде викладене нижче. Одним із завдань вивчення будь-якої сукупності, є

отримання статистичних (або, як іноді кажуть, біометричних) характеристик, або показників. Вони дозволяють робити висновки про цю сукупність в цілому, про відмінності в ній, і про відмінність її від інших, подібних до неї або близьких до неї сукупностей. *Сукупність стає статистичною саме тоді, коли до її опису вноситься кількісний метод.* Застосування кількісного методу вивчення сукупності, дозволяє отримувати для неї низку статистичних показників, і через це ми отримуємо основну інформацію про сукупність. Щоб вибіркова сукупність якнайповніше відображала генеральну, необхідно враховувати такі основні ознаки:

1. Вибірка має бути цілком представницькою, або типовою, тобто щоб у її склад входили переважно варіанти, які найбільш повно відображають генеральну сукупність. Перед тим як розпочати обробку вибіркових даних, їх уважно переглядають і видаляють явно нетипові варіанти. Наприклад: при вимірюванні довжини колосків не можна включати у вибірку такі, що зіпсовані сажковими грибами або обірвані колоски, оскільки вони нетипові для такого роду вибірки. При вивченні ходу росту дерев у висоту треба виключити дерева, що зламані буреломом, всохлі, фаутні, пошкоджені вогнем тощо.

2. *Вибірка має бути об'єктивною.* При формуванні вибірки не можна робити обліки довільно, включати до її складу тільки ті варіанти, які здаються типовими, а решта бракувати. Вибірка проводиться без упереджених думок, за методом жеребкування або лотереї, коли жоден з варіантів генеральної сукупності не має жодних переваг перед іншими потрапити або не потрапити до складу сукупної вибірки. Вибірка повинна проводитися за принципом випадкового відбору, без будь-яких суб'єктивних вилучань. Наприклад, якщо ми хочемо для визначення середньої висоти виміряти 20 дерев сосни звичайної в умовах закладеної пробної площі, то не можна їх вибирати на свій смак чи розсуд, виключаючи, скажімо, низькі пригнічені стовбури. Потрібно вимірювати висоту у кожного 10 або 20-го дерева сосни звичайної і т.д. При цьому треба враховувати обмеження, згадані вище в п.1, наприклад, виключати дерева, що зламані вітром.

3. *Вибірка має бути якісно однорідною.* Не можна включати до складу однієї і тієї ж вибірки дані, що отримані від обліку особин різної статі, виду, віку або фізіологічного стану, оскільки свідомо відомо, що ці фактори по-різному позначаються на величині і функціональному стані ознак за якими може бути утворений неоднорідний за складом матеріал, а це в свою чергу не дає вірної інформації про досліджувані явища. Наприклад, не можна об'єднувати в одну пробну площу деревостани різних типів лісу, хоча би вони росли поруч, скажімо сосняк брусничний і сосняк вересовий. Всіх цих умови може дотриматися лише фахівець, який

добре знає не лише біометрію, а й предмет свого дослідження. В нашому випадку це лісівник. Емпіричні, або вибіркові, сукупності можуть мати різний обсяг. *Залежно від кількості спостережень прийнято розрізняти малі вибірки, що містять не більше 30 варіант, і вибірки більше > 30, які включають до свого складу до 100-200 одиниць сукупності і більше.* Верхня межа тут не обмежена. Принципової різниці між великою і малою вибірками немає. Розрізняти їх доводиться лише з тієї причини, що порівняльна оцінка біометричних показників, обчислюваних на малих вибірках, залежить від від числа спостережень про що буде розказано нижче.

Схематично сукупності можна виразити наступною схемою (рис. 2.1).

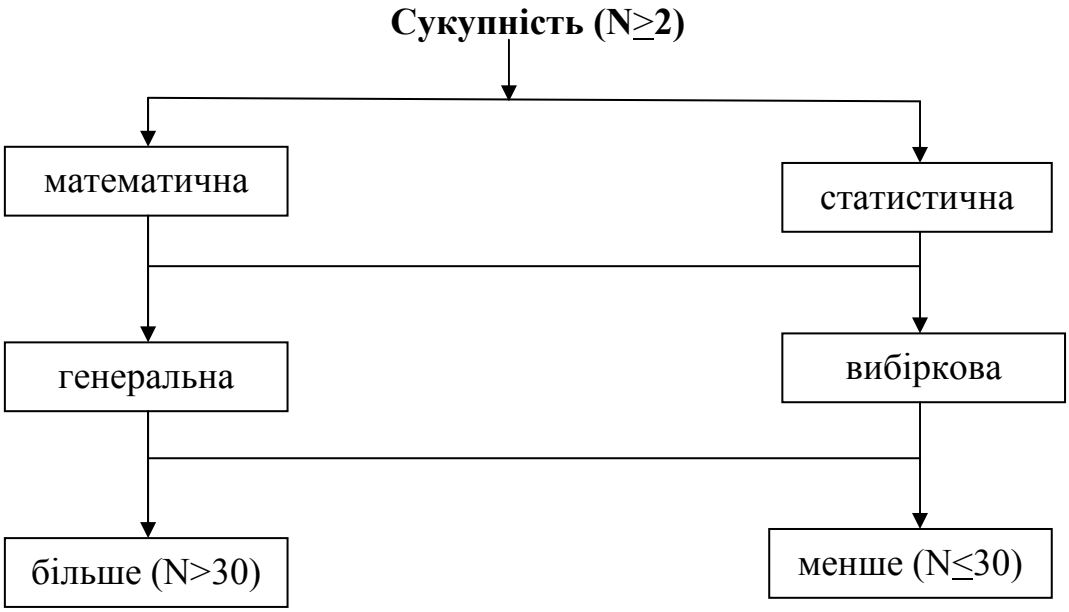


Рис. 2.1 Схема розподілу сукупностей

Якісна однорідність сукупності визначається саме метою дослідження. Варіювання за обліковою ознакою визначається одиницею відліку, яка має бути не більше розміру класу, розряду. В лісівництві розміри розрядів або класів зазвичай приймаються рівними одній дванадцятої амплітуди ряду розподілу, але допускаються в межах 1/8 – 1/16.

Таким чином, можна зробити висновок, що статистична сукупність - сукупність предметів, явищ, речей, якісно однорідних, і варіюють за обліковою ознакою. Звідси і назва - варіаційна статистика.

Генеральна статистична сукупність - якась філософська категорія, загальна сукупність предметів, явищ, речей, що може розглядатися як кінцевою, так і нескінченною, але обов'язково якісно однорідної та варіюючої за обліковою ознакою.

Вибіркова (часткова) статистична сукупність – це частина генеральної сукупності, що відповідає вимогам репрезентативності.

2.3 Методи збору та обробки інформації в лісовій біометрії.

Вибір правильних методів збору та обробки інформації визначає результат дослідження. Найбільш повно вони розроблені К. Є. Нікітіним та О. З. Швиденко, й ми в цьому виданні дотримуватимося цих методів. Певне наукове чи виробниче завдання, яке ставиться відповідно до чергових потреб і планів розвитку галузі, може бути вирішене на основі накопиченої інформації або для її вирішення може знадобитись збір (повний або частковий) нової інформації. В обох випадках у більшості лісівничих завдань на різних етапах використовують нормативно-довідкову інформацію. Іноді частка її у загальній кількості інформації та вплив на кінцеві результати дуже суттєві. Якщо наявної інформації достатньо, то на її основі формулюють відповідну гіпотезу та розробляють модель, яку емпірично перевіряють за допомогою статистичних методів. Якщо ж такої інформації недостатньо, то приймають рішення щодо конкретних шляхах дослідження, що визначаються метою роботи, фінансовими та трудовими можливостями, наявними засобами збору та обробки інформації. Одним із основних моментів при організації спостереження є розробка (вибір) правильної методики. Потрібно пам'ятати, що помилки, допущені у методиці збору первинного матеріалу не можуть бути потім виправлені ніякою камеральною обробкою та викличуть або неточності, або помилки у висновках.

Як відомо, існує два основних способи спостережень щодо охоплення одиниць об'єкта, що вивчається:

- ❖ суцільне обстеження всіх одиниць вивченої сукупності;
- ❖ часткове обстеження, коли спостереженню піддається лише частина одиниць досліджуваної сукупності.

В лісовому господарстві зазвичай обмежуються частковим (вибірковим) обстеженням. Вибірку роблять для отримання характеристики цілого, що підлягає вивченню.

Та сукупність, що підлягає вивченню, називається загальною, або генеральною, а відібрані з неї одиниці спостереження представляють часткову, чи вибіркову, сукупність.

Відносна частина, яку становить число одиниць з даним значення певної ознаки від загального числа одиниць сукупності називається в генеральній сукупності - частиною, а у вибірковій сукупності - часткою.

Як повне, так і не повне спостереження за способом дослідження може бути:

- ✓ по способу зв'язку з об'єктом - безпосереднє, експедиційне, кореспондентське та звітне;
- ✓ за часом дослідження – безперервне, в міру виникнення явища (наприклад, фенологічне); періодичне (повторне) – через певні проміжки часу та одноразове;

✓ за джерелами - власне спостереження, усний та письмовий (анкетне) опитування та за документальними матеріалами - літературними, службовими та архівним даним.

За особливостями відбору досліджуваних одиниць в досліді:

- відбір середніх типових одиниць, що застосовується у лісовій таксації;
- відбір випадкових одиниць, що проводиться за спеціальним планом, який є у статистиці основним;

За видом самого відбору:

➢ без повторне, коли відібрана одиниця після проведення спостережень щодо неї не повертається у генеральну сукупність;

➢ повторне, коли обстежена одиниця повертається до генеральної сукупність і може знову потрапити до відбору.

За принципом взяття одиниць - випадковий, типовий (за групами, однорідним за якоюсь ознакою) та механічний.

Відбір можна робити або з усієї сукупності, а також з обмеженої сукупності. У разі одиниці з крайніми значеннями ознаки не беруться до уваги, відбір може вестись і з окремих сукупностей, на які загальна сукупність розбита на основі кількісних (наприклад, ступені товщини) або якісних (наприклад, класи зростання дерев), ознак. Часткове спостереження методом випадкового відбору, при якому кожна одиниця об'єкта, що вивчається має однакову з іншими можливістю потрапити у вибірку, забезпечує повну об'єктивність спостереження (досліду).

При дослідженнях 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 лише один раз. Принцип вибіркового методу теоретично обґрунтовано видатним математиком, академіком П. Л. Чебишевим (1821 - 1894), якій довів, що при досить великій вибірці, вибіркова середня може бути як завгодно близька до генеральної середньої вибірки. Звідси випливає, що частину ознаки у вибірковій сукупності може бути як завгодно близька до частки цієї ознаки у генеральній сукупності. Іншими словами – достатньо велика вибірка правильно відтворює особливості та властивості генеральної сукупності, і тим краще, що відносно більше вибірка.

Відбір одиниць можна проводити за методом випадкової вибірки, коли із сукупності випадково (наприклад, за жеребом) відбирають необхідне число одиниць. Це досить складно, а в умовах лісу практично не застосовується. Тому матеріал спостереження під час досліджень у лісовій справі збирають шляхом механічного відбору. Сутність цього способу полягає в тому, що всю сукупність механічно розбивають на число частин, однакових за розміром чи кількістю одиниць які відповідають числу спостережень, та в кожній частині навмання вибирають одиницю для спостереження. Можна розбити і на число частин, у кілька разів менше кількості

спостережень. У цьому випадку з кожної частини треба взяти випадково не по одній одиниці, а в стільки разів більше, скільки потрібна для спостереження кількість одиниць більше від числа механічно відібраних частин чи груп. Наприклад, щодо якості насіння (плодів) всю партію насіння можна механічно розбити на кілька частин чи груп, однакових за кількістю або за вагою насіння, та з кожної такої частини чи групи вибрати навмання партію насіння для дослідження. При відборі модельних дерев, всю сукупність стовбурів на пробній площі (200-300 дерев) ділять по ступенях товщини (через 2-4 см) і відбір роблять всередині цих ступенів.

Існує кілька широко застосовуваних способів відбору.

1. Спосіб смуг. Поперек обстежуваної площі (наприклад, лісосіки), через однакову відстань закладають смуги однієї і тієї ж ширини, наприклад 1, 2 або 10 м з суцільним обстеженням кожної. Площа всіх смуг, виражена у відсотках від всієї площі обстеження, дасть відсоток вибірки. Величина відсотка вибірки залежить від величини обстежуваної площі (чим більша площа, тим менше може бути відсоток вибірки) та ступінь коливання досліджуваної ознаки: чим більший ступінь мінливості ознаки, тим більше має бути вибірка. Тому кількість смуг у різних випадках може бути різною. Отже, до складання методики потрібно мати уявлення про розміри обстежуваних площ та орієнтовну мінливість ознаки. Останнє визначається або за літературними даними, або за матеріалами лісовпорядкування або обстежень, або шляхом попередньої закладки пробної площі. Число одиниць спостереження N (у даному випадку - число смуг, що забезпечують результат з наміченою точністю) можна визначити за формулами:

$$N=V^2/P^2 \text{ або } N=\sigma^2/m^2$$

де: V-коефіцієнт мінливості;

σ - середнє квадратичне відхилення значень ознаки одиниць від їхнього середнього значення;

m – задана точність в одиницях вимірювання ознаки;

p - задана точність, або точність досліду, %.

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

Знаючи необхідну кількість смуг і довжину обстежуваної площі, відстань між центрами смуг можна отримати діленням довжини облікової площі на запроектоване число смуг. За неможливості попередньо підрахувати необхідне число смуг, треба заздалегідь брати не менше 5-10 смуг, і встановити відстань між ними шляхом розподілу довжини обстежуваної площі на число цих смуг. При цьому кожну смугу доцільно розділити на три-п'ять частин, однакових по довжині, що й будуть складовими одиницями спостереження, які перебувають з низки одиниць сукупності (обстежуваних дерев).

При проведенні досліджень в лісових культурах, обстеженні методом смуг, за смуги можна приймати ряди культур: один, два і більше у смузі. ***Кількість відібраних рядів дерев у смугах, виражене у відсотках від всього числа рядів на обстежуваній площі, становить відсоток вибірки.***

2. Спосіб - площ. Площі суворо однакового розміру й форми закладають через одну і ту ж відстань одна від одної. Ширина і довжина майданчиків в різних випадках може бути різною, тобто площі можуть бути квадратними (наприклад, 1х1,2х2 м), прямокутними (наприклад 1х2,2х3,2х4 м), круговими і т.д. Площі розміщують за площею чи рядами, коли вони лежать в один ряд, тобто на одній лінії як у поздовжньому так і в поперечному напрямку, або ж шаховому порядку. Це вже залежить від бажання самого дослідника. Відстань вимірюється від центру площі. Такий спосіб застосовують наприклад для обліку природного поновлення. При обліку культур за площі та облікові майданчики приймають посадкові чи посівні місця. Останні для спостереження можна відбирати або цілими смугами, розташованими через однакові відстані один від одного, або ж окремими майданчиками, коли в межах смуги обстежують в повному обсязі майданчики, лише через одне й ту саму їх кількість, наприклад - кожну п'яту, кожну десяту і т.д., незалежно від того, якою за своїм станом вона виявиться. Якщо число майданчиків N необхідних для отримання результатів з заданою точністю вже відомо, та відстань L між центрами майданчиків в метрах визначається за формулою:

$$L=\sqrt{П/N}$$

де: П - розмір облікової площі, $м^2$.

Намітивши візири (ходові лінії обстеження), розташовані один від одного на обчисленій відстані L, кожен з них розбивають на відрізки завдовжки теж в L метрів, в кінці цих відрізків (або тільки справа, або тільки ліворуч, або лише на самому візирі) закладають майданчики, де б у натурі вони не припали, хай навіть на між квартальній смузі або на прогалині. Пересувати їх не можна. Також можна закладати пробні закопки по визначенню заселеності грунтів, наприклад личинками травневого хруща, площі на стовбурі дерева для вивчення ступеня заселеності та пошкодження шкідниками-комахами тощо. Обстежувану ділянку попередньо можна розбити на типові частини, однорідні за якоюсь ознакою, і кожну типову частину обстежити окремо (типова вибірка). Але розбиття цілого на типові частини для їх характеристики можна робити реально лише 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 й ту ж відстань один від одного прокладають візири. На кожному візирі облікові одиниці (майданчики, екземпляри тощо) для спостереження можна відбирати двома способами.

В кожну енну (наприклад, десяту, двадцяту і т. д.) по ходу, але тільки завжди або зліва, або тільки праворуч, незалежно від того, якою за своїм станом виявиться відібрана одиниця спостереження. Слід зазначити, що принцип відбору енної (п'ятої, десятої тощо) за порядком одиниці спостереження можна застосовувати, наприклад, і для відбору шпильок дерев для вивчення ураження хворобами та в ряді інших випадків, наприклад при суцільних переліках - відбір для детальнішого вивчення кожного енного дерева по ходу маршруту обстеження. Так відбирають модельні та облікові дерева.

Інколи облікові одиниці намічають через кілька метрів, наприклад - 10; 20; 50 м і т.д. залежно від величини обстежуваної площі. Тут обстежується та одиниця, яка виявиться саме наприкінці кожного такого відрізка візиру або у найближчій до цієї точки. При цьому дотримується правило відбору, зазначене у попередньому пункті. Число відрізків, на яке будуть розбиті всі візири, дорівнює числу одиниць спостереження, що відбираються. Отже, знаючи необхідне число одиниць спостережень, задавшись відстанню між візирами, можна підрахувати число та загальну довжину візирів та розділити останню на прийняту кількість одиниць спостереження. В результаті отримаємо необхідну відстань по візиру між одиницями, що відбираються, або інакше кажучи – довжину відрізків наприкінці яких і проводиться спостереження. Відібрані таким шляхом дерева називаються модельними, якщо вони зрубуються, і обліковими – якщо вони не зрубуються.

Описаним прийомом можна відбирати дерева для вивчення відсотка виходу ділової деревини, відсотка заражених дерев і т.д. при окомірному спостереженню (опису), скажімо кожне десяте дерево. Для детального вивчення зі звалюванням дерева можна брати наприклад кожне двадцяте, сорокове і т.д. виходячи з того, яка кількість дерев потрібно для детального обстеження.

Завданням спостереження (досліду) може бути виявлення впливу на об'єкт спостереження, наприклад на дерево, умов навколишнього природного або антропогенного середовища, скажімо зімкнутої біогрупи, відкритого місця, лісостану і т. д. Ці особливості середовища можуть також впливати на характер відновлення, склад і стан живого надгрунтового покриву, на плодоношення різних рослин, на ріст та розвиток лісу тощо. Тоді облікові майданчики одного і того ж розміру закладають у цих типових умовах, вплив яких потрібний 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 зіставляти результати інших дослідників; а тому необхідно аналізувати не окремі факти, а всю сукупність фактів, що відносяться до висвітленого питання без жодного виключення у тому зв'язку, і без виривання окремих фактів і цифр із загальної зв'язку явищ.

При проведенні спостережень, важливим є розробка форми запису даних спостереження з чітким переліком всіх ознак, що підлягають обліку та зазначенням одиниць, точності вимірювання. Від якості та форми запису, залежить і повнота отриманих відомостей. Форма запису може бути ***облікова***, коли до відомості заносяться дані спостережень по кожній одиниці спостереження у порядку обстеження, та ***карткова,*** коли всі дані про кожну одиницю спостереження заносяться до окремої картку чи бланку спостережень. Характерним прикладом тут буде картка таксації модельних дерев. Карткова система запису зручніша в тому відношенні, що сильно полегшує камеральну обробку при зведенні та угрупованні матеріалу за різними ознаками В даний час все частіше для запису використовують носії інформації для комп'ютерів, які вставляють у спеціальні пристрої. У цьому випадку інформація одразу вводиться в ПК, виключаючи її набір, що прискорює процес обробки інформації.

Перед початком досліджень розробляється методика проведення спостереження: уточнюється одиниця спостереження, якщо вона складна, визначаються її розміри та форма (смуги, пробні майданчики та площі); встановлюється точність кінцевого результату та необхідна кількість одиниць спостереження, спосіб проведення спостереження, відбору одиниць; оформлення одиниць спостереження у натурі; терміни спостережень та одиниці вимірів. Складається календарний план наукових робіт та кошторис витрат.

Узагальнюючи викладене щодо проведення спостережень, можемо зробити такі висновки:

❖ спостереження є дослідною основою статистичного дослідження;

❖ для того, щоб за даними вибірки можна було б з впевненістю робити висновки про сукупність, вибіркове спостереження має бути правильно організовано.

Тут вирішують два основні питання:

- скільки спостережень є достатнім;

- які одиниці сукупності мають бути обрані для спостереження, тобто, що (хто) складатиме вибірку.

Перше питання може бути вирішене за допомогою таблиці достатньо великих чисел, що наводиться у багатьох посібниках зі статистики, або із застосуванням спеціальних формул. Вони будуть описані нижче. Вибір одиниць для спостереження може бути спланований різним чином в залежності від складу сукупності та

відомостей про неї. Якщо сукупність варіює не надто в широких межах і якщо вибірка становить щонайменше 20% обсягу сукупності, то застосовують простий випадковий відбір одиниць чи просте вибіркове спостереження. Для цього зручно користуватися таблицею випадкових чисел.

В лісовому господарстві скористатися таблицею випадкових чисел у натурі технічно важко та практично часто неможливо. Тому тут користуються систематичним вибірковим спостереженням. Наприклад, якщо належить взяти 10% - ву вибірку дерев з 800 штук, то випадковим порядком вибирають перше, покладемо 5 дерево, і після цього беруть кожне наступне через десять номерів, тобто за номерами: 15, 25, 35 тощо, закінчуючи номером 795. Якщо є відомості про те, що сукупність у своїх частинах неоднакова, наприклад з вищим рівнем явища в одних частинах ніж у 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 розділяють на групи за віком. Отримаємо шари сукупності, з кожного з яких беруть незалежну вибірку та обчислюють її характеристики. Статистична обробка матеріалів досліду пошарової вибірки дещо складніше. Але лише спостереженнями та їх статистичною обробкою не обмежується збір інформації у лісівничих та інших дослідженнях. Найчастіше на додаток до спостережень ставиться експеримент. У сукупності спостереження та експеримент практично вичерпні джерела первинної інформації в лісовому господарстві. Спостереження зазвичай не вимагають втручання у нормальне функціонування об'єкта. У багатьох лісівничих дослідженнях вони є єдино можливими, наприклад, фенологічні спостереження, вивчення зростання дерев та деревостанів, приживання лісових культур та ін. Проте певна «пасивність» спостереження щодо до об'єкта дослідження не передбачає відсутності плану або системи: спостереження як метод наукового пізнання передбачає наявність суворого плану. Вдалою ілюстрацією запланованих спостережень є вибіркові методи інвентаризації лісових ресурсів на великих територіях, що проводяться у багатьох країнах. Експеримент передбачає активний та цілеспрямований вплив на об'єкт, що вивчається, або явище, певну керованість умов його проведення. Співвідношення ролі спостереження та експерименту достатньо складне. У науковому пізнанні завдання спостереження зазвичай більше скромне і зводиться частіше до опису та аналізу явищ, що спостерігаються та процесів. В експерименті сильніша теоретична сторона, рівень осмислення факторів, що спостерігаються; експеримент має у своєму розпорядженні засоби активного втручання у перебіг подій. Однак у лісовому господарстві, особливо при вивченні природних об'єктів, спостереження часто відіграє

важливішу роль, ніж експеримент. Для зручності класифікації можна виділити звичайний модельний експеримент та математично спланований чи екстремальний. Звичайний модельний експеримент відрізняється виділенням досліджуваних зв'язків та ізоляцією їх від зовнішніх впливів; при цьому він може бути однофакторним та багатофакторним. Наприклад, беремо сіянець, поміщаємо його у штучне середовище та досліджуємо вплив на його зростання деякого добрива. Якщо ж цей сіянець спостерігатиме в природних умовах, то треба враховувати і опади, і температуру та інше, тобто багато факторів, а не одне добриво.

Математичне планування експерименту (багатофакторне), дозволяє оптимізувати процес дослідження: заздалегідь вибрати найкращу (з точки зору мети роботи) математичну модель, застосувати послідовну стратегію та скоригувати напрямки досліджень після кожного етапу та ін. За будь-якого методу збору інформації її обробку та використання будують за схемою: інформація - гіпотеза - модель - перевірка відповідності моделі вихідної інформації об’єкту, для якого розроблена модель. Слід наголосити на важливості останнього етапу, що нерідко недооцінюється. Можливість застосування моделі у конкретній ситуації вимагає обов'язкового доказу який може бути ймовірносно статистичним, якщо у процесі дослідження не порушувалися основні статистичні передумови організації збору інформації, або емпіричним, тобто на основі додатково зібраної контрольної інформації.

В основі отримання первинної числової інформації лежить як правило процес вимірювання - знаходження значень фізичної величини дослідним шляхом за допомогою спеціальних технічних засобів вимірювань.

В даний час існує два підходи до вимірювального процесу: класичний та інформаційний. У більшості завдань лісової справи виконуються основні передумови класичного підходу до вимірів: вимірювана величина передбачається незмінною протягом часу вимірів і характеризується одним значенням, для якого можна зазначити інтервал невизначеності, тобто помилку вимірів, час виміру практично не обмежено; зовнішні умови та фактори, що впливають результат вимірювання, враховані повністю. Інформаційна модель вимірювального процесу трактує вимір як випадковий процес, тобто, дозволяє оцінювати якість вимірювання величин, що змінюються в часі.

Вимір може бути прямим і непрямим. У першому випадку, досліджувану величину вимірюють безпосередньо, у другому – спостерігають не досліджувану величину, а іншу, яка з нею прямо пов'язана і яку простіше виміряти. Так об’єм дерев, що ростуть, зазвичай визначають виміром їх діаметрів та висоти, а об'ємний приріст знаходиться шляхом вимірювання радіального приросту,

енергії росту тощо. Перехід до величини, що є предметом вивчення, відбувається за допомогою математичних моделей зв'язку. Докладніше про це буде викладено нижче.

2.4 Дедуктивний та індуктивний методи у лісовій біометрії.

В лісовій біометрії застосовують як *дедуктивний (від загального до конкретного)*, так і індуктивні методи досліджень.

Дедуктивний метод застосовується, коли хоча б орієнтовно, відомі загальні закономірності зміни випадкової величини. Так, ми знаємо центральну граничну теорему, доведену професором А. Ляпуновим (1857 – 1918) в 1901 році, яка говорить, що розподіл суми незалежних випадкових величин (i=1, 2, ..., n) прагне до нормального розподілу при необмеженому збільшенні критерія N якщо всі величини мають кінцеві середні та дисперсії і одна з них за своїм значенням різко не відрізняється від інших. Керуючись цією теоремою, можна розглядати розподіли, скажімо діаметрів стовбурів в деревостані сосни звичайної у віці 85-90 років, повнотою 0,8, середньою висотою 15,5 м, II класу бонітету, закладеної в 2022 році пробної площі №2 групою вчених у складі кандидата с.-г. наук, доцента Левченко В. Б., здобувача освітнього ступеня «Бакалавр» Ткаченко Марини Володимирівни, та старшого наукового співробітника Поліського природного заповідника Бєльської Ольги Валеріївни в лісорослинних умовах B_{2-3} Перганського природоохоронного науково-дослдіного відділення Поліського природного заповідника, як окремий випадок прояву названої закономірності, та використовувати криву нормального розподілу для прогнозу продуктивності та складу деревостану, спираючись на результати вивчення річних приростів за результатами дендрохронологічного аналізу, визначаючи параметри конкретного деревостану за проведеними спостереженнями.

Але в лісовому господарстві частіше доводиться використовувати індуктивний метод досліджень, тобто від конкретного до загального. Вище вже згадувалося, що при статистичних спостереженнях в біології практично завжди мають справу з вибірками, та за результатами їх роблять висновки про сукупність. Таким чином, варіаційна статистика застосовує метод індукції - коли узагальнення роблять, вивчивши окремі випадки. Правомірність цього методу базується на використанні найважливіших понять та положень теорії ймовірностей. Як приклади можна навести вже згадані залежності між діаметрами та висотами в насадженні пробної площі №2 Перганського ПНДВ Поліського природного заповідника. Зробивши аналіз низки вибірок з деревостанів різних деревних порід, що відрізняються також і віком, ми прийдемо до висновку, що в одному випадку для доказу та прогнозу ймовірності росту дерев у висоту та формування їх приросту слід використовувати рівняння параболи 2 порядку, в

іншому випадку 3 порядку, в третьому деяку більш складну криву. Названу закономірність ми отримуємо, проаналізувавши ряд окремих випадків (окремих насаджень), тобто йдемо від конкретного до загального, застосовуючи індуктивний метод. Індуктивний висновок, як загальний логічний процес, що йде від великого та малого покликання на кінцевий результат, має таку форму:

Більше покликання: це дерева виду сосна звичайна, що мають хороший ранній та пізній прирости.

Мале покликання: ці дерева знаходяться в лісорослинних умовах $B_{2\text{-}3}$ закладеної пробної площі Поліського природного заповідника.

Висновок: всі обстежені дерева на пробній площі є сосна звичайна, що дає щорічний позитивний ранній та пізній дендрохронологічний прирости.

Очевидно, що висновок зроблений з індуктивною аргументацією ширший ніж покликання. У висновку додається щось нове, що розширює знання про явище яке вивчається. Це потенційне розширення знань вимагає обережності. Воно може бути плідним, але існує деяка небезпека отримати необґрунтовані та хибні висновки. Логічним підґрунтям індуктивного висновку є припущення про однаковість у системі фактів, які стосуються покликань і висновків. Це припущення, що називають по-різному, - одноманітністю в природі, статистичною стійкістю досліду, обмеженням незалежної варіації в природі, - завжди представляє хіба що невисловлене покликання індукції наукового результату. Якби однаковість у природних процесах не виявлялася, природі був би властивий повний хаос. При цьому жодне нагромадження фактів було б не можливо виправдати індукцію. Не можна було б нічого сказати про умови поза дослідом. Але в природі існує певна однаковість у поведінці окремих одиниць, що становлять те чи інше масове явище. Однак ця одноманітність в природі не настільки сувора, щоб можна було зробити точну оцінку масового (загального) явища складових одиниць, що спостерігаються. Тому статистичні висновки про властивості генеральних сукупностей за вибірковими завжди мають ймовірнісний характер, тобто робляться з певним ступенем безпомилковості і ніколи не робляться з повною достовірністю. Слід зазначити, що конструкція вибіркових оцінок виявляється більш кращою навіть у тих випадках, коли всі одиниці, що становлять те чи інше явище, можна виміряти, тобто вони відносяться до обмежених генеральних сукупностей. Це ситуація, що торкнулася різних видів генеральних сукупностей, потребує більш широкого пояснення. Насправді зустрічаються обстежувані генеральні сукупності кінцеві та нескінченні. Прикладом першої може бути вибіркове обстеження, припустимо вибірки ранніх та пізніх приростів при проведенні дендрохроноіндикаційного аналізу відібраних кернів в умовах закладеної пробної площі Поліського

природного заповідника. З нескінченними сукупностями мають справу при різних експериментальних дослідженнях, коли питання полягає не в тому, щоб отримати точний результат у даному експерименті, а головним чином в оцінці того, якими будуть результати масового застосування цього процесу (у % від обстежених одиниць) - біологічного, технологічного або економічного.

Припустимо, що проводиться оцінка ступеню пошкодження пристигаючих та стиглих деревостанів сосни звичайної під час низових пожеж в умовах Перганського та Копищанського природоохоронних науково-дослідних відділеннях (ПНДВ) Поліського природного заповідника. У цьому випадку генеральна сукупність нескінченна, бо для оцінки не так важливо, скільки пошкоджено вогнем та ступінь підгару на конкретних деревах, як те, - які пошкодження вогнем можливі при подібних лісорослинних умовах, що не досліджувались у досліді. Тут науковий експеримент стає хіба що «механізмом» отримання випадкової вибірки.

Можливі обставини, коли корисно вдатися до конкретної логічної конструкції - гіпотетичної генеральної надсукупності. Іноді ми можемо мати відомості навіть суцільного обстеження реально існуючої сукупності, і все ж таки корисно розглядати ці дані як вибірку з деякої надсукупності. Так роблять, коли не тільки потрібно отримані факти, а й необхідно виявити загальну закономірність, для якої статистичний матеріал представляється лише окремим випадком.

Припустимо, що із статистичних обстежень за 2005 – 2024 роки кількість лісових пожеж, що виникли в лісорослинних умовах Поліського природного заповідника склала 52% від лісопокритої площі. Цей матеріал отримано шляхом суцільного обстеження та характеризує явище однозначно. Однак, якщо нас цікавить результат за межами обстежених років або перевіряється висновок про те, що низові пожежі в умовах Поліського природного заповідника траплялись частіше, тоді отримані дані слід розглядати як вибірку з деякою нескінченною надсукупності різних можливих пропорцій пожеж за кількістю їх виникнення. На основі таких відомостей, користуючись методами статистики, надається можливим дослідити, чи прийнятне припущення про частіше виникнення низових (надземних пожеж). Зауважимо, що така надсукупність не обмежена ні чисельністю, ні територією згідно якої зроблено експеримент. З наведених прикладів видно, що в біометрії (в т.ч. у лісовій біометрії) застосовуються обидва методи (індуктивний та дедуктивний), але переважає індуктивний.

РОЗДІЛ 3. ГРУППУВАННЯ 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, явища суспільного життя і т. п.), їх аналіз може бути якісним або кількісним. Якісний аналіз явища, що вивчається, або процесу полягає у виділенні деяких його характерних властивостей, особливостей, ознак, що якісно відрізняються між собою всередині аналізованої сукупності. Наприклад, ми вивчаємо змішане насадження, що складається із сосни звичайної та берези повислої. Заклали пробну площу, на якій нарахували 180 дерев сосни звичайної та 110 берези повислої. Для подальшого використання цього матеріалу необхідно зробити попередній якісний аналіз. В нашому випадку розділити дерева за головною якістю – належністю до різних ботанічних (дендрологічних) видів, тобто на сосну звичайну та березу повислу. Наведений приклад простий. Насправді буває, що зробити якісний аналіз важко, для чого застосовуються спеціальні методи, що розглядаються нижче. Але якісного аналізу часто буває недостатньо, щоб зрозуміти деяке явище чи процес, дати йому коректний математичний опис. І тут необхідно використовувати кількісні методи дослідження.

Проведення якісного та кількісного дослідження починається з планування та постановки експерименту. У лісовому господарстві дослідження часто полягає у проведенні вимірювань на деяких виділених ділянках лісу (пробних площах) або у вимірі частини дерева. Можливі також інші варіанти експериментів, що ми вже розглядали вище і до чого ще повернемося. При науковому чи практичному дослідженні деякого явища або процесу потрібно з'ясувати його природу, закономірності зміни часу або зв'язок із деякими параметрами. Для цього недостатньо провести спостереження чи поставити експеримент. Вивести пошукові закони та закономірності явища, що вивчається, отримати правильні висновки зі спостережень можна тільки в тому випадку, якщо буде зроблено коректний кількісний та якісний аналіз проведених спостережень, або, як їх ще називають випадкових явищ. Найчастіше для аналізу використовуються кількісні методи. В цьому випадку звертають увагу на наявність подібності чи відмінності, користуючись певними числовими критеріями.

3.2 Систематизація та угруповання вихідних даних.

Будь-який аналіз проведених спостережень починається із систематизації спостережень. Першим її етапом є угруповання вихідних даних чи варіантів. При постановці експерименту, угруповання передбачають вже на етапі збору експериментального матеріалу, тобто на етапі спостережень. Наприклад, вимірюючи висоту в 6-8 - річних культурах сосни звичайної, записують окремо кожен вимір висоти в метрах: 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, у поданому вигляді ускладнений. При великих масивах інформації аналіз окремих вимірів переростає у велику проблему. Для її вирішення результати спостережень, зазвичай, систематизують. Систематизація полягає в групуванні виміряних величин: товщини або висоти дерев, ваги тварин, розміру та ваги насіння тощо. *Для угруповання спостережень виділяють класи (при вимірах дерев їх називають ступені)*, за якими розносять виміряні величини. У наведеному прикладі доцільно виділити такі класи висот: 0 – 0,50; 0,51 – 1,0; 1,01 – 1,50; 1,51 – 2,00; 2,01 – 2,50; 2,51 – 3,0; 3,01 – 3,50. Тоді запис вимірювань можна буде звести до таблиці (таблиця 3.1). У таблиці не обов'язково показувати сам інтервал, достатньо навести значення його середини.

Таблиця 3.1

Розподіл висот за класами в культурах сосни звичайної

Ступінь висоти, м	Кількість дерев, шт.
0,25	1
0,75	2
1,25	3
1,75	4
2,25	5
2,75	6
3,25	7
Всього:	15

Результати, зведені в таблицю дають наочніше уявлення про висоту культур сосни звичайної на досліджуваній пробній площі. До того ж обробку матеріалу легко проводити коли маємо систематизовані дані. Тому саме така таблична форма найчастіше використовується при проведення досліджень у лісовому господарстві.

При впорядкуванні (систематизації) отриманих даних легко обробити їх математично і вивести статистичні показники, які будуть вичерпно характеризувати сукупність, що вивчається. Проблема систематизації та угруповання займає велике місце у статистиці. Помилкове групування даних може призвести до

неправильних висновків про сутність досліджуваного явища. Найбільш просте групування при якісному аналізі. Так, якщо кора осики відрізняється за забарвленням, то розподіл дерев з різним забарвленням кори може бути виражено у відсотках від загальної кількості виміряних дерев, як це показано у таблиці 3.2.

Таблиця 3.2

Розподіл дерев за забарвленням кори

Забарвлення кори	Кількість дерев, шт.	% від загальної кількості дерев
темно - сіра	160	40
світло - сіра	200	50
зелена	40	10
всього:	400	100

Окремим випадком якісної варіації є альтернативна, коли в сукупності можна виділити лише дві групи. В однієї групи присутня певна якість (або ознака), у членів іншої групи її немає. Наприклад, при дослідженні культур сосни звичайної в умовах певного кварталу, що уражені кореневою губкою, ми ділимо 8-річні дерева за альтернативною ознакою: здорові та хворі.

3.3 Складання рядів та таблиць розподілу.

Під час проведення спостережень в лісовому господарстві найчастіше мають справу з безперервним дискретним (перервним) розподілом досліджуваної величини. Так, вимірюючи товщину дерева, ми проводимо вимірювання кожного дерева. Їх сукупність є деяким рядом розподілу, у якого є мінімальна та максимальна величина. Наприклад в таблиці 3.3 наведено результати проведення вимірів 120 дерев сосни звичайної II класу бонітету в лісорослинних умовах $B_{2\text{-}3}$ в процесі відбору приростних кернів при допомозі буру Пресслера для проведення дендрохроноіндикаційного аналізу в типі лісу «сосняк моховий» віком 100 років. Для того, щоб надати дослідним матеріалам певну наочність та отримати з них необхідну статистичну інформацію про досліджувану ознаку, матеріали спостереження піддають групіюванню. Згруповані таким чином матеріали називають статистичними таблицями або статистичними рядами.

Статистичним рядом або рядом розподілу називають ряд значень ознаки, розміщених у порядку зростання чи спадання, з вказаним числом повторень.

Таблиця 3.3

Результат вимірювання діаметра 120 дерев сосни звичайної на висоті 1,3 метра в умовах в лісорослинних умовах B_{2-3} пробної площі №3 Перганського ПНДВ Поліського природного заповідника (за даними Левченко В. Б, Ткаченко М. В., Бєльської О. В., середнє за 2022 – 2025 рр.)

№ дерева	Діаметр, см	№ дерева	Діаметр, см	№ дерева	Діаметр, см	№ дерева	Діаметр, см	№ дерева	Діаметр, см
1	23	25	28	49	24	73	25	97	44
2	32	26	40	50	28	74	32	98	32
3	19	27	32	51	34	75	12	99	42
4	28	28	15	52	40	76	18	100	18
5	24	29	24	53	41	77	16	101	24
6	19	30	40	54	36	78	24	102	26
7	22	31	42	55	32	79	26	103	22
8	24	32	34	56	24	80	30	104	15
9	27	33	28	57	28	81	34	105	10
10	26	34	34	58	30	82	20	106	9
11	20	35	20	59	34	83	22	107	18
12	19	36	19	60	22	84	24	108	16
13	24	37	24	61	26	85	28	109	24
14	39	38	20	62	18	86	30	110	20
15	36	39	26	63	32	87	26	111	30
16	27	40	38	64	38	88	24	112	32
17	34	41	42	65	42	89	22	113	42
18	28	42	46	66	40	90	26	114	44
19	29	43	34	67	22	91	30	115	32
20	32	44	32	68	18	92	24	116	34
21	36	45	24	69	19	93	28	117	32
22	3924	46	20	70	24	94	32	118	24
23	22	47	34	71	32	95	51	119	20
24	20	48	18	72	36	96	18	120	22

Значення ознаки, зведеної в ряд, називають класовими варіантами, а число повторень їх у класах – чисельністю або частотами класів.

При проведенні вимірювань дерев в лісі, класи зазвичай називають ступнями товщини або ступенями висоти. Статистичний ряд значень виміряної ознаки набувають шляхом визначення величини класу або інтервалу, розміщення класів та розподілу в них усіх одиниць спостереження.

В якості k, приймають ціле число, що найближче до отриманого конкретного значення. При цьому дійсна кількість класів визначиться як конкретна величина від розмаху варіант (Xmax-Xmin) на заокруглене значення інтервалу. Як отримане конкретне значення, приймають також ціле (заокруглене) число. Округлення

робиться завжди в більшу сторону. Величину інтервалу (k) визначають за формулою:

$$k = \frac{X_{max} - X_{min}}{t}$$

де: Xmax та Xmin - відповідно найбільше та найменше значення ознаки чи варіанта; t; - число прийнятих класів.

Під час проведення біометричних досліджень, оптимальною кількістю класів за наявності великої вибірки вважається 12. Допустимо збільшувати або зменшувати цю кількість залежно від обсягу спостережень. Якщо ряд спостережень відносно невеликий (40-60 вимірів), а різниця між Xmax і Xmin не велика, його доцільно звузити, т. к. в іншому випадку ряд розподілу виявиться розмитим, і в деяких класах може не виявитися виміряних величин.

Зазвичай число класів, що рекомендується, дорівнює 12±3, тобто коливається від 9 до 15. В окремих випадках допустимо зменшити кількість класів до 8, як виняток - до 7. Меншу та більшу кількість класів приймати не рекомендується, якщо це не пов'язано з специфічними особливостями експерименту.

Межі та серединні значення класів краще встановлювати наступним чином. Як середнє значення першого класу приймають число, кратне класовому проміжку k найближче до найменшої (в зростаючому ряді), або найбільшою (у спадному ряду) варіантів ряду розподілу. Середні значення наступних класів одержують шляхом послідовного додавання величини інтервалу. Нижні межі класів визначають шляхом віднімання половини величини інтервалу із серединних значень кожного класу, а верхні межі – шляхом додавання цієї половини.

З метою виключення перекриття верхньої межі попереднього класу з нижньою межею наступного класу, що входять до першого та другий класи, нижні межі класів збільшують на деяку величину, що рівна точності виміру ознаки. Можна також верхні межі класів зменшити на ту саму величину, але перший варіант використовують частіше. Саме так отримуємо значення меж класів. Наприклад, при вимірювання діаметрів стовбурів дерев сосни звичайної у вище наведеному прикладі по 4 см класам (ступеням) товщини або висоти, величина збільшення (зменшення) зазвичай дорівнює 0,1 см, для висот - 0,1 м. Наприклад, ступінь товщини, що дорівнює 24 см, матиме межі від 22,1 см до 26,0 см. Як варіант, може бути від 22,0 см до 25,9 см, але перший приклад краще.

Проілюструємо викладене на конкретному прикладі. Візьмемо дані вимірювання 120 діаметрів стовбурів сосни звичайної, наведені в таблиці 3.3. У прикладі найменший виміряний діаметр дорівнює 9 див. (дерево №106), найбільший – 51 см (дерево №95). Різниця

складає 42 см. Величина класового інтервалу для оптимального випадку становитиме k = 42/12 = 3,5 см. Округливши його в більшу сторону, отримаємо величину класового інтервалу (ступінь товщини) рівну 4 см. Керуючись викладеними правилами встановлення першого та останнього класів, одержуємо відповідно перший клас, рівним 8 см (найближче число до 9, кратне 4) і останній 52 см - найближче число до 51 см. Так як величини діаметрів, що лежать на межі класового проміжку можна віднести до будь-якого із сусідніх класів (наприклад, дерево діаметром 22 см можна віднести до класового проміжку із серединою класу та 20 см, і 24 см), то доцільно нижні межі класів збільшувати на 0,1 см. Можна аналогічно верхні межі класів зменшувати на 0,1 см, але, як сказано вище, перший варіант буде зручнішим. В цьому випадку класовий інтервал із серединою 20 см дорівнюватиме 18,1-22,0 см, а із серединою в 24 см відповідно 22,1-26,0 см. У цьому випадку дерево завтовшки 22 см однозначно буде віднесено до ступеня товщини 20 см. Після встановлення класів, розносимо виміряні діаметри за класами або, як їх називають у лісівництві - ступенями товщини. При цьому застосовують символічні позначки для врахованих дерев від 1 до 10 – запис здійснюють за принципом «конверта». Вигляд цього символічного запису показано в таблиці 3.4.

Варіаційний ряд, розглянутий нами як приклад є одновершинним за розподілом. Це означає, що він має один модальний клас. Можливі випадки, коли у варіаційному ряді виявляється кілька модальних класів, і тоді варіаційний ряд є багатовершинним. Найбільш простою причиною багатовершинності, особливо при дуже розтягнутих рядах, є недостатня кількість варіант у вивченої сукупності.

При малій кількості особин у деяких класах варіаційного ряду може взагалі не бути жодної варіанти. Варіаційний ряд виявиться переваним, а варіаційна крива – розірваною на частини. Однак, якщо і при великій кількості особин в сукупності, що вивчається спостерігається дво- або багатовершинність, причину цього треба шукати в самому експериментальному матеріалі.

Розподілені варіанти (чисельності) товщин дерев із зазначенням класів показано в таблиці 3.5.

Він в такому разі може бути змішаний двох якісно різних сукупностей вибірки, яка або перебувала в різко умовах зовнішньої середовища, або належить до різних типів, наприклад, до деревостанів різних поколінь. З'єднання в одному ряду особин різних деревостанів може дати зовнішню картину дво- або багатовершинності. Відомо, наприклад, що абсолютно різновікові дерева модрини європейської дають багатовершинний розподіл. Тому в один варіаційний ряд поміщають лише дерева одного покоління. Щоправда можливі випадки, хоча вони відносно

нечисленні, коли дво- або багатовершинність визначається властивостями самих досліджуваних ознак і тому характеризує цілком однорідний матеріал. Визначення цієї однорідності чи неоднорідності – прерогатива фахівця - лісівника чи біолога.

Таблиця 3.4

Символічний запис варіантів під час виміру діаметру стовбурів сосни звичайної в умовах пробної площі №3 Перганського ПНДВ (за даними Левченко В. Б, Ткаченко М. В., Бєльської О. В., середнє за 2022 – 2025 рр.)

Величина варіант	1	2	3	4	5	6	7	8	9	10	11	12	13	…	20 й т. д.
Символи	.	. .	. . .	. . . .	—	⌐	⌐	□	⊘	⊠	. ⊠	. . ⊠	. . . ⊠	...	⊠ ⊠

Таблиця 3.5

Таблиця розподілу діаметрів 120 стовбурів сони звичайної в умовах пробної площі №3

Ступені товщини (класи), X_i (см)	Кількість (число дерев), n_i, шт.
8	1
12	3
16	8
20	14
24	20
28	27
32	17
36	12
40	9
44	5
48	3
52	1
Всього (Σ):	120

3.4 Прогнозування випадкової величини.

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

Теорія ймовірності – це одна з дисциплін математики. Одним із основних понять цієї теорії є ймовірність.

Ймовірністю події А називають відношення числа випадків, що сприяють появі даної події до всіх можливих випадків.

Позначимо ймовірність літерою Р із зазначенням у дужках індексу події, в нашому разі події А. Її визначають за формулою:

$$P(A)=n/N$$

де: n – число випадків, що сприяють події А;
N – загальна кількість випадків.

Так, якщо в певному сосуді міститься 5 однакових перемішаних куль, причому 2-і з них чорні, а 3 - білі, то ймовірність вийняти навмання білу кулю дорівнює Р(А)=3/5=0,6, а можливість вийняти чорну кулю Р(В)=2/5=0,4. Ймовірність змінюється від нуля до одиниці. Ймовірність, що дорівнює нулю показує, що подія є неможливою, а ймовірність, що дорівнює одиниці означає, що подія єдина можлива чи достовірна.

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

У вказаному прикладі події А та 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.

Ймовірності таких подій однакова. В лісо-біологічних роботах ймовірність найчастіше всього встановити неможливо, оскільки вся досліджувана генеральна сукупність та її склад невідомі, наприклад, кількість дерев різної товщини на великій ділянці ліси. У таких випадках отримують аналог ймовірності на основі досліду, тобто у вибірковій сукупності. Для цього підраховують кількість проб, в яких подія практично з'явилася, і відносять її до спільного числа проб. Це відношення називають відносною частотою події та виражають формулою:

$$W(A)=\frac{n}{N}$$

де: n – число появи подій;
N – загальна кількість проб.

Тривалі спостереження показали, що за однакових умов випробувань і досить великому їх числі, відносна частота в різних дослідах змінюється мало, причому тим менше, чим більший обсяг вибірки. Вона коливається (варіює) біля певного постійного числа.

Ця властивість повторюваності конкретних значень називається стійкістю відносної частоти, або - статистичною стійкістю.

Дуже характерним є приклад, що пов'язаний з закладанням ранніх та пізніх приростів у сосни звичайної. Раніше ми наводили приклад із формування ранніх та пізніх приростів в сосновому деревостані і наведемо його як приклад стійкості відносної частоти.

По місяцях за деякий рік, починаючи з березня формування раннього приросту характеризується такими значеннями: 0,486; 0,489; 0,471; 0,478; 0,482; 0,462; 0,484; 0,485; 0,491; 0,482; 0,473. По пізньому приросту, величини залежатимуть від закладання раннього приросту, так як сумарна ймовірність закладання раннього та пізнього приросту в сосни звичайної дорівнює 1,0. Постійне число, біля якого варіюють відносні частоти, є ймовірністю появи події. Таким чином, якщо дослідним шляхом встановлена відносна частота, то отримане число можна прийняти за наближене значення ймовірності. Відносні частоти у вказаному прикладі коливаються близько числа 0,482, яке можна сприйняти за наближене значення ймовірності закладання раннього приросту. Слід звернути увагу, що таке судження про ймовірність в основі відносної частоти тим надійніше, чим більше число досліджень або обсяг вибірки. Найбільш простим та переконливим прикладом для підтвердження цього положення є кидання монети. Багаторазові досліди з монетою, яких підраховували кількість появи герба, дали наступний результат:

Кількість кидань: 4040; 12000; 24000;

Відносна частота: 0,5069; 0,5016; 0,5005;

Ймовірність появи герба дорівнює 0,5.

Неважко уявити чи випробувати на досліді, що за малої кількості спостережень, наприклад, при 6 киданнях, таке наближення відносної частоти до ймовірності, як правило не отримати. Провівши аналіз випадкових величин необхідно виконувати низку вимог.

<u>Вимоги репрезентативності:</u>

1) Вибіркова сукупність має характеризувати собою генеральну з певною точністю.

2) Вибіркова сукупність має бути вільною від суб'єктивних уявлень про генеральну. Інакше висловлюючись, - задовольняти вимоги репрезентативності – отже всебічно та відповідно характеризувати генеральну сукупність.

Залежно від кількості спостережень (К) вибіркова сукупність може називатися великою або малою.

Межею тут є N, що дорівнює 30 спостереженням.

<u>Основні етапи вивчення статистичних сукупностей:</u>

1. Складання програми, найважливішими елементами якої є мета та завдання дослідження.

2. Складання методики дослідження, яка містить вибір та обґрунтування місця, часу та облікової ознаки на об'єкті дослідження. Способи обліку, необхідна кількість спостережень, форма запису, вибір інструментів та одиниці виміру, способи подальшої обробки матеріалів дослідження – усе це предмет методики досліджень.

3. Проведення спостережень, вимірів чи обліку.

4. Первинна обробка результатів спостереження.

5. Моделювання досліджуваного явища.

6. Додаткове проведення спостережень, доведення моделі та дослідження за допомогою моделі, повторне опрацювання результатів дослідження.

7. Систематизація та аналіз отриманих даних.

Ця послідовність загалом дотримується як у наукових дослідженнях, так і під час вирішення виробничих завдань. У цій послідовності необхідно виконувати індивідуальне завдання, не втрачаючи з уваги мети та сенсу результатів дослідження, які зумовлені завданням.

Правила обчислення результатів.

Правила обчислення результатів представлені згідно з принципом Крилова-Брадиса (П. М. Крилов - 1879-1955 – радянський математик) і наводяться у скороченні.

Правила 1-5. При додаванні та відніманні, множенні та діленні, піднесенні до квадрату або кубу, знайденні кореня квадратного або кубічного, при використанні логарифмів – в результаті потрібно зберігати стільки десяткових знаків після коми, скільки їх має сума із найменшою кількістю десяткових знаків.

Правило 6. Для проміжного результату, одержуваного за правилами 1-5, необхідно зберегти одну додаткову «запасну» цифру, як в кінцевому результаті після осмислення результату її можна відкинути.

Правило 7. Якщо вихідні дані мають різну кількість десяткових знаків чи значущих цифр, їх треба попередньо округлити зі збереженням однієї «запасної» цифри.

Правило 8. Якщо результати мають бути отримані з n – значними цифрами, то вихідні дані слід брати із n+1 значущою цифрою.

Однією з істотних умов правильно і добре організованого обчислювального процесу, є акуратність і ретельність проведення записів.

РОЗДІЛ 4. СЕРЕДНІ ЗНАЧЕННЯ

4.1 Статистичні показники варіаційного ряду.
4.2 Середні величини.
4.3 Середні арифметичні та способи їх обчислення.
4.4 Інші види середніх величин.

4.1 Статистичні показники варіаційного ряду.

Важливим показником статистичного ряду є його розмах. Це найпростіший показник, що показує різницю між найбільшими і найменшими величинами (їх ще називають лімітами) у досліджуваному варіаційному ряді:

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

де: L – розмах ряду.
X_{max}, X_{min} – максимальна та мінімальна величини (ліміти).

Простота обчислення розмаху ряду розподілу, й очевидність сприяли широкому використанню цього показника в багатьох дослідженнях в лісовому господарстві. Так, у лісовій таксації щодо будови деревостану розмах низки розподілу - це обов'язковий показник. У біометрії розмах ряду та ліміти визначаються при зведенні даних спостережень чи вимірів у статистичні сукупності, тобто у варіаційні ряди. Кожен варіаційний ряд та його графічне зображення - це як би «згущення» вихідного фактичного матеріалу, перетворення його у наочну формулу. Однак для повного аналізу спостереження явища, або для оцінки деревостану цього недостатньо. Справа в тому, що розмах ряду і ліміти схильні до значних коливань від однієї часткової сукупності до іншої. Тому використання цього показника обмежено. 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 у дерево по товщині в одному насадженні має розмах від 8 до 44 см, а в іншому від 12 до 52 см, то здавалося б можна зробити висновок про якість дерев другого насадження: в другому насадженні вони товстіші. Однак ліміти не вказують на те, як розподіляються за вивченим ознакою окремі члени сукупності. Ось чому для характеристики сукупності потрібні такі показники, які б відображали властивості всіх її членів. Варіаційні ряди можуть відрізнятися за значенням ознаки, навколо якого концентрується більшість варіантів, тобто за величиною модального класу або

ступню. Значення цієї ознаки відображає центральну тенденцію, що типова для цього ряду. Але частоти в ряді відрізняються за рівнем варіації, тобто за величиною відхилення від центральної тенденції низки. Відповідно до цього, статистичні показники поділяються на дві групи: *показники, що характеризують центральну тенденцію, або рівень ряду, та 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 варіансу, дисперсія, коефіцієнти асиметрії та ексцесу. Існують ще й інші показники.

4.2 Середні величини.

Виміряні значення різних біологічних сукупностей, в т. ч. і в лісовому господарстві, є варіюючими математичними величинами. Щоб отримати точну та об'єктивну характеристику щодо варіюючої величини, вдаються поряд із побудовою статистичних таблиць, графіків та діаграм до так званих статистичних показників. Серед них найбільше поширення та застосування знаходить величина середнього значення досліджуваної ознаки. Вона дає сумарну характеристику будь-якої ознаки, вказуючи на те типове і стійке в явищі, що найбільш повно виражає його зміст. Прийнято говорити про середній діаметр і середню висоту насадження, про середню вагу насіння, середній розмірі мисливських тварин і т. п. Тому не завжди потрібно вникати в глибокий зміст, який містить концепція середньої величини. Ряди розподілу кількостей, що були наведено раніше показують, які види концентруються у деякого центрального їх значення. Отже, можна знайти таке значення варіанта або абстрактну середню кількість, яка буде найбільш представницькою характеристикою даної статистичної сукупності. Показники центральної тенденції характеризуються різними середніми величинами: середньої арифметичної, середньої квадратичної, середньою геометричною, середньою гармонійною, модою та медіаною.Сенс середніх величин полягає в тому, щоб відобразити якусь одна властивість сукупності, наприклад, середню висоту, середній діаметр, середній запас деревини на 1 га ділянки лісу, що вивчається.

Та ознака або та властивість сукупності, яка залишається незмінним при заміні індивідуальних значень їх середнім значенням, називається визначальною ознакою.

Середня повинна відобразити визначальну властивість так, щоб утворена за її допомогою вибіркова абстрактна числова сукупність при рівності чисел за величиною певних властивостей не відрізнялася від загальної сукупності. З цієї вимоги до середньої випливає таке її загальне визначення.

Середня - величина ознаки, яка характеризує об'єкти досліджень (спостережень) в абстрактній порівнянній сукупності, заміщує реальну сукупність, але при цьому зберігає незмінним її визначальні властивості: загальну довжину, загальну масу, загальний об'єм тощо.

До цього визначення ми звертатимемося кожного разу, коли будемо обговорювати реальний зміст різних середніх. Пояснимо сказане прикладом. Припустимо, що ми бажаємо дізнатися середній об'єм дерева в сосновому деревостані II класу бонітету у віці 50 років. Для цього вибираємо деяку вибіркову часткову сукупність (закладаємо одну або кілька пробних площ), де визначаємо середні значення цих вибіркових сукупностей. При цьому методично все необхідно зробити так, щоб середні величини вибіркової сукупності відповідали середнім величинам досліджуваних соснових деревостанів у генеральній сукупності. Методи коректного визначення цих середніх величин розглянуто нижче.

4.3 Середні арифметичні та способи їх обчислення.

Середня арифметична - найчастіше вживаний статистичний показник. Вона є свого роду центром тяжіння будь-якого розподілу. Середня арифметична була б значенням величини в точці рівноваги, кривою чисельностей, якби модель такої кривої була виконана у вигляді масивної форми. Середню арифметичну генеральну сукупність зазвичай позначають M, та її вибіркову оцінку, тобто середню арифметичну вибіркову спостережень - $\overline{X}$ (або $\bar{I}$). Вона має те ж значення, що й варіанти. Середня арифметична виходить від поділу суми чисельностей всіх варіантів (n_1, n_2, n_n) на їхньої число (N), тобто:

$$\overline{X} = (n_1 + n_2 + \ldots + n_n)/N = (\Sigma n)/N,$$

де: N – сума чисельності варіант, Σ – сума варіантів.

Тут $\overline{X}$ (і в подальшому його застосування) без вказівки меж суми означає, що має бути проведено додавання всіх виміряних (спостережених) варіант ряду від 1 до N. Для варіант (припустимо, що це висота сходів сосни звичайної, см) 3, 4, 4, 4, 5.

$$\overline{X} = (3 + 4 + 4 + 4 + 5)/5 = 20/5 = 4 \text{ см}$$

Реальний зміст середньої арифметичної та її головне призначення краще усвідомити, якщо приведене вище визначення для всіх середніх застосувати до прикладу її розрахунку. Завдяки отриманій середній можливу реальну сукупність висоти саджанців сосни звичайної з n_1, n_2, n_3, n_4, n_5, (N=5), замінити абстрактною вирівняною сукупністю з $\overline{X}$, $\overline{X}$, $\overline{X}$, $\overline{X}$, $\overline{X}$ (N = 5), не змінюючи при

цьому визначального властивості, що виражається ΣX і також NX, звідки отримуємо: X=(Σn)/N.

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

$$\overline{X} = (n_1 x_1 + n_2 x_2 + \ldots + n_n x_n)/(n_1 + n_2 + \ldots n_n) = (\Sigma n x_i)/N$$

де: x_1, x_2, x_n – класові варіанти (серединні значення класів); n_1, n_2, n_n – частоти відповідних класів; N - загальна кількість варіант (обсяг ряду) чи загальна кількість спостережень.

Групуючи варіанти розглянутого прикладу за їх величиною, отримаємо наступний ряд:

x_i : 3 4 5

n_i: 1 3 1

$\overline{X} = (1 \cdot 3 + 3 \cdot 4 + 1 \cdot 5)/5 = 4$ см

Якщо випадкова величина не виражена у вигляді простого набору чисел або у вигляді таблиці, а задана аналітично, то застосовуються такі формули обчислення середньої арифметичної:

Для безперервної середньої величини: $\overline{O} = \int_{-\infty}^{\infty} x_i f(x) dx$

Для дискретної: $\overline{X} = \sum_{i=1}^{k} x_i f(x_i) = \sum_{i=1}^{k} x_i p_i$

Надалі розглянемо інші формули обчислення арифметичної середньої, що базуються на використанні її основної властивості. Ця властивість полягає в тому, що сума відхилень всіх варіантів від арифметичної середньої дорівнює нулю. Воно випливає зі змісту середньої арифметичної як центру математичного статистичного тяжіння ряду. Сума варіантів, які більше середньої $\overline{X}$, дорівнює сумі варіант, які менші за неї.

Значення середньої арифметичної та її сутність.

Середня арифметична, як і деякі інші середні, відомо давно. Вона широко використовується при дослідженні сукупностей у науці, техніці, біології та лісовому господарстві. Середня арифметична є узагальнюючою величиною, яка вбирає у собі всі особливості досліджуваної сукупності чи низки розподілу. Вона відображає рівень усієї сукупності в цілому, дає вільну, узагальнену характеристику ознаки, що досліджується. Цифрове значення середньої арифметичної як такої може не зустрітися в жодному конкретному випадку в сукупності. Може виявитися, що жодна варіанта нерівна їй. Наприклад, ми вимірюємо діаметр дерев із заокругленням до 1 см: 8, 15, 16, 17...30 см. Середня арифметична вимірюваної сукупності може скласти дробне число, наприклад, 25,6. У виміряній сукупності жодного такого виміру немає. Ще

більш виразний приклад можна навести, якщо проаналізувати приплід (кількість) цуценят, народжених у вовків. В середньому може виявитися, що там 4,2 цуценяти. Зрозуміло, що число вовченят дрібним бути просто не може. У цьому сенсі середня арифметична є абстрактною величиною. Але в водночас вона й конкретна. Середня арифметична виражається у тих самих одиницях виміру, як і варіанти ряду. При її визначенні відхилення зі знаком (+) та (-) взаємозгладжуються, відкидаються випадкові коливання, відхилення від центральної тенденції, від рівня варіаційного ряду та виступає загальний закон явища. Розкривається те типове, що притаманно всієї сукупності загалом.

У той же час слід застерегти від можливих помилок у розумінні середньої арифметичної. Середня арифметична характеризує всю сукупність загалом, а не окремі члени сукупності. Середнє число цуценят у вовків 4,2 відноситься тільки до всієї групи вивчених тварин. Кожна ж окрема вовчиця приносить ціле число вовчинят у приплоді. Зазвичай їх буває від 2 до 6-8. Слід пам'ятати, що середня арифметична має сенс лише по відношенню до якісно однорідної сукупності. Так, не можна обчислювати середню вагу тварин різного віку або середній об’єм дерев різних деревних видів. Під час вивчення лісової таксації буде показано, як окремі вчені помилялися у визначенні закономірностей виміру форми стовбурів. Вони вважали, що форма стволів залежить від віку дерева. Цей висновок був отриманий через об'єднання в одну групу молодняків та перестиглих насаджень. Насправді виявилося, що форма стовбура залежить від віку лише до 40-50 років, а далі при незмінній висоті залишається відносно стабільною. Помилка була доведена видатним вченим-таксатором Ф. П. Моісеєнко (1894-1979). Він показав, що треба вивчити кожну вікову групу окремо і для них обчислювати їх середнє значення. Оскільки середня арифметична відноситься до конкретної сукупності, переносити її на явища, що виходять за рамки цієї сукупності не можна. В окремих випадках, якщо таке все ж потрібно, то повинен бути зроблений спеціальний аналіз явища, що вивчається, і лише за його результатами слід ухвалити рішення про доцільність такого перенесення. Надалі ми побачимо, що особливе місце у варіаційній статистиці займає питання про те, яким чином на основі даних про ту або іншу окрему сукупність можна робити висновки про інші сукупності подібного роду. Врешті-решт, *середня арифметична відноситься лише до окремих досліджуваних ознак і не може бути автоматично перенесена на їх суму.*

4.4 Інші види середніх величин.

Середня геометрична.

При вивченні середнього темпу зростання ознаки, що вивчається, середня арифметична не підходить. Замість неї

обчислюють середню геометричну Qg (або Xg). Її визначаємо за формулою:

$$\overline{\tilde{O}_g} = \sqrt[n]{\tilde{O}_1 \tilde{O}_2 \tilde{O}_{3\ldots} \ldots \tilde{O}_n},$$

де: Q_1, Q_2, Q_3… Q_n - темпи зростання (величини, що показують, у скільки разів збільшувалася ознака від періоду до періоду); n – число періодів.

При n>2 формулу зручніше застосовувати у логарифмічному вигляді:

$$\lg \overline{\tilde{O}_g} = \frac{1}{n}(\lg \tilde{o}_1 + \lg \tilde{o}_2 + \ldots + \lg \tilde{o}_n).$$

Якщо дані, для яких обчислюють середню геометричну, представлені різними чисельностями (n_i) у межах виділених класів (x_i), то застосовується формула:

$$\lg \overline{\tilde{O}_g} = (n_1 \lg \tilde{o}_1 + n_2 \lg \tilde{o}_2 + \ldots + n_n \lg \tilde{o}_n) / N$$

Виходячи із змісту передостанньої та останньої формули, середню геометричну називають також середньою логарифмічною, тому що її логарифм є арифметична середня логарифму складових величин. Застосування середньої геометричної пояснимо наступним прикладом. Припустимо, що на лісокультурну площу в сосняку брусничному висаджено 1-річний сіянець сосни звичайної. Виміряємо його об'єм в $см^3$ в момент посадки (1 рік), а також на 5, 10, 15 та 20 рік. Нехай ці об'єми складуть відповідно 8, 560, 2800, 11200 та 22400 $см^3$. Тоді відношення об'ємів сіянців через сукупні рівні проміжки часу (в 5 років) будуть наступні:

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

Число розглянутих періодів в нас дорівнює 4, тобто. N = 4. За формулою обчислимо середню геометричну:
$\overline{\tilde{O}_{ã}} = \sqrt[4]{70 \cdot 5 \cdot 4 \cdot 2} = \sqrt[4]{2800} \approx 7{,}274.$

Це означає, що об'єм нашого сіянця сосни звичайної від 1 до 20 років збільшується в середньому за кожний період у 7,274 рази. Дійсно, використовуючи середню геометричну, отримуємо в 20 років обсяг стволика $X_5 = 8 * 2800 = 22400$ $см^3$.

Якби ми вирахували тут середню арифметичну з наших 5 сіянців, то вона склала б 7594 $см^3$ і характеризувала б об'м стовбура у віці приблизно 15 років, що відповідає суті досліджуваного процесу, так як об'єм стовбура до кінця 4-го періоду, що дорівнює 7594 $см^3$ * 5 = 30376 $см^3$, що на 36% більше фактичних даних. Для середньої геометричної характерна рівність похідних з початкових

даних вимірювань (X_1, X_2, X_n) та з геометричних середніх X_g, X_g, X_g … X_g, представлених n раз. Згадаймо, що для середньої арифметичної величини характерна сталість суми варіант.

Середня квадратична.

Основна мета вимірювання діаметрів дерев у древостані - це визначення запасу деревини. Для обчислення запасу треба знати суму площ поперечних перерізів виміряних дерев (g_i), а потім скористатися відповідною формулою, куди Σgi входить як співмножник. Ця формула описана в курсі лісової таксації. Суму площ перерізів дерев в деревостані можна визначити, помноживши кількість дерев на площу перерізу середнього дерева. Відомо, що площа перерізу дерева на висоті 1,3 м дорівнює площі кола, і її знаходять за формулою $g_i=\pi d_i^2$, де π= 3,14159, d - діаметр на висоті 1,3 метра. Нам потрібно з'ясувати, яку величину має представляти середній діаметр, щоб він був репрезентативним показником задля досягнення зазначеної мети. Аналіз, зроблений шляхом використання тут $\overline{X}$, показує, що застосування середньої арифметичної величини у цьому разі неприйнятно, а необхідно обчислити середню квадратичну. Її знаходять за такою формулою:

$$\overline{x}_{кв} = \sqrt{\frac{1}{n}\sum x_i^{\,2} n_i} \ .$$

Як ілюстрацію сказаного розглянемо визначення об’м середнього дерева в деревостані сосни віком 100 років (II класу бонітету), який виміряний нами раніше та наведено у таблицях 3.1, 3.3. Результати розрахунків показано в таблиці 4.1. У цій же таблиці представимо вихідні дані для знаходження середньої арифметичної: $\Sigma n_i X_i / \Sigma n$.

Середній арифметичний діаметр буде дорівнювати:

$$D_a = \frac{\sum n_i x_i}{\sum n_i} = \frac{3440}{120} = 28{,}7 \text{ (см)}$$

Середній квадратичний діаметр визначаємо через середню площа перерізу, за правилами, прийнятими у лісовій таксації. Формула для його знаходження відповідає середньої квадратичної:

$$g_{ср} = \frac{D_{сд}^{\,2} \cdot \pi}{4}; \qquad D_{ср} = 2\sqrt{\frac{g_{сд}}{\pi}}$$

Середня площа поперечного перерізу дорівнює:

$$g_{ср} = \frac{\sum g_i n_i}{\sum n_i} = \frac{85016{,}3}{120} = 708{,}5 \text{ см}^2$$

Тоді:

$$D_{ср} = 2\sqrt{\frac{708{,}5}{3{,}1416}} = 2\sqrt{225{,}5} = 2 \cdot 15 = 30 \text{ см}$$

Таблиця 4.1

Визначення середнього діаметра для 120 дерев сосни звичайної в умовах пробної площі №2 Перганського ПНДВ

Ступені товщини (класи) X_i, см	Число дерев n_i, шт	Площа середнього ступня товщини $g_i=x_i^2*\pi/4$, см	g_in_i, см^2	x_in_i, см
8	1	58,27	58,3	8
12	3	113,10	339,3	36
16	8	201,86	1614,9	128
20	14	314,16	4398,2	280
24	20	452,39	9047,8	480
28	27	615,75	16625,2	756
32	17	804,25	13672,3	544
36	12	1017,88	12214,6	432
40	9	1256,64	11309,8	360
44	5	1520,53	7602,6	220
48	3	1809,56	5428,7	144
52	1	2704,78	2704,8	52
Всього: (Σ)	120	-	85016,3	3440

Як бачимо, середній діаметр, який придатний для знаходження запасу через середню площу перерізу та визначений через середню геометричну, більше, ніж середньоарифметичний. Припустимо, що середнє дерево в цьому древостані має висоту (h) 26 м, а коефіцієнт форми (f), тобто відношення об'єму стовбура до об'єму циліндра, що має з цим стовбуром однакову висоту та діаметр на висоті 1,3 м - 0,46. Об’єм дерева (V) у разі визначається за такою формулою V = ghf. При цьому всі величини слід подати в однаковому розмірі. Оскільки об’м дерева, що зазвичай знаходять у м^3, то й усі складові наведеної формули наводимо в м^3. Тоді об’єм дерева, що має середньоквадратичний діаметр (V_1) буде дорівнювати:

$$V_1 = \frac{3{,}1416 \cdot 0{,}3^2 \text{ì}}{4} \cdot 26\text{ì} \cdot 0{,}46\text{ì} = 0{,}84 \text{ м}^3.$$

При використанні середньоарифметичного діаметра об'єм дерева (V_2) складає:

$$V_2 = \frac{3{,}1416 \cdot 0{,}287^2 \text{ì}}{4} \cdot 26\text{ì} \cdot 0{,}46\text{ì} = 0{,}77 \text{ м}^3.$$

Як бачимо, в останньому випадку цей показник занижений майже на 9%. Справжній запас дослідженого деревостану (М) дорівнює добутку числа стовбурів (Σ n_i= N) об’єму середнього дерева, тобто. М = Vср*N. Для нашого прикладу: М = 0,84 м ^{3} * 120 = 101 м ^{3}. Якщо ж для розрахунків брати середньоарифметичний діаметр, то отримаємо М = 0,77 м^3*120 = 92 м^3, тобто недобір склав 9м^3. За середньої біржової ціни деревини сосни в докризові часи на світовому ринку 20 тис. грн./м^3, недобір становитиме 180 тис. грн.

Це для порівняно невеликої ділянки лісу, тобто для пробної площі 0,5-0,8 га. На 1 га недобір буде вже від 4,5 до 6 млн. грн. Таким чином, для отримання справжнього значення площі поперечного перерізу або обсягів усіх дерев за допомогою середнього дерева та числа дерев, діаметр середнього модельного дерева слід знаходити як середню квадратичну величину. У лісовій таксації саме так і роблять. Тому середня квадратична величина набула широкого поширення у лісовому господарстві.

Середня гармонійна.

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

$$\overline{X_h} = N/(\sum 1/X)$$

де: N – частка певних значень показника.

В практиці лісового господарства ця величина застосовується вкрай рідко.

Мода та медіана.

Модою (Мо) називають варіант, що найбільш часто зустрічається.

У нормально розподілених сукупностях, мода чисельно дорівнює середній арифметичній.

У позитивно асиметричних варіаційних рядах, мода більша за середню арифметичну. ($I_o > o$), а в негативно асиметричних рядах вона менше середньої арифметичної ($I_o < o$).

Знайдемо моду у прикладі, наведеному в таблиці 4.1.

X	8	12	16	20	24	28	32	36	40	44	48	52	
n_i	1	3	8	14	20	27	17	12	9	5	3	1	Σ 120

Мода в наведеному ряду складає найбільш представлений клас 28см. Середня арифметична тут x = 287 см, тобто $I_o < o$. Отже, ряд має невелику позитивну асиметрію. Загальна формула для обчислення моди (Мо) така:

$$M_o = x_{Mo} + \frac{C(n_{Mo} - n_{Mo-1})}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}}$$

де: X_{Mo} - середина модального (найбільш представленого) класу;

С – величина класу;

$n_{мо}$, $n_{мо-1}$, $n_{мо+1}$ - частоти відповідно до модального класу, попереднього модальному класу, а також класу, що йде за модальним.

Ця формула, що враховує величину класового проміжку, більше точна, тому що найбільше дерев може бути зсунуто вправо або вліво від середини класу.

Тоді мода для нашого ряду дорівнює:

$$M_o = 28 + \frac{4(27-20)}{2\cdot 27 - 20 - 17} = 28 + \frac{4\cdot 7}{54-37} = 28 + \frac{28}{17} = 28 + 1{,}7 = 29{,}7 \approx 30\,\text{см}.$$

Медіаною (Ме) називають значення ознаки, що займає середнє місце у ряді, і розподіляє всі розподіли на дві рівні за чисельністю частини.

Серед значень: 5, 6, 7, 8, 9 – медіна (Ме) = 7.

Для варіаційного ряду

$$M_e = X_o + k\,[S_1 - S_2)/n],$$

де: X_0 - значення нижньої межі класу, в якому міститься половина накопичених частот;

k - інтервал;

$S_1 = N/2$;

S_1 - накопичена частота, що передує групі в якій знаходиться медіана.

Для нашого ряду діаметрів сосни звичайної (таблиця 4.1) медіану обчислимо в такий спосіб:

X_i	8	12	16	20	24	28	32	36	40	44	48	52
n_i	1	3	8	14	20	27	17	12	9	5	3	1
$\sum n$	1	4	12	26	46	73	90	102	111	116	119	120

$$M_e = 26 + 4[(73 - 46)/27] = 26 + 4\cdot 1 = 30\ \text{см}$$

В цьому прикладі мода та медіана збіглися. Це звичайне явище для симетричні розподіли. В сильній мірі асиметричні ряди мають не збігаючі значення моди та медіани. Мода та медіана є характеристиками центральної тенденції вибірки. Вони не мають свого аналога в генеральній сукупності та тому сприймаються як показники відносного характеру.

Верхня та нижня середні.

У практиці іноді трапляються інші види середніх. Щодо до лісогосподарських об'єктів, іноді виникає потреба застосування «верхньої» чи «нижньої» середньої. Їх застосовують тоді, коли потрібно оцінити частину явища. У лісовій таксації «верхню» середню обчислюють щодо верхньої висоти, тобто коли треба знайти середню висоту дерев з найбільшим діаметром стовбура. Їх може бути 50, 100, 200 або 5, 10, 20% від всієї сукупності тощо. «Нижню» середню зазвичай обчислюють для визначення запасу, що вирубується під час проведення доглядів низовим способом. Узагальнення описаних середніх величин зроблено у таблиці 4.2, де

наведено формули їх обчислень при малому і великому числі спостережень. Пояснення символів було дано вище. Як видно з таблиці 4.2, середня геометрична (середня логарифмічна) дорівнює кореню n-го ступеня з добутку значень випадкової величини, середня гармонічна є зворотна величина середніх арифметичних величин, зворотня середня квадратична виходить як корінь квадратний із середнього значення квадрата. Як показали вчені-таксатори К.Є. Нікітін та О. З. Швиденко, всі названі види середніх можуть бути отримані з формули:

$$\overline{X} = \frac{1}{n}(x_1^a + x_2^a + \ldots + x_n^a)^{1/a}$$

Тут за а= 1 маємо X - середнє арифметичне, а=2 – квадратичне.

(X_g); а = - 1 - гармонійне (H), при а =0 - геометричне (G), причому за своєю величиною $H \le G \le X \le Xg$.

Таблиця 4.2

Основні види середніх, що застосовуються в лісовому господарстві (за К. Є. Нікітіним та О. З. Швиденко)

	Мала вибірка	**Велика вибірка**
Арифметична:	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i$
Геометрична:	$G = \sqrt[n]{x_1 x_2 \ldots x_n}$ $\lg G = \frac{1}{n}\sum \lg x_i$	$G = \sqrt[n]{x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}}$ $\lg G = \frac{1}{n}\sum n_i \lg x_i$
Гармонічна:	$H = \frac{n}{\sum \frac{1}{x_i}}$	$H = \frac{n}{\sum n_i \frac{1}{x_i}}$
Квадратична:	$\overline{X_g}\sqrt{\frac{1}{n}\sum x_i^2}$	$\overline{X_g}\sqrt{\frac{1}{n}\sum x_i^2 n_i}$
«Верхня»: $X_{i+1} \ge X_i$	$\overline{X_\rho} = \frac{1}{\rho}\sum_{i=n-\rho}^{n} x_i$	$\overline{X_\rho} = \frac{1}{\rho}\cdot\left(\sum_{i=[\rho]+1}^{k} x_i n_i - A\right)$*
«Нижня»: $X_{i+1} \ge X_i$	$\overline{X_f} = \frac{1}{f}\sum_{i=1}^{f} x_i$	$\overline{X_f} = \frac{1}{f}\cdot\left(\sum_{i=1}^{[f]-1} x_i n_i + B\right)$**

* $A = \frac{1}{2}\left[2x_{[\rho]+1} - c - \frac{c}{n_{[\rho]}}\left(\rho - \sum_{i=[\rho]+1}^{k} n_i\right)\right]\left(\rho - \sum_{i=[\rho]+1}^{k} n_i\right)$

** $B = \frac{1}{2}\left[2x_{[f]-1} + c - \frac{c}{n_{[f]}}\left(f - \sum_{i=1}^{[f]-1} n_i\right)\right]\left(f - \sum_{i=1}^{[f]-1} n_i\right)$

Нарешті, «верхні» і «нижні» середні є середніми арифметичними з *p* найбільших (X_{n-p}, X_{n-p+1}, ..., Xn) або f найменших (X_1, X_2, ..., Xn) значень випадкової величини, в окремому випадку, при *p* та f, рівним 1, отримуємо просто найбільше та найменше значення вибірки. У формулах визначення Xp і Xf передбачається,

що вихідні дані записані в порядку зростання значень Xi Для ряду розподілу, де умова Xi+1≥Xi дотримується автоматично, індекси [p] та [f] позначають номери класів ряду розподілу, які потрапляють відповідно n - p-е та f-е спостереження. Вибір того чи іншого виду середнього визначається метою роботи та характером досліджуваного явища. Наприклад, якщо ознака X_i, представляє чисельну характеристику деякого елемента сукупності (обсяг або об'ємний приріст дерева, норму виробітку бригади звалювальників лісу та ін.), а ΣXi - чисельну характеристику сукупності в цілому, то найбільш переважна середня арифметична величина X. З іншої сторони, X у певному сенсі анулює позитивні та негативні відхилення окремих значень x_i від X . Тому, якщо середня величина обчислена для спостережень, що вона несе в собі випадкові похибки, і ці похибки підпорядковуються нормальному закону розподілу (тобто можна припускати, що позитивні та від'ємні відхилення врівноважуються), то знову-таки найкраща - середня X. Якщо вплив похибок несиметричний, то часто більш придатна середня геометрична. Цей же тип застосовують у тих випадках, коли потрібно визначити швидкість зміни (наприклад темпи зростання) досліджуваного показника та ін. Наприклад візьмемо дані, наведені А. З. Швиденко, з аналізу приросту за запасом у деякому древності сосни звичайної. Нехай M_0, M_1..., M_n – запаси деревостану у віці A_o, A_1…, An. Потрібно визначити відносну середню зміну запасу деревстану в періоді від Ao до An років. Створимо величини $X_1 = M_1/M_0$, $X_2 = M_2/M_1$ ..., $Xn = Mn/M_{n-1}$, що виражають відносну зміну запасу в кожному з окремих інтервалів часу [A1 - Ao], ..., [An, A_{n-1}]. Похідні X_1, X_2, ..., Xn дає нам загальну зміну запасу у періоді [Ao, An]. Якщо в цьому періоді відносна зміна запасу за окремими періодами постійна, то має виконуватись рівність: $G^n = X_1 X_2 ...Xn$, звідки випливає необхідність використання середньої геометричної величини, тобто:

$$\Delta M = \sqrt{\frac{M_1}{M_O} \cdot \frac{M_2}{M_1} \cdots \frac{M_n}{M_{n-1}}} - 1 = \sqrt{\frac{M_n}{M_O}} - 1.$$

Іноді доцільно середню геометричну обчислювати для «засмічених» вибірок, тобто для ситуацій, коли одне або кілька вибіркових значень сильно відрізняються від більшості. Середню квадратичну (як і інші види статичних середніх) можна застосовувати у тих випадках, коли в деякій багатовимірній сукупності середній елемент встановлюють по одній (зазвичай найбільш важливій) ознаці, причому для цього елемента потрібно вказати середні за іншими ознаками. Наприклад, в однорідних деревостанах, дереву середнього (арифметичного) об'єму відповідає середній квадратичний діаметр. Це явище пояснюється просто: зміна об'єму пропорційна не зміни діаметра, а його квадрата. «Верхні»

(«нижні») середні застосовують у випадках, коли потрібно оцінити інтенсивність частини явища або процесу. Завдання застосування «верхньої» висоти - це визначення висоти *n* найбільш товстих дерев на 1 га, можуть представляти практикичний інтерес розміри певної частини найдрібніших сіянців в розпліднику (у зв'язку з досягненням стандартних розмірів) і т.д. Як статистику стану, медіану зазвичай застосовують у тих випадках, коли потрібно забезпечити мінімум абсолютної величини відхилень між вихідними даними та використовуваним показником. Наприклад логічно вимагати дотримання такої умови для показника, що відбражає середню тривалість життя дерев у різновіковому ліс. Мода представляє інтерес, коли встановлюють деякі типові властивості, явища або об'єкти: колір жолудів, врожайність насіння на плюсових деревах і т.д.

Квантіли є узагальненням поняття медіани. Якщо поділити частоти розподілу на k рівних частин, то до k=1 значення випадкової величини, що відповідає точкам розподілу, називається квантілами розподілу.

При k=2 єдиний квантіль буде медіаною, при k=4, середня з точок поділу буде медіаною, а перша та третя – нижнім та верхнім квантілями.

Аналогічно дев'ять значень випадкової величини, що ділять частоти розподілу на 10 частин, називають децілями.

Знання трьох-чотирьох квантіл дає точне уявлення про розподіл довільної величини. Квантілем X_p або X_{1-a}, що відповідає заданій ймовірності p, p = 1 - a, називають таке значення випадкової величини, для якого $p(x<xp) = p = 1-a$. Узагальнюючи все вище викладене зазначимо, що у лісогосподарській справі найбільше застосування знаходять середня арифметична та середня квадратична. Інші середні застосовують рідко.

РОЗДІЛ 5. ПОКАЗНИКИ ВАРІАЦІЇ

5.1 Варіація як явище та її джерела.
5.2 Типи варіювання.
5.3 Характеристики варіаційних рядів та їх обчислення.
5.4 Асиметрія, ексцес, коефіцієнт варіації.

5.1. Варіація як явище та її джерела.

Ряд розподілу характеризується кількома параметрами. Однинми з них є розглянуті вище середні значення. Вони характеризують ті значення варіаційно-статистичного ряду, в якому розміщені інші чисельні значення, властиві ряду розподілу випадкової величини. У той же час, існують інші важливі характеристики випадкових рядів величин. В тому числі однією з основних, є розмах (діапазон) випадкових величин від мінімальної до максимальної. Ця величина характеризує варіабельність, чи мінливість, ряду розподілу. Наведемо приклад (таблиця 5.1).

Таблиця 5.1

Ряди розподілу числа дерев сосни звичайної по діаметру на висоті 1,3 м та їх середні значення пробної площі № 2

№ класу	Ступені товщини, см (x_i)	Кількість дерев та їх середнє арифметичне значення за варіантами досліду					
		ряд 1		ряд 2		ряд 3	
		n_i	x_in_i	n_i	x_in_i	n_i	x_in_i
1	8	2	16	-	-	-	-
2	12	3	36	-	-	-	-
3	16	13	208	10	160	-	-
4	20	13	260	20	400	7	140
5	24	23	552	21	504	25	600
6	28	32	896	25	700	33	924
7	32	40	1280	35	1128	65	2080
8	36	24	864	38	1368	42	1512
9	40	19	760	30	1200	22	880
10	44	12	528	16	704	6	264
11	48	7	336	4	192	-	-
12	52	5	260	1	52	-	-
13	56	4	224	-	-	-	-
14	60	3	180	-	-	-	-
Всього, Σn_i	8	200	6400	200	6400	200	6400
Середнє значення, $\overline{x}$		-	32,0	-	32,0	-	32,0

У таблиці 5.1 наведено 3 ряди розподілу числа стовбурів сосни звичайної за діаметром. Їхня кількість однакова - по 200 дерев.

Середні арифметичні всіх рядів теж рівні – по 32 см. Але ряди суттєво відрізняються. Середні показники це виявити не спроможні. Так, середньоквадратичне значення цих рядів дорівнює відповідно 37,5; 37,3; 36,6 см, тобто практично однакові, так як різниця в 0,4 см у деревостанах з діаметрами такого розміру при проведенні вимірювань за допомогою мірної вилки у виробничих умовах лежить у межах точності вимірів. У той же час навіть наочно видно, що ряди суттєво відрізняються між собою. Перший ряд має найбільший розмах розподілу, а третій - найменший. У першому ряду крайні варіанти відхиляються від середньої на 6 класів у нижній бік та на 8 класів у верхній, а всього ряд має 15 класів, або ступенів товщини. У третьому випадку відповідні величини дорівнюють 3 та 3, а розмах ряду відповідає 7 класам. Другий ряд розподілу займає проміжне положення між 1 та 3 рядами. Щоб дати більш повну характеристику рядів розподілу випадкової величини, введено характеристику 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 та інші. Виникає закономірне питання - які причини варіації. Основна причина полягає в тому, що за своєю природою будь-які біологічні об'єкти, навіть ті, що належать до однорідної сукупності, відрізняються один від одного. Дерева одного виду, одного віку, що ростуть в однакових умовах мають різну висоту і товщину, що викликано як генетичними властивостями кожної особини, так і деякими особливостями їхнього територіального розміщення: в тіні великих дерев, на мікропідвищеннях або мікрозниженнях тощо. Подібні приклади можна навести для будь-яких біологічних об'єктів: люди навіть однієї раси та національності відрізняються за ростом, тварини одного виду мають різну вагу тощо. Основна причина цьому, як було зазначено, наявність біологічного (у першу черга генетичного) різноманіття. Мінливість відбувається і через помилки вимірів. Домогтися абсолютної точності вимірів дуже важко, а часто й не потрібно. Практику влаштовує певний рівень точності, його й витримують. Наприклад, при перерахунку дерев їх вимірюють по ступенях товщини. Точність вимірювань може підвищуватися, якщо це потрібно в окремих випадках. Так, в умовах ринку підвищуються вимоги до точності оцінки об’єму деревини, особливо для її найціннішої частини. Під час проведення вимірювань, крім допустимих допусків у вимірах, можливі помилки, особливо при масових вимірах. Хоча помилки вимірювань і є однією з причин варіювання, але ця причина не така значуща, як природна біологічна мінливість.

5.2 Типи варіювання.

Отримані в результаті спостережень значення ознаки, що спостерігаються називають варіантами.

Варіанти у біологічних об'єктах виявляють різноманітність (або варіювання) досліджуваної властивості. Наприклад, дерева відрізняються одине від одного по діаметру, висоті, об'єму, санітарному стану. Залежно від характеру ознаки, що вивчається, розрізняють варіювання: безперервне, перервне (дискретне) та атрибутивне.

Безперервне та дискретне варіювання притаманне кількісним ознакам, а атрибутивне – якісним ознакам.

При неперервному варіюванні окремі значення ознаки виражають мірою довжини, об’єму тощо. Окремі варіанти можуть мати будь-яке, але змінне у певних межах значення. Товщина дерев при варіюванні наприклад від найтоншого до найтовстішого може приймати різні значення. Тільки залежно від мети дослідження (вимірювання) виражають її у класах товщини: у кілька сантиметрів, у цілих сантиметрах, у десятих чи сотих частках сантиметра.

При дискретному варюванні окремі значення ознаки виражають абстрактними числами, найчастіше цілими. Наприклад, кількість сходів сосни звичайної на обліковій площі має дискретне варіювання, так як вони так само як і кількість насіння в наважці виражаються цілими числами.

При атрибутивному варіюванні значення ознаки виражають у якісних показниках. Це може бути ступінь забарвлення, консистенції, пошкодженість чи стійкість, і навіть форма, вигляд тощо. 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,*** у якому значення ознаки розглядають у альтернативній формі, тобто протиставляючи здорові хворим, сильні - слабким, забарвлені - незабарвленим, присутні – відсутнім тощо. В альтернативній формі можна уявити і кількісні ознаки, протиставляючи, наприклад високі дерева деревам низьким, домінуючі дерева - пригніченим, здорові - сухим або сухим.

5.3 Характеристики варіаційних рядів та їх обчислення.

Основними показниками варіаційного ряду, крім середнього значення є абсолютне та відносне значення його меж, що виражається величиною дисперсії та коефіцієнта варіації. *Одним із показників амплітуди варіації є так звані ліміти (від латинського Limes - межа, кордон), тобто значення мінімального та максимального варіанту вибіркової сукупності.* Цей показник (Lim) показує фактичні межі варіабельності ознаки. Тому його зазвичай

наводять поряд з іншими біометричними показниками в зведених статистичних даних таблиць. Значення лімітів полягає в їх конкретності. *Величина варіації може бути оцінена і по різниці між значеннями максимальної та мінімальної варіаційної сукупності. Цей показник отримав назва розмаху варіації.* Наприклад, межі першого розподілу (n_1), який щойно розглядався (таблиця 5.1, варіант 1) рівний: 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 у третьому ряду (n_3), розмах якого дорівнює 24 одиницям: 44-20 = 24.

Розглянуті показники варіації цілком об'єктивні та прості. Але через властиві їм недоліки вони мало придатні для вимірювання варіабельності ознак. Справа в тому, що ці показники нестійкі: вони залежать від багатьох випадкових причин та при повторних вибірках можуть різко змінювати своє значення. Головний недолік зазначених показників полягає в тому, що вони не відображають суттєві риси варіювання. Зі сказаного випливає, що ліміти та межі варіації, хоч і дають певне, конкретне уявлення про величину мінливості ознак, але не можуть служити основним мірилом варіабельності біологічних величин. Тому тут застосовують інші показники, що описуються нижче.

Середнє лінійне відхилення.

Для вимірювання варіації можна використати центральний момент першого порядку, як одну з характеристик варіаційного ряду, що представляє суму відхилень варіант від середньої арифметичної, віднесену до загального числа варіант цієї сукупності.

$$\Delta = \frac{\sum \left| (x - \overline{X}) \right|}{n}$$

Цей показник, що називається середнім лінійним відхиленням може мати значення лише за умови, що відхилення варіант від середньої арифметичній беруться без урахування знаків, тому що в іншому випадку Σ (x-X)=0. Використаємо цей показник характеристики взятого нами прикладу. У таблиці 5.2 показано обчислення Δ для третього ряду розподілу (n_1), що має найменший розмах. Отримане значення (Δ=1,36) необхідно порівняти з іншим подібним показником, інакше його важко оцінити. Обчислимо Δ для першого ряду (n_1), що має найбільший розмах варіації. Хід обчислення показаний у таблиці 5.3. Згадаймо, що середнє значення для наших рядів дорівнює 32 см.

Таблиця 5.2

Обчислення лінійного відхилення ряду розподілу числа стовбурів по діаметру в сосновому деревостані, третій варіант пробної площі №2

Ступені товщини (варіанти, x_i)	Кількість дере (частота, n_i)	$x_i n_i$	$\lvert x_i\rvert-\overline{x}$	$n_i(x_i-\overline{x})$	Обчислення
20	7	140	12	84	$\Delta = \frac{\sum n_i \lvert(x_i - \overline{x})\rvert}{\sum n_i} = \frac{822}{200} = 4{,}11$
24	25	600	8	200	
28	33	924	4	132	
32	65	2080	0	0	
36	42	1512	4	168	
40	22	980	8	176	
44	6	264	12	72	
Сума:	200	6400	-	822	

Таблиця 5.3

Обчислення лінійного відхилення ряду розподілу числа стовбурів по діаметру в сосновому деревостані - перший варіант

Ступені товщини (варіанти, x_i)	Кількість дере (частота, n_i)	$x_i n_i$	$\lvert x_i\rvert-\overline{x}$	$n_i(x_i-\overline{x})$	Обчислення
8	2	16	24	48	$\Delta = \frac{\sum n_i \lvert(x_i - \overline{x})\rvert}{\sum n_i} = \frac{1252}{200} = 6{,}26$
12	3	36	20	60	
16	13	208	16	208	
20	13	260	12	156	
24	23	552	8	184	
28	32	896	4	128	
32	40	1280	0	0	
36	24	864	4	96	
40	19	760	8	152	
44	12	528	12	144	
48	7	336	16	112	
52	5	260	20	100	
56	4	224	24	96	
60	3	180	28	84	
Сума:	200	6400	-	1252	

Порівнюючи перший результат (4,11) з другим (6,26), бачимо, що ряд, у якого більша амплітуда мінливості, має і більший показник варіації, що виражається за величиною лінійного відхилення. При меншому розмаху варіаційного ряду, показник варіації виявляється меншим. У той же час описаний показник має суттєві недоліки. Так,

для ряду №3 від середньої арифметичної відхиляються три класи, а в першому ряду вже вісім. Отже, амплітуда мінливості першого ряду у 2,66 рази більше, ніж третього. Якщо середнє відхилення для третього ряду дорівнює 4,11, то першої сукупності воно має бути в 2,66 разів більше, тобто. 4,11 - 2,66 = 10,9. Насправді ж цей показник дорівнює 6,26. Різниця становить 10,9 – 6,26 = 4,64 або 74,1%, тобто вона досить велика. Отже, середнє лінійне відхилення не може бути досить точним показником варіації, не кажучи вже про те, що цей показник втрачає всякий сенс якщо брати не модулі відхилень, а враховувати знаки відхилень варіант від середньої арифметичної, оскільки сума відхилень буде близька до нуля.

Середнє квадратичне відхилення.

Щоб подолати недоліки лінійного відхилення, прийнято відхилення варіант від середньої арифметичної зводити у квадрат та суму квадратів відхилень відносити до загальної кількості спостережень, тобто до обсягу вибірки. Отриманий таким чином показник є центральним моментом другого порядку, він характеризує дисперсію чи розсіювання.

Випадок при середній арифметичній. Цей показник називається дисперсією, або варіансою, позначається грецькою літерою σ² (сигма мала у квадраті)[1] і виражається такою формулою:

$$\sigma^2 = \frac{\sum(\tilde{o}_i - \bar{x})^2}{\sum n_i} = \frac{1}{\sum n_i} \cdot \sum(\tilde{o}_i - \bar{\tilde{o}})^2$$

Для безперервної випадкової величини, заданої аналітично, дисперсію знаходять за формулою:

$$\sigma^2 = \int_{-\infty}^{\infty} (x_i - \bar{x})^2 \, fx(dx)$$

При зведенні відхилень варіант від середньої арифметичної в квадрат, їх сума не перетворюється на нуль, так як і позитивні й негативні відхилення отримують один і той же позитивний знак. Крім того, великі відхилення від середньої, будучи зведені у квадрат, отримують і більшу «питому вагу», більш впливаючи на величину показника варіації. Однак, зводячи відхилення варіант від середньої арифметичної в квадрат, ми, таким чином штучно збільшуємо і показник варіації. Щоб подолати це, замість дисперсії беруть корінь квадратний із зазначеного відношення:

$$\sigma = \pm\sqrt{\frac{\sum(x_i - \bar{x})^2}{\sum n_i}} = \pm\sqrt{\frac{1}{x_n}\sum(x_i - \bar{x})^2}$$

Отриманий таким чином показник називається середнім квадратичним відхиленням.

Іноді його називають основним відхиленням, або просто (для стислості) сігмою. Знаки + та -, що поставлені перед радикалом, вказують на те, що даний показник у рівній мірі характеризує відхилення варіант від середньої арифметичної як у бік більших (+), так і у бік менших (-) значень. Надалі з метою економії, ці знаки опущені, так як мається на увазі, що вони стоять попереду цього показника. Середнє квадратичне відхилення, як і середня арифметична, відноситься до величин іменованих і виражається в тих же величинах, як і ознака. Вибірка, у якій розсіювання варіант у середньої арифметичної більше, характеризується і більшою величиною середнього квадратичного відхилення, і навпаки, за меншої варіабельності ознаки, середнє квадратичне відхилення виявляється меншим. В лісовому господарстві найчастіше використовують σ замість σ^2. Перевага середнього квадратичного відхилення проти дисперсії пояснюється практичною зручністю: у разі використання ми маємо міру розсіювання, виражену у тих самих величинах, як і середнє значення. Порівняно із середнім лінійним відхиленням, середнє квадратичне відхилення точніше характеризує варіабельність ознак. Для підтвердження, наведеного твердження обчислимо величину σ для 1 і 3 ряду розподілу числа стовбурів по діаметру, які показані в таблиці 5.1, і порівняємо її з лінійним відхиленням. Методика цього обчислення показано у таблиці 5.4. На основі даних таблиці 5.4 виконаємо обчислення за формулою:

$$\sigma_1 = \sqrt{\frac{1}{200} \cdot 21878} = \sqrt{109{,}39} = 10{,}5\,(\text{см}); \quad \sigma_2 = \sqrt{\frac{1}{200} \cdot 6240} = \sqrt{31{,}2} = 5{,}6\,(\text{см}).$$

Порівнюючи між собою σ_1 і σ_2, бачимо, що вони реальніше відображають варіацію рядів, ніж Δ_1 та Δ_2. З теоретичної статистики відомо, що варіація генеральної сукупності більше варіації вибірки, взятої з цієї генеральної сукупності, в середньому в $\Sigma Ï_i/\Sigma_{ni-1}$ разів.

Для спрощення символів позначимо $\Sigma n_i=N$. На цій підставі у формулу слід внести поправку, взявши як множник підкореневого виразу величину n/n-1. В результаті формула перетворюється так:

$$\sigma = \pm\sqrt{\frac{1}{N-1} \cdot \sum (x_i - \bar{x})^2 \cdot n_i}$$

Таблиця 5.4

Обчислення середньоквадратичного відхилення для ряду розподілу числа стовбурів за діаметром. Середнє значення (σ) дорівнює 32,0 см

Ступені товщини (класи)	Обчислення по варіантам							
	Ряд 1				Ряд 2			
	n_i	$x_i-\overline{x}$	$(x_i-\overline{x})^2$	$n_i(x_i-\overline{x})^2$	n_i	$x_i-\overline{x}$	$(x_i-\overline{x})^2$	$n_i(x_i-\overline{x})^2$
8	*2*	-24	576	1152	-	-	-	-
12	3	-20	400	1200	-	-	-	-
16	13	-16	256	3328	-	-	-	-
20	13	-12	144	1872	7	12	144	1008
24	23	-8	64	1472	25	-8	64	1600
28	32	-4	16	512	33	-4	16	528
32	40	0	0	0	65	0	0	0
36	24	4	16	384	42	4	16	832
40	19	8	64	1216	22	8	64	1408
44	12	12	144	1728	6	12	144	864
48	7	16	256	1782	-	-	-	-
52	5	20	400	2000	-	-	-	-
56	4	24	576	2880	-	-	-	-
60	3	28	784	2352	-	-	-	-
Всього:	200	-	-	21878	200	-	-	6240

Розмір (N-1) називається числом ступенів свободи. Вона показує, що в обмеженій сукупності (а будь-яка вибірка завжди має обмежений обсяг), всі варіанти свободи можуть приймати будь-які значення, крім однієї, значення якої визначається різницею між сумою решти варіант і обсягом вибірки. У таких випадках кажуть, що один варіант не має ступеня свободи. Так, якщо $\Sigma n_i = a$, то $\Sigma n_i = b$, одна з варіант дорівнює $n_k=\Sigma n_i-\Sigma n_{i-1}$.

Наприклад, якщо чотири якихось значення варіюють необмежено, їх кількість ступенів свободи 4-0=4. Коли ж варіація цих значень обмежена яким-небудь обсягом, наприклад величиною, що дорівнює 100, то три варіанти (n_1, n_2, n_3) можуть приймати будь-які значення, скажімо, 27, 16, 15, або 59, 3, 37, то четверта варіанта (n_4) буде мати тільки одне значення, а саме n_4 =100 - (27+16+15) = 100-58 = 42, або 100-(59 +3 +37) = 100-99 = 1, тобто вона немає ступінь свободи. У цьому випадку залишається лише три ступеня свободи: 4-1 = 3. У будь-якій емпіричній сукупності, завжди є один член, що не має свободи варіації. Тому число ступенів свободи для будь-якої вибірки дорівнює (N-1). У великих сукупностях різниця між N та (N-1) невідчутна, тобто вона помітно не позначається на величині варіації та середньому квадратичному відхиленні. А на вибірках малого обсягу ця різниця позначається на величині зазначених показників.

Способи обчислення середньоквадратичного відхилення.

Способів обчислення середньоквадратичного відхилення варіаційного ряду як і інших його показників (статистик) є кілька. В даний час для обчислення статистик розроблено комп'ютерні програми. Вони відразу дають результат, але дослідник, який працює з варіаційними порядами, повинен розуміти суть досліджуваного явища, знати алгоритм обчислень, що ми зараз і розглянемо. Одним з найпростіших способів є, так званий, прямий або довгий. Його раціонально використовувати для невеликої кількості спостережень, не згрупованих у варіаційний ряд певним чином. Після того, як обчислена середня арифметична, необхідно визначити відхилення від неї кожного варіанта, тобто визначити значення x_1-$\overline{X}$, x_2-$\overline{X}$, x_3-X і т.д. у квадрат і, якщо варіанти повторюються, квадрати відхилень множаться на відповідні частоти, та результати підсумовуються. Одержана сума $\Sigma\sigma^*(\sigma\text{-}Q)^2$ ділиться на загальну кількість спостережень без одиниці (n-), і з отриманого значення добувається квадратний корінь. Візьмемо для прикладу невелику сукупність та обчислимо для неї середнє квадратичне відхилення. Наприклад, враховуючи природне поновлення на 10 закладених пробних площах, ми нарахували наступне число сіянців сосни звичайної: 91, 82, 76, 65, 54, 102, 94, 78, 88, 96. Їх сума (Σn_i) дорівнює 750. Середня арифметична цих значень становитиме 750/10=75 сіянців. Обчислення σ для даного ряду показано в таблиці 5.5.

Таблиця 5.5

Обчислення середнього квадратичного відхилення для кількості сіянців на обстежених пробних площах. Середнє значення дорівнює 75 штук.

№ п/п	x_i	x_i-$\overline{x}$	$(x_i\text{-}\overline{x})^2$	Обчислення
1	91	-16	256	$\sigma^2 = \frac{1}{750} \cdot 2516 = 3{,}35$ $\sigma = \sqrt{3{,}35} = 1{,}83$ шт.
2	82	-7	49	
3	76	+1	1	
4	65	-10	100	
5	54	-21	421	
6	102	+27	729	
7	94	+19	361	
8	78	+3	9	
9	88	+13	169	
10	96	+21	421	
Всього:	750	-	2516	

При обчисленні на комп'ютері, абсолютна величина чисел не має значення. При розрахунках вручну, дуже великі або дуже дрібні

величини зручніше збільшувати чи зменшувати на 1-3 чи більше порядків, а результат відповідно коригувати.

За наявності великих сукупностей обчислення статистичних показників варіаційного ряду найчастіше здійснюють за допомогою допоміжних величин, які називають моментами.

Поняття про моменти розподілу.

Моментом називають середнє відхилення класових варіантів від середньої величини чи від будь-якого числа.

Моменти називають початковими, якщо вони обчислювалися від умовного початку, і центральними, якщо обчислювалися від середнього ряду $\overline{X}$. Початкові моменти позначають літерою m з індексами, що вказує на порядок моменту чи ступінь відхилень: m_0 - нульовий, m_1 - перший, m_2 – другий, m_3 – третій та m_4 – четвертий – початкові моменти. Це означає відповідно: середнє відхилення нульового, першого ступеня, середній квадрат, середній куб відхилень і т. При чому $m_0=1$ так як всі відхилення в нульовому ступені дорівнюють одиниці, а отже сума похідних їх на частоти повторюваності дорівнює загальному числу частот.

Центральні моменти позначають буквою μ з тими самими індексами: μ_0, μ_1, μ_2, μ_3, μ_4 й т.д. При чому $\mu_0=1$, $\mu_1=0$, що легко перевірити, користуючись даним поняттям моменту. Моменти розподілу на сьогоднішній день розраховуються з використанням табличного редактора Microsoft Excel, Пакет аналізу, тому алгоритм розрахунку моментів за складовими авріаційних рядів ми не наводимо.

5.4 Асиметрія, ексцес, коефіцієнт варіації.

Центральні моменти використовують для обчислення основних статистичних показників ряду розподілу: середнього значення (ð), середнього квадратичного відхилення (σ), коефіцієнта варіації (υ), показників асимметрії (а) та ексцесу (Е). Асиметрію та ексцес знаходять через основні моменти. Останні визначають за формулами:

$$rh = \frac{\mu h}{\sigma h}$$

де: rh – основний момент з показником h;

μh – центральний момент з показником h;

σh – основне відхилення в ступені h.

З формули випливає, що основний момент дорівнює відношенню центрального моменту того чи іншого порядку до основного відхилення у відповідному ступеню. Для розподілу 120

дерев сосни звичайної по діаметру за нашим дослідом, ми визначили $m_0 = 1$; $m_1 = 0$; $m_2 = 4{,}623$; $m_3 = 2{,}698$; $m_4 = 61{,}378$, а $\sigma = 1{,}830$. Тоді згідно обчислення за вище згаданою формулою, враховуючи значення μ_0, μ_1, отримаємо:

$$r_0 = \frac{\mu_0}{\sigma_0} = \frac{1}{(1{,}830)^0} = \frac{1}{1} = 1;$$

$$r_1 = \frac{\mu_1}{\sigma_1} = \frac{0}{1{,}830} = 0;$$

$$r_2 = \frac{\mu_2}{\sigma_2} = \frac{4{,}623}{(1{,}830)^2} = \frac{4{,}623}{3{,}349} = 1{,}380;$$

$$r_3 = \frac{\mu_3}{\sigma_3} = \frac{2{,}698}{(1{,}830)^3} = \frac{2{,}698}{6{,}128} = 0{,}440;$$

$$r_4 = \frac{\mu_4}{\sigma_4} = \frac{61{,}378}{(1{,}830)^4} = \frac{61{,}378}{11{,}215} = 5{,}473.$$

Третій основний момент (r_3) є зкошеність (зміщення) ряду розподілу та називається асиметрією, яка позначається a = r_3.

Четвертий основний момент служить для знаходження крутизни ряду розподілу, який називається ексцесом і позначається E=r_4-3. Для нашого прикладу а=0,440, Е=2,473.

Середньоквадратичне відхилення, як і середня величина ряду, є іменованою величиною, що виражається у величинах виміру ряду. Для вирішення багатьох теоретичних та практичних питань лісової біометрії потрібні відносні величини, що характеризують загальні особливості розмаху низки розподілів. *Таким показником є коефіцієнт варіації, який у літературі з лісового господарства називають ще показником мінливості таксаційних ознак лісових насаджень.* Коефіцієнт варіації є показником мінливості ознаки, що вивчається, виражений у відносних одиницях, і зазвичай записується в відсотках. Він визначається за формулою:

$$V = \frac{\sigma}{\overline{X}} \cdot 100\%$$

де: V – коефіцієнт варіації;
σ – середнє квадратичне відхилення;
X – середнє значення.

Оскільки коефіцієнт варіації не залежить від прийнятих одиниць виміру (при розподілі σ на Q одиниці виміру взаємно скорочуються), то він застосовується при порівняльній оцінці варіювання різних ознак. В лісовому господарстві це може бути

діаметр дерева, його висота, об'єм, приріст і т.д. Значення коефіцієнта варіації використовують для обчислення точності досліджень (Р) за формулою:

$$P = \frac{V}{\sqrt{N}}$$

де: N – об’єм вибірки.

Для прикладу обчислимо υ і Р за даними таблиці 5.5.

Для обчисленої кількості сіянців (таблиця 5.5) σ = 1,83 шт., N=10, середнє значення (Q) = 15.

Коефіцієнт варіації сіянців на майданчиках дорівнює^

$$V = \frac{1,83}{10} \cdot 100\% = 18,3\%$$

Тоді точність досліду складає: $P = \frac{18,3}{\sqrt{10}} = \frac{18,3}{3,162} = 5,8\%$

Для виміряних 120 діаметрів сосни значення υ і Р наступні:

Q = 28,7 см, (дивись. Розділ 4, таблиця 4.1); σ = C$\sqrt{\mu^2}$ = 8,83, де С – величина ступеня товщини (4 см); N = 120 шт.

В такому випадку:

$$V = \frac{8,83\ см}{28,7\ см} \cdot 100\% = 30,8\%$$

$$Đ = \frac{30,8}{\sqrt{120}} = 2,8\% .$$

Отримані величини варіювання діаметрів сіянців відповідають даним, наведеним у дослідженнях з лісової таксації. Точність 2,8% достатня при проведенні вимірювань у практиці лісового господарства та низька для наукових досліджень, де потрібна точність ну хоча б - 1,5-2%. З формули можна визначити необхідну кількість спостережень (N) за заданої точності (Р).

Якщо: $P = \frac{V}{\sqrt{N}}$

$$P\sqrt{N} = V;$$

$$\sqrt{N} = \frac{V}{P};$$

$$N = \frac{V^2}{P^2}.$$

Для наших 120 дерев, де υ = 30,8%, необхідна кількість спостережень за різної точності складе:

Для $P = 10\%$ $N = \frac{(30,8)^2}{10^2} = \frac{949}{100} \approx 10$

Для $P = 5\%$ $N = \frac{(30,8)^2}{5^2} = \frac{949}{25} \approx 40$

Для $P = 2\%$ $N = \frac{(30,8)^2}{2^2} = \frac{949}{4} \approx 240$ дерев.

Отже, щоб вивчити наш деревостан з прийнятною точністю, необхідно виміряти 240 стовбурів дерев. Узагальнюючи все вище викладене у цьому розділі, наведемо формули визначення статистичних показників низки розподілу через моменти:

Середня арифметична: $\overline{\tilde{O}} = M' + \overline{\tilde{N}} m_1$

Середнє квадратичне відхилення:

а) в одиницях інтервалу (дисперсія) $\sigma = \sqrt{\mu_2}$

б) в одиницях виміру $\sigma = \tilde{n}\sqrt{\mu_2}$

Коефіцієнт варіації: $V = \frac{\sigma}{\overline{\tilde{O}}} \cdot 100\%$

Показник асиметрії: $\alpha = r_3$

Показник ексцеса: $E = r_4 - 3$

Показник точності дослідження: $P = \frac{V}{\sqrt{N}}$

В наведених формулах умовні позначення(x, M’, m_1, μ_2, σ, r_3, r_4, V, a, E, C, P) вказані вище.

РОЗДІЛ 6. ФУНКЦІЇ РОЗПОДІЛУ. НОРМАЛЬНИЙ РОЗПОДІЛ

6.1 Поняття про види розподілу.
6.2 Емпіричні функції розподілу.
6.3 Функція нормального розподілу та її параметри.
6.4 Обчислення теоретичних частот кривої нормального розподілу.

6.1 Поняття про види розподілу.

Масиви випадкових величин зазвичай розподіляються не хаотично, а по деяким законам. Ці закони розподілу призводять до різних видів розподілів. Усі розподіли можна зобразити графічно, що переважно й робили у 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, так як застосування комп'ютерів зробило таку роботу простою, швидкою та доступною. У той же час ми повинні чітко розуміти суть того явища, що досліджується. Цього не досягти, якщо дослідник використовує комп'ютер тільки як користувач. Розподіли можуть бути дискретними та безперервними. Так, якщо ми розглядаємо деякий розподіл, наприклад, дерева в деревостані, що 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 тощо.

Все, що може бути виміряно або обчислено у живій природі, називають величиною постійної чи змінної.

Залежно від обставин, ці величини можуть набувати різних значень. Змінну величину вважають визначною, якщо наперед до проведення досліду можна вказати її значення.

Якщо ж в одних і тих самих умовах змінна величина може приймати різні значення, які заздалегідь вказати не можна, вона називається випадковою величиною.

Поняття випадкової величини, як і поняття випадкової події, належить до фундаментальних теоретичних ймовірностей. Випадкові величини бувають залежними та незалежними.

Випадкові величини називають незалежними, якщо ймовірність будь-якого значення однієї величини (X) не залежить від того, які значення набуває інша величина (У).

В іншому випадку ці величини залежать одна від одної і називаються взаємозалежними випадковими величинами. Наприклад, ми вивчаємо деяку ділянку лісу. Там є певний ґрунт, 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 вид дерев сьогодні тут росте. Таким чином, при розгляді системи: грунт-дерево, характеристики грунту, скажімо, відсоток фізичної глини, буде величиною незалежною. У той же час висота дерева й сам породний склад лісової ділянки визначається ґрунтовими умовами, тобто показники зростання дерев залежать від ґрунтових характеристик та є залежною величиною.

Якщо ми розглядатимемо діаметр і висоту дерева то виявимо, що зі зміною першого показника, змінюється і другий. Це величини взаємозалежні. Існують різні типи випадкових величин. Для лісівника та біолога найбільш істотне значення мають дискретні та безперервні випадкові величини. З ними ми вже зустрічалися вище під час розгляду емпіричних розподілів. Дискретна випадкова величина приймає лише окремі числові значення, якими можна вказати ймовірності, чи частоти. Безперервна випадкова величина може приймати будь-які значення в деякому заданому інтервалі. Вказати ймовірності чи частоти її окремих значень взагалі неможливо, тому вони відносяться до тих значень, які ця величина приймає в інтервалі (від - до), причому цей інтервал може бути будь-яким – як великим, так і малим.

6.2 Емпіричні функції розподілу.

В лісовій біометрії ми, як правило, маємо справу з емпіричними функціями розподілу. Це означає, що виконуючи деякі виміри випадкової величини, наприклад діаметрів дерев у лісі, отримуємо розподіл та визначаємо вид цього розподілу, керуючись графіком деяких функції розподілу. Дуже важливо знати розподіл ознак, що найчастіше зустрічаються в лісових дослідженнях. Це дозволяє застосовувати різні способи обробки вибіркових даних не наосліп, а осмислено, а також правильно оцінювати результати досліджень та спостережень, об'єктивно та точно порівнювати їх між собою.

В емпіричних розподілах, тобто тих, які спостерігаємо, проводячи вимірювання в лісі, впадає в око одна важлива особливість – переважне накопичення варіант у центральних класах та поступове зниження їх числа в міру віддалення від середньої арифметичної варіаційної ряду. Ця особливість, що становить одну з характерних рис варіювання біологічних ознак взагалі та лісогосподарських зокрема, - факт фундаментального значення, що мають широке поширення в природі. Таку картину отримуємо, якщо проаналізуємо розподіл людей по росту, диких тварин, наприклад зайців за вагою, розмір взуття у дорослих людей однієї статі і т. д. На малюнку 6.1 як класичний приклад показано розподіл розмірів чоловічого взуття серед жителів Східної Європи, що наводиться у більшості підручників з біометрії.

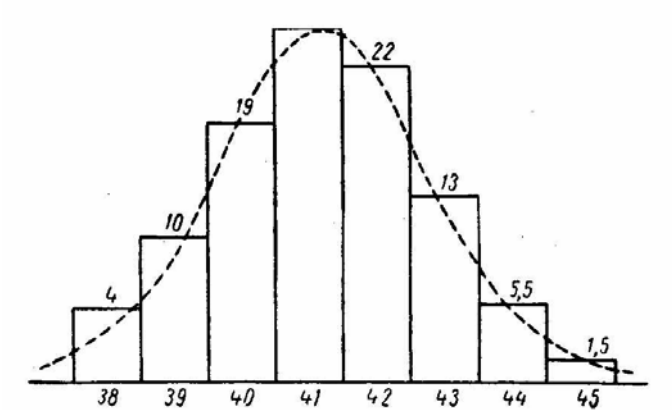


Рис. 6.1 Гістограма розподілу розмірів чоловічого взуття серед населення у Східній Європи (на осі абсцис – номери взуття, на осі ординат – відсотки)

Саме такі закономірності розподілу є основними для замовлень взуттєвим фабрикам на пошиття взуття певного розміру, щоб повністю задовольнити попит та уникнути втрат від нерозпроданих екземплярів. Таку закономірність, тобто концентрацію найбільших чисельностей в середині низки розподілу, вперше описав бельгійський статистик А. Кетле в 1835 році, який досліджував розподіл кількох тисяч солдатів американської армії за ростом.

Дерева в лісі розподіляються за товщиною, висотою та іншими ознаками аналогічно. Якщо побудуємо графік за даними перерахунку на 120 досліджуваних раніше нами ствбурів сосни звичайної то побачимо, що він одновершинний, найбільша концентрація дерев спостерігається у середніх ступенях товщини (рис. 6.2).

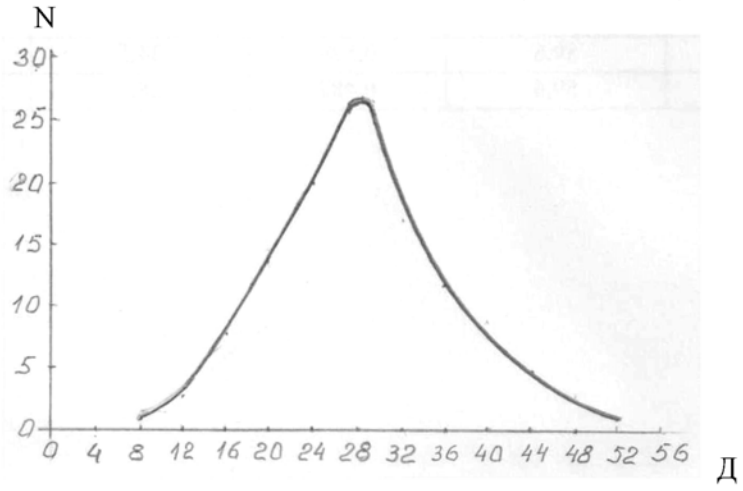


Рис. 6.2 Розподіл 120 дерев сосни звичайної в дерев віком 100 років на пробній площі Перганського ПНДВ Поліського природного заповідника.

6.3 Функція нормального розподілу та її параметри.

Вище показано, що розподіл випадкових величин в біологічних, у тому числі і лісових спільнотах носить закономірний характер. Є багато функцій, якими описують названі розподіли. Найбільш універсальною та вживаною найчастіше за інших, є рівняння кривої нормального розподілу чи *функція Гаусса-Лапласса.*

Її суть полягає в тому, що частота відхилення окремих варіантів від середньої арифметичної цієї сукупності є функція їх величини. Ймовірність частоти тієї чи іншої варіанти в генеральній сукупності й визначається цією функцією. Графічно функція нормального розподілу схожа з графіками на рис. 6.1. та 6.2. У той же час графік, який відповідає рівнянню кривої нормального розподілу, виглядає так (рис. 6.3.).

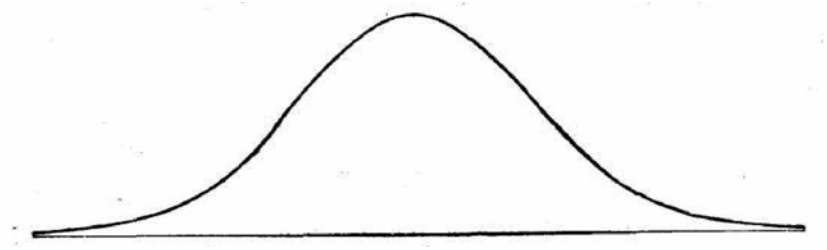


Рис. 6.3 Крива нормального розподілу

Рівняння кривої нормального розподілу виражає залежність теоретичних чисельностей *f(x)* або *y* від значень *x*, що є безперервно випадковою величиною, яка розподіляється. Є різні форми вираження цієї кривої. Основна форма написання цього виразу:

$$y = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2}$$

Тут *y* - ордината чи висота кривої будь-якій відстані від $\bar{x}$, тобто центр розподілу. Праворуч від цього центра, випадкова величина x_i має позитивні, а вліво – негативні значення σ або середньоквадратичне відхилення характеризує амплітуду коливання окремих значень випадкової величини близько середньої арифметичної ($\bar{x}_i$-x) - відхилення варіанта від середньої арифметичної, що є серединою ряду;

$\frac{1}{\sigma\sqrt{2\pi}}$ - максимальна ордината, що відповідає точці $\bar{x}$.

В міру віддалення цієї точки, тобто центру розподілу, щільність значень випадкової величини падає, і крива асимптотично наближається до осі абсцисс.

Так як π - 3,141593 і *e* - основа натуральних логарифмів дорівнює 2,7183, є постійними величинами, отже величина x-x/σ = t є не що інше, як нормальний розподіл.

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

Якщо прийняти (σ=1, то рівняння матиме такий вигляд:

$$У = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(x_i - \bar{x}\right)^2}$$

Крива, що описується цим рівнянням, отримала назву стандартизованої кривої розподілу, чи кривою Гаусса-Лапласса.

Вона виражає закон нормального розподілу з площею під кривою, що дорівнює одиниці. Щоб ордината *У* виражала не ймовірнісні, а абсолютні числа, тобто частоти варіант, потрібно в чисельник правої частини вище наведеного рівняння ввести як додаткові множники N - загальну кількість варіантів даної сукупності та *i* - величину класового інтервалу (якщо варіація розбита на класи).

Тоді рівняння набуває наступного вигляду:

$$У = \frac{N \cdot i}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}t^2}$$

Тут *У* позначає теоретичну частоту варіаційного ряду, що відповідає нормованого відхилення I.

Крива нормального розподілу має такі властивості:

•однозначно визначається двома параметрами: *X* - середнім значенням та σ - середньоквадратичним відхиленням;

•крива симетрична щодо середнього значення ($\overline{X}$) і має дзвоноподібну форму, яка залежить від величини с, що є параметром масштабу, а положення визначається $\overline{X}$;

•крива має один максимум, рівний $\frac{1}{\sigma\sqrt{2\pi}}$ і дві точки перегину на $\pm$ σ від від $\overline{X}$.;

•гілки кривої асимптотично наближаються до осі абсцис на відстань ±∞.

Отже, нормальний закон розподілу виражає функціональну залежність між величиною ознаки та 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 у цій сукупності. Отже, ймовірність відхилення будь-якої варіанти від середньої арифметичної є функція нормованого відхилення. Ця функція виражається за допомогою асимптотичних, тобто наближених формул, а також графічно (див. рис. 6.3) та у формі таблиць. Таблиця значень ймовірності, що відповідають різним значенням нормованого відхилення t наведена у додатку А. Користуючись цією таблицею, можна за двома параметрами - X та σ побудувати теоретичний варіаційний ряд, що має значення при порівняльній оцінки емпіричних розподілів. Вище було показано, що нормальний розподіл є широко поширеною закономірністю у живій природі, в тому числі й у лісових насадженнях. Такому явищу має бути певна переконлива основа. Його навів відомий математик та механік А. М. Ляпунов (1857 - 1918), довівши в 1901 граничну теорему Ляпунова, яка відноситься до галузі теорії ймовірностей. Враховуючи фундаментальність нормального закону розподілу та його велику практичну значимість, наведемо короткий виклад теореми Ляпунова в інтерпретації відомого математика А. К. Митропольського.

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 приблизно нормальним.

Умови теореми Ляпунова є настільки широкими, що у 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 практично мізерний, і якщо навіть розподіл складових величин нам невідомі, можна часто заздалегідь бути впевненим, що величина яка вивчається має нормальний розподіл. Завдяки цьому стає зрозумілим, що, наприклад, розподіл випадкових помилок при вимірах буде нормальним. Так само нормальним буде розподіл фізичних ознак людей, розподіл механічних властивостей матеріалів тощо. Цей висновок повністю підтверджується численними дослідженнями.

Таким чином, теорема Ляпунова дає пояснення тому важливому явищу, що у багатьох випадках, величини мають нормальний розподіл. Причину такої відповідності нормальної кривої отриманим під час спостереження рядів розподілу можна бачити у виконанні тих самих умов, на підставі яких з'являється теоретично нормальна крива. Можна припустити, що окремі значення є результатом незліченної множини дуже незначних незалежних між собою причин, кожна з яких може зробити дуже мале позитивне або негативне відхилення від середнього значення досліджуваної величини.

Для наочного з'ясування тих умов, за яких виникає нормальний розподіл, побудований прилад Гальтона. Цей прилад представляє скриню, зображену на рис. 6.4. Вгорі приладу влаштовано отвір у вигляді лійки, що веде у вузький простір між дошкою та склом. Внизу приладу розміщені перегородки, що утворюють кілька відділень. Весь простір між лійкою та відділеннями зайнятий рядами голок, розставлених у шаховому порядку. Якщо прилад поставити у похилому положенні і сипатимуть у лійку дрібний дріб, то голки ділитимуть цей дріб на окремі потоки. Внаслідок цього дробинки розташуються не рівномірно, а утворюють у своїй сукупності нормальну криву.

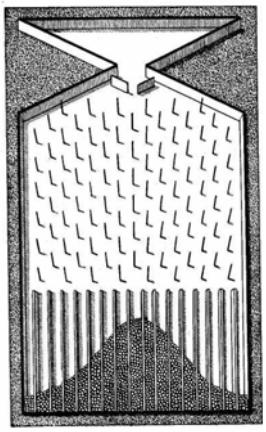


Рис. 6.4 Прилад Гальтона

Прилад, що розглядається, ясно показує ту дію, яку надає множинність причин виникнення явища. Насправді, при відсутності голок в прилад,і весь потік дробу йшов би без відхилень вниз, і в результаті дріб накупичувався б між двома — трьома сусідніми вертикальними смужками проти отвору лійки. Однак перешкоди у вигляді голок відхиляють дробинки в різні боки, причому для більшої частини дробинок ці відхилення взаємно врівноважуються, і весь розподіл приймає характерний вигляд: у відділенні проти лійки накопичується найбільша кількість дробу, а в інших відділеннях буде дробу тим менше, чим далі вони знаходяться від середнього відділення, так як для того щоб відхилитися від його далеко, необхідно зустріти багато односторонніх перешкод, а це трапляється тільки з невеликою кількістю дробинок.

Картина розподілу, подібна до описаної вище, спостерігається при вивченні випадкових величин, взятих із будь-якої галузі. Прикладів, коли застосуємо закон, що розкритий теоремою Ляпунова, безліч. Проаналізуємо дію цього закону у лісі. Так, на деревостан, зростаючий без втручання людини діє безліч абіотичних і біотичних факторів. Ресурси зростання дерева визначаються спадковими властивостями конкретного насіння, місця, де воно росте (лісорослинні умови мають певну неоднорідність навіть у межах одного таксаційного виділу), сусідніми деревами, випадковими пошкодженнями (снігом, тваринами, блискавкою і т.д.)

та багатьма іншими причинами. Кожна з причин носить випадковий характер та 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, про що буде сказано нижче. Нормальний розподіл можемо спостерігати, якщо проаналізуємо відхилення довжини реальних сортиментів, заготовлених у лісі, від заданої їх довжини. Наприклад, ми зрізаємо соснові стовбури завдовжки 6,5 м. За чинним стандартом тут допустиме відхилення у великий бік (припуск) до 1%. У нашому випадку це становитиме 6,5 см. При реальній розкрижовці через різні випадкові причини, допуски виходять неоднакової довжини. Середній допуск тут відхилятиметься від середньої величини (приблизно 4 см) на ± 3 см, а весь масив чисел, що характеризують відхилення від 6,5 см буде розподілено нормально.

Основні властивості нормального розподілу. Теоретично випадкова величина X, розподілена за нормальним законом, може приймати будь-які значення, змінюючись від - ∞ до + ∞. Насправді ж, як це видно з графіка нормальної кривої (рис. 6.3), значення ймовірності по мірі віддалення від центру розподілу X швидко зменшуються. Якщо на обидві сторони від X відкласти відрізки, рівні 3σ, вийдуть точки X-3σ та X+3σ; їх можна виразити у нормованій формі:

$$t_1 = \frac{(\bar{x} - 3\sigma) - \bar{x}}{\sigma} = -3 \text{ и } t_2 = \frac{(\bar{x}+3\sigma) - \bar{x}}{\sigma} = +3$$

Зробимо аналіз ймовірності того, що значення випадкової величини опиниться у цьому інтервалі, тобто. між X-3σ та X+3σ. Як стверджує теорія ймовірностей, знайдена ймовірність приблизно дорівнює:

$$P(x_1 \le X \le x_2) \approx \frac{1}{2} [\Phi(t_2) - \Phi(t_1)]$$

де вираз $P(x1 \le X \le x_2)$ означає ймовірність (P) того, що випадкова величина X знаходиться між заданими межами (x_1 - x) і (x_2 - x); та t_1, t_2 - нормоване відхилення, тобто: $t_1 = x_1 - x/\sigma$; $t_2 = x2 - x/\sigma$.

Символ Ф(t), читається як «фі від t", називається інтегральною функцією Лапласа.

Одна з важливих властивостей цієї функції полягає в тому, що вона прагне одиниці, якщо t необмежено зростає. Значення цієї функції, відповідні різним значенням t, і становить зміст таблиці, наведеної у додатку А.

Підставимо в рівняння взяті значення t_1 та t_2:

$$P(\bar{x} - 3\sigma) \le X \le (\bar{x} + 3\sigma) = \frac{1}{2}[\Phi(3) - \Phi(-3)] = \Phi(3)$$

Обчислення показують, що Ф(3) = 0,9973 = 1. ***Це означає, що випадкова величина, розподілена за нормальним законом, практично не відхилиться від центру розподілу X, тобто середньої арифметичної генеральної сукупності більш ніж на 3σ. Цей важливий для нас висновок відомий у біометрії під назвою «правила плюс-мінус трьох сигм».*** Наприклад, ми вивчили розподіл за вагою самців різновікової популяції дикого кабана. Це поділ буде близьким до нормального. Припустимо, що середня вага кабана (х) у цій популяції становив 100 кг, а середньоквадратичне відхилення (а) дорівнює 30 кг, тобто мінливість (Коефіцієнт варіації (V)) складе 30%. Тоді 68% кабанів матимуть вагу від 70 до 130 кг. Для 95% цієї популяції коливання становитимуть від 40 до 160 кг. При аналізі 99,9% цієї популяції, вага окремих особин може коливатися від 10 до 190 кг. Тільки дуже мала їх кількість (окремі особини) можуть мати більшу вагу, що ми й спостерігаємо у природі.

З таблиці у додатку А видно, між межами від -t до +t перебуває 0,6828, або 68,28%, усієї площі, що покрита під кривою ймовірності. Площа цієї кривої, обмежена межами від -2t до +2t, становить0,9545 частка одиниці, або 95,45%, а в межах від -3t до +3t знаходиться 0,9973, або 99,73% усієї площі нормальної кривої. Висловлюючись більш конкретним мовою, правило «плюс-мінус трьох сигм» говорить, що в межах х ±1σ знаходиться 68,28% всіх варіантів емпіричної сукупності, що розподіляється за нормальним законом; у межах х ±2σ укладено 95,45%, а в межах х ±3σ міститься 99,73% всіх варіантів сукупності. Таким чином, нормальну криву можна розділити на три ділянки або «сигмальні зони», кожна з яких містить певну кількість варіант. Межі цих «зон» приблизно збігаються з найбільш помітними вигинами нормальної кривої (рис. 6.4.).

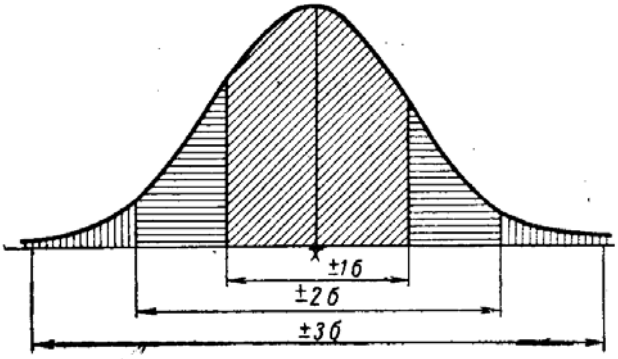


Рис. 6.4 Графічна ілюстрація «правила плюс - мінус трьох сигм»

Узагальнюючи викладене, можна сказати, що, як зазначають відомі вчені-лісівники К. Є. Нікітін та А. З. Швиденко, у процесі

статистичного аналізу лісівничої інформації, яка стосується деякої випадковій величині, теорію розподілів застосовують у двох основних напрямках:

•як основу статистичних висновків, зокрема, оцінки параметрів та перевірки статистичних гіпотез;

•як засіб та метод подання вибіркових розподілів.

У першому випадку часто основну роль відіграє нормальний закон розподілу, у другому - як модель можна застосовувати різні типи розподілів. При цьому у подібних практичних ситуаціях щодо однієї і тієї ж величини можливе використання різних теоретичних схем, що пояснюється неповнотою відповідності реальній ситуації теоретичним передумовам, необхідним для формування того або іншого розподілу, та обмеженістю обсягу вибірки. Останнє визначає наближений характер розв'язання задачі, необхідність статистичної оцінки її результатів і пояснює зазвичай застосовування термінів «апроксимація розподілів», або «підгонки».

6.4 Обчислення теоретичних частот кривої нормального розподілу.

Обчислення вирівнюючих частот за допомогою нормальної кривої розподіли починають із знаходження статистик імовірнісного ряду: Q, σ, a, E. Потім аналізуємо величини асиметрії та ексцесу. Теоретично для кривою Гаусса-Лапласса вони дорівнюють 0. Але на практиці таке спостерігається рідко. У той же час, якщо a та E відносно невеликі, то апроксимацію виправдано робити з допомогою нормального розподілу. За відносної асиметрії ($a \geq 0$), розподіл скошено вліво, тобто довша гілка кривої розташована зліва і навпаки - при $a \leq 0$ розподіл скошено праворуч. Знак показника ексцесу (крутизни) ряду розподілу характеризує ступінь зосередження частот у центральній частині розподілу. При $E>0$ вершина кривої буде вища і гостра, тобто більше варіант зосереджено в центральних розрядах і навпаки, при $E<0$ крива виглядає більш полгою. Для оцінки можливості використання нормального розподілу, коли a і e не дорівнюють 0, можна оцінити, використовуючи такі ознаки. Віднесення розподілу до нормального можна оцінити за допомогою обчислення t-критерія Стьюдента для a та E й порівняння його з табличними значеннями цього критерію за певної кількості ступенів свободи V n = N-1 (додаток E). Зазвичай нормальний розподіл застосовують, якщо модуль $a \leq 0.3$ та $E \leq 0,4$. Розрахунок теоретичних частин (n_i) проводять за формулою:

$$(n_i) = \left[(cN)/\left(t\sqrt{2\pi}\right)\right]e^{-\frac{t}{2}}$$

де: N - обсяг ряду розподілу, c – величина інтервалу; t - нормоване відхилення класових варіантів x_i від середнього значення X.

Інші позначення (t, π, e) показані у формулах, що наведено вище.

При розрахунках за допомогою моментів $t = (x_k - m_1)/\bar{t}$ де:

Xk - мови відхилення;

m_i - перший початковий момент;

t - середнє квадратичне відхилення в одиницях інтервалу:

$$\bar{t} = \sigma i = \sqrt{\mu^2}$$

Для класу, в якому спостерігається найбільша кількість досліджуваних величин там де $x_i = x$, а t = 0, теоретична частота дорівнює:

$$n_0 = [(cN/\sigma)] * 0.39894 * e^0 = (0.4 * c * N)/\sigma$$

Множник f(t) називається функцією нормованого відхилення. Ця функція показує значення ймовірностей розподілу величини x_i в залежності від t. Значення $f(t)=0,39894 * e^{-\frac{t^2}{2}}$ одержуємо зі спеціальних таблиць (додаток Б). В результаті знаходять наступний робочий вид рівняння кривою нормального розподілу:

$$n = [(c * N) * \sigma] * f(t)$$

Приклад розрахунку вирівнюючих частот по кривій Гаусса-Лапласса наведено у таблиці 6.1.Обчислення X, σ, a, E і m зроблено за формулами, наведеними у розділі 5.

Наведений в таблиці 6.1 ряд розподілу характеризується високою концентрацією у кількості в середніх ступенях товщини, про що свідчить значний від'ємний ексцес E=-0,51. Ряд трохи зкошений (зсунутий) праворуч, тобто праворуч від середнього значення налічується 75 дерев, а справа лише 72 про що свідчить невелика величина a = -0,07.

Наведений розподіл характерний для пристигаючих та стиглих деревостанів, яких незначною мірою торкнувся антропогенний вплив. Дерева сосни звичайної, що виростають на низинних болотах відносяться саме до таких насаджень. Для інших лісових насаджень характерні скошені (зсунуті) розподіли. Біометричні методи, вивчення їхньої будови, тобто розподілу числа стовбурів за деякими таксаційними показниками, наводяться нижче.

Таблиця 6.1

Обчислення вирівнюючих частот кривої нормального розподілу для деревостану сосни звичайної у віці 60 років на прикладі Перганського ПНДВ

x_i ступені товщини	n_i кількість	x*k	$\frac{x*k-m}{\sigma}$	f(x)	$\tilde{n}_j$	Заокруглення $\tilde{\tilde{n}}_i$
8	2	-5	-2,459	0,0196	1,74	2
12	5	-4	-1,967	0,0570	5,05	5
16	15	-3	-1,476	0,1392	12,32	12
20	22	-2	-0,984	0,2468	21,85	22
24	28	-1	-0,443	0,3530	31,25	31
28	33	0	0	0,3989	35,5	35
32	31	1	0,443	0,3530	31,25	31
36	23	2	0,984	0,2468	21,85	22
40	15	3	1,476	0,1392	12,32	12
44	5	4	1,967	0,0570	5,05	5
48	1	5	2,454	0,0196	1,74	2
Всього:Σ	180	-	-	-	179,74	180

Графічно наведений в прикладі розподіл зображено на рис. 6.5.

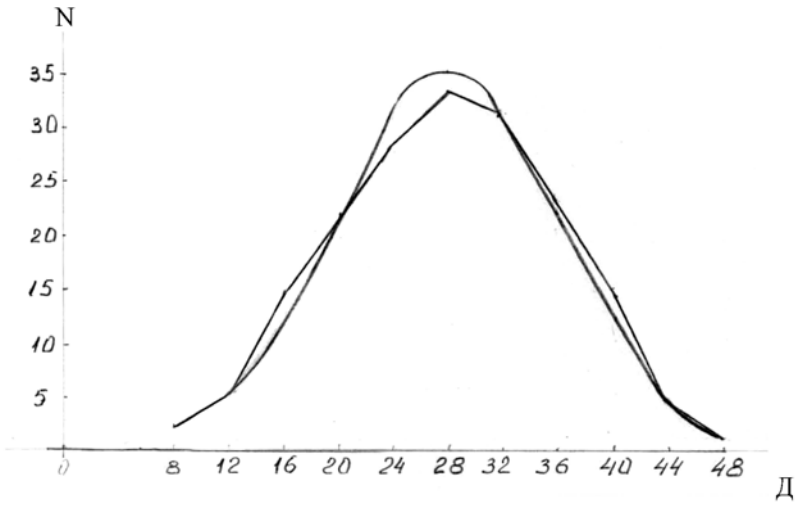


Рис. 6.5 Експериментальний та вирівняний розподіли древостані сосни звичайної у віці 60 років на прикладі пробної площі №2 Перганського ПНДВ

РОЗДІЛ 7. БІНОМІАЛЬНИЙ РОЗПОДІЛ. РОЗПОДІЛ ПУАССОНА

7.1 Поняття про біноміальний розподіл.
7.2 Біноміальний розподіл як прояв подій із двома наслідками.
7.3 Розподіл Пуассона як окремий випадок біноміального розподілу.
7.4 Обчислення вирівнюючих частот біноміального та пуасонівського розподілів.

7.1 Поняття про біноміальний розподіл.

Нормальний розподіл поширений у природі. Але варіаційні ряди розподілу різних біологічних і лісівничих об'єктів дуже різноманітні, й не можуть бути описані тільки нормальним розподілом. Хоча більшості варіаційних рядів властива велика кількість варіант в середині цього ряду та зменшення їх по мірі віддалення від центру кривих, що описують це явище. Однім із досить поширених видів подібних розподілів є біноміальний. Наприклад, зробимо аналіз появи сходів на пробній площі, де створено лісові культури сосни звичайної. Припустимо, є ділянка шириною 100 м, де було 50% сосни звичайної і 50% берези повислої, що у лісівництві записується як формули складу 5Сз5Бп. Стіна лісу, що оточує облікову ділянка має такий самий породний склад. По ділянці плугом ПКЛ-70 нарізали борозни, і наступний рік з'явився самосів сосни звичайної та берези повислої. Нас цікавить яка частка сосни звичайної у загальній кількості природного поновлення. Для цього заклали 150 облікових пробних площ розміром 1х1 м, на яких врахували всі сходи сосни звичайної та берези повислої. На кожному майданчику обчислили частку сосни у відсотках. Для зручності частку сосни звичайної на площі округлим до 10%. Результати показані у таблиці 7.1. Розподіл ймовірностей, показаний у таблиці 7.1, описується біноміальною кривою. Дамо пояснення цьому терміну. Якщо ймовірність появи самосів сходів сосни звичайної на облікових пробних площах виразити графічно, то отримаємо варіаційну криву або полігон розподілу ймовірностей (рис. 7.1). Крива рис. 7.1. носить назву біноміальної, тому що її чисельності відповідають розкладанню бінома:

$$(a+b)^n$$

З таблиці 7.1. видно, що для отримання ймовірнісні чисельності різних результатів при деякій їх кількості N, потрібно ймовірності x помножити на N. Сума ймовірностей завжди дорівнює 1. У прикладі N - це загальна кількість майданчиків, тобто 150. Оскільки кількість варіантів у нас 11, то маємо справу з біномом $(a+b)^{11}$. Теоретичне та практичне розкладання цього виразу буде викладено нижче.

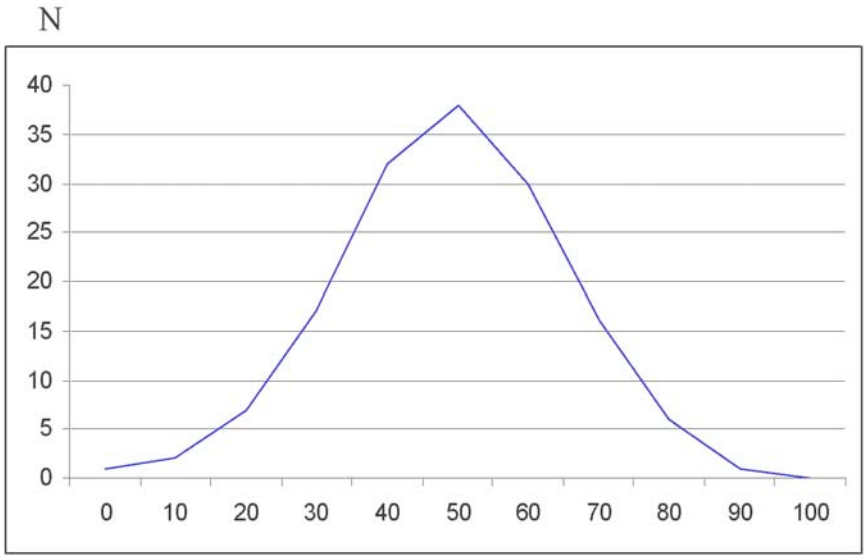


Рис. 7.1 Полігон розподілу ймовірностей появи сходів сосни звичайної на облікових майданчиках пробних площ під час обліків природного поновлення

7.2 Біноміальний розподіл як прояв подій із двома наслідками.

Біноміальний розподіл зазвичай застосовують, коли необхідно провести дослідження подій, які можуть наступити чи вже наступили. Сутність біноміальної кривої пояснимо на такому прикладі. Відомо, що кількість річних простів у стовбура сосни звичайної віком 100-120 років приблизно однакова. Тут ми не розглядаємо аномальні місцевості, де пройшли наприклад лісові пожежі. Візьмемо середній вік стиглої сосни звичайної 100 років, де співвідношення закладених ранніх та пізніх приростів у розрізі стовбура вважатимуться рівним 1:1.

Припустимо, ми відбираємо при допомозі бура Пресслера приростні керни в деревостану сосни звичайної віком 100 років, II класу бонітету, середньої висотою 18 м, середнім діаметром 64. Всі відібрані нами керни в своїй структурі мають ранній та пізній приріст, що об'єднують пару приростів і формують річне кільце. Ці пари можуть мати такі варіанти: ранній+пізній приріст (РП), ранній приріст (Р), пізній приріст (П), пізній + ранній приріст (ПР). Ймовірність раннього приросту позначимо літерою *a*, а пізнього - b. Ймовірність відбору раннього та пізнього приросту одночасно в структурі керна однакова, тобто a = b = ½. Ймовірність появи один за одним ранніх приростів або двох пізніх приростів відповідно до теорії ймовірності дорівнює a x a = a^2 або bxb= b^2.

Таблиця 7.1

Ймовірність появи сходів сосни звичайної на облікових майданчиках пробних площ

Частка сходів сосни звичайної в % від загальної кількості самосіву на обліковій площ, Хі		0	10	20	30	40	50	60	70	80	90	100	Σ
Кількість мйданчиків, (фактично n_i)		1	2	7	17	32	38	30	16	6	1	0	150
Вириняна, (теоретична кількість) майданчи -ків	без округлення	0,15	1,46	6,59	17,58	30,76	36,55	30,76	17,58	6,59	1,46	0,15	150
	з округленням	0,1	7	18	31	36	31	18	7	1	1	0	150
Ймовір ність	фактичне	0,07	0,013	0,047	0,113	0,213	0,253	0,200	0,107	0,040	0,07	0	1
	теоретичне без заокруг-лення	0,01	0,010	0,044	0,118	0,205	0,244	0,205	0,118	0,044	0,010	0,001	1
	теоретичне з заокруг-ленням	0	0,007	0,047	0,120	0,207	0,240	0,207	0,120	0,046	0,007	0	1

У нашому випадку вона дорівнює 0,5 2 = 0,25, тобто це один випадок з 4. Поєднання появи один за одним раннього та пізнього приростів дорівнює ab + ab = 2ab. Таким чином, розглядаючи ймовірність появи двох рівноймовірних подій, отримуємо їх наступний розподіл:

$$(a+b)^2 = a^2 + 2ab + b^2$$

Розглядаючи 3,4....n поєднань різних рівноймовірних випадків, приходимо до висновку, що їх ймовірності описуються вищенаведеною формулою, тобто біном Ньютона. Звідси розподіл отримав назву біноміального.

Наведимо докладний опис бінома Ньютона. Останній виражається формулою:

$$(a+b)^n = a^n C_n^1 a_b^{n-1} + C_n^2 a^{n-2} b^2 + \ldots + C_n^k a^{n-k} b^k + \ldots + C_n^{n-1} a b^{n-1} + C_n^n b^n$$

де: C_n^k – кількість комбінацій з n елементів по k.

$$C_n^k = \frac{A_n^k}{P_k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{1*2*3\ldots k} = C_n^{n-k}$$

Наприклад, при n=8, k=5, рівняння буде мти вигляд:

$$C_8^5 = \frac{A_8^5}{P_5} = \frac{8*7*6*5*4}{1*2*3*4*5} = 56$$

$$C_8^5 = C_8^3 = \frac{8*7*6}{1*2*3} = 56$$

Після підстановки виразів, отримуємо робочу формулу обчислення бінома Ньютона:

$$(a+b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{1*2} a^{n-2}b^2 + \ldots + \frac{n(n-1)\ldots(n-k+1)}{1*2*\ldots*k} * a^{n-k}b^k \ldots + nab^{n-1} + b^n$$

Наприклад:

$$(a+b)^5 = a^5 + C_5^1 a^4 b + C_5^2 a^3 b^2 + C_5^3 a^2 b^3 + C_5^4 a b^4 + C_5^5 b^5 =$$
$$= a^5 + 5a^4 b + 10a^3 b^2 + 10a^2 b^3 + 5ab^4 + b^5$$

Коефіцієнти розкладання бінома Ньютона можна отримати за допомогою трикутника Паскаля. В ньому величини будь-якого ряду є сумою двох цифр, що розташовані вище обчислювального ряду, що видно на рис. 7.2.

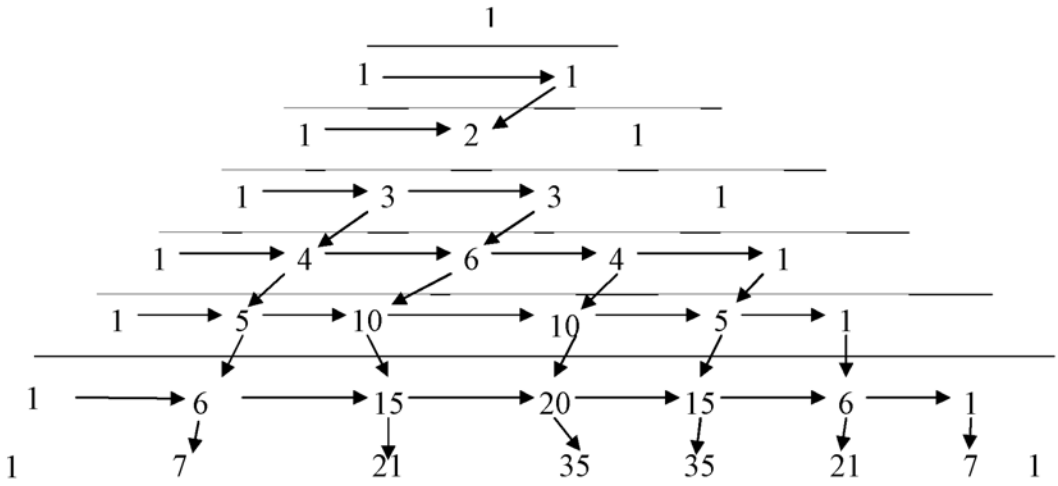


Рис. 7.2 Трикутник Паскаля

Зі шкільного курсу відомі квадрат і куб суми двох чисел, що є окремим випадком загального розкладання бінома Ньютона. Далі розподіл здійсюється в наступному порядку:

$$(a+b)^1 = a+b$$

$$(a+b)^2 = a^2+2ab+b^2$$

$$(a+b)^3 = a^3+3a^2b+3ab^2+b^3$$

$$(a+b)^4 = a^4+4a^3b+6a^2b^2+4ab^3+b^4.$$

Крива на рисунку 7.1. носить назву «кривої біноміального розподілу», оскільки відповідає розкладу бінома Ньютона. У розглянутому варіаційному ряді, як і при нормальному розподілі, який описано в розділі 6, більшість варіант розміщуються поблизу центральної частини полігону ймовірностей. Біноміальний розподіл, як сказано вище, використовується для математичного опису подій, де є ймовірність протилежних подій a та b.

Розглядаючи цей розподіл із загальних позицій теорії ймовірностей у припущенні, де ймовірність появи або відсутності деякогої події дорівнює 1,0, - тобто a + b = 1, ця проста схема називається схемою Бернуллі.

Вона описана відомим швейцарським математиком Я. Бернуллі (1654 – 1705). Опублікована була лише у 1713 р. Зі схеми Бернуллі випливає наступна однойменна теорема: «В умовах схеми Бернуллі ймовірність $a_{n,m}$ того, що подія F з'явиться у всіх *n* випробуваннях дорівнює $_m$ разів, дорівнює коефіцієнту при t^m у розкладанні бінома $(at+b)^n$ за ступенем 1:

$$a_{n,m} = C_n^m d^m b^{n-m} = \frac{n!}{m!(n-m)!} d^m b^{n-m}$$

де: a=1-b; C_n^m - число комбінацій з n елементів по m; ! – знак факторіала, тобто: a,b,c,…..,n! = a*b*c*….*n!

Наприклад: 5!=1*2*3*4*5=120.
Теорема справедлива для m=0 тобто m=n.
Вище ми розглянемо випадок рівноймовірних подій (випадкова зустріч лише раннього або лише пізнього приростів) тобто коли a = b = 0,5. Але при проведенні біометричних досліджень так трапляється далеко не завжди. Тому треба зробити аналіз інших варіантів. При біноміальному розподілі $(a+b)^2$, (a і b - ймовірності прояву (непрояву) деякої події або ознаки, щодо числа класів), можливі різні значення a і b, наприклад: a = 0,6 та b = 0,4 або a = 0,2 і b = 0,8 і т. д. При цьому змінюється форма полігону розподілу. В міру збільшення відмінностей між a і b, полігон стає все більш зкошеним, асиметричним. Однак в міру збільшення N навіть за значної різниці між a і b, ступінь симетрії полігону знову посилюється. Як і для інших розподілів, параметрами для біноміального розподілу є середня арифметична (x) та середня квадратична відхилення (σ), які можна визначити за допомогою наведених вище формул (розділ 4) для будь-якого конкретного емпіричного ряду. Теоретично їх значення визначаються значеннями ймовірностей a та b, і навіть значенням Σn, тобто числа незалежних подій, розподіл яких вивчається. Середня арифметична при цьому розподілі $\overline{\times} = k*a$ й середнє квадратичне відхилення $\sigma = k*a*b$ де k – показник ступеня бінома k=N-1.
Ці формули дають можливість зв'язати певні X та σ, обчислені на основі даного конкретного матеріалу, з ймовірностями a та b.
Вище сказане пояснимо прикладом.
Нехай на деякій ділянці було посіяно жолуді дуба в ділянки по 4 шт. Через 5 років провели облік сіянців, що вижили. Для цього підрахували кількість дерев на кожній ділянці, а всього в облік включили 100 ділянок. Результати показані в таблиці 7.2.

Таблиця 7.2

Розподід сіянців дуба звичайного, що прижились на 100 ділянках

Кількість сіянців дуба звичайног, що прижились, x_i	Кількість ділянок з сіянцями, n_i	x_i*n_i	$x_i^2*n_i$
0	7	0	0
1	24	24	24
2	37	74	148
3	26	78	234
4	6	24	96
Всього (Σ)	100	200	502

Обрахуємо x та σ за загально прийнятою методикою:

$$\overline{X} = \frac{\sum x_i n_i}{\sum n_i} = \frac{200}{100} = 2$$

$$\sigma = \sqrt{\frac{\sum x^{\alpha} n_i - \frac{\left(\sum x_i n_i\right)^2}{\sum n_i}}{\sum n_i}} = \sqrt{\frac{502 - 400}{100}} = \sqrt{1.02} = 1.01$$

Оскільки у нас збереглося 200 сіянців із 400, то ймовірність обліку екземплярів, що збереглися і відпали однакова, тобто 20/400 = 0,5, тобто a=b=0,5. Обчислимо X та σ за вище наведеними формулами де k=4.

$$\overline{X} = k * a = 4 * 0.5 = 2;\ \sigma = \sqrt{k * a * b} = \sqrt{4 * 0.5 * 0.5} = \sqrt{1} = 1.$$

Величини x і σ, що обчислені різними способами, майже однакові, але повного збігу немає. Це сталося тому, що ряд в експерименті (таблиця 7.2) дещо зкошений. Якщо його вирівняти та отримати теоретичні величини, то за наявності суворого біноміального розподілу значення x та σ, що обчислені різними способами збігатимуться. Оскільки під час проведення досліджень ніколи немає впевненості, що експериментальний розподіл строго відповідає теоретичному, то надійніше обчислити x і σ безпосереднім способом, тобто за схемою, поданою в таблиці 7.2. У наведеному прикладі виходимо з того, що a=b=0,5, тобто вони будуть точно встановлені теоретичні можливості. Але можливі і інші значення ймовірностей a та b. Наведемо приклад, описаний відомим вченим в галузі біометрії П. Ф. Рокицьким. Так, наприклад, було отримано такий фактичний розподіл самок у 103 приплодах з 4 мишенятами в кожному приплоді (таблиця 7.3).

Таблиця 7.3

Розподіл самок в приплоді мишей

Кількість самок	0	1	2	3	4
Кількість приплодів	8	32	34	24	5

Тоді: $\overline{\times} = \frac{8 \cdot 0 + 1 \cdot 32 + 2 \cdot 34 + 3 \cdot 24 + 3 \cdot 24 + 4 \cdot 5}{103} = 1{,}864$

Але так як: $\overline{\times} = kp,\ a\ k = 4,\ то\ p = 1{,}864 / 4 = 0{,}47.$ Це ймовірність появи самок. Ймовірність появи самців: q=0,53.

Виходячи з формули $\sigma^2 = kpq$,

можна обчислити $\sigma^2 = 4\ .\ 0{,}47\ .\ 0{,}53 = 1{,}0.$

Так як даний ряд є ряд розкладання бінома $(0,54 +0,47)^4$ при n=103, то легко обчислити, скільки особин слід очікувати у кожному класі. Вийдуть такі цифри для частот кожного класу (таблиця 7.4).

Таблиця 7.4

Очікувана кількість самок у приплоді мишей

Кількість самок	0	1	2	3	4
Кількість приплодів	8	29	38	23	5

У загальному вигляді, біноміальний розподіл описується формулою Бернуллі. Її доведення наведено вище, а саму формулу повторимо ще раз.

$$P_{m,n} = C_n^m P^m (1-P)^{n-m} = \frac{n!}{m!(n-m)!} P^m (1-P)^{n-m}$$

Тут: P - ймовірність настання події під час проведення n незалежних випробувань; m - число подій, що настали; C_n^m - число поєднань із n елементів по m.

Формула Бернуллі має дуже важливе значення в теорії ймовірностей, оскільки вона пов'язана з повторенням випробувань в однакових умовах. Узагальнюючи викладене, приходимо до наступного висновку. Альтернативні (тобто протилежні), дискретно варіюючі ознаки розподіляються так, що ймовірні чисельності їхньої появи можуть бути знайдені за формулою бінома Ньютона.

$$N(a+b)^n = N\left(a^n + na^{n-1}b + \frac{n(n-1)}{1\cdot 2}a^{n-2}b^2 + \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}a^{n-3}b^3 + \cdots b^n\right)$$

де: де n – число незалежних результатів в одному випробуванні;
a - ймовірність сприятливого результату одного випадку;
b - ймовірність несприятливого результату;
N – загальна кількість випробувань (результатів);
Так, при n=5 можливі 2^5=32 результати. При рівній ймовірності альтернатив, тобто: a=b=0,5 за отримаємо такі числові величини:

$$32(0,5+0,5)^2 = 32\left(0,5^5 + 5*0,5^4*0,5 + \frac{5*4*0,5^3 0,5^2}{1*2} + \frac{5*4*3*0,5^2*0,5^3}{1*2*3} + \frac{5*4*3*2*0,5*0,5^4}{1*2*3*4} + 0,5^5\right) =$$

$$= 32 = 32*(0,03125 + 5*0,03125 + 10*0,03125 + 10*0,03125 + 5*0,3125 + 0,3125) = 32*(0,03125 + 0,15625 +$$

$$+ 0,3125 + 0,3125 + 0,03125) = 32*(0,03125*2 + 2*0,15625 + 2*0,3125) = 32*(0,06250 + 0,31250 + 0,62500) =$$

$$= 32*1 = 32$$

Відкладаючи значення числа сприятливих наслідків «m» по осі абсцис, а по осі ординат ймовірні чисельності «т», отримаємо

багатокутник чисельностей розподілу. ***Ламана лінія, що з'єднує точки на графіці називається кривою розподілу.***

Як сказано вище, ймовірність події у, яка проявляється «m» разів в «n» незалежних випробуваннях описується формулою Бернуллі. Біноміальний розподіл визначається двома параметрами: середньою величиною μ=N*a та дисперсією σ^2=nab або середньоквадратичним відхиленням a=√nab. В нашому випадку;

$$\overline{\sigma} = N * a = 5 * 0.5 = 2.5; \quad \sigma^2 = 5 \cdot 0{,}5 \cdot 0{,}5 = 1{,}25.$$

7.3 Розподіл Пуассона як окремий випадок біноміального розподілу.

У лісогосподарській науці та практиці часто трапляються випадки, коли з двох (або більше) явищ, що спостерігаються, одне зустрічається рідко. Наприклад, природне поновлення на лісосіці, де зрубаний деревистий має склад 6Бп30с1Сз (60% берези повислої, 30% осики та 10% сосни звичайної), йде в основному за рахунок самосіву: насіннєвого для Сз, Бп, Ос і одночасно порослевого у берези повислої та осики. Дерева сосни тут рідкісні, а до 4-5 років, якщо не вести догляд, їх практично повністю заглушить береза та осика. Тому ймовірність зустріти через 6-7 років на такій лісосіці дерева сосни звичайної дуже мала.

В умовах радіоактивного забруднення великих територій після Чорнобильської катастрофи збільшилася кількість різноманітних мутацій серед молодих дерев. Ці мутації виражаються у зміні форми хвої, викривлення пагонів і т. д. Але їх частота (у всіх зонах радіоактивного забруднення в межах Чорнобильського сліду), хоч і залежить від рівня радіоактивності, але за абсолютним значенням мала і належить до рідкісних явищ. Подібні приклади рідкісних явищ часто зустрічаються у фізиці, біології та інших науках. Для опису названих та інших подібних розподілів зазвичай застосовують розподіл Пуассона.

Розподіл Пуассона. Розподіл Пуассона, або пуассонів розподіл, подібно до біноміального, відноситься до дискретної або переривчастої мінливості. Він має самостійне значення, хоча його можна розглядати і як граничний випадок біноміального. При біноміальному розподілі значення a і b можуть бути близькими одине до одного, при пуассоновому, дуже мало, - тобто події здійснюються дуже рідко, а b наближається до 1.

Розподіл окремих рідкісних спостережень є при цьому частіше всього асиметричним, але симетрія зростає зі збільшенням X. При збільшенні *a* розподіл наближається до біноміального. Пуассоновий розподіл характеризується по суті лише одним параметром – середньою арифметично X, оскільки σ^2 в цьому випадку зазвичай дорівнює X чи близька їй за значенням. Саме з цієї

рівності X та σ^2 найлегше визначити, що даний розподіл є пуассоновим.

Середня арифметична для пуассонового розподілу дорівнює *na*, де: a - ймовірність виявлення даної ознаки;

n - кількість фактично проведених спостережень.

Згадаймо, що величина a може бути дуже малою.

$$\overline{\times} = \lambda = na = \sigma^2$$

У формулі: середнє значення показано як X та λ, так як більшість дослідників та авторів підручників у розподілі Пуассона X позначають як λ. Частота розподілу Пуассона виражена наступним рядом:

$$n/e^{\lambda}; \frac{n\lambda}{e^{\lambda}}; \frac{n\lambda^2}{2e^{\lambda}}; \frac{n\lambda^3}{(2)\ (3)e^{\lambda}}; \frac{n\lambda^4}{(2)\ (3)\ (4)e^{\lambda}}$$

(нульова складова)

В цьому випадку:

n - загальна кількість варіантів, e - основа натурального логарифму (2,71828...) та λ -середня арифметична.

Конкретні пуассонівські ряди є кінцевими через їх обмеженість кількості спостережень. Але теоретично вони можуть продовжуватись до нескінченності. Таким чином в загальному вигляді розподіл Пуассона описується формулою:

$$y = C_n^m p^m q^{n-m} = \frac{n(n-1)\ldots(n-m+1)}{m!} * \frac{\lambda}{n^m}\left(1-\frac{\lambda}{n}\right)^{n-n}$$

де: $\lambda = np\ ;\ p = \lambda / n$.

Оскільки чисельник першого дробу має m співмножників, а знаменнику стоїть n^m, то кожен із співмножників можна розділити на n. Отримаємо вираз:

$$y = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{m-1}{n}\right) * \frac{\lambda^m}{m}\left(1-\frac{\lambda}{n}\right)^{n-m}$$

При $n\to\infty$ межа будь-якого дробу $(1-\lambda/n)$ є 1, а межа $(1-\lambda/n)^{n-m2} = e^{-\lambda}$

За цих умов $y = (\lambda^m/m!)*e^{-\lambda}$.

Вираз $y = (\lambda^m/m!)*e^{-\lambda}$ називається функцією розподілу ймовірностей у розподіл Пуассона.

У цьому виразі m - частота очікуваної події в n випробуваннях, що складає e=2,7183; параметр $\lambda = np$ відповідає математичному очікування або найімовірнішій частоті події, тобто μ, а також дисперсії σ^2. Доведення цієї рівності тут опускаємо. Вона

розглядається у багатьох підручниках та навчальних посібника з практичної статистики.

7.4 Обчислення вирівнюючих частот біноміального та пуассонівського розподілів.

Обчислення вирівнюючих (теоретичних) частот біномного розподілу проводять у такий спосіб.

- знаходять величину k, тобто ступінь бінома: k =N-1;
- проводять розподіл бінома за формулою Ньютона. Тут зручніше скористатися наведеним вище трикутником Паскаля.
- за розподілом бінома Ньютона обчислюємо (i) ймовірності для кожного значення i=N*a*b, де N – об'м ряду розподілу ($\Sigma x_i n_i$);
- чисельності (n_i) знаходимо, помноживши знайдені ймовірності для кожного x_i, на N, тобто $n = \Sigma x_i * n_i * a * b$.

Розглянемо обчислення вирівнюючи частот на прикладі (таблиця 7.3), взявши вихідні дані із таблиці 7.2.

Таблиця 7.3

Обчислення вирівнюючих частот біномінального разподілу для приживлених 5-річних сіянців дуба звичайного

Кількість сіянців дуба звичайног, що прижились, x_i	Кількість ділянок з сіянцями, n_i	Ймовірності a*b*k	$\tilde{n} = \Sigma n_i * a * b$		$x_i * \tilde{n}_i$	$x_i^k * \tilde{n}_i$
			без заокруглення	з заокругленням		
0	7	0,0625	6,25	6	0	0
1	24	0,25	25,0	25	25	25
2	37	0,375	37,5	38	35	152
3	26	0,25	25,0	25	75	225
4	6	0,0625	6,25	6	24	96
Σ	100	1,0	100	100	200	498

Розподіл $(a+b)^4$ описується наступним рівнянням:

$$a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$$

Тоді:

$$0.5^4 + 4*(0.5)^3 * 0.5 + 6*(0.5)^2(0.5)^2 + 4*0.5*(0.5)^3 + 0.5^4 =$$
$$= 0.0625 + 0.25 + 0.375 + 0.25 + 0.0625 = 1$$

Чисельності (n_i) дорівнюватимуть добутку ймовірностей з вище вказаної формули на $N(\Sigma n_i)$ З таблиці 7.3 бачимо, що:

$$\overline{X} = \frac{200}{100} = 2\,;\ \sigma = \sqrt{\frac{498 - \frac{200^2}{100}}{100}} = \sqrt{0.98} = 0.99 \approx 1.0$$

Тоді: $\overline{X} = 4 * 0.5 = 2.0\,;\ \sigma = 4 * 0.5 * 0.5 = 1.0$

Наведені розрахунки показують правомірність використання біноміального розподілу. Для обчислення вирівнюючих частот розподілу Пуассона використовують його якості, тобто. $X = \sigma = \mu_3 = \lambda$.

Враховуючи, що цей розподіл поодиноких подій за схемою Бернуллі (імовірність 0,1), тобто це ймовірність того, що подія настане 0,1, ..., k раз у серії з N випробувань. При $n \to \infty$ ймовірність $p \to 0$, то $np = const = \lambda$, тобто єдиним параметром λ розподілу Пуассона. Це дає можливість обчислювати ймовірність p_m розподілу Пуассона за складеною таблицею значень функції при різни

$$p_m = \frac{\lambda^m}{m!} * e^{-\lambda}$$

значеннях λ, які наведен у додатку В. Ця таблиця дана для значень X від 0 до 20. Для великих величин λ відповідна таблиця наведена у книзі А. К. Митропольського. Внаслідок того, що в лісовому господарстві великі величини λ зустрічаються рідко, ми опускаємо цю таблицю.

Наведемо приклад обчислення вирівнюючих частот під час розподілу Пуассона, який описаний К. Є. Нікітіним та А. З. Швиденко (таблиця 7.4). Обчислимо частоти, що вирівнюють для ряду розподілу числа дерев на пробних площах розміром 0,002 га. Статистики для цього ряду X = 1,508, λ_2 = 1,5189, λ_3 = 1,5434, що передбачає близькість емпіричного розподілу закону Пуассона. Такий самий висновок випливає і з логіки аналізованого явища. Для обчислення вирівнюючих частот ймовірності $f(k)$,отримані за формулою:

$$f(k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!} = \frac{(np)^k \cdot e^{-np}}{k!}$$

слід помножити обсяг ряду розподілу n=224. Задамо a=1,508. Зауважимо, що 0!=1, а 1g k! обчислюють безпосередньо або за таблицями логарифмів факторіалів Для обчислення за схемою таблиці 7.4 формулу попередньо логарифмують, тобто:

$$\lg f(k) = k \lg a - a \lg e - \lg k! = 0{,}0178k - 0{,}6549 - \lg k!.$$

В цілому нині модель добре відбиває емпіричний ряд. Наступні обчислення чисельностей щодо розподілу Пуассона покажемо з

прикладу обліку сходів сосни на зрубі. Для цього заклали облікові пробні площі, площею 10м (5х2).

Таблиця 7.4

Схема обчислення вирівнюючих частот розподілу Пуассона

k_i	n_i	$0{,}178k_i$	$k_i!$	-lg ki!	lg *f(k)*= -0,6549 + (3)+(5)	*f(k)*	$\overline{n_i}$
1	2	3	4	5	6	7	8
0	50	0	1	0	1,345	0,221	50
1	74	0,1780	1	0	1,523	0,333	75
2	57	0,3562	2	0,3010	1,400	0,251	56
3	27	0,5343	6	0,7782	1,101	0,126	28
4	12	0,7123	24	1,3802	2,577	0,048	11
5	3	0,8905	120	2,0792	2,156	0,014	3
6	1	1,0686	720	2,8573	3,557	0,004	1
Σ	224	-	-	-	-	-	224

Всього закладено 8 пробних площ. Результати розрахунків показано в таблиці 7.5. При цьому для розрахунків використані такі формули:

$$\tilde{n} = \left[\frac{\overline{X}^{-m}}{m!\, e^{-\overline{X}}} \right] N \quad y = \left(\frac{\lambda^m}{m!} e^{-\lambda} \right)$$

де: x, λ – середні значення;
N – об’єм ряду розподілу;
m – варіанти досліду;
e – 2,71 – арифметична стала.

Таблиця 7.5

Обчислення вирівнюючих частот за кривою Пуассона для пробних площ Перганського ПНДВ

n	n_i	mn_i	y(m)	$\tilde{n}$ без округ.	$\tilde{n}$ округ.	n_i %	$\tilde{n}$ %	$\overline{X}-n_i$	$(\overline{X}-m_i)$	$(\overline{X}+m_i)^2 \cdot n$	$(\overline{X}+m_k)^{2N}$ ·
1	2	3	4	5	6	7	8	9	10	11	12
0	549	0	0,3329	384	385	47,5	33,3	1,1	1,21	664,2	465,9
1	271	271	0,3662	422,96	423	23,5	36,6	ОД	0,01	2,7	4,2
2	137	274	0,2014	232,62	233	11,9	20,2	0,9	0,81	111,0	188,7
3	110	330	0,0738	85,24	25	9,5	7,4	1,9	3,61	397,1	306,8
4	56	224	0,0203	23,45	23	4,8	2,0	2,4	8,41	470,0	193,5
5	22	110	0,0045	5,20	5	1,9	0,4	3,9	15,27	335,9	76,4
6	9	54	0,0008	0,92	1	0,8	ОД	4,9	24,01	216,1	24,0
7	1	7	0,0001	0,11	0	ОД	0	5,9	34,81	34,8	0
Σ	1155	1270	1,0000	1158	1155	100	100	-	88,14	2231,8	1259,6

Значення функції Пуассона із середнім значенням (X або λ) та показником ступеня m, як показано вище, виписується із спеціальної таблиці (Додаток В). Графи 9-12 введені у таблиці 7.5 для обчислення σ. Її близькість до X (σ) показує, що розподіл описується рівнянням Пуассона. Тоді для емпіричного розподілу:

$$\sigma_1 = \sqrt{\frac{(\bar{X} - m_i)^2 n_i}{N}} = \sqrt{\frac{2232}{1155}} = \sqrt{1{,}93} = 1{,}39$$

Для теоретичного розподілу:

$$\sigma_2 = \sqrt{\frac{1260}{1155}} = \sqrt{1{,}10} = 1{,}05$$

Близькість σ, σ^2 до X показує правомірність використання для вирівнювання розподілу Пуассона. Для практичних розрахунків, коли знаходять теоретичні ординати розподілу, тобто чисельності розподілу випадкової події X, вище наведений вираз множать на N - загальну кількість спостережень, замість натомість n_i приймають експериментальну середню кількість випадків, що спостерігаються. Формула для n буде мати такий вигляд:

$$\tilde{n} = \left[(\bar{x}^m / m!) e^{-\bar{x}} \right] N$$

Розподіл Пуассона зі збільшенням середньої λ наближається до біномінального. Розподіл Пуассона описує багато явищ у техніці та біології. В техніці він знаходить широке застосування при контролі якості продукції, для апроксимації розподілу дефектних виробів.

В лісовому господарстві його застосовують як модель розподілу числа домішок в контрольних наважках при аналізі насіння, припустимо, сосни звичайної, при обстеженні проведенні аналізів плодів, скажімо, жолудів дуба звичайного, пошкоджених шкідником. Ним описують розподіл чисельності природного поновлення деревостанів, коли розмір елементарних облікових пробних площ дуже малий або умови пробної площі несприятливі, тож ймовірність позитивного результату мала. В лісовому господарстві до розподілу Пуассона вдаються, коли досліджують мутації при проведенні генетико-селекційних досліджень, при аналізі появи альбіносів серед лісових звірів та птахів, для оцінки виживання прижившуваності сосни звичайної під наметом лісу, при аналізі інших не систематичних подій.

РОЗДІЛ 8. СТАТИСТИЧНЕ ОЦІНЮВАННЯ

8.1 Помилки вибіркових статистичних показників та їх теоретичне пояснення.
8.2 Основні завдання статистичного оцінювання. Зміщені та незміщені оцінки.
8.3 Помилки статистик та їх визначення. Довіряючий інтервал.
8.4 Помилка суми чи різниці середніх значень.

8.1 Помилки вибіркових статистичних показників та їх теоретичне пояснення.

Значення основних статистичних показників (ð, σ, a, E) та їх розподіл дуже важливі для характеристики статистичної вибірки, але недостатньо для її повної оцінки. Якби ми працювали з чисельностями генеральної сукупності, наприклад, заміряли всі дерева, що ростуть в умовах природного заповідника, окремої області, держави, то обчислених значень середнього, середньоквадратичного відхилення та інших показників було б достатньо. Але в переважній більшості випадків, виміри проводять для вибіркової сукупності: пробної площі, на дослідній ділянці і т.д. Тому нам треба знати, на скільки наші статистичні показники відповідають аналогічним даним із генеральної сукупності, тобто чи достовірні вони та які їх помилки.

Таку оцінку нам дають обчислені помилки статистики. В основі оцінок статистик лежить знання про їх розподіл. Так, якщо з однієї генеральної сукупності взяти кілька вибірок, і для кожної з них визначити статистики, наприклад, про ð та σ, то можна виявити розподіл цієї статистики, яка також є випадковою величиною. Знання закону розподілу досліджуваної статистики дозволяє робити її ймовірні оцінки. Отже, оцінку добутку статистичного параметру можна зробити лише з певною ймовірністю. Теоретичну основу таких оцінок дає теорія ймовірності та описаний нею закон великих чисел. Тут важливими для вирішення наших завдань є теореми Маркова, Чебишева, Пуассона та Бернуллі. Їх докладний виклад з наведенням доказів є у підручниках з теорії ймовірностей, а також у книзі А. К. Митропольського. Враховуючи обмеженість курсу біометрії для фахівців лісового господарства, наведемо тут лише самі теореми без їхнього доведення у тому порядку, в якому вони взаємопов'язані між собою й статистичними оціками генеральної сукупності під час проведення статистичних досліджень на науково-дослідних об’єктах лісового фонду, в тому числі й природно-заповідного. Докази названих теорем базуються на використанні теоретичних гіпотез Маркова. Тому наведемо її визначення та доказ.

Теорема Маркова формулюється в такий спосіб:

якщо випадкова величина x може набувати лише позитивних значень, то ймовірність P того, що ця величина не перевершить свого математичного очікування M(x), помноженого на деяке позитивне число t^2 більше різниці між одиницею та числом, зворотним даному позитивному числу.

Позначаючи можливість співвідношення як P{x}, теорему Маркова можна записати наступним виразом:

$$P\left\{\tilde{o} \le M(x)t^2\right\} > 1 - \frac{1}{t^2}$$

Для доказу припустимо, що *x* приймає лише деякі позитивні значення, тобто $x_1, x_2, \ldots x_k$ з ймовірностями $P_1, P_2, \ldots P_k$. З раніше викладеного матеріалу ми знаємо, що: $\Sigma\delta_i = 1$.

Тоді M(x) буде дорівнювати:

$$M(x) = \sum P_i x_i$$

Якщо візьмемо будь-яке позитивне число t^2 (при зведенні у квадрат будь-яке число буде позитивним), припустимо, що перші і значень цього ряду не більше $M(x)\ t^2$, а решта більше $M(x)\ t^2$.

Тоді за теорією добутків ймовірностей маємо:

$$P\left\{(x) \le M(x)t^2\right\} = \sum P_i \text{ и}$$

$$Q\left\{x > M(x)t^2\right\} = P_{i+1} + \ldots \mathrm{P_k}$$

Тобто P+Q=1.

Так як ймовірності представляють числа не від'ємні, то, опускаючи в правої частини рівності члени з індексами $1, 2, \ldots, j$, $M(X) \ge p_{j+1}x_{(j+1)} + \ldots + p_k x_{(k)}$ Оскільки далі всі значення $x_{(j+1)}, \ldots, x_{(k)}$ більше $M(X)t^2$, то, підставляючи цю останню величину, замість кожного із значень, отримаємо більш сувору нерівність:

$$M(X) > M(X)t^2[p_{j+1} + \ldots + p_k],$$

$$M(X) > M(X)t^2 Q.$$

Розділивши ліву та праву частини цієї нерівності на позитивну величину $M(X)t^2$, знаходимо: $\frac{1}{t^2} > Q$

Так як на основі того, що P+Q=1, маємо:

$$Q=1-P,$$

$$\frac{1}{t^2}>1-P,$$

$$P\left\{X \le M(X)t^2\right\}>1-\frac{1}{t^2},$$

що й виражає теорему Маркова.

Зауважимо, що хоча теорема Маркова має місце за будь-якого позитивного числа t^2, однак, через те, що величина ймовірності не може бути менше 0, ми розглядаючи цей вираз констатуємо, що має сенс брати лише ті t2, які не менше 1. При $t^2=1$, маємо рівняння:

$$P\left\{X \le M(X)\right\}>0.$$

Чим більше P, тим більшою буде ймовірність того, що:

$$X \le M(X)t^2.$$

З теореми Маркова знаходимо, що:

$$Q\left\{X > M(X)t^2\right\} \le \frac{1}{t^2}$$

тобто ймовірність нерівності $X>M(X)t^2$ не більше $\frac{1}{t^2}$.

Описуючи рівність у вигляді: $P\left\{X \le M(X)t^2\right\}>1-\frac{M(X)}{M(X)t^2}$ й поклавши $M(X)t^2=\tau$, маємо $P\left\{X \le \tau\right\}>1-\frac{M(X)}{\tau}$.

Тому теорема Маркова може бути виражена наступним чином:

Якщо випадкова величина X може приймати лише позитивні значення, то ймовірність того, що вона отримає значення, що не перевищують деякого позитивного числа τ, більше

$$1-\frac{M(X)}{\tau}.$$

Наведемо геометричну ілюстрацію теореми Маркова (рис. 8.1).

Ліва частина нерівності виражає ймовірність того, що величина X не перевищує деякого позитивного числа τ. За визначенням, ймовірність змінюється від нуля, для неможливих подій, до одиниці

- для подій достовірних (можливих). Так як за умовою величина X може приймати лише позитивні значення, то за $\tau = 0$ очевидно, що:

$$P\{X \leq \tau\} = 0$$

як ймовірність неможливої події. При зростанні τ ймовірність P зростатиме, прагнучи до 1 при прагненні τ до ∞. На рисунку ймовірність P буде зображена лінією, яка при $\tau \leq 0$ збігатиметься з віссю абсцис, а при $\tau \geq 0$ підніматиметься над цією віссю, прагнучи злитися з прямою, паралельною цієї осі, і проходить від неї з відривом, рівному 1 (ри. 8.1).

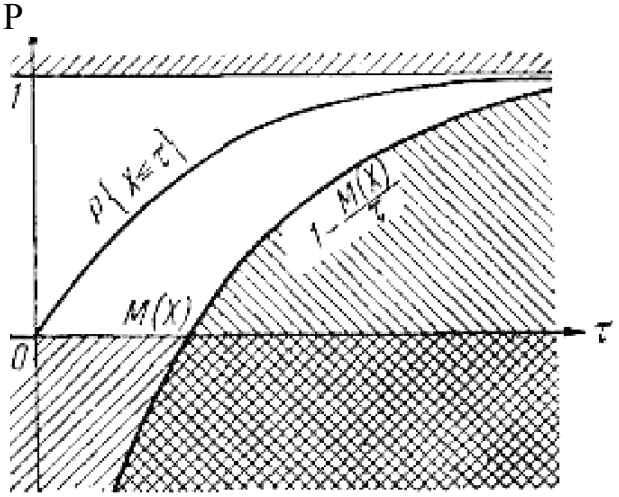


Рис. 8.1. Геометрична ілюстраці теореми Маркова.

На підставі нерівності, крива ймовірностей P розташована над гілкою гіперболи. Отже, ця крива не може проникати в області, заштриховані на рисунку. Теорема Маркова є основною пропозицією статистичного обчислення. Чудовою властивістю цієї теореми є її незалежність від природи розподілу позитивної випадкової величини. В теоремі Маркова закладено доказ багатьох теорем статистичного обчислення і зокрема, найважливішої з цих теорем – закону великих чисел.

Нерівність Чебишева.

Теорема Маркова дає можливість знайти ймовірність того, що позитивна випадкова величина набуде значення, що не перевищує деякого даного числа; при цьому потрібно лише знання математичного очікування цієї величини. Певний висновок про випадкову величину дає також нерівність Чебишева, яке прикладається до будь-яких (не обов'язково позитивних) випадкових величин, причому потрібне лише знання математичного очікування та дисперсії випадкової величини.

Нерівність Чебишева: Якщо випадкова величина X може приймати й позитивні, й негативні значення, то яким би не було позитивне число τ

$$P\{-\tau \le X \le \tau\} > 1 - \frac{M(X^2)}{\tau^2}$$

При встановленні цієї нерівності застосовується теорема Маркова. Доказ її ми провели вище. Вказана вище нерівність дає нижню межу.

$$1 - \frac{\sigma^2}{\tau^2}$$

Ймовірність того, що відхилення значень випадкової величини від її математичного очікування не перевищить певного заданого числа ± τ. Якщо дисперсія σ^2 зменшується, то нижня межа ймовірностей цих відхилень зростає. Це показує, що значення випадкової величини тим більше зосереджуються щодо її математичного очікування, чим менше дисперсія. Таким чином, з'ясовується сенс дисперсії як міри розсіювання значень випадкової величини у її математичного очікування.

Слід вважати, що нерівність: $\tau = t\sigma$.

Маємо: $P\{-t\sigma \le x - M(x) \le t\sigma\} > 1 - \frac{1}{t^2}$

Й відповідно рівність має такий вигляд: $Q\{|x - M(x)| > t\sigma\} \le \frac{1}{t^2}$

Ці нерівності справедливі для будь-якого розподілу випадкової величини із кінцевою дисперсією. При постійному основному відхиленні σ про нерівність показує, що якщо t буде збільшуватися, то ймовірність того, що значення випадкової величини будуть знаходитися в проміжку, $(-t\sigma+M(x) \quad t\sigma+M(x))$, що збільшується. Зокрема, якщо t=2, то

$$P\{-2\sigma \le x - M(x) \le 2\sigma\} > 0{,}75,$$

Якщо t=3, то:

$$P\{-3\sigma \le x - M(x) \le 3\sigma\} > 0{,}889$$

Нехай тепер величина t буде постійною. Тоді при зменшенні в основному відхиленні σ, зменшується проміжок $(M(x)-t\sigma,\ M(x)+t\sigma)$, нижня межа ймовірності значень x-M(x), що полягають в цьому проміжку, залишатиметься постійною. Звідси знову таки випливає, що чим менше основне відхилення, тим більші окремі значення випадкової величини зосереджуються в її

математичному очікуванні. Таким чином, основне відхилення є мірою розсіювання значень довільної величини.

Після засвоєння теореми Маркова та нерівності Чебишева пвдтвердимо (без доказу) теорему Маркова.

Теорема Маркова виражається рівнянням:

$$P=\left\{-v<\frac{\sum_{s=1}^{n}x_s}{n}-\frac{\sum_{s=1}^{n}a_s}{n}<v\right\}>1-w$$

де: v, w, - деякі довільно задані позитивні числа;

a_s – математичне очікування.

Теорема Маркова викнується при умові: . $\frac{\sigma^2}{n^2}\to 0;\ n\to\infty$.

Теорема Маркова є найбільш загальним виразом закону великих чисел. Її окремим випадком є теорема Чебишева. Вона формулюється в такий спосіб.

Теооема Чебишева.

Якщо число *n* попарно незалежних $x_1, x_2, \ldots x_n$ випадкових величин можна збільшувати безмежно, а математичні суми їх квадратів не перевершують одного і того ж постійного числа, то при достатньо числі цих величин буде близькою до достовірності ймовірність того, що їхнє середнє арифметичне відрізняється мало від середнього арифметичного їх математичних ймовырностей:

$$P\left\{-\varepsilon<\frac{\sum_{s=1}^{n}x_s}{n}-\frac{\sum_{s=1}^{n}a_s}{n}<\varepsilon\right\}>1-\eta$$

Одним із важливих наслідків теореми Чебишева є застосування її у випадку, коли всі попарно незалежні величини $x_1, x_2, \ldots x_n$ мають одну й те саму математичну ймовірність, тобто коли всі $a_s=a \qquad (s=\overline{1,n})$, а крім того всі $M(x_s^2)=b_s=b \qquad (s=\overline{1,n})$ до того ж *b* існує, тобто воно кінцеве.

Інакше кажучи, незалежні величини $x_1, x_2, \ldots x_n$ можна розглядати як значення, отримані в n незалежних випробуваннях щодо однієї і тієї ж випадкової величини x. У такому разі, згідно з теоремою Чебишева маємо вираз:

$$P\left\{-\varepsilon < \frac{\sum_{s=1}^{n} x_s}{n} - a < \varepsilon\right\} > 1-\eta$$

Таким чином, з теореми Чебишева виходить як наслідок важлива теорема:

Якщо з величиною *x*, яка має кінцеву дисперсію знаходиться достатньо велике число незалежних досліджень, то з ймовірністю близькою до достовірності можна встановити, що середнє арифметичне дослідних значень величини *x* мало відрізнятиметься від її математичного очікування.

Теореми Пуассона та Бернуллі ми розглянули раніше, коли описували біноміальний розподіл. Таким чином, розглянуті теореми доводять, що чим більший обсяг вибірки, тим більше середній результат, тобто вибіркова середня (Х) меншою мірою відхиляється від середньої арифметичної (М) генеральної сукупності, і навпаки, що менше вибірка, то менше і шансів на те, що вибіркова середня збігається за величиною із середньою арифметичною генеральної сукупності.

Дія цього закону основана на властивості самих випадкових величин, негативні та позитивні значення яких здатні компенсувати один одного і тим повніше, чим більшій кількості випробувань піддається випадкова величина. На цій властивості випадкових величин компенсувати один одного і основано відносну стійкість середніх значень. *Тому описані у попередньому розділі закономірності розподілу, що спостерігаються в ранжованих сукупностях варіантів слід розглядати як прояв найбільш загального закону поведінки випадкових величин – закону великих чисел.*

Закон великих чисел стверджує, що практично малоймовірно значне відхилення середньої арифметичної вибіркової сукупності (Х) від середньої арифметичної генеральної сукупності (М), якщо кількість спостережень досить велика.

8.2 Основні завдання статистичного оцінювання. Зміщені та незміщені оцінки.

Ми розглянули теоретичні аспекти статистичного оцінювання.

Тепер розглянемо його практичні приклади. Статистичне оцінювання інформації включає три основні завдання:

- знаходження за вибіркою найімовірніших значень оцінок деяких параметрів («точкове» оцінювання);
- оцінку інтервалів щодо меж, відносно яких з певною ймовірністю можна стверджувати, що вони складають невідомий параметр (інтервальне оцінювання);

- перевірку справедливості тих чи інших тверджень щодо явища, що вивчається (перевірка статистичних гіпотез).

Ці завдання тісно взаємопов'язані, але можна вирішувати кожне окремо чи все одночасно залежно від мети статистичного аналізу. Наприклад, ми, визначивши запас деревини на 1 га дуба звичайного, середнього віку 65-70 років знайшли його рівним 250 $м^3$/га. Цю величину слід оцінити наступним чином:

- наскільки точно значення запасу (% помилки), та які допустимі відхилення від неї в обидві сторони: точкове оцінювання та оцінка інтервалу.
- наскільки знайдена величина відповідає середньому запасу дубових деревостанів у цьому віці за певного класу бонітету.

Оцінки параметрів одержують різними методами. Тому, зазвичай вибирають ті, що дають найкращі результати. З цією метою вводять поняття незміщеності, спроможності, ефективності та достатності оцінок.

Оцінку називають незміщеною, якщо вона не дає систематичної помилки при оцінюванні деякого параметра 0, іншими словами, якщо середнє значення оцінки, отримане за безліччю вибірок, практично збігається з 0.

У принципі може бути кілька незміщених оцінок одного й того ж параметра; наприклад, як оцінку середнього значення генеральної сукупності за деяких умов можна взяти вибіркову середню або вибіркову медіану. І тут доцільно брати оцінку з меншою мінливістю.

Оцінку називають об'єктивною, якщо зі збільшенням обсягу вибірки вона все більше наближається до параметра, який оцінюється.

Об'єктивну оцінку із найменшою дисперсією називають ефективною.

Зрештою, достатність оцінки розуміють у тому сенсі, що не існує іншої оцінки параметра 0, обчисленої на підставі даної вибірки, яка містить додаткову інформацію про цей параметр. Достатність оцінки тягне за собою її ефективність та спроможність. Для знаходження оцінок із заданими властивостями, існує ряд методів, з яких найчастіше застосовують - метод максимальної ймовірності та метод моментів.

Метод максимальної ймовірності передбачає використання як оцінки нульового параметра такого значення, яке (за даними вибірки) найбільш ймовірно з точки зору можливості заміни ним нуля.

Цей метод дає ефективні, але не завжди не зміщені оцінки, а для вибірки великого обсягу оцінки має нормальний розподіл.

Метод моментів забезпечує об'єктивні, але не завжди ефективні оцінки. В цьому методі, використаному вище при апроксимації вибіркового розподілу, параметри розподілу представлені через

моменти. За вибіркою обчислюють оцінку моментів та підставляють їх в отримані рівняння, якими знаходять невідомі параметри. Доцільність практичного використання оцінок з тими чи іншими властивостями має обговорюватися в конкретних завданнях. Найчастіше «позитивні» характеристики оцінок збігаються. Так, вибіркове середнє завжди є не зміщеною та об'єктивною оцінкою середньої генеральної сукупності, а за нормальності вихідної сукупності - ще й достатньої, в той час як медіана, що є вибіркова оцінка середньої сукупності, такою не є, оскільки її дисперсія більша за дисперсію вибіркового середнього в розрізі сукупності варіант досліду. У деяких випадках краще мати оцінку дещо зміщену, але об'єктивну; іноді незміщеність можна усунути.

Так, при знаходженні дисперсії, знаменник n-1 пояснюється саме тим, що вибіркова дисперсія $\hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \overline{X})^2$ є зміщеною оцінкою дисперсії генеральної сукупності, а множник n/(n - 1) цей зсув усуває.

8.3 Помилки статистик та їх визначення. Довіряючий інтервал.

Для практичного використання нам потрібно знати помилки основних статистик розподілу: середнього значення, середнього квадратичного відхилення, коефіцієнта варіації, асиметрії та ексцесу. В літературі часто зустрічається вираз «помилка репрезентативності». Тому розпочнемо розгляд саме з неї.

Помилки репрезентативності.

Розбіжність між величиною середньою арифметичної (X) вибірки та величиною середньої арифметичної генеральної сукупності (M) прийнято називати помилкою репрезентативності, тобто помилкою, яка допускається не в самому процесі вимірювальної та обчислювальної роботи, а в результаті випадкового відбору варіант з генеральної сукупності під час утворення вибірки.

Репрезентативна помилка – це не технічна, а статистична помилка. Вона вказує на величину відхилення вибіркової середньої (X) від середньої (M) генеральної сукупності.

Величина репрезентативної помилки визначається по різниці між середніми величинами вибірки та генеральної сукупності, тобто як (X)-M Однак цей показник у практиці використовувати неможливо, так як середня арифметична генеральної сукупності зазвичай залишається невідомою. Якщо ж середня (M) генеральної сукупності відома, то вказана різниця (X-M) втрачає своє значення. Тому помилки репрезентативності визначаються не прямим шляхом - через відхилення варіант від вибіркової середньої.

Помилка окремо взятої варіанти.

Якщо робити висновки про величину статистичної помилки окремо взятої варіанти, то вона дорівнює середньому квадратичному відхиленню, тому що будь-який емпіричний розподіл, що прямує за нормальним законом розподілу практично вкладається в межах плюс-мінус трьох сигм, (X±3σ). Через це помилки репрезентативності називають середньою квадратичною помилкою, або просто середньою помилкою. Будемо її позначати через m, вказуючи при цьому її характеристику, яку вона супроводжує. Таким чином, середня квадратична помилка окремо взятої варіанти виражається у вигляді:

$$m_{x_i} = \pm\sigma$$

Середнє квадратичне відхилення має подвійне значення:

- по-перше, воно основне мірило мінливості, показник варіабельності ознак;
- по-друге, середнє квадратичне відхилення служить як статистична помилка окремо взятої варіанти.

Помилка середньої арифметичної.

Математична статистика стверджує, що вибіркова середня ($\bar{X}$) відхиляється від математичної ймовірності або середньої арифметичної (M) генеральної (теоретично розрахованої) сукупності менше в $\sqrt{n}$ разів порівняно з окремими варіантами цього розподілу. Звідси випливає, що середня квадратична помилка вибіркової середньої (X) дорівнює сумі від поділу середнього квадратичного відхилення на корінь квадратний в складі всіх варіант цієї сукупності, тобто:

$$m_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Наведемо приклад. Візьмемо розподіл числа дерев за товщиною стовбура на висоті 1,3 м., і зробимо необхідні нам обчислення (таблиця 8.1).

Таблиця 8.1

Обчислення середнього значення та його помилки на прикладі розподілу діаметрів в сосновому 80-річному деревостані II класу бонітету пробної площі №2 Переганського ПНДВ (середнє за 2022 – 2025 рр).

Ступені товщини (варіанти, x_i)	Кількість стовбурів, (частоти, n_i)	x_i*n_i	$a=x_i-X$	a^2	n_ia_2
12	4	48	14	196	784
16	8	128	10	100	800
20	26	520	-6	36	936
24	43	1032	-2	4	172
28	31	868	+2	4	124
32	22	704	+6	36	792
36	11	396	+10	100	1100
40	4	160	+14	196	784
44	1	44	+18	324	324
Всього:	150	3900	-	-	5816

За даними таблиці 8.1 обчислимо статистики розподілу та їх основні помилки.

$$\overline{X} = \frac{\sum x_i n_i}{\sum n_i} = \frac{3900}{150} = 26$$

$$\sigma = \sqrt{\frac{\sum x_i a_i^2}{\sum n_i}} = \sqrt{\frac{5816}{150}} = \sqrt{38{,}77} = 6{,}23 \approx 6{,}2$$

$$m_{\overline{X}} = \frac{\sigma}{\sum n_i} = \frac{6{,}23}{\sqrt{150}} = \frac{6{,}23}{12{,}247} = 0{,}509 \approx 0{,}51 \approx 0{,}5$$

Прийнято записувати середню величину разом із її основною помилкою, тобто X = 26 ± 0,5. Середня помилка вказує найбільш ймовірні межі, в яких можливі випадкові коливання величини середньої арифметичної залежності від обсягу вибірки. При збільшенні числа випробувань середня помилка зменшується. Коли кількість спостережень необмежено зростає, середня помилка прагне нулю, тобто при N →∞ і m → 0. Отже, середня помилка є мірою точності, або відносною достовірності, нашого судження про можливі коливання середніх показників варіюючих величин. Оскільки весь варіаційний ряд нормально розподіляється, випадкова величина X практично вкладається в межах між X+3σ та X-3σ на 99,9 %, тобто можна сказати, що генеральна середня (M) таких

розподілів не виходить за межі потрійного значення помилки середньої арифметичної будь-якої вибірки, взятої з цієї генеральної сукупності, тобто вона завжди вкладена між межами від $X-3m_x$ до $X+3m_x - \sigma$ або в межах $X \pm 3m_x$.

Тому потрійне значення середньої квадратичної помилки називається граничною помилкою середньої арифметичної вибіркової сукупності.

Вираз $X \pm 3m_x$ містить у собі зміст так званого «правила потрійної помилки» та правила трьох сигм. При обчисленні помилки середньої арифметичної на малих вибірках, число спостережень (N) береться з «числом ступенів свободи», і формула набуває такого вигляду:

$$m_{\bar{x}} = \sqrt{\frac{\sum (x - \bar{X})^2}{N(N-1)}} = \frac{\sigma}{\sqrt{N-1}}$$

При великому N, різниця між N і N-1 несуттєва, і формулу можна записати у вигляді:

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

Коли середня арифметична обчислюється прямим способом на матеріалі, не згрупованому в класи, помилку можна визначити за наступною формулою:

$$m_{\bar{x}} = \sqrt{\frac{\left(\frac{\sum x^2}{N} - \bar{X}^2\right)}{N-1}}$$

Наприклад, є наступні десять варіант, що виражають товщину дерев на пробній площі в культурах сосни звичайної віком 10 років:

$$5, 6, 6, 7, 4, 4, 2, 5, 5, 5, 6$$

Обчислення тут будуть проведені за таблицею 8.2.

Якщо ж середня арифметична визначається на несгрупованом в класи матеріалі коротким способом моментів (способом умовної середньої), то середня помилка обчислюється за формулою:

$$m_{\bar{x}} = \sqrt{\left[\frac{\sum a^2}{N} - \left(\frac{\sum a}{N}\right)^2\right] / (N-1)}$$

a=x-A, тобто відхилення варіант від умовної середньої. Продемонструємо застосування цієї формули на попередньому прикладі (таблиця 8.3).

Таблиця 8.2

Обчислення помилки середньої при несгрупованих варіантах

№ п/п x_i	Товщина дерев (варіанти, n_i, см)	n_i^2	Розрахунок:
1	5	25	$\bar{x} = \frac{\sum n_i \cdot x_i}{n_i} = \frac{50}{10} = 5$см Из уравнения (9.14) $m_{\bar{x}} = \sqrt{(\frac{268}{10} - 25) + 9} = \sqrt{1,68 + 9} = \sqrt{10,68} = 3,27$ñì $\bar{\sigma} = \sqrt{\frac{\sum x_i a_i^2}{\sum n_i}} = \sqrt{\frac{268}{50}} = \sqrt{5,36} = 2,31$
2	6	36	
3	6	36	
4	7	49	
5	4	16	
6	4	16	
7	2	4	
8	5	25	
9	5	25	
10	6	36	
Всього (Σ):	50	268	

Таблиця 8.3

Схема обчислення помилки середнього

№ п/п x_i	Кількість, n_i,	Відхилення від умовної середньої а	a^2	Розрахунок за формулою $m_{\bar{x}} = \sqrt{\left[\frac{\sum a^2}{N} - \left(\frac{\sum a}{N}\right)^2\right]/(N-1)}$
1	5	+1	1	$\bar{x} = 4 + \frac{10}{10} = 5$ñì $m_x = \sqrt{\frac{28}{10} - (\frac{10}{10})^2 + 9} = \sqrt{2,8 - 1 + 9} = \sqrt{10,8} =$ $= 3,29$
2	6	2	4	
3	6	2	4	
4	7	3	9	
5	4=к	0	0	
6	4	0	0	
7	2	2	4	
8	5	+1	1	
9	5	+1	1	
10	6	2	4	
Всього (Σ):	50	10	28	

З наведених розрахунків випливає, що отриманий результат практично (з урахуванням заокруглення) ідентичний до попереднього.

Помилка середнього квадратичного відхилення.

Поняття помилки репрезентативність відноситься не тільки до середньої арифметичної, але та до інших середніх показників, зокрема до середнього квадратичного відхилення, що характеризує варіювання ознаки в даній сукупності. Середня помилка середнього квадратичного відхилення обчислюється за формулою:

$$m_\sigma = \frac{\sigma}{\sqrt{2N}}$$

$$m_{\sigma^2} = \frac{\sigma^2}{\sqrt{\frac{N}{2}}} \quad \text{или} \quad m_{\sigma^2} = \sigma^2\sqrt{\frac{2}{N}}$$

В нашому прикладі (таблиця 8.1), де N=150, σ=6,2, помилка та становитиме:

$$m_\sigma = \frac{6,2}{\sqrt{300}} = \frac{6,2}{17,32} = 0,356\,.$$

Отже, m лежить у межах (з достовірністю 99,9%) 6,2 +3*0,356 = = 6,2 +1,068.

Для прикладу, розрахованого за таблицею 8.2 $m_\sigma = \frac{2,31}{\sqrt{20}} = \frac{2,31}{4,47} = 0,517$ Тоді границі m_σ складають: 2,31 +1,55.

Помилка коефіцієнта варіації.

Середня помилка коефіцієнта варіації (V) визначається за формулою:

$$m_V = \frac{V}{\sqrt{2N}} \cdot \sqrt{1+\left(\frac{V}{100}\right)^2}$$

Формулу можна також подати в наступному вигляді:

$$m_v = V\sqrt{\frac{0.5 + 0.0001V^2}{N}}$$

Обчислимо за даними таблиці 8.1 коефіцієнт варіації та її помилку.

$$V = \frac{\sigma}{\overline{X}} \cdot 100\% = \frac{6,2}{26} \cdot 100\% = 23,8\% \approx 24\%$$

В такому випадку: $m_V = \frac{23,8}{\sqrt{300}} \cdot \sqrt{1+\left(\frac{23,8}{100}\right)^2} = \frac{23,8}{17,32} \cdot \sqrt{1+0,113} = 1,37 \cdot 1,05 = 1,44$

$$m_v = 23.8\sqrt{\frac{0.5 \cdot 0.0001 \cdot 23.8^2}{150}} = 23.8 \cdot 0.0609 = 1.449 .$$

Значення m_v за обома формулами з урахуванням округлення практично однакові. Отже, коефіцієнт варіації генеральної сукупності, якій належить наша вибірка не вийде за межі т m_v=23,8±4,332.

Довірчі ймовірності та рівні значущості.

В математичній статистиці, а також і в біометрії прийнято суттєвість того або іншого результату оцінювати за значеннями трьох ймовірностей, близьких до достовірності: P_1=0,95, або 95%, P_2=0,99, або 99%, P_3=0,999, або 99,9%. ***Ці ймовірності отримали назву довірених.***

Ймовірності ж, якими вирішено нехтувати, тобто. P_1=0,05, або 5%, P_2=0,01, або 1%, та P_3=0,001, або 0,1%, ***отримали назву рівнів заначущості, або рівнів суттєвості.***

Всі ймовірності позначаються символами як P_{095} або $P_{0.05}$ і т.д. Кожній довіряючій ймовірності відповідає певне значення нормованого відхилення (t):

Ймовірності	P_1=0,95	Відповідає	t_1=1,96
Ймовірності	P_2=0,99	Відповідає	t_2=2,50
Ймовірності	P_3=0,999	Відповідає	t_3=3,30

Довіряючий інтервал та межі довіри.

Вище було сказано, що величина помилки вибіркової середньої визначається по різниці між цією середньою (X) та середньою генеральною сукупністю (M), тобто як X-M. Чи можна за емпіричними даними визначити найймовірніші межі, в яких перебуває середня (M) генеральної сукупності? Математична статистика дає на це питання позитивну відповідь.

Інтервал, в якому із заданою ймовірністю чи рівнем значущості вкладена середня арифметична генеральна сукупність, називається довіряючим інтервалом.

Межі цього інтервалу отримали назву довіряючих меж, або граніць довіряючого інтервалу. Як же визначити довіряючий інтервал та його межі? Це досягається нормуванням відхилення варіант, або вибіркової середньої від середньої генеральної сукупності. Так, якщо взяти нормальне відхилення варіантів від

$$t = \frac{x_i - \overline{X}}{\sigma}$$

вибіркової середньої , то можна перетворити його так: x, - X=tσ. Аналогічно відхилення вибіркової середньої X від середньої генеральної сукупності (M) виражається через:

$$\overline{X} - M = t\frac{\sigma}{\sqrt{N}}, \text{ або } \overline{X} - M = t\, m_{\overline{X}}.$$

Величина цього відхилення залежить, від ступеня варіабельності ознаки, а також від рівня ймовірності, з якою визначається довіряючий інтервал. Замінивши в цьому рівнянні знаки на зворотні та переставивши X у праву частину рівняння отримаємо:

$$M = \overline{X} - t m_{\overline{X}}$$

Оскільки X може бути і більше і менше M, то зазначений вираз можна записати у такому вигляді: $M = X \pm tm_x$. Звідси довіряючий інтервал для середньої арифметичної генеральної сукупності виражається нерівністю:

$$\overline{X} - tm_{\overline{X}} \le M \le \overline{X} + tm_{\overline{X}}$$

де: $\overline{X} - tm_{\overline{X}}$ та $\overline{X} + tm_{\overline{X}}$ - межі довіряю чого інтервалу.

З викладеного випливає багато важливих практичних складових для ведення лісового господарства Наприклад, потрібно, щоб матеріально-грошова оцінка лісосік проводилась досить точно на кожній із протаксованих ділянок, тобто достовірність тут повинна дорівнювати 3σ. Прийнятна технологія таксації лісосік забезпечує достовірність в 3σ при точності ±10%. Правда, 1-2 лісосіки з 1000 можуть не вкластися в точність ±10%, але на це довелося піти, тому що більш висока точність обліку вимагає іншої, набагато дорожчої технології, і економічно не виправдана. Але вже сукупність 3-5 лісосік за відсутності систематичної помилки має таксуватися з точністю 5-6%, а за наявності 10 лісосік точність має становити 3-4%.

8.4 Помилка суми чи різниці середніх значень.

Часто нам треба знати, чи між собою різняться середні величини. Наприклад, через n років після внесення добрив, середній діаметр досліджуваного деревостану становив 24 см, а контрольного - 22 см. Виникає питання, наскільки ця різниця суттєва, і є результатом проведених заходів, чи залежить вона від випадкових причин. Для оцінки таких відмінностей використовують t -критерій Стьюдента.

t-розподіл Стьюдента.

Перш ніж розпочати розгляд питань, пов'язаних з методикою оцінки достовірності відмінностей, що спостерігаються між вибірковими середніми, необхідно розглянути ще один бік нормованого відхилення, яке має пряме відношення до статистики малої вибірки В даному випадку мається на увазі закон розподілу

величин нормованого відхилення вибіркової середньої (X) від середньої арифметичної генеральної сукупності (M), відкритий англійським математиком Вільямом Госсетом в 1908 році.

Цей вчений друкувався під псевдонімом Стьюдент (Student). Вільямом Госсетом (Стьюдент) встановив, що можливість нормованого відхилення X-M: σ = t виражається наступним рівнянням:

$$P(t) = C\left(1+\frac{t^2}{N-1}\right)^{-\frac{1}{2}N}$$

яке носить назву розподілу Стьюдента.

Тут P(t) позначає ймовірність зазначеного нормованого відхилення, а C – деякий множник, що залежить лише від обсягу вибірки (N).

В практиці (за незалежних x_i) для оцінки достовірності різниці між середніми, використовують такі прийоми. Тож якщо x_i, незалежні, то, наприклад, для $y=x_1-x_2$ маємо вираз:

$$m_y = m_{x_1-x_2} = \sqrt{m_{x_1}^2 + m_{x_2}^2}$$

Наведений вираз – основна помилка різниці двох випадкових величин. Цей показник можна застосовувати для оцінки значущості різниці між середніми двох вибірок, наприклад, для висновку про те, чи можна вважати, що ці вибірки належать до однієї генеральної сукупності. Для цієї мети обчислюють величину:

$$t^* = \frac{\left|\tilde{x}_1 - \tilde{x}_2\right|}{m_{x_1-x_2}} = \frac{\left|\tilde{x}_1 - \tilde{x}_2\right|}{\sqrt{m_{x_1}^2 + m_{x_2}^2}}$$

Якщо t*>2, то з ймовірністю 0,95, а при t*>3 з ймовірністю практично не відрізняється від 1 можна стверджувати, що різниця між середніми значуща.

Наведемо приклад практичного використання t - критерію Стьюдента в лісогосподарської практиці. Нехай на певній ділянці лісових культур сосни звичайної віком 30 років II класу бонітету внесли мінеральні добрива Через 10 років виміряли дослідні та контрольний ділянки, що 10 років тому не відрізнялись за приростом, висотою, тобто мали середні діаметри 10,9 (контрольний варіант) та 11 см (дослідний варіант), а через десять років відповідно 14,0 і 15,1 см. Помилка середнього значення становила в 40 років 0,3 та 0,4 см.

Використовуючи формулу знайдемо:

$$t = \frac{15,1 - 14,0}{\sqrt{0,3^2 + 0,4^2}} \frac{1,1}{\sqrt{0,09 + 0,16}} = \frac{1,1}{\sqrt{0,25}} = \frac{1,1}{0,5} = 2,2$$

Таким чином, з ймовірністю 0,95 ми можемо стверджувати, що добрива дали позитивний ефект. В практиці лісового господарства і особливо під час проведення наукових досліджень, t-критерій застосовується досить часто. Критичні значення t - критерія Стьюдента (Вільяма Госсета) за певної кількості ступенів свободи для різних рівнів ймовірності наведено в додатку Е.

РОЗДІЛ 9. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

9.1. Статистичні гіпотези. Прості та складні гіпотези.
9.2 Параметричні методи оцінки гіпотез.
9.3 Непараметричні методи оцінки гіпотез.
9.4 Перевірка статистичних гіпотез в практиці лісового господарства та їх використання а в лісовому господарстві.

9.1 Статистичні гіпотези. Прості та складні гіпотези.

Гіпотеза – це науково обґрунтоване припущення про ймовірність деякої події, явища, закону тощо.

Гіпотезою може вважатися не будь-яке припущення, а тільки таке, що має деяке наукове обґрунтування, хоча ще недостатньо доведене та перевірене. Тому гіпотеза може як підтвердитись, так і бути відкинутою. Але попереднє наукове обґрунтування зробити треба, щоб не виконувати явно непотрібну роботу, перевіряючи припущення, які ні на чому не базуються. Гіпотеза, яка знаходить підтвердження та обґрунтування, переростає в теорію, закон, закономірність тощо. Прикладом тих, хто підтвердивли гіпотези, що потім стали теоріями та законами є періодична система елементів Д. І. Менделєєва (1834-1907), атомна будова матерії, будова атома, наявність кварків. В лісовому господарстві теорія будови деревостоїв, що була висунута в 19 столітті була лише гіпотезою, яка потім підтвердилася роботою відомих лісівників А. В. Тюріна (1882-1979), В. К. Захарова (1882-1966), Н. В. Третьякова (1880-1957), Ф. П. Моїсеєнко (1894-1979), П. С. Погрбняк 1900 – 1976) та інших. Ми розглядатимемо не всі гіпотези, а лише статистичні, тобто ті, що стосуються галузі математичної статистики. Для їхньої перевірки існує стандартна процедура. Для того, щоб її пояснити, візьмемо простий приклад із відібраними дослідником в сосновому стиглому деревостані за допомогою бура Пресслера кернами по визначенню раннього та пізнього приростів в структурі річних кілець. Ми вже бачили, що ймовірність поява майже однакових за приростом ранніх і пізніх кілець в сосни звичайної становить: Рр = Рп = 0,5. Таке співвідношення спостерігатиметься, якщо загальна кількість відібраних кернів спостерігачем досить велика > 30, а краще > 100. У співвідношенні ранніх та пізніх приростів немає асиметрії, тобто дотримано симетрію. Припустимо, що така симетрія порушена, скажімо проводимо спостереження в умовах пробної площі, де пройшли низові лісові пожежі, але висота підгару стовбурів дерев становить до 50 см. У цьому випадку співвідношення між закладеними раннім та пізнім приростами порушиться, що свідчить про некоректність досліду.

При аналізі гіпотез спочатку необхідно зупинитися на важливому факті: результат експерименту, що підтверджує справедливість висунутої гіпотези, майже ніколи не може служити

основою для ухвалення цієї гіпотези. У той же час результат несумісний з висунутою гіпотезою, цілком достатній для відхилення її як хибної.

Звичайно помилки завжди можуть мати місце, проте висловлене положення є важливим та потребує ретельного обґрунтування. Причина того, що результат, який підтверджує висунуту гіпотезу, не обов'язково може бути основою її прийняття тому, полягає в тому, що даний результат може бути спільним також з іншими гіпотезами і отже, не обов'язково може бути доказом справедливості цієї гіпотези проти інших висунутих альтернатив.

Наприклад, проведення замірів 53 ранніх приростів у 100 відібраних дендрохронологічних кернах сосни звичайної спільно з гіпотезою про те, що кількість продуктивних ранніх приростів в умовах досліджуваного деревостану приблизно однакове. Цей результат спільний і з припущенням, що пізніх приростів з достатньою шириною, яка забезпечує високу продуктивність сосни звичайної не набагато, але значно більше за ранні. Таким чином, результат, спільний з спочатку висунутою гіпотезою, не є стовідсотковим доказом її достовірності. Навіть визначення 50 пізніх продуктивних приростів не може бути підставою для висновку про те, що в деревостані сосни звичайної є строго однакова кількість продуктивних ранніх та пізніх приростів. З іншого боку, обстеження 90 пізніх приростів при 100 обстежених могло б на практиці стати причиною для спростування гіпотези про однакову кількість продуктивних ранніх та пізніх приростів у досліджуваному сосновому деревостані. У наступному прикладі припустимо, що середній коефіцієнт продуктивності дубового деревостані становить 100. Результат вибірки приростних кернів показав, що середній показник дорівнює 102, сумісний з висунутою нами гіпотезою. Однак цей результат спільний також і з припущенням, що середній коефіцієнт продуктивності дорівнює 101 або 99, і звичайно ж спільний з гіпотезою про те, що середній показник вихідної сукупності становить 102. Отже, даний результат аж ніяк не може бути свідченням перевагу гіпотези, відповідно до якої середній коефіцієнт дорівнює 100. Допустимо тепер, що деяка вибірка дала середній коефіцієнт продуктивності, що дорівнює 135. Припустивши, що обсяг вибірки був досить великим, ми могли б показати таке: якщо вихідна гіпотеза є достовірною, ми практично ніколи не отримали б такого результату. На підставі цього висновку отриманий результат цілком обґрунтовано може бути використаний нами як свідчення помилковості висунутої гіпотези, причому ризик помилки в рази був би мінімальним. Все вищесказане ґрунтується на вже висловленому факті, що зазвичай результат експерименту, спільний з висунутою гіпотезою, виявляється також спільним і з іншими гіпотезами. В результаті подібний результат не може бути прийнятий як обґрунтування переваги деякої гіпотези перед іншими

гіпотезами. Однак ми завжди можемо отримати результат, що розходиться з висунутою гіпотезою, який може викликати суттєві сумніви щодо її достовірності. Гіпотезу можна порівняти зі свідченнями обвинуваченого у суді. Він неспроможний довести істинності своїх слів. Водночас деякі наведені ним факти залишають відкритою можливість для висування припущень у тому, що він міг діяти не так, як описував. І прокурор може поставити під сумнів точність його розповіді, показавши, що можна інтерпретувати наведені факти інакше. Водночас для доказу вини обвинуваченого має бути надано незаперечні докази. У всіх цивілізованих країнах діє принцип «презумпції невинності». Це означає, що всі сумнівні гіпотези трактуються на користь обвинуваченого з метою виключення можливості засудження невинного, тобто дотримується правило, що для суспільства менш шкідливо не засудити винного, ніж покарати невинного.

В біометрії для доказу деякого твердження часто застосовують метод, відомий у математиці, як «доказ від зворотнього». Для цього як робочий інструмент використовують так звану «нульову гіпотезу». Пояснимо її суть.

Нульова гіпотеза.

Коли ми не можемо відкинути гіпотезу, ми цим визнаємо, що ця гіпотеза може виявитися вірною. З іншого, якщо ми можемо відкинути висунуту гіпотезу, то тим самим робимо цілком певний висновок про її хибність. Останнє є дуже важливим. При перевірці гіпотез ми можемо зробити остаточний висновок лише в тому випадку, коли можемо остаточно відкинути висунуту гіпотезу. Отже, мета проведеного нами експерименту має полягати в спростуванні висунотої гіпотези. Це означає, що в якості гіпотезу ми повинні сформулювати припущення, альтернативне тому, в що ми віримо. Наприклад, якщо треба показати, що дерева сосни звичайної загалом вищі ніж дерева берези повислої, то висунемо гіпотезу про відсутність відмінностей у їхньому зростанні. Потім спробуємо відкинути цю гіпотезу. В іншому випадку, щоб довести, що між анатомічною будовою деревини сосни та берези є істотні відмінності, потрібно перевірити гіпотезу, що між ними немає відмінностей. І знову ми повинні спробувати цю останню гіпотезу, щоб цим встановити істинність нашого вихідного припущення.

Гіпотеза, відповідно до якої відсутні відмінності між різними сукупностями, називається нульовою гіпотезою.

Гіпотеза, яку ми зможемо перевірити, не може бути сформульована на основі будь-якого судження. Ми це вже спостерігали з прикладу дослідження ранніх та пізніх приростів в структурі відібраного дендрохронологічного керна. Судження про те, що пізніх приростів та ранніх приростів однакової продуктивності при відбрі кернів з 50 ранніми та стільки ж пізніми приростами недостатньо, щоб на його основі можна було

сформулювати деяку певну гіпотезу. Подібні випадки звичайні в практиці досліджень. Зі сказаного слід, що експериментатор повинен сформулювати альтернативу тому, що він намагається довести, у вигляді цілком певної гіпотези. Тільки якщо це можливо, він може спробувати відкинути її для того, щоб довести справедливість своїх вихідних припущень.

Таким чином, перший крок, зроблений експериментатором, має полягати у формулюванні статистичної гіпотези, яку він сподівається спростувати для того, щоб показати істинність свого вихідного припущення. Після цього він зможе застосувати процедуру перевірки гіпотези.

В біометрії (статистиці) застосовуються досить конкретні гіпотези, пов'язані із проведенням числових обчислень. З цього випливає, що поняття статистичної гіпотези вужче, ніж поняття наукової гіпотези взагалі, і передбачає можливість статистичного експерименту для об'єктивного підтвердження (або відхилення) припущення, що розглядається. *Інакше кажучи, статистичні гіпотези відносяться до статистичних моделей.*

Прикладами гіпотез такого роду є припущення щодо параметрів розподілу – середнього, дисперсії тощо (параметричні гіпотези), або щодо типу розподілу чи зв'язку – непараметричні гіпотези. Так, параметричними гіпотезами є твердження: середнє значення в деякій генеральній сукупності дорівнює числу а (позначається H_0:X = a), середнє дисперсії двох вибірок рівні (не рівні) між собою (H_0:X_1 = X_2, H_0:σ^2_1 = σ^1_2). Непараметрична гіпотеза, наприклад, одна з наступних: розподіл діаметра даного деревостану підпорядковується нормальному закону зростання деревостану в висоту є експонентною кривою тощо.

Гіпотези називають простими, якщо вони відносяться до конкретного значення параметра (числа); складні гіпотези представляють поєднання простих.

З експерименту, тобто вибіркових даних вирішують питання: прийняти чи відкинути гіпотезу, тобто свідчать отримані дані «за» або «проти» випробуваної гіпотези. Для вирішення цього питання мало розглядати лише перевірену гіпотезу H_0, що перевіряється. Необхідно знати і область «проти» гіпотези H_0. Проте гіпотез може бути кілька, що виключає її альтернативну гіпотезу H_a. Для перевірки, необхідно вибрати статистичну характеристику критерію - показник, що розділяє зоні, кожена з яких свідчить на користь гіпотезі H_0 або H_a. Якщо йдеться про параметричні гіпотезах, то в якості статистичної характеристики зазвичай використовують певні значення аналізованого параметра, а основою для висновків служить розподіл статистики, що оцінює цей параметр. Тому перевірка гіпотез тісно пов'язана з інтервальним оцінюванням, але дозволяє робити глибші висновки. Розглянемо як приклад гіпотезу про те, що кількість опадів в травні – липні впливає на поточний приріст

деревостанів сосни звичайної в умовах Перганського ПНДВ Поліського природного заповідника. Для цього розглянемо поточний приріст сосни звичайної в лісорослинних умовах B_{2-3} пробної площі №2 Перганського ПНДВ за різні роки з різною кількістю опадів, що випали за досліджуваний період, наприклад, 200 мм і 600 мм. Ми вважаємо, що при 600 мм опадів приріст буде вищим. Перевірці підлягає затвердження, що середнє значення поточного приросту (обумовленого впливом опадів) в генеральній сукупності дорівнює 0, тобто випробуваною є гіпотеза Но: $X = 0$ проти альтернативної На: $X \neq 0$. Вибір такої альтернативи говорить про те, що нас цікавлять як позитивні відхилення від гіпотези, що перевіряється (опади в кількості 600 мм збільшують приріст), і негативні. В цьому випадку перевірку гіпотези називають двосторонньою. Якби нас цікавили відхилення в один бік - лише позитивні (тоді альтернативна гіпотеза На: $X>0$) або від'ємна (На:$X<0$), то перевірка була б односторонньою. В даному випадку випробувана гіпотеза проста, а всі три альтернативні - складні. Основні ідеї перевірки гіпотез розглянемо з прикладу середнього значення. Нехай перевіряється гіпотеза Но : $X = Xo$ та розподіл статистики X відомий (рис. 9.1).

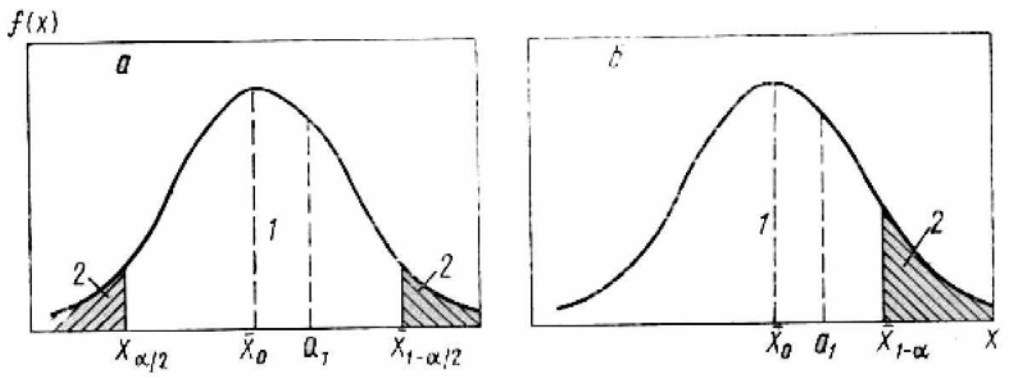


Рис. 9.1 Зони прийняття (1) та відхилення (2) гіпотез за рівня значущості а:
а – двостороння перевірка; б – одностороння.

На підставі вибірки отримано вибіркове середнє $X = a_1$. Якщо a_1 не відрізняється сильно від Хо, то природно вважати, що експериментальні дані не суперечать гіпотезі, що перевіряється, в іншому разі її відхиляють. В оцінку величини відмінності вкладається більш конкретний зміст, а саме: ще до отримання вибірки задаються деякою ймовірністю *а*, яка поділяє розподіл статистики на дві зони. У першу - відносять всі значення статистики, які вважаються практично можливими (область допустимих значень), іншу - ті значення статистики, поява яких в окремому випробуванні (на основі однієї вибірки) визнаються практично неможливими за умови, що гіпотеза яку перевіряють вірна. Тому величина *а* повинна бути досить мала, наприклад 0,05 та 0,01. При двосторонній перевірці критична область має вигляд:

$$p(|\tilde{X}|>x_{1-\alpha/2}) = p(\tilde{X}<x_{\alpha/2}) + p(\tilde{X}>x_{1-\alpha/2}) = \alpha/2+\alpha/2=\alpha$$

При лівосторонній та правосторонній односторонніх перевірках відповідно:

$$p(\tilde{X}<x_{\alpha}) = \alpha; \qquad p(\tilde{X}>x_{1-\alpha}) = \alpha$$

де: $\tilde{X}$ - вибіркове значення статистики;

x_a та x_{1-a} - відповідні квантили розподілу даної статистики (статистичні характеристики критерія).

Далі обчислюють конкретне вибіркове значення X. Якщо воно потрапляє в область допустимих значень – гіпотеза не відхиляється, якщо в критичну - гіпотеза відхиляється, у зв'язку з чим ці дві області називають відповідно областю прийняття та неприйняття гіпотез.

Число α називають рівнем значущості критерію. Від його величини залежить рішення щодо випробуваної гіпотези. Якщо гіпотеза Но вірна, то α дає нам ймовірність того, що статистика потрапить в критичну область, і правильна гіпотеза помилково буде відкинута: при α=0,001 - в одному випадку з 1000, при α=0,05 - у п'яти випадках зі 100 і т.д. Слід розрізняти рівні значущості (α), рівень достовірності (р) та довірчий коефіцієнт (t). Між ними є тісний зв'язок, яка видно з таблиці 9.1

Таблиця 9.1

Співвідношення між різними критеріями оцінки статистичних величин

Рівень значущості, α	Рівень достовірності, P	Довірчий коефіцієнт, t
32%	68%	1,00
5%	95%	1,96
1%	99%	2,58
0,1%	99,9%	3,39

Дамо пояснення до таблиці 9.1. Рівень значущості (α) – це значення ймовірності, яка показує, що відмінності між середніми значеннями можна вважати несуттєвою. Рівень достовірності (Р) – це випадкова величина, для якої відомий закон її розподілу. Зазвичай використовують його критичні значення для певного рівня значущості (α) та числа ступенів свободи (γ).

Наприклад, t = 1 - критичне значення t-критерію Стьюдента.

Методи оцінки достовірності поділяються на параметричні та непараметричні, про що йтиметься нижче.

Чим менший рівень значущості, тим менша ймовірність помилкового відхилення правильної гіпотези. Однак зменшення величини рівня значимості не завжди доцільно. Якщо нульова гіпотеза неправильна (наприклад, середня генеральна сукупність насправді відрізняється від Хо), то зі зменшенням α, зменшується критична область, і зростає область допустимих значень, тобто статистика при дуже малих α потрапить в область допустимих значень перевіреної гіпотези $H_0 : X = X_0$ і остання не відхиляється, будучи насправді хибною. Тому мало перевірити гіпотезу H_0, одночасно потрібно відчувати альтернативну гіпотезу На. Тільки в такому разі можна оцінити ризик відхилення гіпотези, коли вона вірна, або прийняття гіпотези, коли вона невірна, а вірна альтернативна.

Отже, в оцінці гіпотез можливі помилки двох типів:

1) гіпотеза вірна, але відкидається - ймовірність цієї помилки оцінюється рівнем значущості та дорівнює α. Величина 1-α дає нам можливість прийняти гіпотезу, якщо гіпотеза вірна;

2) гіпотеза не вірна, але приймається - якщо позначити ймовірність помилки другого роду β , то 1-β (потужність критерію) є ймовірністю відхилити гіпотезу, якщо вона не вірна, а вірна альтернативна.

Співвідношення між помилками першого та другого порядку ілюструє рис. 9.2 стосовно середнього значення Х.

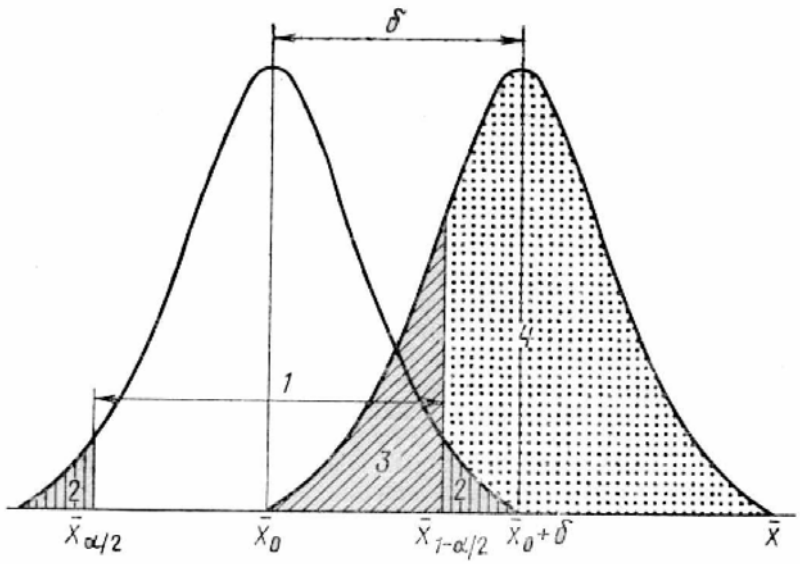


Рис. 9.2 Помилки першого та другого порядку:

1 – зона прийняття H_0; 2 - можливість помилки 1-го порядку;
3 – ймовірність помилки 2-го порядку; 4 – потужність критерію.

Якщо гіпотеза правильна, то зона 1-α дає можливість прийняти гіпотезу H_0 α=α/2+α/2 - рівень значимості чи ймовірність помилки 1-го порядку. Нехай насправді середня генеральна сукупність дорівнює Х+ð. Крива розподілу статистики Х не змінюється, але

центр розподілу зміщується на величину β. Тоді заштрихована площа β, що відповідає області допустимих значень перевіреної гіпотези, дасть можливість прийняти гіпотезу H_0 X = Xo, тоді як насправді вірна гіпотеза H_a: X = X0 + β, а площа 1-β дає величину потужності критерію.

Критичну область для гіпотези, що перевіряється, вибирають так, щоб забезпечувалася максимальна потужність критерію, який використовується. В такому разі, при заданому рівні значимості α гарантована мінімальна ймовірність помилки 2-го порядку β. Критерії, для яких забезпечується ця умова, що називають найбільш потужними.

З рис. 9.2 видно, що потужність критерію за інших рівних умов є функція ð, а величини α і β взаємопов'язані. Для вибірки фіксованого обсягу N (від якого також залежить розподіл даної статистики), зменшення ймовірності однієї з помилок веде до збільшення ймовірності іншої. Зменшити одночасно обидві ймовірності можна лише шляхом збільшення обсягу вибірки. Але це пов'язано з технічними та економічними проблемами, і тому не завжди можливо. Тому з урахуванням того, що практичні наслідки помилок 1 та 2-го порядку неоднакові, їх знаходять в такий спосіб. Якщо, наприклад практичний ризик, що пов'язаний з помилками 1-го порядку, більший, ніж ризик, пов'язаний із помилками 2-го порядку, то слід зменшити α за рахунок збільшення 1-β. Якщо ж видається можливість оцінити наслідки помилкових рішень чисельно (в кубометрах деревини, у вартісному вираженні або якось інакше), то співвідношення величин α та β може бути встановлено на цій основі.

Так, під час перевірки гіпотези про вплив добрив на приріст деревостанів помилка 1-го порядку призводить до відхилення гіпотези про те, що добрива не впливають на збільшення поточного приросту, хоча насправді це може бути не так, тобто впливу немає або він незначний. Ця помилка спричиняє невиправдані витрати на внесення добрив, які не збільшують продуктивність деревостанів. Помилка 2-го порядку призводить до втрати деякої додаткової кількості деревини. Очевидно, що наслідки помилки 1-го порядку істотніші. Співвідношення помилок 1 і 2-го порядку можна оцінити з урахуванням, з однієї сторони - вартості добрив та витрат на їх внесення, а з іншого - вартості додатково отриманої деревини з урахуванням того, що ця деревина може бути використана лише через А років, а А досягне 30 або 60 років. Звичайно, далеко не завжди вартість визначає рівень допустимих помилок. Наведемо дещо абстрактний приклад. Якщо у системі ППО перевіряють гіпотезу про наявність в зоні оборони ворожої крилатої ракети, то помилка 1-го порядку призведе до пропуску ракети до цілі, а помилка 2-го порядку - до оголошення помилкової тривоги, але обидві помилки небажані, хоча наслідки першої помилки найбільш значущі й приведуть до ураження та руйнувань.

9.2 Параметричні методи оцінки гіпотез.

Параметричні методи оцінки достовірності статистичних гіпотез базуються з урахуванням аналізу деяких параметрів вибіркової сукупності. Для застосування таких оцінок обчислюють середнє значення (X), середньоквадратичне відхилення (σ) або дисперсію (σ^2). Найчастіше вживаним методом параметричної оцінки є вже згаданий вище критерій Стьюдента. Цей критерій завжди позначається латинською літерою t, в інтерпретації автора критерію. Стьюдент встановив, що закон розподілу випадкової величини залежить від обсягу вибірки та основного відхилення. Опускаючи достатньо складні описи розподілу ймовірностей, які вивів Стьюдент, оскільки це виходить за межі щодо невеликого курсу лісової біометрії, значення 1 можна визначити за формулою:

$$t = \frac{\overline{X} - M}{\sigma} * \sqrt{N}$$

де: X – середнє значення вибіркової сукупноті;

σ – стандартне (середньоквадратичне) відхилення;

N – об'єм ряду розподілу;

M – середнє значення в генеральній сукупності.

В практиці найбільше значення t-критерія Стьюдента вибирають з спеціальних таблиць (додаток Е), де критична величина t-критерію визначається рівнем значущості (P) та числом ступенів свободи υ, де $\upsilon = N_1 + N_2 - 2$ де N_1 і N_2 - величина вибірок. Критерій Стьюдента (t) використовують для порівняння суттєвості різниці між середніми значеннями двох вибірок у таких варіантах:

1. При $N_1 = N_2$

$$|t| = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{m_1^2 + m_2^2}}$$

2. При $N_1 \neq N_2$

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma^2 * \frac{N_1 + N_2}{N_1 - N_2}}}, \text{ где}$$

$$\sigma^2 = \frac{\sum (X_i^1 - \overline{X}_1)^2 + \sum (X_i^2 - \overline{X}_2)^2}{N_1 + N_2 - 2}$$

де: N_1, N_2 - обсяг відповідних вибірок;

m_1 і m_2 - основні помилки середніх значень (X_1 та X_2) досліджуваних двох вибірок;

σ^2 – об'єднана дисперсія двох вибірок.

$$m_{\bar{X}_i} = \frac{\sigma_i}{\sqrt{N}}$$

Всіх, хто цікавляться описом рівнянь, виведених Стьюдентом, можуть знайти їх у книзі Митропольського А. К. «Техніка статистичних обчислень», а також в інших посібниках, наприклад, М. П. Горошко, С. І. Міклуш, П. Г. Хам'юк «Біометрія», які наведені у списку літератури. Поряд з перевіркою нульової гіпотези рівності середніх величин в генеральній сукупності, виконується перевірка рівності середньоквадратичної відхилення (σ) та коефіцієнта варіації (V). Це викликано тим, що при $X_1 = X_2$ σ_1 та σ_2, а, відповідно V_1 та V_2 можуть відрізнятися і доцільно знати, наскільки ці відмінності є суттєвими. Для цього використовують такі формули:

$$|t| = \frac{\sigma_1 - \sigma_2}{\sqrt{\frac{\sigma_1^2}{N} + \frac{\sigma_2^2}{N}}}$$

$$|t| = \frac{V_1 - V_2}{\sqrt{\frac{V_1^2}{N} + \frac{V_2^2}{N}}}$$

Критичні значення t при заданому рівні значимості беруть з вже згаданих спеціальних таблиць (додаток Е) для υ ступенів свободи, де $\upsilon = N_1 + N_2 - 2$. Порівнюючи обчислені та табличні величини t, обираємо нульову (Но) (нульову) або альтернативну гіпотезу, а саме:

$$t_i \le t_{кр} = H_0 ; \; t_i > t_{кр} = H_p$$

Критерій Фішера.

Крім критерію Стьюдента, інколи перевірка нульової гіпотези проводиться за критерієм Фішера. Цей критерій вважається більш точним при оцінці рівності дисперсій в генеральній та вибірковій сукупності або двох генеральних сукупностей.

Р. Фішер відкрив закон F-розподілу, який описав спеціальною F-функцією. Відзначимо, що функція Ф. Фішера (F) є безперервною і залежить від кількості ступенів свободи.

При вибірках малого розміру (n>30) важливість відмінності між стандартними відхиленнями σ_1 та σ_2 оцінюють за допомогою виразу:

$$t=(\sigma_1 - \sigma_2) / \sqrt{\sigma^2_{\sigma_1} + \sigma^2_{\sigma_2}}$$

де: $\sigma_{\sigma 1}$ і $\sigma_{\sigma 2}$ - помилки стандартного відхилення, що визначаються за формулою $p = (\sigma_x/x) * 100\%$.

При вибірках малого обсягу, різниці стандартних відхилень мають розподіл, що відрізняється від нормального, та розглянутий метод оцінки цих різниць (за довіряючими межами чи перевіркою H_0 на підставі t -критерію) є неточним. Тут можна скористатися формулою:

$$t=\frac{\bar{x} - a}{\sigma}\sqrt{n}$$

яка підпорядковується t-розподілу Стьюдента з k=n-1 ступенями свободи.

Р. А. Фішер запропонував замість різниці σ_1 та σ_2 оцінювати різницю $Z=\ln \sigma_1 - \ln \sigma_2$, що має нормальний розподіл та при вибірці середнього обсягу. При обчисленні Z можна користуватися десятичними логарифмами $Z = 1{,}15131 \lg (\sigma_1^2/\sigma_2^2)$. Критерій Фішера (F) для оцінки різниці між вибірковими дисперсіями зазвичай застосовують у вигляді, який запропонував Д. Снедекор:

$$F = \sigma_1^2 / \sigma_2^2$$

В рівнянні наведеному вище, значення $\sigma_1^2 > \sigma_2^2$. Критичні значення F для різних рівнів значимості в практиці визначають за спеціальними таблицями в залежно від числа ступенів свободи υ_1 та υ_2. При цьому $\upsilon_1 = N_1 - 1$, а $\upsilon_2 = N_2 - 1$. Першою сукупністю (N_1) буде та, де величина σ_2 більша ($\sigma_1^2 > \sigma_2^2$). Для рівнів достовірності 0,95, 0,99 та 0,999 критичні значення F наведено в додатку Ж. Після порівняння обчисленого та критичного (табличного) значень F вибирають нульову (Но) або робочу (альтернативну) - Нр-гіпотезу. В такому вразі вона має вигляд:

$$F_i \le F_{кр} = H_0;\ F_i > F_{кр} = H_p$$

Значимість відмінностей якісних ознак.

Якісні ознаки, що розподіляються за моделлю біноміального розподілу, оцінюють на основі часток. Методи оцінки аналогічні вищерозглянутим для середніх, виражених кількісно.

Помилка різниці вибіркових часток p_1 і p_2 визначається за такою формулою:

$$\sigma_{p1\text{-}p2} = \sqrt{\frac{p_1(1-p_1)}{N_1} + \frac{p_2(1-p_2)}{N_2}}$$

Для критерія 1; з числом ступенів свободи $\upsilon = N_1\text{-}N_2\text{-}2$ маємо вираз:

$$t=(p_1\text{-}p_2) \,/\, \sigma_{p1\text{-}p2}$$

Коли є одна вибірка, значення її середньої частки може бути оцінено шляхом порівняння з гіпотетичною (теоретичною) часткою. Наприклад, щодо теоретичної частки закладання продуктивних ранніх приростів під час дендрохронологічного моніторингу соснових деревостанів в умовах пробної площі №2 Перганського ПНДВ може бути висунута нульова гіпотеза $N_0 : P = 0{,}5$. В цьому випадку критерій

$$t=P\text{-}p \,/\, \sqrt{[p(1-p)] \,/\, N}$$

де: P – теоретична частка або ймовірність;
p – вибіркова частка;
N – чисельність вибірки.
При $N_0 : P = 0$,

$$t=p \,/\, \sqrt{[p(1-p)] \,/\, N}$$

9.3 Непараметричні методи оцінки гіпотез.

Непараметричні критерії оцінки статистичних гіпотез не вимагають обчислення показників (X, σ, υ) у вибірковій сукупності. Вони не базуються на нормальному розподілі випадкових величин в сукупності, і часто застосовують інші закони розподілу. У ряді випадків для визначення цих оцінок використовують умовні значення, порядкові номери і т.д. В сучасній практиці математичної статистики з непараметричних критеріїв, найчастіше застосовують такі критерії: x - критерій Ван-дер-Вардена, T-критерій Уайта, Z - критерій знаків та W - критерій Вілкоксона. Не наводячи доведень відповідних теорем, дамо формули для їх застосування у практиці лісового господарства.

x-критерій Ван-дер-Вардена знаходять за формулою:

$$\bar{x} = \sum \psi\left(\frac{R}{N_1 + N_2 + 1}\right)$$

де: R – порядковий номер (ранг);
N_1, N_2 – об'єми вибірок;

Ψ - значення функції, визначене за спеціальною таблицею залежно від величини $R/(N_1+N_2+1)$ (додаток 3).

Х - критерій Ван-дер-Вардена, використовується для вибірок з непов'язаними парними варіантами. Для цього варіанти обох рядів ранжують у міру їх зростання. В результаті ранжування кожне значення x_i отримує порядковий номер (ранг). Потім знаходять відношення:

$$K = \frac{R}{N_1 + N_2 + 1}$$

Критичне значення х-критерію знаходять за спеціальною таблицею (Додаток І) для різних рівнів значущості (5%, 1%) з числом ступенів свободи ($Y \neq N_1+N_2-2$), й за різниці між обсягами виборок N_1-N_2. По х-критерію вибираємо нульову (H_0) та робочу (Нр) гіпотезу.

$$x \leq x_{кр} = H_0$$

$$x > x_{кр} = H_p$$

Пояснимо викладене прикладом. Нехай ми виміряємо висоти в двох соснових деревостанах II класу бонітету у віці 80-85 років, в типі лісу – бір чорничнo-брусничний (А ч-б), а також бір сфагново-брусничний (А с-б) (таблиця 9.2).

Таблиця 9.2

Результати вимірів висот у двох деревостанах сосни звичайної пробної площі №2 Перганського ПНДВ

Типи лісу	Висота виміряних дерев (x_i), м										Середнє значення (х)
А ч-б	22,0	22,5	21,7	17,9	20,8	28,4	25,6	19,6	23,6	20,9	22,3
А с-б	26,3	18,8	20,6	21,9	17,5	22,4	22,6	28,3	17,8	-	21,8

Проведемо ранжування даних таблиці 9.2 і знайдемо К та Ψ (таблиця 9.3).

В таблиці 9.3 ми обчислили $K_1 = R/N_1+N_2+1$ для другої (меншої) сукупності (N_2). Потім, використовуючи таблицю в додатку І, де наведено величини х, Ψ в залежності від величини R, знайшли

критерій Ван-дер-Вардена для наших дерев сосни звичайної. Він дорівнював - 1,2. За таблицею у додатку І знайшли критичні значення критерію х при N_1-N_2= 1. Він дорівнює (при 1% рівні значущості) для Y = 19 (У = N_1+N_2) 4,77. Так як у нас 1,20 < 4,77, то приймається нульова гіпотеза, тобто, що різниця при різниці в середніх значеннях 0,5 м незначні. Це означає, що обидві вибірки належать до однієї статистичної сукупності висот сосни звичайної II класу бонітету, а тип лісу не надав суттєвого впливу на величину середньої висоти.

Таблиця 9.3

Розрахунок х-критерію Ван-дер-Вардена

Висоти (м) за типами (x_i) лісу		Ранг дерева, (R)	$K_1 = \frac{R}{N_1 + N_2 + 1}$	$\psi(\frac{R}{N_1 + N_2 + 1})$
А ч-б	А с-б			
-	17,5	1	0,050	-1,64
-	17,8	2	0,100	-1,28
17,9	-	3	-	-
-	18,8	4	0,200	-0,84
19,6	-	5	-	-
-	20,6	6	0,300	-0,53
20,8	-	7	-	-
20,9	-	8	-	-
21,7	-	9	-	-
-	21,9	10	0,500	0,00
22,0	-	11	-	-
-	22,4	12	0,600	0,25
22,5	-	13	-	-
-	22,6	14	0,700	0,52
23,6	-	15	-	-
25,6	-	16	-	-
-	26,3	17	0,850	1,04
-	28,3	18	0,900	1,28
28,4	-	19	-	-
N =10	N_2=9	-	-	-1,20

Критерій Уайта.

Цей критерій також використовують для оцінки різниці між середніми значеннями (x_1 та x_2) двох вибірок з попарно незв'язаними варіантами. Схема обчислення по Т-критерію Уайта показана у таблиці 9.5. Для прикладу скористаємося наведеними вище даними вимірювання висот сосни звичайної (таблиця 9.2). Наведемо парні величини висот у порядку їх зростання та покажемо обсяг кожної вибірки (таблиця 9.4).

За підсумками таблиці 9.4 побудуємо таблицю 9.5.

В таблиці 9.5 ми виписали ранги дерев по мірі зростання висот:

X (ранги) рівні від 1 до 19, так як $N_1 + N_2 = 19$. Потім проти кожного дерева вказуємо номер вибірки, з якої його взято (графа 2). У графі 3 сумуємо ранги дерев (графа 1), які розміщені попарно в таблиці 9.4.

Таблиця 9.4

Попарні виміри висот (Hi) сосни звичайної пробної площі №2 Перганського ПНДВ в міру зростання

№ ряду	Висоти, м										Об'єм вибірки,	Сума, x_i	Середнє начення
1	17,9	19,6	20,8	20,9	21,7	22,0	22,5	23,6	25,6	28,4	10	223	22,3
2	17,5	17,8	18,8	20,6	21,9	22,4	22,6	26,3	28,3	-	9	196,2	21,8

Наприклад, ранг дерева з висотою 20,8 м у першому ряду дорівнює 7, а парного йому дерева у ряду 2 (18,8 м) - 4. Середній ранг цієї пари буде (7+4)/2=5,5. У графи 5 та 6 виписуємо середні ранги, що належать першій та другій вибіркам.

Таблиця 9.5

Порядок розрахунків для знаходження Т-критерію Уайта

Ранги	№ вибірки	Висота, м	Сумісний ранг	Ранги для вибірки 1	Ранги для вибірки 2
1	2	3	4	5	6
1	2	17,5	2	2	2
2	2	17,8	3,5	3,5	3,5
3	1	17,9	2	5,5	5,5
4	2	18,8	5,5	7,0	7,0
5	1	19,6	3,5	9,5	9,5
6	2	20,6	7,0	11,5	11,5
7	1	20,8	9,5	13,5	13,5
8	1	20,9	9,5	16	16
9	1	21,7	11,5	17	17
10	2	21,4	11,5	19	85,5
11	1	22,0	13,5	104,5	6
12	2	22,4	13,5	5	2
13	1	22,5	16	2	3,5
14	2	22,6	16	3,5	5,5
15	1	23,6	17	5,5	7,0
16	1	25,6	19	7,0	9,5
17	2	26,3	173	9,5	11,5
18	2	28,3	4	11,5	13,5
19	1	28,4	2	13,5	16
Всього (Σ):	-	-	3,5	16	17

Перевірка правильності обчислень проводиться за формулою:

$$\sum R = \frac{(N+1)*N}{2}$$

де $N=N_1+N_2$ В нашому прикладі ΣR – 104,5+85,5-190.
N+1/2=(20*19)/2=380/2=190.

Таким чином, необхідну рівність дотримано, тобто обчислення зроблено правильно. За значення Т-критерію Уайта приймається менша сума рангів. У нашому прикладі це 85,5. За таблицею у додатку К знаходимо критичне значення Т-критерію для більшого (N_1) та меншого (N_2) обсягів вибірки при 1% рівні значимості. Для N_1| =10 та N_2 = 9, Т-критерій Уайта дорівнює 58. Порівнявши обчислену нами величину (R_{min} = 85,5) з критичним значенням по Т-критерію при 1% рівні значущості, бачимо, що $T_{факт} > T_{крит}$, тобто 85,5>58. Таким чином, Т-критерій Уайта також підтверджує, що вимірні висоти сосни звичайної у віці 80-85 років з типів лісу - бір чорничнo-брусничний (A_{1-2} ч-б), а також бір сфагново-брусничний (A_{2-3} с-б) належать до однієї статистичної сукупності, яка визначається другим класом бонітету.

Крім вище зазначених методів оцінки гіпотез, існують Z-критерий знаків, W-критерий Вілкоксона. Але в данму розділі ми їх не розглядатимемо, так як вони не знайшли широкого застосування у практиці статистичної обробки числових даних в лісовому господарстві.

9.4 Перевірка статистичних гіпотез в практиці лісового господарства та їх використання а в лісовому господарстві.

В практиці лісового господарства, особливо під час проведення наукових досліджень, часто виникають питання оцінки ефекту від проведених лісогосподарських заходів або від заподіяної шкоди: рубки догляду, застосування добрив, селекція, меліорація, шкідники та хвороби лісу, лісові пожежі тощо. Приклади таких оцінок за допомогою критеріїв Стьюдента та Фішера, а також при використанні непараметричних оцінок наведено вище. Найчастіше, як зазначалося вище, для досягнення цієї мети використовують параметричні оцінки як найбільш об’єктивні. Тут ми опишемо типову методику проведення таких оцінок, яка принципово застосовується до більшості оцінок гіпотез у лісовому господарстві.

У дослідах найчастіше виникають проблеми оцінки ефектів, наприклад, між висотами, діаметрами, приростами дерев, які отримали різні дози добрив, що залишаються при проведенні рубок догляду різної інтенсивності, утворюються парні спостереження, де одне з них відноситься до першого варіанту досліджень (зазвичай це контроль), а інше до другого. Різниці між значеннями ознак по парах утворюють вибірку, аналізуючи яку за допомогою t-критерію

Стьюдента, F-критерію Фішера або непараметричних критеріїв, роблять відповідний висновок. Різниці у дослідах можуть бути наслідком досягнутого ефекту, але бувають через випадкові причини, які зазвичай залишаються невідомими. Якщо б діяли лише випадкові причини, то за законами теорії ймовірності вони мали б різні знаки, і їхня середня в одній вибірці дорівнювала 0. Якщо ж середня тут не дорівнює 0, її значущість потрібно оцінити. Методику такої оцінки покажемо на штучній моделі двох вибірок із однієї сукупності. Припустимо, закладено пробну площу насаджень сосни звичайної у віці 95 років II класу бонітету на площі 1 га - 100 х 100 м. в лісорослинних умовах B_{2-3}. На цій пробній площі виміряно 238 діаметрів та висот дерев сосни звичайної (таблиця 9.6). Дані виміри приймемо як генеральну сукупність та знайдемо її статистичні оцінки. Для спрощення зробимо угруповання діаметрів ступенями товщини через 4 см (таблиця 9.6), а висоти по ступенях висоти через 2 м. На основі розподілу згрупованих даних обчислимо статистичні показники.

Узагальнюючи викладене у цьому розділі, наведемо орієнтовний алгоритм перевірки статистичних гіпотез, що використовуються у лісовому господарстві.

При проведенні досліджень, визначенні цілей та конкретних шляхів його проведення, слід розглядати можливість статистичного підтвердження результатів, тобто дослідження в результаті має зводитися до перевірки (однієї чи кількох) статистичних гіпотез явно сформульованих. Для лісівничих досліджень попередній аналіз такого роду тим паче важливий, так як у багатьох ситуаціях математична формалізація досить складна. Крім зазначених вище труднощів переходу від об'єкта до моделі (часто перевірка гіпотез стосується саме щодо обґрунтування можливості подання даного об'єкта за допомогою певної статистичної моделі), подальші труднощі випливають з необхідності знання вибіркових розподілів, що використовуються статистикою. Тут для простих випадків можуть бути запропоновані стандартні методи, наприклад, використання передумови про нормальність розподілу статистик, обчислених виходячи з великих вибірок. Перевірку статистичних гіпотез у дослідженнях з лісового господарства зазвичай проводять готових результатів спостережень (досліджень).

Порядок роботи при цьому наступний:

✓ вибирають статистичну характеристику критерію, з'ясувавши відповідність структури завдання передумов, що лежать в основі застосовуваних критеріїв. Зокрема, це можуть бути вибіркові розподіли статистик, незалежність спостережень та ін.

✓ вибирають і формулюють H_0 та На – перевіряється альтернативна гіпотеза. Нульову гіпотезу зазвичай приймають таким чином, щоб помилки 1-го порядку були суттєвішими, ніж результати помилки 2-го порядку.

✓ на основі результатів помилок 1 та 2-го порядку встановлюють допустимий рівень значущості α (односторонній або двосторонній залежно від типу альтернативної гіпотези) та обчислюють критичні значення статистичної характеристики, тобто ті значення, які розділять розподіл статистичної характеристики на область допустимих значень та критичну.

Таблиця 9.6

Відомість вимірювання діаметрів (D) та висот (H) сосни звичайної
в умовах 10 закладених пробних площ Перганського, Копищанського, Селезівського ПНДВ
Поліського природного заповідника
(за результатми досліджень 2022-2025 рр.)

№ дер.	D, см	H, м	№ дер.	D, см	H, м	№ дер.	D, см	H, м	№ дер.	D, см	H, м	№ дер.	D, см	H, м	№ дер.	D, см	H, м	№ дер.	D, см	H, м
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	44,3	25,4	35	22,6	22,1	69	16,3	17,1	103	40,0	28,1	137	11,8	16,5	171	45,9	28,8	205	28,2	24,4
2	23,6	22,7	36	25,4	24,1	70	21,5	21,0	104	28,1	23,3	138	40,5	28,3	172	25,9	24,0	206	32,7	25,8
3	18,2	17,4	37	24,4	25,7	71	28,8	26,0	105	44,0	28,3	139	40,0	27,5	173	32,0	24,6	207	24,4	25,3
4	36,6	18,7	38	10,5	29,3	72	28,0	26,2	106	33,5	24,6	140	36,3	26,6	174	24,9	24,1	208	38,8	27,6
5	30,5	26,0	39	36,7	28,4	73	24,3	20,5	107	36,4	26,7	141	33,0	25,7	175	39,6	27,3	209	24,0	23,1
6	33,4	26,5	40	30,1	24,7	74	30,2	25,8	108	28,7	24,1	142	44,1	29,6	176	33,8	25,6	210	36,2	25,6
7	40,1	29,1	41	33,0	26,6	75	36,0	27,4	109	32,4	25,8	143	30,7	25,8	177	28,8	26,1	211	27,0	24,2
8	29,3	24,7	42	32	27,1	76	34,1	28,1	110	45,6	29,3	144	40,5	30,2	178	26,1	23,3	212	30,3	27,0
9	28,4	24,4	43	27,6	22,3	77	33,2	27,5	111	29,4	25,6	145	28,4	25,1	179	44,3	27,2	213	31,9	27,2
10	27,5	24,2	44	28,1	24,6	78	38,5	27,7	112	29,6	25,5	146	42,6	30,8	180	31,7	26,0	214	27,4	25,0
11	15,5	16,0	45	31,4	26,4	79	40,9	28,8	113	24,3	23,3	147	18,7	20,3	181	24,8	23,3	215	43,3	28,5
12	32,2	25,5	46	16,3	25,1	80	49,9	30,6	114	18,1	20,7	148	36,0	26,1	182	39,0	26,3	216	23,3	24,0
13	30,6	25,3	47	21,7	20,2	81	45,7	30,1	115	44,8	30,0	149	42,7	30,3	183	28,0	23,0	217	31,1	26,6
14	28,5	24,7	48	34,5	26,3	82	12,2	16,1	116	26,1	27,1	150	33,1	25,7	184	30,6	24,0	218	22,1	22,7
15	27,1	24,2	49	37,7	26,8	73	14,7	15,5	117	24,5	20,9	151	34,0	25,6	185	24,0	23,5	219	36,0	25,8
16	26,6	23,7	50	19,5	19,5	84	26,6	24,8	118	40,6	27,3	152	36,2	26,8	186	48,1	31,0	220	23,5	23,0
17	36,8	27,1	51	18,2	18,4	85	32,0	26,3	119	25,6	23,5	153	38,3	27,0	187	32,0	25,1	221	30,6	25,1

Продовження таблиці 9.6

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
18	34,1	26,5	52	24,4	22,3	86	33,4	26,4	120	32,0	25,0	154	33,4	26,6	188	23,5	23,1	222	28,7	25,0
19	32,5	28,8	53	41,2	30,5	87	34,5	26,8	121	39,7	28,7	155	36,1	26,9	189	36,9	26,2	223	45,7	28,5
20	30,6	25,4	54	40,1	29,3	88	37,8	26,6	122	26,8	24,4	156	27,7	24,4	190	30,1	21,4	224	24,0	22,8
21	33,8	26,0	55	20,5	18,4	89	42,2	29,3	123	36,0	27,8	157	44,0	30,0	191	32,8	25,8	225	33,3	26,1
22	23,0	21,7	56	20,4	18,9	90	33,5	27,2	124	19,4	21,5	158	32,1	25,0	192	35,1	24,3	226	28,8	26,0
23	17,7	17,0	57	32,5	27,0	91	16,6	19,1	125	40,1	27,1	159	25,0	22,1	193	20,6	20,5	227	40,3	28,2
24	31,5	25,6	58	31,9	25,2	92	36,1	26,5	126	20,0	18,8	160	36,0	27,3	194	22,1	22,0	228	33,8	27,6
25	11,3	14,6	59	30,6	25,8	93	18,0	20,3	127	33,7	24,7	161	29,1	25,3	195	18,1	20,0	229	32,2	26,8
26	39,8	28,8	60	29,5	24,1	94	36,5	32,0	128	36,1	25,8	162	32,3	25,4	196	40,2	19,1	230	29,5	26,2
27	22,5	20,5	61	26,1	23,5	95	30,8	25,4	129	22,0	22,0	163	37,0	26,7	197	25,7	24,1	231	40,6	28,5
28	20,0	21,6	62	27,0	24,0	96	40,6	27,5	130	40,1	28,8	164	32,0	24,4	198	28,0	23,4	232	43,9	30,4
29	16,4	18,0	63	28,1	24,7	97	26,6	18,2	131	20,5	20,4	165	31,6	24,6	199	19,9	18,6	233	34,0	27,7
30	37,2	28,4	64	34,6	27,4	98	36,0	25,7	132	34,4	26,5	166	38,0	26,7	200	34,9	25,3	234	32,1	27,6
31	35,4	28,1	65	36,6	28,2	99	37,2	28,9	133	40,8	27,4	167	28,2	25,0	201	24,0	25,0	235	38,8	23,5
32	31,1	26,5	66	40,8	29,0	100	16,2	19,7	134	20,7	20,8	168	32,0	25,0	202	26,3	22,4	236	40,1	29,4
33	30,8	25,8	67	23,0	22,4	101	15,1	16,8	135	19,4	17,6	169	36,2	25,5	203	22,8	20,5	237	28,4	26,6
34	29,1	25,5	68	28,5	26,6	102	44,8	29,5	136	20,2	19,6	170	28,3	24,4	204	36,1	26,7	238	36,3	27,7

✓ обчислюють вибіркове значення статистичної характеристики і перевіряють гіпотезу. Якщо вона не відхилена (отримане вибіркове значення належить області допустимих значень), то обчислюють зв'зок критерію 1-β; за достатньої міцності цього зв'язку гіпотезу приймають, інакше висновок залишається невизначеним через, що в такому разі потрібне збільшення обсягу вибірки.

При плануванні вибірки об'ємом N можна визначити потужність критерію або відповідну цій потужності різницю між параметрами, які критерій може вловити. Якщо потужність виявляється недостатньою, то підвищують об'єм вибірки, що забезпечує необхідну потужність. При рівні значущості, що дорівнює або менше величини помилки 2-го порядку, невідхилення гіпотези означає її прийняття, бо в такому разі критична область альтернативної гіпотези при цьому рівні значущості відповідає області допустимих значень перевіреної гіпотези. Необхідно наголосити, що термін «прийняття» гіпотези не є синонімом абсолютної її істинності H_0, оскільки Н0 і На, як правило не вичерпують всіх впливів та обставин, що позначаються на досліджуваному явищі. Вони справедливі лише тому, що дані експерименту не суперечать (або суперечать) гіпотезі, що перевіряється. Якщо для визнання невірним будь-якого положення достатньо навести один приклад, то в такому разі відхилення гіпотези (з ймовірністю наявності помилки 1-го порядку) є свідченням того, що вона - хибна. У цьому випадку будь-яка кількість підтверджень може виявитись недостатнім для того, щоб гіпотеза була справедливою.

Під час проведення досліджень часто обмежуються лише частковим використанням вищенаведених положень, застосовуючи один із методів порівняння. Але при складних випадках, і за високої практичної важливості, дослідженням необхідно робити повний цикл перевірки статистичних гіпотез.

РОЗДІЛ 10. КРИТЕРІЇ УЗГОДЖЕННЯ. СТАТИСТИЧНА ОЦІНКА В ЛІСОВОМУ ГОСПОДАРСТВІ

10.1 Критерії узгодження.

Статистичні критерії є найважливішою частиною оцінювання рядів розподілу. У багатьох випадках на підставі деяких гіпотез або якихось даних робиться припущення про вид законів розподілу випадкової величини X. Ми часто маємо дослідні дані у вигляді емпіричних дискретних рядів розподілу за якими будуємо криву розподілу нормального типу А, β- розподілу, логнормальну, Пуассонівську і т.д. Але тут без додаткових досліджень не можна стверджувати, що застосована для визначення розподілу крива відповідає існуючому закону розподілу досліджуваної випадкової величини X. Наприклад, зробивши перелік діаметрів дерев, ми залежно від їх віку, деревостану, його походження, повноти, лісорослинних умов можемо (побудувати) апроксимувати емпіричний розподіл різними кривими. Заздалегідь не можна з повною впевненістю заявити, який розподіл відповідає нашому ряду розподілу вимірюваних діаметрів. Тому виникає потреба перевірити відповідність теоретичного розподілу дослідним даним. З огляду на обмежену кількість спостережень, теоретична крива буде певною мірою відрізнятися від фактичного розподілу дослідних даних, навіть якщо припущення про закон розподілу зроблено правильно. У зв'язку з цим, виникає необхідність вирішувати таке завдання: чи є розбіжність між дослідним законом розподілу та передбачуваним законом розподілу наслідком обмеженої кількості спостережень, тобто, чи залежить від випадкових причин, які істотно не впливають на характер розподілу, або воно є суттєвим і пов'язане з тим, що є дійсне за результатми досліджень (спостережень).

Розподіл випадкової величини відрізняється від передбачуваного. Для вирішення поставленого завдання є критерії узгодження. Ідея цих критеріїв полягає в тому, що на підставі певного статистичного матеріалу вони дозволяють перевірити гіпотезу H, яка складається у цьому, що випадкова величина X має функцію розподілу F(x). Для того, щоб прийняти або спростувати гіпотезу H, розглядатимемо випадкову величину Y, що характеризує ступінь розходження теоретичного та статистичнго розподілів. Величину Y можна вибирати у різний спосіб. Наприклад, як Y можна взяти максимальне відхилення статистичної функції розподілу F(x). Очевидно, закон розподілу випадкової величини Y

залежить від закону розподілу випадкової величини X, над якою проводили досліди, та від числа дослідів n.

Припустимо, закон розподілу випадкової величини нам відомий. Тоді нехай в результаті проведених n дослідів над випадковою величиною X величина Y прийняла деяке значення. Постає питання, чи можна пояснити прийняте значення Y=y випадковими причинами, або це значення занадто велике, і вказує на наявність суттєвої різниці між теоретичним та статистичним розподілами, тобто непридатність гіпотези H.

Для відповіді на це питання припустимо, що вірна гіпотеза H, і обчислимо ймовірність того, що випадкова величина Y за рахунок випадкових причин, пов'язаних з обмеженим обсягом дослідного матеріалу, набуде значення не менше, ніж спостерігається значення y, тобто обчислимо ймовірність P(У≥y). Якщо ця ймовірність мала, то гіпотезу H слід спростувати як малоправдиву, а якщо ж ця ймовірність досить значна, то експериментальні дані не суперечать гіпотезі H. Для обчислення ймовірності P(У≥y) необхідно знати закон розподілу випадкової величини, який, як ми вже зазначали раніше, залежить від закону розподілу випадкової величини X (функції розподілу F(x)) та від числа досліджень N. Виявляється, що за деяких способів вибору випадкової величини Y, її закон розподілу при досить великому N практично не залежить від закону розподілу випадкової величини X. Саме такими розбіжностями і користуються в математичній статистиці та біометрії в якості критеріїв узгодження, за якими оцінюють відповідність розподілу, отриманого в досліді теоретичному значенню. Критерії узгодження можна розділити на дві групи: в першій не використовують значення вибіркових статистик; в другій – критерії будують з використанням властивостей генеральної сукупності, тобто оцінок останніх на основі статистик. З критеріїв другої групи найцікавішими є критерії Колмогорова-Смирнова та χ^2 (критерій Пірсона) порівняння вибіркової і теоретичної функцій розподілу. Однак використання статистик замість параметрів, які невідомі в задачах апроксимації, зазвичай призводить до завищення (іноді досить значному) ймовірностей узгодження. З параметричних критеріїв найбільш загальновживаним та науково обґрунтованим є критерій узгодження Пірсона.

10.2 Критерій узгодження Пірсона.

Цей критерій було запропоновано К. Пірсоном в 1900 році. Він базується на визначенні певної статистики, яку автор визначив як χ^2. Опустимо докази для обґрунтування величини χ^2, які зводяться до дослідженого розподілу випадкової величини χ^2 з ν-ступенями свободи, через стислість курсу лісової біометрії. Тут же зазначимо, що критерій χ^2 представляє суму відношень між квадратами різниць

емпіричних та обчислених або очікуваних частот до всіх очікуваних частот розподілу:

$$\chi^2 = \sum \frac{(p - p')^2}{p'}$$

Тут Σ – знак суми; p – емпірична частота;
p’ - очікувана чи теоретично обчислена частота.

Якщо різницю між емпіричними та обчисленими частотами позначити через d, тобто прийняти p - p' = d, то формула набуває більш простий вигляд:

$$\chi^2 = \sum \frac{d^2}{p'}$$

Щоб визначити критерій χ^2 - квадрат, необхідно кожне відхилення емпіричного ряду від його очікуваного значення звести в квадрат і розділити на величину очікуваного значення, потім отримані результати скласти.

Коли емпіричні та обчислені чисельності повністю збігаються один з одним, різниця p - p ' = d дорівнює 0. Чим більша різниця між спостереженнями та очікуваними чисельностями, тим більше і зазначена різниця, отже й величина критерію χ^2, яка може зростати нескінченно. Оскільки відмінності між очікуваними та емпіричними частотами зводяться у квадрат, то значення критерію χ^2 може бути лише позитивними. Тому при встановленні різниці p - p' = d, знаки можна не враховувати. Переважне значення цього критерію полягає в тому, що він застосовується до оцінки досліджуваних і очікуваних значень у різних випадках, як при порівнянні даних по одному, так і кількох незалежних ознаках, а також і для порівняння даних досліду та контролю. Особливо часто критерій χ^2 використовується в генетиці для порівняльної оцінки результатів розщеплення. Знаходить своє широке застосування цей критерій у багатьох галузях лісового господарства, зокрема в лісовій таксації. Критерієм χ^2 часто користуються для оцінки вибіркових розподілів зі своїми теоретично обчисленими частотами.

Критерій узгодження χ^2 Пірсона для порівняння рядів розподілів визначають за однією з еквівалентних формул, першу з яких застосовують при порівнянні емпіричних n_i і теоретичних $\tilde{n}_i$ частот, другу – частот p_i^1 та ймовірностей p_i:

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(n_i - \widetilde{n}_i\right)}{\widetilde{n}_i} = N \sum_{i=1}^{m} \frac{\left(p_i - \widetilde{p}_i\right)^2}{\widetilde{p}_i}$$

де: N – загальна кількість спостережень;

m - кількість порівнюваних груп частот.

Критичні значення χ^2 беруть з таблиці (додаток Н) з урахуванням числа ступенів свободи (V). Якщо правильна нульова гіпотеза H_0, яка полягає в тому, що розподіл в генеральній сукупності відповідає обраному теоретичному закону розподілу, тобто. $H_0 : \chi^2 > 0$ (проти альтернативної чи робочої Нσ : $x^2 = 0$ - перевірку проводять односторонню. Так як χ^2 не може бути від'ємним, то величина χ^2 асимптотично підпорядковується розподілу χ^2 с v = m-*l*-1 ступенями свободи, де *l* - число параметрів, використаних під час розрахунку теоретичного розподілу. Кількість параметрів (*l*) залежить від виду розподілу. Для розподілів, які найчастіше застосовуються в лісовому господарстві, число параметрів таке:

- розподіл Пуассона – 1(ð);
- нормальне, логарифмічне нормальне, біноміальне – 2 (ð, σ);
- узагальнене нормальне (типу А або Грамма-Шарльє), Вейбула, система кривих Пірсона або Джонсона - 4 (ð, σ, а, Е).

З різниці m-*l* віднімають одиницю, оскільки один додатковий зв'язок накладається на частоти, щоб суми частот вибіркового та теоретичного розподілів були рівними. Висновок розподілу критерію χ^2 отримують за умови, що всі порівнювані частоти не дуже малі, тому при використанні частоти n_i має бути більше 5, тобто $n_i, > 5$. У зв'язку з цим, в практичних завданнях щодо одновершинного розподілу, подібного до нормального, частоти крайніх інтервалів рядів розподілу необхідно об'єднувати. Звідси випливає, що одним із недоліків критерію χ^2 є його мала чутливість до відхилень частот для крайніх значень. Другий недолік - залежність від способу угруповання частот. В деяких роботах з математичної статистики стверджується, що найкращі результати дає використання замість інтервалів рівної довжини, інтервалів рівних ймовірностей. Останній спосіб поки не прижився у практиці через ускладнений синтез матеріалів так як необхідно розглядати впорядковану величину за зростаючими значеннями випадкових величин вибірки. Однак, незважаючи на наближений характер критерію χ^2, він зручний і достатньо надійний для отримання обґрунтованих висновків в завданнях лісового господарства.

Порядок використання χ^2 звичайний. Обчислюють статистичну характеристику критерію $\chi^2_{\text{обчисл.}}$ і порівнюють її з табличним (критичним) значенням χ^2_{1-a} (у) за рівня значущості α. Якщо

$\square^2_{\text{обчис}} <, \square^2_{1\text{-а}}$ (у), то гіпотезу приймають, тобто можна вважати, що емпіричний ряд підпорядковується обраному закону розподілу.

При виборі величини слід перевіряти гіпотезу. Але на той предмет, що між вибірковим і теоретичним рядами немає істотних відмінностей, тобто помилка 1-го порядку полягає в тому, що насправді існуюча узгодженість визнається хибною. Тоді помилка 2-го порядку зводиться до того, що хоча насправді ряд розподілів не підпорядковується цьому теоретичному закону, критерій визнає узгодженість. У цьому випадку помилка 2-го порядку найважливіша, ніж помилка 1-го прядку, тобто при формулюванні гіпотези про узгодження ми відступили від правил, викладених вище. Однак тут «поміняти місцями» H_0 й H_p неможливо, тому що існують різні теоретичні закони розподілу, за якими можна отримати практично однакові частоти - значення $\square^2$.Тому обчислення потужності критерію узгодження немає великого практичного значення і зазвичай не робится, а рівень значущості вибирають більшим, щоб зменшити ймовірність помилки 2-го проядку.

Для завдань у лісовому господарстві можна рекомендувати α=0,1 або навіть 0,2.

Наведемо приклад застосування критерію узгодження $\square^2$. Візьмемо розподіл діаметрів у дерев сосни звичайної (таблиця 9.6), який спробуємо апроксимувати кривою нормального розподілу. Обчислимо частоти, що вирівнюють, для цього ряду розподілу (схема обчислень описано у розділі 5). У таблиці 10.1 наведено схему знаходження $\square^2$ за даними про чисельність - фактичним та апроксимованим кривою нормального розподілу, і показано схему знаходження $\square^2$.

Таблиця 10.1

Схема обчислення $\square^2$ для розподілу 238 діаметрів дерев сосни звичайної в умовах пробних площ Поліського природного заповідника (за результатами досліджень 2022-2025 рр.)

Ступені Товщини x_i	Кількість		$n_i - \tilde{n}_i$	$(n_i - \tilde{n}_i)^2$	$\chi^2 = \frac{(n_i - \tilde{n}_i)^2}{n_i}$	Перевірка $\frac{n_i^2}{\tilde{n}_i}$
	фактичне, n_i	теоретичне, $\tilde{n}_i$				
1	2	3	4	5	6	7
> 16	12	10	2	4	0,4	14,40
20	21	18	3	9	0,4	24,50
24	30	33	3	9	0,3	27,27
28	44	45	1	1	0,0	43,02
32	54	48	6	36	0,7	60,75
36	35	37	2	4	0,1	33,11
40	23	24	1	1	0,1	22,04
>44	19	20	1	1	0,1	18,05
Σ	238	235	3	69	2,1	243,14

У таблиці 10.1 ми бачимо величини чисельності ≤5 об'єднали із сусідніми значеннями, про що було сказано вище. Обчислений критерій χ^2 дорівнює 2,1. Колонка 7 служить для контрольної перевірки розрахунків. За умови, якщо N = ñ, то:

$$\chi^2 = \sum_{i=1}^{m} \frac{n_i^2}{\tilde{n}_i} - N$$

повинен дорівнювати обчисленому за схемою таблиці 10.1. Бо в нас N ≠ ñ, то слід внести поправки, тоді формула набуде вигляду:

$$\chi^2 = \sum_{i=1}^{m} \frac{n_i^2}{\tilde{n}_i} - N - (N - \tilde{N})$$

Для нашого прикладу:

$$\chi^2 = 243{,}14 - 238 - 3 = 2{,}1$$

Порівняємо знайдений критерія χ^2=2,1 з його табличним критичним значенням (α) для різних рівнів значущості (α), взятих із спеціальної таблиці (додаток Н). Число ступенів свободи у нас рівне: m-*l*-1 = 8-2-1= 5. Для γ = 5 маємо $\alpha_{0,9}$ = 9,24;$\alpha_{0,8}$ = 7,29; $\alpha_{0,5}$ = 4,35; $\alpha_{0,2}$ = 2,67; $\alpha_{0,1}$= 1,62; $\alpha_{0,05}$=1,15 і т.д. Отже, за критерієм χ^2 наш ряд розподілу діаметрів у дерев сосни звичайної на пробній площі №2 в лісорослинних умовха $A_{2\text{-}3}$ Перганського, Копищанського, Селезівського ПНДВ Поліського природного заповідника відповідає нормальній кривій Гауса-Лапласа з ймовірністю понад 80% (приблизно 85-86%). Розглянуті вище критерії (t-й критерій Стьюдента, χ^2) вводилися в гіпотезі, що вибірки взяті з нормально розподіленої генеральної сукупності, а самі дослідження є незалежні. Тим часом у багатьох практичних випадках розподіл випадкової величини невідомий, а в деяких – відомо те, що воно не відповідає кривій нормального розподілу. Тому нам треба знати, які відхилення від нормальної не спотворюють висновків, одержуваних за допомогою критеріїв. Не наводячи доказів з причин, що зазначені вище, висловимо лише наступне. В цілому критерії, що відносяться до середньої генеральної сукупності (t-критерій Стьюдента), досить не чутливі до відхилень від нормальності розподілів, особливо якщо останні не дуже асиметричні. Критерії, що відносяться до дисперсій χ^2 і F, навпаки - дуже чутливі, при цьому відхилення від нормального ексцесу відіграють набагато більшу роль, ніж відхилення від симетричної форми. Застосування критеріїв σ^2 дисперсії вимагає обережності та перевірки вихідного розподілу на нормальність, особливо при малих вибірках. Якщо не підтверджується гіпотеза про нормальність, то позитивні результати дає застосування перетворення розподілів у нормальні. Так, якщо потрібно вирівняти дисперсії вибірок з приблизно рівними

коефіцієнтами мінливості, то хороші результати дає логарифмічне перетворення y=lnx. Перетворення Джонсона є гарним прикладом перетворення до нормального вигляду. Для розподілів близьких до біномінального, хороші результати дає перетворення y = arcsin√x, для пуассонівського y = √x або y = √x + 0,5 (для малих значень), для розподілів зі значною лівою асиметрією перетворення Фішера:
y = 1/2ln [(1 + x) / (1 - x)] і т.д.

Одним із критеріїв для перевірки відхилення від нормальності (у разі великих вибірок), можуть бути основні помилки асиметрії та ексцесу. Відповідно до стандартних методів розподіл слід вважати приблизно нормальним, якщо $\alpha/m_a < 2$ і $E/m_e < 2$. Для більш точного судження про значимість критеріїв α та E залежно від рівня значущості та числа спостережень, можна використовувати спеціальні таблиці для α (додаток О) та E (додаток П). При користуванні цими таблицями треба мати на увазі, що розподіл α симетрично, тобто (1-α) = -(α) при даному N. Так, якщо α = 0,05, а вибіркові значення α < 0,389 та E < 0,77 при N = 100, то гіпотезу про нормальності розподілу приймають. Для наведеного вище прикладу з розподілом 238 стволів сосни звичайної по діаметру та висоті (розділ 9) ми отримали такі величини. Для ряду розподілу діаметрів:

$$\bar{\delta}=30,7;\ \sigma=1,9;\ \bar{\sigma}=7,6;\ v=24,7\%;\ \alpha=-0,01;\ E=-0,44;$$
$$m_{\bar{x}}=0,49;\ m_\sigma=0,09;\ m_{\bar{\sigma}}=0,35;\ m_\alpha=0,39;\ m_E=0,78.$$

Для ряду розподілу висот:

$$\bar{\delta}=25,0;\ \sigma=1,7;\ \bar{\sigma}=3,4;\ v=13,6\%;\ \alpha=-0,79;\ E=0,35;$$
$$m_{\bar{x}}=2,21;\ m_\sigma=0,078;\ m_{\bar{\sigma}}=1,56;\ m_\alpha=0,39;\ m_E=0,78.$$

В нашому випадку:

$$\frac{\alpha_d}{m_{\alpha D}}=\frac{0,01}{0,39}=0,03<2;\ \frac{\alpha_í}{m_{\alpha í}}=\frac{-0,79}{0,39}=2,02\approx 2;$$
$$\frac{E_d}{m_{ED}}\frac{0,44}{0,78}=0,56<2;\ \frac{E_í}{m_{Åí}}=\frac{0,35}{0,78}=0,45<2.$$

На основі проведених обчислень можемо стверджувати, що розподіл діаметрів відповідає нормальному закону (це підтверджується критерієм узгодження χ^2), а розподіл висот цьому розподілу не відповідає. Величина критичних значень α та E (додатки О, П, Р) для розподілу діаметрів і висот (обсяг сукупності, від якої залежать β_1 і β_2 однакові) $\beta_1 = 0,246$; $\beta_2 = 0,55$. Величини β_1 і β_2 підтверджують сформульований вище висновок, про відповідність

кривої нормального розподілу ряду розподілу по діаметру і відхилення від нього (α) для ряду висот.

10.3 Критерій узгодження Колмогорова-Смирнова.

Порівняльну оцінку двох однорідних варіаційних рядів, а також зіставлення частот емпіричного та обчисленого розподілів, можна зробити і за допомогою так званих непараметричних, або порядкових критеріїв. На відміну від критерію χ^2-квадрат та критерію t - Стьюдента, застосування яких ґрунтується на використанні вибіркових характеристик (параметрів) x і σ, при обчисленні непараметричних критеріїв це не потрібно. Але для їх застосування необхідне впорядкування у вигляді кумуляції емпіричних та теоретичних розподілів, тобто отримання рядів накопичених частот.

Для непараметричних критеріїв характерно те, що вони однаково придатні для оцінки вибіркових розподілів будь-якого виду, тоді як застосування параметричних критеріїв виходить із положення про нормальність розподілу оцінюваних рядів.

Один з найпростіших і зручніших при зіставленні емпіричних сукупностей великого об'єму - критерій, що запропонований фундаментальними математиками О. М. Колмогоровим (1903-1987) та Н. В. Смирновим (1900-1966). Цей непараметричний показник, що позначається грецькою літерою λ (лямбда), являє собою максимальну різницю (d_{max}) між значеннями накопичених частот емпіричного та обчисленого рядів (без урахування знаків d), віднесену до кореня квадратного із суми всіх варіантів сукупності:

$$\lambda = \frac{d_{\max}}{\sqrt{n}}$$

На відміну від критерію χ^2 критерій λ не тільки простий конструкційно, але не вимагає і спеціальних таблиць, хоча такі таблиці є (додаток С), і застосовуються при уточнених розрахунках, але користуються ними рідко. Для спрощеної оцінки критерію Колмогорова-Смирнова використовують граничні значення критерію λ, що відповідають трьом рівням довіряючої ймовірності - $P_1 = 0,95$, $P_2 = 0,99$ та $P_3 = 0,999$, які відповідно дорівнюють 1,36; 1,63 та 1,95. Цей висновок випливає з наступного розрахунку:

$$\lambda = \sqrt{\frac{1}{2}\ln\frac{2}{P}}$$

де: P – відповідний рівень занчущості. Якщо прийняти $P_1 = 0,05$,

то:

$$\lambda = \sqrt{\frac{1}{2}\ln 40} = 1{,}36$$

При P_2=0,01, λ=1,63 й т.д. При обчисленні критерія λ, відпадає необхідність визначення числа ступенів свободи. У тих випадках, коли порівнюються два емпіричні розподіли, що взяті з однієї і тієї ж генеральної сукупності, але мають різний об'єм, критерій λ обчислюється за такою формулою:

$$\lambda = d_{max}\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

Тут $d_{max} = \sum\frac{p_1}{n_1} - \sum\frac{p_2}{n_2}$

Це максимальна різниця між значеннями першого: P_1/n_1 та другого: P_2/n_2 рядів накопичених часто.

Обчислення критерію Колмогорова-Смирнова продемонструємо на приклад розподілу 238 діаметрів в сосновому деревостої (таблиця 9.6). Результати наведено у таблиці 10.2

Таблиця 10.2

Обчислення критерію λ для розподілу 238 діаметрів сосни звичайної в умовах пробних площ Поліського природного заповідника (за результатми досліджень 2022-2025 рр.), апроксимованих кривою нормального розподілу

Ступені товщини, (розряди) (x_i)	Кількість		Накопичення частоти (P_i)		$d=P_ф-P_т$
	фактичне (n_i)	вирівняне	фактичне ($P_ф$)	теоретичне ($P_т$)	
12	3	3	3	3	0
16	9	7	12	10	2
20	21	18	33	28	5
24	30	33	63	61	2
28	44	45	107	106	1
32	54	48	161	154	7
36	35	37	196	191	5
40	23	24	219	215	4
44	17	16	236	231	5
48	2	4	238	235	3
Всього:	238	235	-	-	-

Максимальна величина різниці $d=P_ф-P_т$ (без урахування знаків) дорівнює 7.

Тоді:

$$\lambda = \frac{d}{\sqrt{N}} = \frac{7}{\sqrt{238}} = \frac{7}{15,7} = 0,45$$

Отримана величина λ (0,45) значно менша за його граничного значення (1,36) для P = 0,05. Таким чином, можна стверджувати, що розбіжності між теоретично обчисленими частотами (по кривій нормального розподілу) і фактичними мають випадковий характер, і наш ряд розподілу можна описати за допомогою кривої Гаус-Лапласа (нормального розподілу).

Тепер розглянемо приклад, коли зіставлено два ряди розподілу, і необхідно оцінити, чи вони належать до однієї генеральної сукупності. Для цього порівняємо дані вимірювань діаметрів на 2 пробних площах, що біли закладені у двадцятирічних соснових насадженнях, в типі лісу сосняк моховий II класу бонітету, різних за походженням: природний деревостан та лісові культури. При цьому обсяги вибірок різняться (таблиця 10.3).

Обчислення λ здійснюється за формулою:

$$\lambda = d_{max}\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

$$\lambda = 0,232 \cdot \sqrt{\frac{320 \cdot 434}{320 + 434}} = 0,232 \cdot \sqrt{\frac{138880}{754}} = 0,232 \cdot \sqrt{184,2} = 0,232 \cdot 13,57 = 3,149 .$$

Обчислити λ можна також за формулою:

$$\lambda^2 = d^2 \frac{N_1 \cdot N_2}{N_1 + N_2} = 0,232^2 \cdot \frac{138880}{754} = 0,05382 \cdot 184,2 = 9,9136 ;$$

$$\lambda = \sqrt{9,9136} = 3,15 ,$$

тобто ми отримали однакові величини в межах точності заокруглень.

Таблиця 10.3

Обчислення критерію Колмогорова-Смирнова для двох пробних площ, закладених у соснових молодняках

Ступені товщини, (x_i)	Кількість		Частоти		Накопичені частоти			
	n_i^1 лісові культури	n_i^2 природне поновлення	$P_1 = \frac{n_i^1}{\sum n_1}$	$P_2 = \frac{n_i^2}{\sum n_2}$	ΣP_{i1}	ΣP_{i2}	$d_i = \lvert P_1 - P_2 \rvert$	Max d_i
2	-	86	0,191	0,198	-	0,148	-	-
4	61	98	0,237	0,225	0,191	0,423	0,232	0,23
6	76	102	0,284	0,235	0,428	0,658	0,230	2
8	91	69	0,144	0,150	0,712	0,816	0,104	-
10	46	35	0,091	0,080	0,856	0,896	0,040	-
12	29	27	0,053	0,062	0,947	0,958	0,011	-
14	17	12	1,000	0,027	-	0,985	-	-
16	-	5	0,191	0,015	-	1,000	0	-
Σ(N)	320	434	0,237	1,000	-	-	-	-

Аналіз обчисленого значення λ показує, що порівнювані сукупності належать до різних генеральних сукупностей, і описуються різними кривими: λ=3,15>2,1, тобто достовірність відмінностей перевищує 99% рівень. Це відповідає матеріалам, що наводяться багатьма вченими, що досліджували склад штучних та природних соснових молодняків.

З досвіду застосування критеріїв λ та λ^2 випливає, що критерій Колмогорова-Смирнова менш чутливий, ніж χ^2. У спірних випадках (при граничних значеннях) λ зазвичай підтверджує відповідність теоретичному розподілу, а χ^2 може це твердження спростовувати. У цьому випадку дослідник через свою класифікацію, особливості завдання, її важливість та складність приймає нульову чи альтернативну (робочу) гіпотезу. При застосуванні критерію λ, неодмінною умовою застосування повинно бути відносно велике число (не менше 100) спостережень. Тому простий за своєю методкою проведення розрахунків, цей метод не може бути використаний при оцінці малочисленних сукупностей.

10.4 Застосування статистичного оцінювання в лісовому господарстві.

Статистичні критерії знаходять широке застосування в лісовому господарстві, особливо під час проведення наукових досліджень. Про застосування t-критерія Стьюдента та F-критерія Фішера вже

описано у розділі 9. Критерії узгодження використовують при численних дослідженнях рівнів товарності деревостанів. Товарність деревостану залежить від багатьох факторів (їх вивчає лісова таксація та деревинознавство), але одним з головних є закономірності розподілу числа стовбурів за ступенями товщини залежно від середнього діаметра деревостану.

Для того щоб переконатися, що розподіл обрано правильно, застосовують критерії узгодження χ^2 та λ. Помилка у виборі правильного розподілу може дорого коштувати в прямому значенні цього терміну, так як товарність деревостану визначає ціну 1 м^3 та вартість деревостану у виділі. Критерії узгодження використовують також для оцінки відносної однорідності різних вибірок. Це часто має значення для доказу належності деревостанів на різних пробних площах до однієї чи різних генеральних сукупностей як у прикладі, поданому вище. Таке порівняння важливе при підборі серії пробних площ, що закладаються з різною метою в деревостанах приблизно однакового віку, але відмінних за походженням, повнотою та густотою, доглядом, проведеними меліоративних заходів тощо.

Закінчуючи розгляд статистичного оцінювання, наведемо основні статистичні оцінки в компактному вигляді для зручності їх використання.

Помилка середнього значення:

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

де: σ – середнє квадратичне відхилення.

$$P\left(\left|x_i - \bar{x}\right| \le m\right) = 0{,}683$$

$$P\left(\left|x_i - \bar{x}\right| \le 2m = 0{,}95\right)$$

$$P\left(\left|x_i - \bar{x}\right| \le 3m = 0{,}99\right)$$

Помилка середнього квадратичного відхилення:

$$m_\sigma = \frac{\sigma}{\sqrt{2n}}$$

Помилк коефіцієнта варіації:

$$m_V = \frac{V}{\sqrt{2N}} \cdot \sqrt{1 + \left(\frac{V}{100}\right)^2} \quad \text{або} \quad m_V = V \cdot \sqrt{\frac{0.5 + 0.0001V^2}{N}}$$

Помилка асиметрії:

$$m_\alpha = \sqrt{\frac{6}{n}}$$

Помилка ексцесса:

$$m_E = 2m_\alpha$$

Точність досліду:

$$P = \frac{V}{\sqrt{n}}$$

Кількість спостережень при заданій точності:

$$n = \frac{V^2}{P^2}$$

Об'єктивність (достовірність) висновку:

$$t = \frac{\bar{x}}{m_x}$$

Помилка суми середніх значень (x)

$$m_{\bar{x}_1 = \bar{x}_2} = \sqrt{m_{\bar{x}_1}^2 + m_{\bar{x}_2}^2}$$

$$m_{\bar{x}_1 + \bar{x}_2 + \ldots \bar{x}_n} = \sqrt{m_{\bar{x}_1}^2 + m_{\bar{x}_2}^2 + \ldots m_{\bar{x}_n}^2}$$

Середня помилка різниць двох середніх велечин при $N_1=N_2$:

$$m_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}$$

При: $N_1 \neq N_2$: $m_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{об}^2 \left(N_1 + N_2\right)} / N_1 N_2$

$$\sigma_{об}^2 = \left[\sum \left(x_{i1} - \bar{x}_1\right)^2 + \sum \left(x_{i2} - \bar{x}_2\right)^2\right] / \left(N_1 + N_2 - 2\right)$$

Помилка суми середніх величин:

$$m_{\bar{x}_1 \cdot \bar{x}_2} = \bar{x}_1 \bar{x}_2 \sqrt{\left(\frac{\sigma_{\bar{x}_1}}{\bar{x}_1}\right)^2 + \left(\frac{\sigma_{x_2}}{\bar{x}_2}\right)^2}$$

Середня помилка частки середніх величин:

$$m_{\frac{\bar{x}_1}{\bar{x}_2}} = \frac{\bar{x}_1}{\bar{x}_2} \sqrt{\left(\frac{m_{x_1}^2}{\bar{x}_1}\right)^2 + \left(\frac{m_{x_2}^2}{\bar{x}_2}\right)^2}$$

Визначення суттєвості різниці між середніми:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{m_{x_1}^2 + m_{x_2}^2}} = \frac{D}{m_D} \qquad |t| \geq 3$$

t – критерій об’єктивності (достовірності).

Об’єктивність різниці між дисперсіями:

$$t = \frac{\sigma^2{}_1 - \sigma^2{}_2}{m^2{}_\sigma} = \frac{D}{m^2{}_\sigma}$$

$$m_\sigma = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}} \qquad \chi^2 = \sum_{i=1}^{n} \frac{(p - p')^2}{p}$$

де: p – імперичні частоти; p’ – теоретичні частоти.

$$\lambda = d_{\max} \sqrt{\frac{N_1 N_2}{N_1 + N_2}}$$

$$d_{\max} = \sum \frac{p_1}{N_1} - \sum \frac{p_2}{N_2}$$

N_i – об’єм ряду розподілу;
p_i – накопичені частоти.

РОЗДІЛ 11. СТАТИСТИКИ ЗВ'ЯЗКУ

11.1 Поняття кореляції.

Досліджуючи лісові насадження ми помічаємо, що в них гармонійно поєднуються різні ознаки. У лісі немає хаосу, проте діють закони та закономірності, за якими лісовий біогеоценоз розвивається і на основі яких він існує. Подібні закономірності властиві як біологічним об'єктам, так і людському суспільству.

Взаємна залежність двох величин (або явищ), коли зміна однієї з них веде до закономірної зміни іншої називається кореляцією.

Це поняття широко використовується в науці, особливо в біології, фізики, хімії. Наприклад, в лісовому господарстві відомі тісні зв'язки між діаметром та висотою дерева. Для однієї породи, наприклад берези повислої при більшій висоті спостерігається і більший діаметр. В свою чергу, висота дерева у певному віці залежить від родючості ґрунту. Зі збільшенням висоти, а також родючістю ґрунту тісно пов'язана продуктивність деревостанів. Форма стовбура залежить від таксаційної характеристики насадження: його густота, висота дерева. Залежності та зв'язки в природі та суспільстві, що мають спільні методи їх статистичного виміру називають кореляцією, зв'язком або залежністю.

Останні два слова – це синоніми терміну «кореляція». Відомі функціональні та кореляційні зв'язки. Останні ще називають стохастичними. До функціональних відносять закони математики та фізики. Наприклад $E = mc^2$; $L = \pi R$; $V = \alpha t$; $S = gt^2/2$ та інші відомі зі шкільного курсу математики та фізики. Тут зміна однієї з величин, завжди веде до обов'язкової, чіткї означеної і завжди певної зміни іншої величини. ***Загальні зв'язки такого роди називаються фізичними чи природними законами. Закон, застосований до деякого окремого випадку, називається закономірністю.***

Наприклад, є загальний фізичний закон – метали при нагріванні розширюються. Щодо конкретного металу, скажімо - міді, конкретні коефіцієнти температурного розширення - це вже закономірність. Щодо лісового господарства можна привести наступний приклад. Зростання дерева у висоту – біологічний закон росту. Параметри зростання (швидкість збільшення висоти дерева з віком) для конкретного деревини, скажімо, берези повислої - це закономірність.

В природі явища розвиваються під впливом різних факторів навколишнього середовища. Тому зв'язок між ознаками

проявляється у вигляді кореляційного зв'язку, чи кореляції. У цьому випадку кожному значенню однієї ознаки, відповідає не одне, а кілька значень іншої ознаки, тобто його розподіл. Одну з ознак, що зазвичай точніше вимірна, приймають за факторіальну, а інший – за результативну. Іноді, в умовному значенні одну називають незалежним, а іншу – залежною від першого.

Статистичне дослідження кореляції зводиться до визначення факту зв'язку, визначення її форми, спрямованості та тісноти. Встановлення факту зв'язку фахівці у певній галузі проводять спочатку з урахуванням загального аналізу явища. Наприклад, можна сказати про наявності кореляції між розмірами дерева: товщиною та висотою ще до їхнього виміру. В інших випадках наявність кореляції між досліджуваними ознаками не можна передбачити так виразно. Наприклад, без вимірювань та подальшого аналізу важко оцінити зв'язок форми стовбура з його висотою. І тут вирішують питання наявності кореляції з урахуванням вимірювання та статистичного аналізу його результатів. Пояснимо вище сказане прикладами. Так ми вже говорили, що зі збільшенням діаметру дерева, збільшується його висота. Нехай у нас є дерева сосни звичайної І класу бонітету з діаметрами: 20, 24, 28, 32 см. Висоти цих дерев дорівнюють: 23, 25, 29, 32м. Але, якщо ми виміряємо скажімо з 20 дерев кожне з названих діаметрів, то виявиться, що висоти коливаються у таких межах:

Діаметр, см	Висота (від-до), м
20	21-25
24	23-27
28	26-32
32	29-35

Загальна закономірність (збільшення висоти з зростанням діаметра) витримується, середнє значення висоти теж, але висота буде відповідати обчисленму значенню (23, 25, 29, 32) з певною ймовірністю (0,68) та середньоквадратичною помилкою. Це і є ймовірнісний або стохастичний зв'язок.

Кореляцію називають простою, якщо вона вимірюється на основі двох ознак, або множинної, якщо зміна результативної ознаки вивчають у зв'язку з впливом або зміною кількох факторіальних ознак. Наприклад, коли ми розглядаємо зв'язок діаметр дерева – висота, то це проста кореляція. Зв'язок, коли висоту дерева вивчають в залежності від ґрунтової родючості, густоти деревостану, деревної породи, то це складна.

За формою розрізняють кореляцію лініну, коли залежність між ознаками відображається прямою лінією, і криволінійною, коли її відображає рівняння якоїсь кривої.

У багатьох випадках форму кореляції можна передбачити ще до закладання досліду. Наприклад, між довжиною і товщиною коріння молодих дерев в деревостані очікується лінійна кореляція. Але не можна очікувати такої ж форми кореляції в дерев, що ростуть у старих деревостанах. Статистичний аналіз дає відповідь про форму зв'язку і в тих випадках, коли на основі біологічного аналізу її встановити важко чи взагалі неможливо.

За спрямованістю розрізняють пряму кореляцію, коли зі збільшенням однієї ознаки в середньому збільшуються і значення іншої, а зі зменшенням - зменшуються, і зворотну, - коли зі збільшенням значень однієї ознаки, значення іншої в середньому зменшуються і навпаки.

Приклад прямого зв'язку наведено вище - діаметр та висота дерева. Зворотній зв'язок ми можемо спостерігати, вивчаючи вплив шкідників (припустимо, звичайного соснового пильщика) на приріст сосни звичайної. Зі збільшенням кількості гусениць (личинок) на одному дереві, приріст зменшується.

Типові картини статистичних зв'язків спостерігаються у двох практично різних випадках:

а) вивчається зв'язок між випадковими величинами;

б) насправді вивчається функціональний зв'язок, але похибки вимірювань породжують змінюватись та створюють видимість статистичної зв'язку.

З погляду статистичного аналізу на етапі з'ясування існування залежності, цю різницю можна не враховувати; на етапі побудови моделей, виникають суттєві відмінності у її інтерпретації та використанні.

Можна дати таке трактування стохастичним зв'язкам. При вивченні зв'язку між двома величинами, не можна гарантувати, що одна величина повністю визначає значення інший, тобто - враховано всі основні чинники, загальні для обох величин. Крім того, може існувати різна частка основних чинників, загальних для обох величин. Крім основних існують і випадкові фактори, що впливають на досліджувані величини, і відводять в тінь існуючу закономірність. Тому чим більше загальних факторів для досліджуваних величин, і чим повніше вони враховані, тим виразніше зв'язок між досліджуваними величинами і тим більше «зона розсіювання», яка в зоні розподілу перетворюється на деяку лінію, функціональну залежність, що відображає процес або значення показника цього процесу.

З наведених вище думок випливає очевидна необхідність в статистичних показниках, що відображають наявність та ступінь тісноти зв'язку. Такі показники (для вибірки) називають статистиками зв'язку.

Важливість статистик зв'язку у моделюванні очевидна: перш ніж конструювати модель, слід з'ясувати в яких зв’язках між собою

знаходяться показники, що цікавлять дослідника. За змістом тут спостерігається повна аналогія зі статистиками розподілів: статистики зв'язку є вибірковими оцінками параметрів зв'язку у генеральній сукупності. Вихідними даними для статистичного аналізу за величиною в числі спостережень, служать таблиці розподілу, що складаються з дотриманням правил, ідентичних правилам для складання рядів розподілу, проте спостереження «розносять» з урахуванням обох ознак. Приклад розподілу такого показано в таблиці 11.1. Наочно подібні розподіли видно на графіках. Наприклад такі графіки наведені на рис. 11.1, де чітко проглядається закономірна залежність досліджуваних ознак, але на рис. 11.16 вона виражена набагато слабкіше.

Таблиця 11.1

Розподіл діаметрів та висот для 238 дерев сосни звичайної в умовах пробних площ Поліського природного заповідника (за результатми досліджень 2022-2025 рр.)

Ступені товщини (середини класів)	Кількість стовбурів, шт., (частоти)									Σ частот	Умовні середні
	Ступені висоти, м., (середини класів)										
12	3	-	-	-	-	-	-	-	-	3	16,0
16	3	3	3	-	-	-	-	-	-	9	18,0
20	2	5	10	4	-	-	-	-	-	21	19,5
24	-	2	2	11	12	3	-	-	-	30	22,8
28	-	-	-	3	28	11	2	-	-	44	24,5
32	-	-	-	-	5	38	11	-	-	54	26,2
36	-	-	-	-	1	22	11	-	1	35	26,7
40	-	-	-	-	-	1	15	7	-	23	28,5
44	-	-	-	-	-	-	7	10	-	17	29,2
48	-	-	-	-	-	-	-	1	1	2	31,0
Σ частот	8	10	15	18	46	75	46	18	2	238	-
Умовні середні	15,5	19,6	19,7	23,8	27,6	32,3	37,2	42,7	42,0	-	-

Таблиці, що складаються подібно до 11.1, використовують потім для розрахунку тісноти зв'язку між обчислюваними величинами. При малій кількості спостережень, тісноту зв'язку можна обчислювати безпосередньо, тобто без зведення вихідних даних у таблиці.

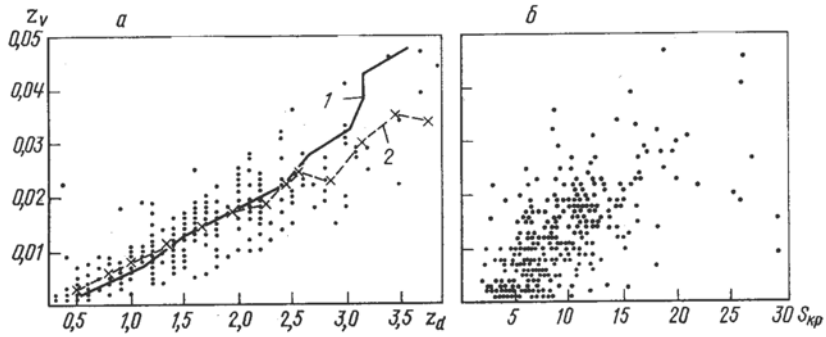


Рис. 11.1 - Точкові діаграми (за К .Є. Нікітіним та А. З. Швиденко):

а - поточного приросту за діаметром z_d та за об'ємом z_v;
(1 – регресія у на х; 2 - регресія х на у)
б - площі проекцій крон $S_{кр}$ та z_v

Потрібно пояснити сенс термінів «зв'язок» та «залежність». Як і у всьому статистичному аналізі, основним тут є з'ясування причинної суті встановлених кореляцій. Так, якщо розглядається зв'язок між діаметром дерева та його висотою, то очевидно, що y_i , (D_i) залежить від x_i (H_i), але можна вважати і навпаки. Тому в деяких завданнях можлива симетрія дії. Але якщо розглядається зв'язок між кількістю опадів, що випадають, і продуктивністю деревостану, то тут може бути лише залежність «в одному напрямку», хоча статистично можна спробувати встановити і залежність опадів від продуктивності деревостанів, і навіть отримати якесь її цифровий вираз, але недоцільність подібних дій очевидна. Тісноту кореляції, чи ступінь зв'язку між значеннями однієї та іншої ознаки, виражають у вигляді абстрактних статистичних характеристик (показників) зв'язку - коефіцієнта кореляції **r** і кореляційного відношення - **η.**

11.2 Коефіцієнт кореляції як міра лінійного зв'язку.

Коефіцієнт кореляції – це статистика, яка є чисельною характеристикою зв'язку між ознаками, коли вона має лінійний характер. Це означає, що зв'язок між величинами х та у виражається загальною формулою $y=a_o+a_1x$, де a_o та a_1. - коефіцієнти, які визначають на основі вибіркових спостережень. Коефіцієнт кореляції чисельно вражає відношення числа факторів, що діють на зміну обох ознак до загального числа факторів.

Вказаний вміст коефіцієнта кореляції досить добре висловлює формула:

$$r_{yx} = \frac{\sum n_{ij}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \cdot \sum (y_i - \overline{y})^2}}$$

Цей вираз також можна записати:

$$r = [\sum(X_i - \overline{X})(Y_i - \overline{Y})] / N\sigma_x \sigma_y$$

В першій формулі x_i і y_i - випадкові (досліджувані) величини; x, y - середні значення для x та y.

Величина σ_x і σ_y у другій формулі - середньоквадратичні відхилення розподілів X, та Y; N - число сформованих пар чи кількість спостережень.

З формули видно, що при незалежному варіюванні ознак, коли будь-яке з відхилень X може поєднуватися з будь-якими Y (як з позитивними, так і з негативними, до того ж однаковими частотами), чисельник її дорівнюватиме 0 або близькій до 0 величині. Отже, й r=0. При сполученому варіюванні відхилення X_i-X поєднуються тільки з деякими відхиленнями Y_i-Y наприклад, позитивні в основному тільки з позитивними (при прямому зв'язку) або позитивні з негативними (при зворотному зв'язку). У цьому випадку сума добутків матиме позитивне (при прямому зв'язку) або негативне (при зворотному зв'язку) значення, до того ж тим більше за своєю величиною (при даному N) зв'язок більший.

Поділом суми добутків відхилень на кількість корелюючих пар, отримують середню величину добутку, а поділом на стандартні відхилення σ_x і σ_y висловлюють цей добуток абстрактним числом, що характеризує тісноту зв'язку. В цілому коефіцієнт кореляції є емпіричний основний змішаний момент, тобто:

$$r = \frac{m_{1/1}}{\sigma_1, \sigma_2}$$

Для його обчислення треба знайти перший початковий змішаний момент $m_{1/1}$ та перші два початкові моменти кожного з рядів розподілу. Обчислення моментів для одномірних рядів розподілу описано раніше (розділ 5). Тут же покажемо обчислення необхідних моментів для двовимірної сукупності.

Емпіричним змішаним початковим моментом порядку (k_1, k_2) двох випадкових величин (x_1, y_2,), які зведені в таблицю та згруповані за певним розрядам, називається сума добутків кожної пари відхилень $x_{1(i2)}$ і $x_{2(j2)}$. Відхилення беруть від початкових значень $x_{1(a)}$ і $x_{2(a)}$ в h_1 та h_2 ступеня і множать на відповідну частку $P_{i1|j2}$.

$$m_{k_1/k_2} = \sum_{i_1=1}^{n_1} \sum_{j_2=1}^{n_2} P'_{k_1/k_2} x^{k_1}_{1(i/1)} x^{k_2}_{2(j/2)}$$

Надаючи величинам k_1 та k_2 з різними значеннями, отримуємо змішані моменти $m_{1/1,}$ $m_{2/1,}$ $m_{1/2,}$ $m_{1/3,}$ $m_{2/2.}$

Співвідношення між визначеними вище змішаними центральними та початковими моментами визначаються формулами:

$$\mu_{1|1} = m_{1|1} - m_{1|0}m_{0|1},$$

$$\mu_{2|1} = m_{2|1} - 2m_{1|1}m_{1|0} - m_{0|1}(\mu_{2|0} - m_{1|0}^2),$$

$$\mu_{1|2} = m_{1|2} - 2m_{1|1}m_{0|1} - m_{1|0}(\mu_{0|2} - m_{0|1}^2),$$

$$\mu_{3|1} = m_{3|1} - 3m_{2|1}m_{1|0} + 3m_{1|1}m_{1|0}^2 - m_{0|1}(\mu_{3|0} - m_{1|0}^3)$$

$$\mu_{1|3} = m_{1|3} - 3m_{1|2}m_{0|1} + 3m_{1|1}m_{0|1}^2 - m_{1|0}(\mu_{0|3} - m_{0|1}^3),$$

$$\mu_{2|2} = m_{2|2} - 2(m_{2|1}m_{0|1} + m_{1|2}m_{1|0}) + 4m_{1|1}m_{1|0}m_{0|1} + m_{1|0}^2(\mu_{0|2} - m_{0|1}^2) + m_{0|1}^2\mu_{2|0}$$

Для перевірки обчислень використовують формули:

$$\mu_{1|1} = m_{1|1} - m_{1|0}m_{0|1},$$

$$\mu_{2|1} = m_{2|1} - 2\mu_{1|1}m_{1|0} - m_{2|0}m_{0|1},$$

$$\mu_{1|2} = m_{1|2} - 2\mu_{1|1}m_{0|1} - m_{0|2}m_{1|0},$$

$$\mu_{3|1} = m_{3|1} - 3\mu_{2|1}m_{1|0} + 3\mu_{1|1}m_{1|0}^2 - m_{3|0}m_{0|1},$$

$$\mu_{1|3} = m_{1|3} - 3\mu_{1|2}m_{0|1} + 3\mu_{1|1}m_{0|1}^2 - m_{0|3}m_{1|0},$$

$$\mu_{2|2} = m_{2|2} - 2(\mu_{2|1}m_{0|1} + \mu_{1|2}m_{1|0}) + 4\mu_{1|1}m_{1|0}m_{0|1} + m_{0|2}m_{1|0}^2 + \mu_{2|0}m_{0|1}^2$$

Емпіричні змішані основні моменти порядку (h_1, h_2) за допомогою центральних:

$$r_{h_1|h_2} = \frac{\mu_{h_1|h_2}}{\sigma_1^{h_1}\sigma_2^{h_2}}$$

І зокрема в розрізі окремих моментів:

$$r_{1|1} = \frac{\mu_{1|1}}{\sigma_1\sigma_2}, \qquad r_{2|1} = \frac{\mu_{2|1}}{\sigma_1^2\sigma_2}, \qquad r_{1|2} = \frac{\mu_{1|2}}{\sigma_1\sigma_2^2};$$

$$r_{3|1} = \frac{\mu_{3|1}}{\sigma_1^3\sigma_2}, \qquad r_{1|3} = \frac{\mu_{1|3}}{\sigma_1\sigma_2^3}, \qquad r_{2|2} = \frac{\mu_{2|2}}{\sigma_1^2\sigma_2^2}$$

Змішаний основний момент першого порядку $r_{1|1}$ називається коефіцієнтом кореляції і позначається через r.

Він визначається за формулою:

$$r = \frac{\mu_{1|1}}{\sigma_1\sigma_2}$$

Змішані моменти різних порядків можуть бути розраховані як за способом добутків, так і за способом сум. За способом добутків обчислюється змішаний момент $m_{1|1}$ порядку (1,1). Для цього застосовується схема, в якій поряд із таблицею розподілу складається ще допоміжна таблиця. Схема обчислення $m_{1/1}$ показана

в таблиці 11.2. Для прикладу взято розподіл 238 дерев сосни звичайної за діаметром та висотою (таблиця 11.1). Схема обчислення $m_{1|1}$ представляє собою таблицю розподілу, в заголовках якої розрядні значення першої величини (діаметр) та розрядні значення другої величини (висота) замінюються відхиленнями $x_{1(i/1)}$ та $ч_{2(j/2)}$ значень від відповідних початкових величин початкових значень.

Таблиця 11.2

Схема обчислення $m_{1/1}$ для розподілу діаметрів та висот за способом добутків для 238 дерев сосни звичайної в умовах пробних площ Поліського природного заповідника (за результатами досліджень 2022-2025 рр.)

Ступені товщини, см	Кількість стовбурів, шт (частоти)								Сума частот
	Ступені висоти, м (середини класів)								
		16	20	22	24	26	28	30	-
		4	2	-1	0	1	2	3	-
12	5	+2 0 3	-	-	-	-	-	-	-	3
16	4	+1 6 3	+8 3	-	-	-	-	-	-	9
20	3	+9 5	+6 10	+3 4	-	-	-	-	-	21
24	2	+6 2	+4 2	+2 11	12	-2 3	-	-	-	30
28	1	-	-	+1 3	28	-1 11	-1 2	-	-	44
32	0	-	-	-	5	38	11	-	-	54
36	1	-	-	-	1	+1 22	+2 11	-	+4 1	35
40	2	-	-	-	-	+2 1	+4 15	+6 7	-	23
44	3	-	-	-	-	-	+6 7	+9 10	-	17
48	4	-	-	-	-	-	-	+12 1	+16 1	2
Сума частот	-	10	15	18	46	75	46	18	2	238

Таблиця 11.3

Допоміжна таблиця для обчислення $m_{1/1}$ в деревостані сосни звичайної в умовах пробних площ Поліського природного заповідника (за результатми досліджень 2022-2025 рр.)

x_2/x	Чверті:				Срядки 1+4	Срядки 2+3	Срядки 5+6	Рядок 0* Рядок 7	Перевірка
	I	II	III	IV					
0	I	2	3	4	5	6	7	8	9
1	1	11	-	22	25	11	14	14	1.127+16+ +46+54- -5=238
2	3	3+	-	11+1	23	5	18	36	
3	11	-	-	-	4	16	4	12	
4	4	-	-	1+15	18	6	18	72	
6	2	-	-	7+7	26	11	26	156	2.14+18+ +4+18+ +26+3+ +15+6+4+ +3=111
8	10+	-	-	-	3	5	3	24	
9	3	-	-	10	15	16	15	135	
12	5	-	-	1	6	6	6	72	
16	2+	-	-	1	4	11	4	64	
20	3	-	-	-	3	5	3	60	
Σ	3	-	-	-	127	16	111	645	

Частоти рядка та стовпця таблиці, розташованих проти відхилень, рівних нулю, будучи помножені на нуль, дадуть нуль. Виділивши ці нульові рядок і стовбець за допомогою жирних стрілочок, ми розділимо всю таблицю розподілу чотирма чвертями.

Допоміжна таблиця (таблиця 11.3) складається із десяти стовпців (0)-(9). У стовпці (0) виписують у зростаючому порядку, не звертаючи уваги на знак і число повторень, всі добутки відхилень $x_{1(j1)}x_{2(j2)}$ позначені маленькими цифрами у клітинах таблиці розподілу. У стовпці (1) проти відповідних абсолютних значень відхилень виписують із I чверті таблиці розподілу всі частоти, з'єднуючи їх знаком плюс. Так само, у стовпцях (2), (3) і (4) виписують всі частоти з II, Ш та IV чвертей таблиці розподілу. У стовпці (5) записують суми чисел кожного рядка (1) та (4) стовпців, а у стовпці (6) – суми чисел кожного рядка (2) та (3) стовпців. Потім знаходять суми чисел стовпця (5) та стовпця (6). У аналізованому прикладі результат стовпця (5) дорівнює 127, а результат стовпця (6) дорівнює 16. Перш ніж робити подальші обчислення, необхідно зробити перевірку правильності попередньої роботи Для цього складають підсумки стовпців (5) та (6) допоміжної таблиці 11.3 з підсумками нульового рядка та нульового стовпця таблиці 11.2 розподілу та з отриманої суми віднімають число, яке стоїть у центральній нульовій клітині таблиці розподілу 11.2, оскільки це число було підраховано двічі. В результаті таких дій має вийти число, що дорівнює обсягу таблиці розподілу. У цьому прикладі маємо:

$$127+16+46+54-5=238$$

Перевірка записується у верхній частині стовпця (9) допоміжної таблиці 11.3. Це перевірка 1. Переконавшись таким чином у правильності виконаної обчислювальної роботи, розраховують числа стовпця (7), які отримують шляхом віднімання чисел стовпця (6) з чисел стовпця (5); при цьому необхідно проставляти знак отриманої різниці. Для перевірки цього кроку виконаної роботи зауважимо, що сума чисел стовпця (7) допоміжної таблиці 11.3 має дорівнювати різниці сум стовпців (5) і (6) цієї ж таблиці. У цьому випадку

$$127-16=111$$

Ця перевірка записується в нижній частині стовпця (9); таблиці 11.3. Це перевірка 2. У нас рівність вийшла. Останній крок роботи при обчисленні змішаного моменту $m_{1/1}$ робиться в стовпці (8) допоміжної таблиці. У цьому стовпці розміщуються добутки чисел стовпців (0) та (7). Сума чисел стовпця (8) представляє чисельник змішаного моменту $m_{1/1}$. Розділивши цю суму на об'єм таблиці розподілу, отримаємо момент. У цьому прикладі маємо:

$$m_{1/1} = \frac{645}{238} = 2{,}170$$

При ручному розрахунку, великий обсяг обчислювальної роботи та відсутність можливості повної перевірки в ході обчислень змішаного моменту $m_{1/1}$ були суттєвими недоліками способу добутків. Для обчислення ж змішаних моментів вищого порядку, ручний метод виявляється зовсім непридатним.

В даний час всі подібні розрахунки автоматизовані і, звичайно практично ніхто вручну коефіцієнт кореляції не обчислює, але фахівець повинен знати суть цього процесу та алгоритм обчислень.

Для ручного розрахунку раніше застосовувався спосіб сум як легший в розрахунках. При комп'ютерних обчисленнях простіше застосовувати спосіб добутків (легше складати програму), тому опис громіздкого способу сум тут опускаємо.

Обчисливши перший змішаний момент, визначаємо коефіцієнт кореляції за наведеною вище формулою. Для цього необхідно обчислити перший змішаний центральний момент $\mu_{1/1.}$ Формула для його обчислення наводилася вище, нагадаємо її:

$$\mu_{1/1} = m_{1/1} - m_{1x} \cdot m_{2y}$$

де: $\mu_{1/1}$ – перший основний змішаний момент;

$m_{1x,}$ m_{2y} - відповідно перші початкові моменти розподілу за x та y. У нас це діаметр та висота. Ці моменти обчислені раніше (розділ 9). Вони дорівнюють $m_{1(D)}$= 0,672; m_{2}(m) = 0,500

Тоді: $\mu_{1/1} = 2{,}71 - 0{,}672 \cdot 0{,}500 = 2{,}76 - 0{,}336 = 2{,}374$

Середньоквадратичне відхилення для рядів розподілу по діаметру та висоті обчислені раніше (розділ 9) та рівні

$$\sigma_D = 1{,}9\ (\overline{\sigma} = 7{,}6),\ \sigma_M = 1{,}702\ (\overline{\sigma} = 3{,}403)$$

Тоді:

$$r = \frac{2{,}374}{1 \cdot 9 \cdot 1{,}7} = \frac{2{,}374}{3{,}23} = 0{,}74$$

При малих вибірках, коефіцієнт кореляції можна визначити безпосередньо за наведеними вище формулами. Приклад обчислення коефіцієнта кореляції, що наводить Н. М. Свалів, для малих вибірок показано в таблиці 11.4. Підставивши в формулу величини, обчислені в 1214, отримаємо:

$$r = \frac{227 - (55 \cdot 40)/10}{\sqrt{(313 - \frac{55^2}{10})(166{,}26 - \frac{40^2}{10})}} = \frac{7{,}0}{\sqrt{10{,}50 \cdot 6{,}26}} = 0{,}86$$

Таблиця 11.4

Обчислення коефіцієнта кореляції між довжиною стебла та довжиною кореня сіянців сосни звичайної

Довжина стебла, X, см	Довжина кореня, Y	Відхилення					Обчислення
		x	y	xy	x^2	y^2	
5	3,5	-0,5	-0,5	+0,25	0,25	0,25	
6	4,0	+0,5	0	0	0,25	0	$\overline{X} = \sum X/N = 55/10 = 5{,}5$ см
5	4Д	-0,5	+0,1	-0,05	0,25	0,01	
7	5,0	+1,5	+1,0	+1,50	2,25	1,00	
6	3,5	+0,5	-0,5	0,25	0,25	0,25	
4	3Д	-1,5	-0,9	+1,35	2,25	0,81	$\overline{Y} = \sum Y/N = 40/10 = 4{,}0$ см
5	3,5	-9,5	-0,5	+0,25	0,25	0,25	
4	3,0	-1,5	-1,0	+1,50	2,25	1,00	$r = \sum xy / \sqrt{\sum x^2 \sum y^2} =$
7	5,3	+1,5	+1,3	+1,95	2,25	1,69	$= 7{,}00 / \sqrt{10{,}50 \cdot 6{,}26} = 0{,}86$
6	5,0	+0,5	+1,0	+0,50	0,25	1,0	
Σ55	40,0	0	0	+7	10,50	6,26	

Оцінка показників зв'язку при малих вибірках. Оцінюючи коефіцієнт кореляції при малій вибірці виникають деякі нові проблеми у зв'язку з тим, що при високих значеннях кореляції в генеральній сукупності (ρ) вибіркові коефіцієнти кореляції (r) мають не нормальний, а позитивний (правосторонній) асиметричний розподіл. В таких випадках на основі вибіркового r та його помилки

σ_i можна застосувати лише оцінку значущості p при гіпотезі p= 0, тобто лише одну форму оцінки із двох.
Для даних таблиці 11.4 при r=0,86 маємо таку його оцінку:

$$\sigma_r = \sqrt{(1-r^2)/(N-2)} = \sqrt{(1-0{,}86)(10-2)} = 0{,}18$$

Потім обчислимо t-критерій і порівняємо його з табличними значеннями (Додаток Е).

$$t_r = r / \sigma_r = 0{,}86 / 0{,}18 = 5{,}33$$

Отримане значення t_r перевищує $t_{0,001}$ =5,0. Отже, r відповідає високому рівні достовірності. Цю форму оцінки не можна застосовувати за інших гіпотез, крім p= 0. Так само критерій t не можна використовувати для побудови довіряючого інтервалу. Р. А. Фішер запропонував для цієї мети z - перетворення величини p.
Величина z обчислюється по формулі:

$$z = 1/2\ [ln\ (1+r) - ln\ (1-r)]$$

Маємо нормальний розподіл з дисперсією:

$$\sigma_z^2 = 1/(N-3)$$

Для сукупності довжин стебел і коріння при r=0,86 z = 1,293.
Помилка цієї величини $\sigma_z = \sqrt{1/(10-3)} = 0{,}378$.

Довірчий інтервал для z_r в генеральній сукупності дорівнює

$$z \pm t_{0,05} S_z = 1{,}293 \pm 2{,}3 \cdot 0{,}378,$$

тобто від 0,424 до 2,162. Тоді інтервал дорівнює від 0,40 до 0,97.
Перетворення z застосовується також при порівнянні двох вибіркових коефіцієнтів кореляції та при знаходженні узагальненого для них коефіцієнта коли виявляється, що вибірки взяті з однієї сукупності. Нехай взято 2 вибірки сходів сосни, ймовірно з однієї сукупності: r_1=0,86, S_{r1} = 0,18, N_1=10 і r_2=0,70; σ_{r2} = 0,25, N_2 = 10. Чи можна вважати різницю в коефіцієнтах значущим? Квадратична помилка різниці z_1 та z_2 наступна:

$$\sigma_{z1-z2} = \sqrt{[1/(N_1-3)] + [1/(N_2-3)]}$$

Для порівняльних вибірок отримуємо:

$$\sigma_{z1-z2} = \sqrt{[1/7 + 1/7} = 0{,}54,$$

$$t_{z1-z2} = (z_1 - z_2) / S_{z1-z2} = (1{,}293 - 0{,}867) / 0{,}54 = 0{,}78 < t_{0,05}$$

Наявну різницю між коефіцієнтами кореляції слід вважати випадковою, а вибірки – взятими з однієї генеральної сукупності. Об'єднання вибіркових коефіцієнтів кореляції доцільно, якщо кореляція між ознаками у вибірках суттєво не відрізняється, як у розглянутому прикладі. Узагальнений коефіцієнт кореляції є більш надійною оцінкою коефіцієнта кореляції р загальної сукупності. Для отримання r також мають значення z, знаходячи з них середню зважену величину. В якості ваги приймають кількість спостережень в кожній вибірці, зменшене на 3 одиниці, тобто:

$$z=[z_1 (N_1 - 3) + z_2 (N_2 - 3)] / [(N_1 - 3) + (N_2 - 3)]$$

Для двох вибірок сіянців сосни звичайної маємо:

$$z = (1{,}293 \cdot 7 + 0{,}867 \cdot 7) / (7 + 7) = 1{,}080$$

Квадратична помилка середнього зваженого z:

$$\sigma_z = 1 / \sqrt{(N_1 - 3) + (N_2 - 3)}$$

Для нашого прикладу $\sigma_z = 0{,}071$, $t_z = z/S_z = 1{,}080/0{,}071 = 15 > t_{0.05}$ і Навіть $t_{0,01}$, тобто z має значно високий рівень значущості. Знайденому значенню z=1,080 відповідає r=0,79, який є найбільш надійною оцінкою r генеральної сукупності. Коефіцієнт кореляції може набувати значення від +1 до -1. При повній прямій кореляції r = +1, при повній зворотній r = - 1. r = 0 або близький до 0 - прямолінійний зв'язок відсутній (криволінійний зв'язок при цьому може бути, наприклад, коло). Зазвичай вважають, при величині до 0,1 – зв'язку немає; r = 0,11-0,30 - зв'язок слабкий, при r = 0,5-0,6 - середній; при r = 0,7-0,9 – сильний або тісний; к = 0,9-1,0 – дуже високий. Таким чином, у нашому прикладі зв'язок між діаметрами і висотами дерев сосни звичайної виявився тісний. Наявність чи відсутність зв'язку між різними явищами чи ознаками можна підтвердити іншими прикладами. Так, немає зв'язку між класом бонітету деревостану та наявністю лісових доріг. Слабкий зв'язок між водозабезпеченістю деревостанів та рівнем залягання грунтових вод на гідроморфних ґрунтах: він проявляється тільки в посушливі роки. Середній зв'язок між величиною поточного приросту ($м^3$/га) та повнотою, так як. тут проходять два взаємно протилежні процеси: зменшення числа дерев і збільшення приросту на їх стовбурах, що залишилися. Сильний зв'язок спостерігається між класом бонітету деревостану та рівнем ґрунтової родючості. Дуже високий зв'язок існує в дерев між діаметрами і висотами, висотами та видовими числами. ***Детальні дослідження теорії кореляції показали, що ступінь зв'язку випадкових величин x і y більш точно описує квадрат коефіцієнта кореляції, який отримав назву коефіцієнта детермінації (d), який визначається як $d=r^2$.*** Коефіцієнт детермінації, виражений у відсотках, показує ту частину мінливості

залежної змінної (у), яка викликана впливом незалежної змінної (х), тобто він чіткіше виражає залежність у = f(х).

11.3 Кореляційне відношення як міра криволінійного зв'язку.

Зв'язок між х і у далеко не завжди виражається лінійно. ***Тому для будь-якої форми зв'язку, К. Пірсон запропонував статистику залежності між випадковими величинами х та у, яку назвав кореляційним відношенням.***

Статистичне обґрунтування цього показника має певну складність і тут не наводиться. Найбільш важливо знати обчислення кореляційного відношення. Особливо якщо за допомогою коефіцієнта кореляції встановлено, що зв'язок слабкий або середній (г <0,7-0,8). У цьому випадку зв'язок може бути навіть сильним, але мати криволінійний характер. Наприклад, для залежності у =f(х), що має вигляд кола, г = 0, хоча зв'язок є і дуже тісний. Тому і було запропоновано новий показник – кореляційне відношення, - який з успіхом відображає як лінійний, так і нелінійний зв'язок. Кореляційне відношення виражають грецькою літерою ета (іта)- η. Попередню оцінку криволінійного зв'язку можна зробити, побудувавши графік функції у =f(х), на якому вид зв'язку (прямолінійний або криволінійний) буде видно. Для обчислення кореляційного відношення будують таблицю розподілу. Подібну ми складали щодо коефіцієнта кореляції (таблиця 11.5)

Таблиця 12.5

Загальний вид таблиці вихідних даних для обчислення кореляційного відношення

х	у	γ
x_1	y_{11} y_{12} y_{13} … y_{1n}	γ_1
x_2	y_{21} y_{22} y_{23} …y_{2n}	γ_2
x_3	y_{31} y_{32} $y_{зз}$ …y_{3n}	γ_3
…	…	…
x_k	Yk1 Yk2 Yk3 … Ykп	γ_k

За даними таблиці 11.5 обчислюємо наступні суми квадратів:

$$\sigma_y^2 = \sum (y_i - \overline{\gamma_i})^2$$

де: γ_i – умовне середнє, обчислене для кожного рядка.

$$\sigma_{xy}^2 = \sum (\overline{\gamma_i} - \overline{\gamma})^2$$

де: γ – середня по γ для всієї сукупності.

$$\sigma_0^2 = \sum (\overline{\gamma_i} - y_i)^2$$

де: σ_0 - залишкова сума квадратів, обумовлена іншими причинами, ніж регресія у на х.
Тоді:

$$\sigma_y^2 = \sigma_{yx}^2 + \sigma_0^2$$

$$\text{а} \quad \eta^2 = 1 - \frac{\sigma_0^2}{\sigma_y^2}$$

Після елементарних математичних перетворень отримуємо:

$$\eta^2 = \frac{\sigma_y^2 - \sigma_0^2}{\sigma_y^2} = \frac{\sigma_{yx}^2}{\sigma_y^2}$$

Вираз є відношенням суми квадратів відхилень, обумовленої залежністю у від х до загальної суми квадратів відхилень. Для лінійного зв'язку це збігається з визначенням коефіцієнта детермінації, але форма зв'язку можлива будь-яка.

Пояснимо сказане прикладом обчислення кореляційного відношення. Візьмемо вже раніше аналізований розподіл 238 дерев сосни звичайної по діаметру та висоті (таблиця 11.1). Для зручності користування представимо її у вигляді таблиця 11.6, так як нам потрібно ввести додаткові графи, а в тому вигляді, якою вона є, всі графи важко розмістити на одному аркуші. В таблиці 11.6 знаком y^2 позначені групові середні, обчислені в таблиці 11.1. Загальне середнє (для сукупності) позначено у . За даними таблиці 11.6 визначаємо:

$$\overline{x} = \frac{5952}{238} = 25$$

Загальна дисперсія:

$$\sigma^2 = \frac{2758}{238-1} = \frac{2758}{237} = 11{,}64$$

Дисперсія згрупованих середніх:

$$\sigma_{yx}^2 = \frac{2305{,}1}{237} = 9{,}73$$

Таблиця 11.6

Обчислення кореляційного відношення для зв'язку діаметрів та висот сосни звичайної

Ступені товщини, x_i	Ступені висоти, y_i	n_i	$y_i n_i$	$\bar{y}_2$	$y_i - \bar{y}$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2 \cdot n_i$	$\bar{y}_2 - \bar{y}$	$(\bar{y}_2 - \bar{y})^2$	$(\bar{y}_2 - \bar{y})^2 n_i$
12	16	3	48	16	-9	81	243	-9,0	81,0	243,0
16	16	3	48	18,0	-9	81	243	-7,0	49,0	441,0
	18	3	54		-7	49	177			
	20	3	60		-5	25	75			
	Σ	9	162	-	-	-	465			
20	16	2	32	19,52	-9	81	162	-5,48	30,03	630,1
	18	5	90		-7	49	245			
	20	10	200		-5	25	250			
	22	4	88		-3	9	36			
	Σ	21	410	-	-	-	693			
24	18	2	36	22,8	-7	49	98	-2,2	44,4	133,2
	20	2	40		-5	25	50			
	22	11	242		-3	9	99			
	24	12	288		-1	1	22			
	26	3	78		+1	1	3			
	Σ	30	684	-	-	1	272			
28	22	3	66	24,54	-3	1	27	-0,46	0,21	9,2
	24	28	672		-1	1	28			
	26	11	286		+1	1	11			
	28	2	56		+3	1	18			
	Σ	44	1080	-	-	1	74			
32	24	5	120	26,22	-1	9	5	+1,22	1,49	88,5
	26	38	988		+1	49	38			
	28	11	308		+3	-	99			
	Σ	54	1416	-	-	81	142			
36	24	1	24	26,74	-1	81	1	+1,74	3,03	106,1
	26	22	572		+1	49	22			
	28	11	308		+3	25	99			
	32	1	32		+7	-	49			
	Σ	35	24	-	-	81	171			
40	26	5	26	28,52	+1	1	1	+3,52	12,39	285,0
	28	15	420		+3	9	35			
	30	7	210		+5	25	175			
	Σ	23	656	-	-	-	311			
44	28	7	196	29,18	+3	9	63	+4,18	17,47	297,0
	30	10	300		+5	25	250			
	Σ	17	496	-	-	-	313			
48	30	1	30	31,0	+5	25	25	+6,0	36,0	72,00
	32	1	32		+7	49	49			
	Σ	2	64	-	-	-	74			
Σ	-	238	5952	25,0	-27	-	2758	-	-	2305,1

Кореляційне відношення:

$$\eta_{yx} = \sqrt{\frac{9,73}{11,64}} = \sqrt{0,836} = 0,914$$

Оцінка достовірності кореляційного відношення.
Основна помилка:

$$m_\eta = \sqrt{\frac{1-n^2}{N-2}} = \sqrt{\frac{0,164}{236}} = \sqrt{0,0007} = 0,03$$

Критерій Стюдента:

$$t = \frac{0,914}{0,03} = 30,5$$

Для γ = N-2 = 236 критичне значення t; (додаток Е) навіть за достовірності 99,9% дорівнює 3,29<30,5, тобто наше кореляційне відношення достовірне із високим рівнем значущості.

11.4 Інші статистичні показники кореляції. Використання кореляції у лісовому господарстві.

Під час проведення досліджень не завжди обмежуються знаходженням коефіцієнта кореляції та кореляційного відношення. Буває, що під час спостереження трапляються випадки якісно непорівнянні або ті з них, де проводилася лише фіксація наявності або відсутності ознаки. Тоді роблять ранжування ознак, тобто надають їм певні номери та визначають показник кореляції рангів. Насправді є приклади, коли характер розподілу невідомий, але він вочевидь відрізняється від нормального. У разі, якщо показники піддаються ранжуванню, застосовуються рангові коефіцієнти кореляції. Опишемо знаходження показника кореляції рангів у редакції А. К. Митропольського. Ранг вказує на те місце, яке займає дана одиниця часткової сукупності з-поміж інших її одиниць. Якби кожна з цих одиниць відрізнялася щодо аналізованої ознаки від інших одиниць сукупності, тоді ранги мали б порядкові номери від 1 до n, рівного обсягу часткової сукупності. Якщо ж деякі з одиниць сукупності виявляються щодо аналізованої ознаки однаковими, тоді ранг усіх цих одиниць приймається рівним середньому з відповідних номерів. Показник кореляції рангів р дорівнює:

$$\rho = 1 - \frac{6\sum_{h=1}^{n} d_h^2}{n(n^2-1)}$$

де: величини d_h представляють різниці між рангами одиниць, взятих разом з двох загальних сукупностей.

Подібно до коефіцієнта кореляції r, показник кореляції рангів ρ змінюється від -1 до +1. Чим тісніше буде зв'язок між величинами, тим ближче до одиниці за своєю абсолютною величиною буде показник кореляції рангів; знак показника вказує, чи є зв'язок прямий або зворотної.

Як приклад обчислимо показник кореляції рангів, виражальний зв'язком між об'ємною вагою $X_1(\gamma$, г/см^3) та межею міцності при стисканні X_2 (σ_B, кг/см^3) деревини дуба (таблиця 11.7, яку наводить) А.К. Митропольський).

Для перевірки обчислень зауважимо, що сума рангів повинна дорівнювати сумі натуральних чисел від л1 до n, тобто n(n+1)/2. В розглянутому прикладі:

$$\frac{19(19+1)}{2} = 190$$

Крім того, сума від'ємних різниць повинна дорівнювати сумі додатніх різниць. Застосовуючи формулу, знаходимо:

$$\rho = 1 - \frac{6 \cdot 171{,}5}{19(19^2 - 1)} = 0{,}85$$

Таблиця 11.7

Обчислення показника кореляції рангів

№	X_1	X_2	h_1	h_2	d_h	d_h^2
1	758	676	19	19	0	0
2	739	646	17	13	4	16
3	719	642	13	12	1	1
4	717	635	10	9,5	0,5	0,25
5	678	567	5	5	0	0
6	613	476	2	2	0	0
7	753	635	18	9,5	8,5	72,25
8	724	665	14,5	18	3,5	12,25
9	718	661	12	17	5	25
10	693	610	7	7	0	0
11	658	531	4	3	1	1
12	737	649	16	14	2	4
13	707	635	8	9,5	1,5	2,25
14	717	635	10	9,5	0,5	0,25
15	653	544	3	4	1	1
16	724	650	14,5	15	0,5	0,25
17	717	653	10	16	6	36
18	685	578	6	6	0	0
19	612	462	1	1	0	0
Σ	-	-	190	190	-17,5 17,5	171,5

Ранговий коефіцієнт кореляції Спірмена викладено тут у редакції М. П. Горошко із співавторами. Рівняння коефіцієнта кореляції рангів має вигляд:

$$r_s = 1 - \frac{6\sum d_i^2}{N^3 - N}$$

де:d - різниця між рангами;
x_i, і y_i, - ранги значень x_i, і y_i ($1 \le x_i, \le N$; $1 \le y_i \le N$);
N – об'м вибірки.

Наведений коефіцієнт є непараметричним показником зв'язку та використовується при вивченні як кількісних, так і якісних значень. При цьому закон розподілу цих величин та форма зв'язку нам можуть бути невідомі, як і звичайний коефіцієнт кореляції r_s може мати величину від -1 до +1. При незалежності величин x, y r_s розподіляється ассимптотично нормально при N →□, а ð = 0, де дисперсія $\sigma^2 = \frac{1}{N-1}$.

Як і інші вибіркові показники, коефіцієнт кореляції рангів це оцінка параметра в генеральній сукупності. Його достовірність встановлюють на основі нульової гіпотези, яку приймають для умови, що $r_s=0$ у генеральній сукупності. Ця гіпотеза перевіряється шляхом обчислення t-критерія Стьюдента та зіставлення його з критичною величиною $t_{кр}$

$$t = r_s \sqrt{\frac{N-2}{1-r_s^2}}$$

Зазвичай використовують 95% чи 99% рівень достовірності, де застосовують такі формули.

Для рівня достовірності 99%:

$$t_{кр\%} = \frac{2{,}58}{\sqrt{N-1}}\left(1 - \frac{0{,}69}{N-1}\right)$$

Для рівня достовірності 95%:

$$t_{кр\%} = \frac{1{,}96}{\sqrt{N-1}}\left(1 - \frac{0{,}16}{N-1}\right)$$

При $t < t_{таб}$ приймають нульову гіпотезу (Но), тобто r_s недостовірний. При $t > t_{таб}$ приймається робоча (альтернативна) гіпотеза (На) та r_s є достовірним.

Достовірність r_s можна оцінити і за спеціальними таблицями критичних значень r_s (додаток У), де він дається в залежності від рівня значущості (5%, 1%) та обсягу вибірки (N). При $r_s < r_{таб}$ приймається нульова гіпотеза, а $r_{таб}$-альтернативна або робоча.

Використання кореляції у лісовому господарстві.

У лісовому господарстві наведені величини, що визначають кореляцію між різними ознаками застосовуються дуже широко. Вже понад 100 років коефіцієнт кореляції та кореляційне відношення постійно обчислюють під час проведення наукових досліджень в лісовому господарстві. Ці показники ми зустрічаємо в працях класиків лісівництва та лісової таксації М. М. Орлова, А. К. Турського, А. В. Тюріна, В. К. Захарова та ін. Застосування кореляції бачимо в класичних працях вченого, лісівника та таксатора Ф. П. Моїсеєнко. Він встановив високий рівень кореляції між формою стовбура та середнім діаметром дерева і довів невелику мінливість середньої форми стовбура в деревостані. Це стало науковою основою для складання об'ємних та сортиментних таблиць за середньою формою стовбура, що у кілька разів спростило складання цих нормативів та користування ними.

В даний час обчислення r і η є обов'язковою умовою проведення досліджень у разі наявності зв'язку між величинами, що вивчаються. Застосування ПК зробило цю роботу простою та рутинною. При дослідженні лісових біогеоценозів часто вплив одних показників в біоценозі на інші не очевидно. Для доказу (або спростування) цього впливу користуються показниками кореляції. Наприклад, чи впливає відсоток фізичної глини у різних ґрунтових горизонтах (A_1, A_2, B, C) на величину класу бонітету без обчислення показників кореляції для кожного типу лісу (лісорослинних умов) сказати важко, так як у ряді випадків вплив та інших факторів, наприклад - зволоження ґрунту суттєво впливає на продуктивність деревостанів.

Водночас використання законів кореляції має супроводжуватись логічним аналізом суті явища, що можуть зробити лише фахівці. В лісовому господарстві це лісівники, в окремих випадках біологи (ботаніки, зоологи). Механічне зіставлення логічно незв'язаних величин з обчисленням показників кореляції може призвести до невірних, а іноді і абсурдних висновків. Наприклад, ми можемо зв'язати величину, що характеризує кількість місячного світла з рівнем продуктивності деревостану і навіть отримаємо якісь величини коефіцієнта кореляції, але результат таких розрахунків не відображатиме реальність. Інший приклад. Ми можемо зв'язати кількість сухостою на 1 га деревостану тільки з родючістю грунту, але ці порівняння будуть некоректними, так як на кількість сухостою (відпаду) на 1 га впливають більш значущо інші фактори: густота насадження, його вік, проведені раніше рубки догляду.

РОЗДІЛ 12. КОРЕЛЯЦІЙНИЙ АНАЛІЗ ЯК ІНСТРУМЕНТ НАУКОВОГО ДОСЛІДЖЕННЯ

12.1 Мета та завдання кореляційного аналізу.

В розділі 11 ми розібрали зв'язки між двома величинами. Встановили, що вони бувають прямолінійні та криволінійні. Для характеристики тісноти зв'язку використовується коефіцієнт кореляції (r) та кореляційне відношення (η). Але, встановивши сам факт залежності однієї величини від іншої, отримали ще недостатню інформацію. Необхідно знати форму цього зв'язку, її числове вираження. Тоді, маючи одну з корелюючих величин, зазвичай ту, яка легше і простіше визначається, можна з певною часткою ймовірності передбачати значення іншої (досліджуваної) величини у: $y=f(x)$. Для вирішення цього завдання використовують кореляційні рівняння або як їх ще називають, кореляційні моделі. Ми вже зазначали, що визначень моделей є багато, але краще інших їх висловлює наступне формулювання, дане В. А. Штоффом:

«Модель - це уявна чи матеріально реалізована система, яка, відображаючи або відтворюючи об'єкт дослідження, здатна заміщати його так, що її вивчення дає нам нову інформацію про цей об'єкт».

Саме це визначення використовують у лісівничих дослідженнях К. Є. Нікітін та А.З. Швиденко. Отже, модель - це деяка абстракція, як можлива на певній стадії вивчення деякого предмета або явища, подальшого його пізнання або використовувана для вирішення практичних задач. Кореляційні рівняння є різновидом стохастичних моделей.

Теорія кореляції розроблена в основному наприкінці XIX і на початку XX століття Карлом Пірсоном та Юлом. Вона дозволяє описати різні зв'язки, але не розкриває причин їх походження. Тут потрібний спеціальний аналіз: біологічний, лісівничий, генетичний тощо. При цьому вивчаючи кореляцію, ми маємо знати причинний зв'язок, так як інакше можемо зробити помилку, знайшовши зв'язок там, де його немає. Про це добре сказав великий англійський письменник Бернард Шоу (1856-1956) ще в 1906 році у передмові до свого видатного твору «Доктор на роздоріжжі»: «Навіть досвідчені статистики часто виявляються не в змозі оцінити, наскільки сенс статистичних даних спотворюється мовчазними припущеннями їх інтерпретаторів... Легко довести, що носіння циліндрів та парасольок розширює грудну клітину, подовжує життя та дає відносний імунітет

від хвороб... Математик, чиї кореляції привели б у захоплення Ньютона, може збираючи дані та роблячи з них висновки впасти в абсолютно грубі помилки на основі таких популярних помилок, як описані вище». В цьому вислові Б. Шоу наголосив, що не самі циліндри та парасольки призводять до описаних наслідків, а спосіб життя їх володарів, якими на той час були багаті люди.

Раніше вже зазначено, що і на початку поширення статистики, а потім у 60-70-ті роки, коли математичні методи стали широко застосовувати в лісовому господарстві, математики ще слабо знались на причинних зв'язках предметів, які вони описували за допомогою кореляційних рівнянь, тому зробили багато помилок. Про цю небезпеку попереджали фундатори вчення про кореляцію. Так, Роберт Юл в 1926 році налякав вчених прикладами високих кореляцій між кількістю самогубств в Англії та приналежністю до англіканської церкви. Причинного зв'язку тут немає, а найвища кореляція є. Справа тут пояснюється просто - переважна більшість жителів Англії в ті роки належало до англіканської церкви.

Тому ще раз нагадаємо про важливість проведення професійного аналізу причинно-наслідкових зв'язків, перш ніж взятися за конкретні обчислення та висновки.

12.2 Множинна кореляція.

При розгляді кореляції часто трапляються випадки, коли дві величини, начеб то взаємозалежні, але більш докладний їх аналіз показує, що ця «взаємозалежність» є відображенням того факту, що вони обидві кореловані з деякою третьою величиною або з сукупністю величин. У природі ми часто зустрічаємось з такими явищами, коли зміна однієї величини (функції) визначається зміною не одного, а кількох аргументів. Наприклад, висота дерева залежить від ґрунтової родючості, кількості вологи в кореневмісному шарі ґрунту, віку деревостану та інше. Розмір суми площ перерізу чи запас деревостану залежить від висоти (H) та повноти насадження. Додавання до цього рівняння діаметр деревостану (D) буде зайвим через високу кореляцію (H-D). Аналогічний приклад можна навести, розглядаючи ***коефіцієнт форми деревного стовбура (q_2) який дорівнює частці від поділу діаметра на половині висоти дерева до діаметру на 1,3 м: $q_2 = D_{0,5м} / D_{1,3м}$***

Розглядаючи кореляцію q_2-D, ми можемо дійти висновку про її наявність. Насправді є кореляція q_2-H та H-D. Такі закономірності призводять до поняття множинної кореляції. Її суть полягає у наступному.

Коефіцієнт множинної кореляції - це показник тісноти лінійного зв'язку між однією залежною величиною та сукупністю незалежних.

Якщо у нас є 3 сполучені величини X, У, Z, то коефіцієнт множинної кореляції визначається з матриці:

$$R_{XYZ} = \begin{vmatrix} 1 & r_{XY} & r_{XZ} \\ r_{YX} & 1 & r_{YZ} \\ r_{ZX} & r_{ZY} & 1 \end{vmatrix}$$

Розв'язання цієї матриці призводить до рівнянь для визначення $R_{X,Y,Z}$; $R_{Y,X,Z}$; $R_{Z,X,Y}$.

$$R_{XYZ} = \sqrt{\frac{r_{XY}^2 + r_{XZ}^2 - 2r_{XY}r_{XZ}r_{YZ}}{1 - r_{YZ}^2}}$$

$$R_{YXZ} = \sqrt{\frac{r_{XY}^2 + r_{YZ}^2 - 2r_{XY}r_{XZ}r_{YZ}}{1 - r_{XZ}^2}}$$

$$R_{ZXY} = \sqrt{\frac{r_{XZ}^2 + r_{YZ}^2 - 2r_{XY}r_{XZ}r_{YZ}}{1 - r_{XY}^2}}$$

де: r_{xy}, r_{XZ}, r_{yz} – парні коефіцієнти кореляції між величинами X, У, Z.

Замість коефіцієнта множинної кореляції, який зазвичай позначають як К на відміну парного - r, часто при моделюванні зручніше використовувати коефіцієнт множинної детермінації - $D=R^2$ (на відміну від парного $d=r^2$). Він вимірює ту частку загальної дисперсії залежною змінною у, яка може бути пояснена впливом зміни аргументів. Значимість коефіцієнта множинної кореляції оцінюють за F-критерієм Фішера.

$$F = \frac{R^2}{1 - R^2}\left(\frac{N - k}{k - 1}\right)$$

де: N – об'єм вибірки;
k - кількість факторів впливу (кількість незалежних змінних).

Критичні значення F беруть із таблиць (додаток Ж) при числі ступенів свободи $\gamma_1 = k\text{-}1;\ \gamma_2 = N\text{-}k$. При $F_{факт} < F_{таб} = H_0$, тобто приймається нульова гіпотеза про відсутність кореляції і, навпаки, приймають альтернативну або робочу гіпотезу, тобто про наявність кореляції, при $F_{факт} \geq F_{таб}$-

Величина К (як і показника r) коливається не більше (-1)-(0)-(1).

Межі (l) , в яких може знаходитись R_{xz}, якщо відомі R_{xy} і R_{xz} визначають за формулою:

$$l = R_{XY} \cdot R_{XZ} - \sqrt{(1-r_{XY}^2)(1-r_{XZ}^2)} < R_{YZ} < R_{XY} R_{XZ} + \sqrt{(1-r_{XY}^2)(1-r_{XZ}^2)}$$

Для того, щоб з'ясувати частку впливу однієї з величин на іншу в загальній системі кількох взаємовпливових показників, введено поняття конкретного коефіцієнта кореляції. Графічно це можна висловити в такий спосіб.

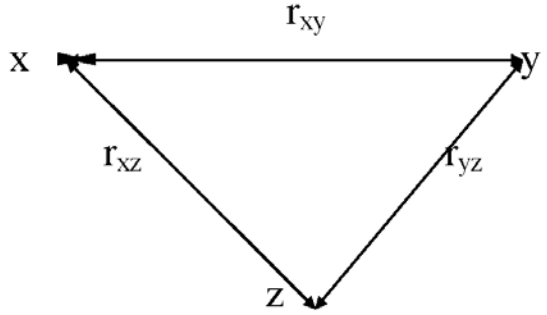


Рис. 12.1 Графічна інтерпретація конкретної кореляції

Якщо існує тісна кореляція між x-z та y-z, то зв'язок між x-y може створюватися за рахунок одночасного впливу на x і y третьої ознаки z. Щоб знайти той вплив, який надає x на y (або y на x), треба досліджувати залежність x-y при постійному z, тобто x і будуть змінюватися, а z= const (тобто z ми елімінуємо).

Обчислений в даному разі коефіцієнт кореляції x-y називається конкретним коефіцієнтом кореляції. Формула його обчислення наступна:

$$r_{xy\text{-}(z)} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

$$r_{xz\text{-}(y)} = \frac{r_{xz} - r_{xy} \cdot r_{zy}}{\sqrt{(1-r_{xy}^2)(1-r_{zy}^2)}}$$

$$r_{zy\text{-}(x)} = \frac{r_{zy} - r_{zx} \cdot r_{yx}}{\sqrt{(1-r_{zx}^2)(1-r_{yx}^2)}}$$

Статистиками множинної нелінійної кореляції є множинні та конкретні кореляційні взаємозв’язки.

Емпіричне множинне кореляційне відношення $\eta_{1,23}$ величини X_1 до величин X_2 та X_3 подано виразом:

$$\eta^2_{1.23} = \sum_{j_2=1}^{k_2}\sum_{j_3=1}^{k_3} p'_{j_2 j_3} \cdot r^2_{1(j_2 j_3)}$$

Або рівнозначними йому виразами:

$$\eta^2_{1.23} = 1 - \frac{\sum_{j_2=1}^{k_2}\sum_{j_3=1}^{k_3} p'_{|j_2|j_3} \cdot \mu_{2|(j_2)|(j_3)}}{\mu_{2/0/0}},$$

$$\eta^2_{1.23} = \frac{\sum_{j_2=1}^{k_2}\sum_{j_3=1}^{k_3} p'_{/j_2/j_3} \cdot (m_{1/(j_2)/(j_3)} - m_{1/0/0})^2}{\mu_{2/0/0}}$$

Емпіричне конкретне кореляційне відношення $\eta_{1,23}$ величини X_1 та X_2 при даному значенні X_3 обчислюється за формулою:

$$\eta^2_{12.3} = 1 - \frac{\sum_{j_2=1}^{k_2} p'_{/j_2/(j_3)} \cdot \mu_{2/(j_2)/(j_3)}}{\mu_{2/0/(j_3)}}$$

або

$$\eta^2_{12.3} = \frac{\sum_{j_2=1}^{k_2} p'_{/j_2/(j_3)} \cdot (m_{1/(j_2)/(j_3)} - m_{1/0/(j_3)})^2}{\mu_{2/0/(j_2)}}$$

де: m_{ij}, η_{ij} – відповідні моменти, що описані у розділі 11;
p_{ij} – частоти, що також були описані у розділі 11.

Як приклад застосування конкретного коефіцієнта кореляції в лісовому господарстві, наведемо встановлені залежності q_2, від d, та h де q_2 – другий коефіцієнт форми, h – діаметр дерева на висоті 1,3 м, h – висота дерева. Коефіцієнт множинної кореляції досягає 0,95-0,96. В той же час конкретні коефіцієнти кореляції виявилися r_{HD} - 0,75-0,90; r_{DH}- 0,73-0,92; r_{Hq2} - 0,90-0,92; r_{Dq2} - 0,16-0,20.

Довгий час вважалося, що q_2 залежить від обох величин d та h. Подібну помилку зробив навіть вчений-таксатор Матвєєв-Мотін. Лише Ф. П. Моїсеєнко, вивчивши конкретну кореляцію q_2-d та q_2,-h довів, що величина коефіцієнта кореляції у зв'язку q_2-d визначається високою кореляцією d-h. При фіксованих h коефіцієнт кореляції q_2-d виявився вище 0,2, тобто цього зв'язку практично не було. Чому ж до Ф. П. Моісеєнко вчені помилялися? Адже як методично проводити дослідження більшість з них знали. Але при цьому потрібен був величезний експериментальний матеріал. Маючи близько 18 000

зрубаних та обміряних дерев, Ф. П. Мойсеєнко такий матеріал зібрав, він зміг обчислити конкретні r, а в інших вчених такої великої кількості вимірів не було. Водночас в роки, коли проводилась ця робота (1936-1941 рр.), подібні обчислення вимагали тривалого часу, тому не всі вчені були до цього готові. Тепер на ПК така робота робиться за лічені хвилини.

12.3 Кореляційні моделі.

Кореляційні моделі є рівняннями, де на основі зв'язку або взаємозалежності двох величин будується рівняння, якому одна з величин виступає як аргумент, а інша - функцією, тобто. y=*f*(x). Побудова таких моделей пов'язана з обчисленням коефіцієнта кореляції, тобто тут передбачається лінійна залежність. З курсу математики знаємо, що лінійна залежність реалізується у вигляді рівняння прямої. Його формула:

$$y = a + вx$$

де: а, в – деякі коефіцієнти.

Тут коефіцієнт а показує величину відступу від початку координат, в - кут нахилу прямої до осі ох.

Наведемо приклад. Для цього скористаємося результатами вимірів діаметрів та висот сосни звичайної в молодому віці (таблиця 12.1).

Таблиця 12.1

Результати вимірювання діаметрів та висот молодих дерев сосни звичайної в умовах пробної площі №3 Селезівського ПНДВ

№ п\п	Діаметр, см	Висота, м	№ п\п	Діаметр, см	Висота, м	№ п\п	Діаметр, см	Висота, м
1	7,6	8,7	6	8,8	9,8	11	8,2	9,2
2	6,4	7,5	7	9,0	10,1	12	4,0	5,6
3	5,2	6,6	8	6,2	7,2	13	5,6	7,0
4	4,1	5,7	9	7,0	8Д	14	4,8	6,5
5	3,7	5,2	10	7,4	8,5	15	6,9	8,0

Для того, щоб оцінити вид моделі та загальну закономірність зміни функції залежно від зміни аргументу, побудуємо графік. На осі абсцис відкладатимемо діаметри, на осі ординат – висоти (рис. 12.2).

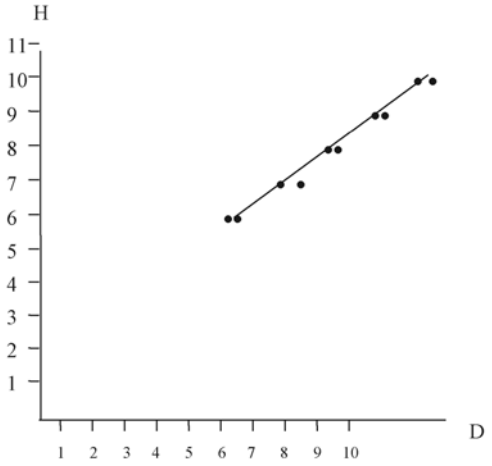


Рис. 12.2 Залежність висот та діаметрів у соснових молодняках пробної площі №3 Селезівського ПНДВ

З рис. 12.2 видно, що точкі, які показують співвідношення D-H лежать на прямій лінії. Раніше в лісівничих дослідженнях обмежувалися графічною побудовою прямих. Переваги графічної побудови очевидні: простота та наочність. Її недоліками є певний суб'єктивізм. Навіть аналізуючи переконливий рисунок 12.2, можна розійтися на думці, як провести пряму лінію, хоча розбіжності в різних виконавців будуть невеликі. Кореляційні рівняння бувають не тільки двомірні (парні), а й багатовимірні. Їхній вигляд наступний:

$$y = ax_1 + bx_2 + cx_3 + \ldots + nx_n$$

Графічне зображення множинних зв'язків (після трьох аргументів) важко відтворити. Тому графічні побудови тут зазвичай не роблять. Обчислення коефіцієнтів кореляційних рівнянь здійснюють за спеціальними методиками, які ми розглянемо пізніше. Тут ще зупинимося на деяких особливостях застосування кореляційних рівнянь. У загальну формулу кореляційного рівняння всі показники проставляються у своїх одиницях виміру: дії проводяться над їх абсолютними величинами, не зважаючи на найменування. Відповідь виходить у одиницях виміру залежної величини. Для перевірки правильності отриманого рівняння, до нього підставляється середнє значення незалежної ознаки в його одиницях виміру, а у відповіді має вийти середнє значення залежного змінного, рівне вихідному.

Кореляційне рівняння застосовується для отримання точного середнього значення залежної змінної, за яке приймається важко вимірна ознака, знаючи точне середнє значення незалежної змінної, за яке приймається швидко і легко вимірна ознака, яка кореляційно взаємопов'язана з першою. Так щодо діаметрів і висот деревостану за незалежну змінну приймають діаметр на висоті 1,3 м, який значно

легше, простіше та точніше вимірюється, ніж висота. У цьому випадку необхідна кількість спостережень для залежної ознаки потрібно значно менше того, яке було б необхідне при самостійному аналізі цієї ознаки, і обчислюється за формулою (якщо точність задана у відсотках):

$$n = \frac{V^2}{p^2}(1 - r^2)$$

А точність досліду обчислюється за формулою:

$$p = \frac{V}{\sqrt{n}} \cdot \sqrt{1 - r^2}$$

Згадаємо, що для одного статистичного ряду: $n = \frac{V^2}{p^2}$.

Нехай потрібно визначити середню висоту з точністю 1%, знаючи точний середній діаметр на відведеній ділянці лісу. За даними, наявними у лісотаксаційній літературі відомо, що коефіцієнт варіації висот у межах стиглого деревостану дорівнює 12 - 15%. Значить для отримання середньої висоти з точністю 1% потрібно виміряти $n = \frac{V^2}{p^2} = 15^2/1^2 = 225$ що досить складно зробити.

Оскільки є високий кореляційний зв'язок між діаметрами та висотами, то для отримання середньої висоти, з тією ж точністю потрібно значно менша кількість виміряних висот. Так, при звичайному коефіцієнт кореляції між D та H рівному 0,90, він складе всього 225(1-0,902) = 225(1-0,81) = 43. Вимірявши у 43 дерев висоти та діаметри, обробляємо отримані дані та складаємо кореляційне рівняння, яке підставляємо точну величину середнього діаметра. Середню висоту потім знаходимо за кореляційним рівнянням із заданою точністю.

Так, нехай за даними виміру висот у 43 стовбурів дерев рівняння зв'язку має вигляд: h=0,40d+14,2м, а точний середній діаметр 32,2см; звідси отримана висота дорівнює h = 0,40-32,2 +14,2 = 12,9 +14,2 = 27,1 м. При цьому виявилося: V = 15%, а r = 0,90. Точність досліду, тобто отриманої середньої висоти, дорівнюватиме:

$$- p = \frac{15}{\sqrt{43}} \cdot \sqrt{1 - 81} = 2{,}27 \cdot 0{,}436 = 0{,}99 = 1\%$$

Кутовий коефіцієнт кореляції рівняння $R = r\frac{\sigma_y}{\sigma_x}$ ***називається коефіцієнтом регресії***.

Він показує, на скільки одиниць в середньому змінюється залежна ознака, якщо незалежна змінилась на одну одиницю.

У наведеному вище прикладі ($\sigma_д$=6,96 і $\sigma_н$=2,69 обчислені раніше) кутовий коефіцієнт отриманий:

$$R = 0{,}90 \cdot \frac{2{,}69}{6{,}96} = 0{,}90 \cdot 0{,}386 = 0{,}35 \text{ м.}$$

Це означає, що при зміні діаметра на 1 см, висота дерев в середньому змінюється на 0,36 м. В загальному вигляді лінійна регресія між двома змінними величинами може бути виражена за допомогою наступного загального рівняння:

$$y = \bar{y} + r \frac{\sigma_y}{\sigma_x} (x - \bar{x}),$$

де: x, y – змінні варіаційних рядів;

σ_x, σ_y - середньоквадратичні відхилення варіаційних рядів;

r – коефіцієнт кореляції.

Основна помилка рівняння кореляційного зв'язку визначається за формулою:

$$m_{y/x} = \sigma_y \sqrt{1 - r^2}$$

12.4 Кореляційні рівняння в лісовому господарстві.

Кореляційні рівняння дуже широко застосовуються як в лісівничих дослідженнях, так і у практичній роботі, особливо в лісовпорядкуванні та під час проведення проектних робіт. Прикладів дуже багато.

Так, зв'язок діаметрів та висот і конкретний вираз цього зв'язку безпосереднім чином використовують при побудові сортиметальних таблиць. Їх зазвичай будують за розрядами висот. Для визначення розряду висот використовують виміри діаметрів та висот в 9-12 дерев. Далі для обчислення об’єму дерев застосовують спеціальні сортиментні таблиці, де закладено рівняння зв'язку H-D, виведене під час проведення наукових досліджень. Раніше, до виведення таких зв'язків були таблиці, що вимагали вимірювання висоти у кожного дерева. Це дуже трудомістко. Використання тут кореляційних рівнянь знизило витрати на відведеннях лісосік в десятки і сотні разів. Продовжуючи тему сорментації зазначимо, що з обчисленням об’ємів стовбурів та виходу сортиментів в сосни після підсочування, коли форма стовбура на висоті 1,3 м деформована карами, з яких збирають живицю, використано зв'язок діаметра на 1,3 м, виміряного по ременях, тобто по непошкодженій частині стовбура та діаметра на 0,5 Н.

Запас деревостану під час проведення лісоінвентаризації інженер-таксатор визначає з таблиці, де основними даними служать висота і повнота, тобто використовується кореляція. Дуже широко

використовуються кореляційні залежності в мисливствознавстві. Наприклад, за формою слідів та їх розмірами досвідчені мисливствознавці визначають стать, вік, стан тварини. Велике значення кореляції в захисті лісу від шкідників. Є рівняння, що описують характер розмноження шкідника, очікуваний збитки від стану популяції в певний момент часу, погодних умов та стану лісового насадження. Використання таких рівнянь дозволить зробити прогнози масового розмноження шкідників та провести профілактичні заходи захисту насаджень.

Знаходження величин у за кореляційними рівняннями з формули y = a+bx називається вирівнюванням чи апроксимацією.

Вирівняні значення функції відображають закономірну її зміну (з деякою основною помилкою), де виключені випадкові впливи на окремі виміряні значення варіаційного ряду.

В наукових дослідженнях необхідно не просто виміряти чи визначити якісь величини, але й знайти закономірні зв'язки між ними, визначити те загальне, що пов'язує функцію та аргументи, виразити це математично. Більш загальні випадки визначення залежності y = f(x) будуть розглянуті далі.

РОЗДІЛ 13. РЕГРЕСІЙНИЙ АНАЛІЗ

13.1 Сутність регресійного аналізу. Регресійні моделі.

Розглянуті у попередньому розділі кореляційні рівняння є окремим випадком більш загального виду ймовірнісних зв'язків, які виражаються методами регресійного аналізу. Пояснимо його сутність. Нехай у нас є деяка функція y = f(x). Вона може набувати різні вирази в залежності від того, що мається на увазі під f(x). Наприклад, вже відомий нам вираз рівняння прямої y = a+bx або рівняння параболи другого порядку $y = a + bx + cx^2$, або гіперболи $y = a + b/x$ і т.д. Для позначення різних зв'язків у біологічній статистиці англійським вченим Ф. Гальтоном запропоновано термін регрєсія. Сенс цього терміна полягає в тому, що корелюючі пари в біологічних об'єктах, які виявляючи в потомствах відхилення від середньої лінії, визначають кореляцію ознак сукупності, мають тенденцію повернення до цієї середньої, якщо діють лише випадкові причини. Надалі ми користуватимемося цим терміном як більш загальним, говорячи про рівняння стохастичного зв'язку між випадковими величинами.

Регресійні моделі – це рівняння стохастичного зв'язку виду y = f(x), де f(x) може бути виражено у будь-якому вигляді. Кореляційні моделі - окремий випадок регресійних, коли зв'язок має прямолінійний характер. Регресійні моделі зазвичай використовують для вираження різного роду зв'язків в лісовій таксації, лісівництві та інших лісових дисциплінах.

Найчастіше вони застосовуються для знаходження загальної залежності за експериментальними даними. У цьому випадку виведене рівняння служить для вирівнювання матеріалу, отриманого під час постановки досвіду. В цьому сутність головної тенденції зміни функції залежно від зміни аргументів, та усуваються випадкові відхилення. Таке рівняння зручно використовувати для отримання величини функції через рівний крок аргументу, хоча в дослідному матеріалі цей рівний крок не завжди витримується. Наприклад, нам зручніше, вивчаючи хід росту деревостану, мати дані через 10 років: у 10, 20, 30, ... 100 років, а наші пробні площі мають вік 9, 22, 29, 44, 57, ... 102 роки. Тому необхідно знайти проміжні значення в 10, 20,

30 років за даними обліків пробних площ, що називають зглажуванням дослідних даних.

13.2 Методи визначення виду регресійних рівнянь та їх параметрів.

При обробці дослідних даних, часто доводиться вирішувати завдання, в якому необхідно досліджувати залежність однієї фізичної чи біологічної величини у від іншої фізичної або біологічної величини х. Наприклад, залежність продуктивності деревостану від кількості опадів та середніх температур, розмірів лося від величини його сліду, об'єму дерева від лісорослинних умов, коефіцієнта форми стовбура від його висоти тощо. Нехай проводиться досід з метою дослідження залежності величини у від величини х, яка в загальному випадку може бути записана у вигляді:

$$y = f(x)$$

Вигляд цієї залежності потрібно визначити з досліду. Припустимо, що ми досліджуємо залежність видового числа деревостою (Fq) з його середньої висоти (Н). В результаті досліду отримано низку експериментальних точок (x_1 y_1), (x_2, y_2), ..., (x_n, y_n), які наведені в таблиці 13.1. За даними таблиці 13.1 побудовано графік зміни змінної величини у(Fq) пр різних незалежних змінних х(Н) (рис. 13.1). На ньому показано зв'язок видового числа, що розглядається (залежна змінна - у) від висоти Н (незалежна змінна – х). Ця залежність зазвичай виражається рівнянням гіперболи.

$$y=a+b/x$$

Таблиця 13.1

Середні видові числа (Fg) деревостану сосни звичайної в залежності від середньої висоти (Н) в умовах пробної площі №2 Перганського ПНДВ

Н	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36
Fg0.001	615	568	541	522	509	495	492	486	482	476	472	469	467	465	464	463

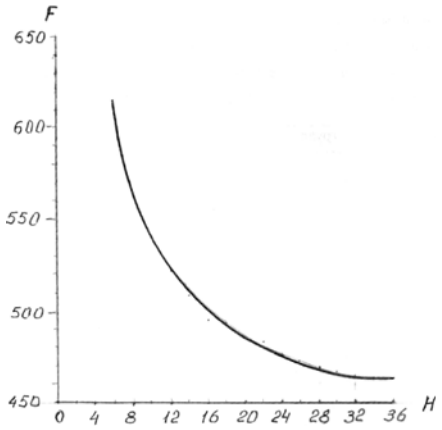


Рис. 13.1 Графік зміни видового числа деревостану сосни звичайної в залежності від середньої висоти

Оскільки виміри, що проводяться в ході досліду, пов'язані з помилками випадкового характеру, то зазвичай експериментальні точки на графіку мають певний розкид щодо загальної закономірності. В силу випадковості помилок виміру, цей розкид або ухилення точок від загальної закономірності також випадкові. ***Отже, завдання полягає в такій обробці експериментальних даних, при якій по можливості точно булаб б відображена тенденція залежності у від х, і можливо повніше виключити вплив випадкових, незакономірних відхилень, пов'язаних з похибками досліду. Таке завдання є типовим для практики і називається завданням згладжування експериментальної залежності.***

Дуже часто буває так, що вид залежності y = f(x) до досліду відомий з фізичних, біологічних чи лісівничих міркувань, пов'язаних зі змістом розв'язуваної задачі. Тоді на підставі дослідних даних потрібно визначити лише деякі параметри зв'язку, які входять до неї лінійно, чи з іншої моделі. Наприклад відомо, що залежність діаметра та висоти в молодняках до 10 - 15 років можна виразити рівнянням прямої. Трапляються і складніші випадки. Так, запас деревини (M) в лісовому насадженні визначають через суму площ перерізів стовбурів на висоті 1,3 м (G), середню висоту (H) та видове число (F) за формулою M=G-H-F. Тільки G досить просто обчислити безпосередньо, вимірявши діаметри стовбурів. Середню висоту визначають після виміру 12-15 дерев (D і H у кожного) у зв'язку H=F(D). Цей зв'язок виражається поліномами різних ступенів, логарифмічної кривої та іншими моделями, які підбирають на вигляд їх графіків. Видове число обчислюють за моделями F=f'(H), що описується простою або ускладненою гіперболою. Подібних прикладів в лісовому господарстві безліч. Звичайно, якщо вид зв'язку хоча б приблизно відомий, це спрощує та полегшує роботу з проведення регресійного аналізу. Якщо ж вид зв'язку невідомий, то його треба встановити хоча б орієнтовно, керуючись логічною

верифікацією та професійними знаннями. У звичайних завданнях лісового господарства, часто буває достатньо побудувати графік і зіставити його з графіками відомих функцій. Останні в більшій кількості представлені в спеціальних альбомах. Іноді (для недостатньо очевидних явищ). доводиться перебирати ряд відомих моделей або виводити нові. Останнє в лісовому господарстві відбувається рідко та доступно лише вченим з фаховою професійною лісівничою та математичною підготовкою. Деякі рівняння зв'язку відносно прості, інші ж відрізняються підвищеною складністю за участю кількох незалежних змінних. Наприклад, для знаходження поточного приросту деревостану як незалежних змінних для певної деревини, повинні використовуватися такі показники як вік, повнота, клас бонітету (висота) та інші аргументи. При вирішенні описаних завдань, коли вид залежності y = f(x) відомий, застосовують різні методи знаходження таких параметрів рівнянь, тобто коефіцієнтів a, Ь, c, ... Найбільш загальний підхід розроблено математиком Чебишевим (1821-1894). Він створив теорію найкращого наближення функції за допомогою багаточленів. Хоча метод Чебишева найбільш підходить для вирішення названих задач, але він складний і використовується рідко, в основному професійними математики. Метод Чебишева простий, але він вимагає рівних інтервалів між дослідними даними, тобто інтервали між:

$$x_1, x_2, x_3, \ldots x_n = k$$

повинні бути однакові (k), що можна забезпечити в досліді дуже рідко. Найчастіше для знаходження коефіцієнтів регресійних рівнянь використовують метод найменших квадратів.

13.3 Метод найменших квадратів.

Метод найменших квадратів застосовується для вирішення різних задач, пов'язаних із обробкою результатів досліду. Найбільш важливим практичним застосуванням цього методу є вирішення задачі згладжування експериментальної залежності, тобто зображення дослідженої функціональної залежності, аналітичної формулі. Але метод найменших квадратів не вирішує питання про вибір загального виду аналітичної функції, а дає можливість при заданому типі аналітичної функції y = f(x) підібрати найбільш ймовірні значення для параметрів цієї функції. Сутність методу найменших квадратів при вирішенні поставленого завдання полягає в наступному.

Нехай отримано n експериментальних точок з абсцисами:

$$x_1, \ x_2, \ \ldots, \ x_n$$

Та відповідними їм ординатами:

$$y_1, \ y_2, \ \ldots, \ y_n$$

Залежність y від x відображується відповідною функцією:

$$y = f(x)$$

яка зазвичай повністю не збігається з експериментальними значеннями y_i у всіх n точках. Це означає, що для всіх чи деяких точок різниця:

$$\Delta_i = y_i - f(x_i)$$

буде відмінна від нуля. Потрібно підібрати параметри функції таким чином, щоб сума квадратів різниць була меншою, тобто потрібно звернути в мінімум вираз:

$$z = \sum_{i=1}^{n} \Delta_i^2 = \sum_{i=1}^{n} [y_i - f(x_i)]^2$$

Таким чином, при методі найменших квадратів наближення аналітичної функції y = f(x) до експериментальної залежності рахується найкращим, якщо виконується умова мінімуму суми квадратів відхилень визначеної аналітичної функції від експериментальної залежності. Слід зазначити, що вказаний вираз є поліном другого ступеня щодо невідомих параметрів. Ці невідомі параметри залежність y = f(x) входять лінійно, і вираз не може набувати негативних значень. Тому існують такі значення невідомих параметрів, за яких функція досягає мінімуму, і цей мінімум в залежності від значень x_i та y_i, буде позитивним чи рівним нулю. При вирішенні багатьох практичних завдань функціональну залежність y від x шукають у вигляді:

$$y = \sum_{k=1}^{m} a_k f_k(x)$$

де: $f_1(x)$, :$f_2(x)$,...., $f_m(x)$ - відомі функції,
a_1 a_2, ..., a_n – невідомі параметри.

Так, наприклад, при дослідженні коливальних процесів функціями f_k (x) (k = 1, 2, ..., m) є тригонометричні функції.

$$f_k(x) = \cos kx, \qquad f_k(x) = \sin kx.$$

При дослідженні у багатьох галузях техніки, а також у лісовому господарстві часто зустрічаються ступеневі функції:

$$f_k(x) = x^{k-1} \quad (k = 1, 2, ..., m)$$

Є й інші види функцій. Зазвичай гіпотезу про вид необхідної функції приймають, аналізуючи за експериментальними даними їх графічне зображення та порівнюючи його з графіками різних

функцій, які наводяться у спеціальних альбомах. Таким чином, f(x) є відомими елементарними функціями аргументу x.

Виходячи з принципу найменших квадратів, ми маємо підібрати такі значення невідомих параметрів a_1, a_2, ..., a_n , при яких зводиться до мінімуму вираз:

$$z = \sum_{i=1}^{n}\left[y_i - \sum_{k=1}^{m} a_k f_k(x_i) \right]^2$$

Вище наведений вираз є функцією невідомих параметрів $a_к$, тому для визначення мінімуму цієї функції, потрібно згідно з правилами диференціального обчислення, знайти похідні функції z за всіма параметрами $a_к$ (к = 1, 2, ..., m), і прирівняти їх до 0:

$$\left.\begin{aligned}
&\frac{dz}{da_1} = 2\sum_{i=1}^{n}\left[y_i - \sum_{k=1}^{m} a_k f_k(x_i) \right]\cdot\left[- f_1(x_i)\right] = 0,\\
&\frac{dz}{da_2} = 2\sum_{i=1}^{n}\left[y_i - \sum_{k=1}^{m} a_k f_k(x_i) \right]\cdot\left[- f_2(x_2)\right] = 0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&\frac{dz}{da_m} = 2\sum_{i=1}^{n}\left[y_i - \sum_{k=1}^{m} a_k f_k(x_i) \right]\cdot\left[- f_m(x_i)\right] = 0.
\end{aligned}\right\}$$

або перетворивши рівняння вище наведене рівняння, отримаємо систему рівнянь:

$$\left.\begin{aligned}
&a\cdot\sum_{i=1}^{n} x_i^4 + b\cdot\sum_{i=1}^{n} x_i^3 + c\cdot\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i,\\
&a\cdot\sum_{i=1}^{n} x_i^3 + b\cdot\sum_{i=1}^{n} x_i^2 + c\cdot\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i,\\
&a\cdot\sum_{i=1}^{n} x_i^2 + b\cdot\sum_{i=1}^{n} x_i + c\cdot n = \sum_{i=1}^{n} y_i.
\end{aligned}\right\}$$

Наведене вище система, є системою трьох лінійних рівнянь щодо невідомих параметрів a, b та c. Вирішуючи цю систему за допомогою визначників третього порядку чи послідовним відкиданням невідомих, ми отримаємо значення параметрів a, b c за методом найменших квадратів. Тут наведено метод найменших квадратів в чіткій математичній формі, що для здобувачів вищої оствіти освітніх ступенів «бакалавр», «магістр» цілком зрозуміло.

Але бажано показати і наочнішу форму обчислення коефіцієнтів рівнянь за допомогою найменших квадратів. Для цього наведемо наступний приклад, де для простоти обчислимо коефіцієнти рівняння прямої y = a + bx. Візьмемо приклад який є залежністю довжини коренів сіянця сосни звичайної від висоти стебла. Виміряні величини зведемо у таблиці. 13.2.

Таблиця 13.2

Залежність довжини коріння від висоти стебла сіянців сосни звичайної

Покзники	Величини, см									
Довжина коріння, (у)	4	4	5	5	5	6	6	6	7	7
Довжина стебла, (х)	3,0	3,1	3,5	3,5	4,1	3,5	4,0	5,0	5,0	5,3

Графічно це зображення на рис. 13.2.

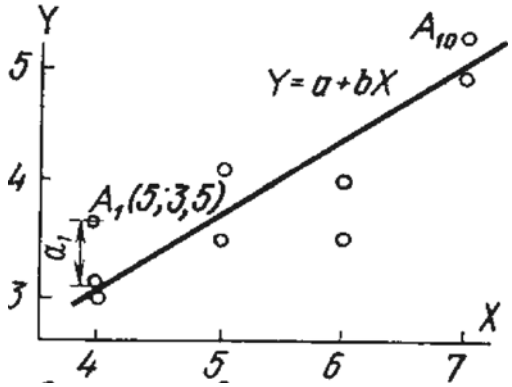


Рис. 13.2 Регресія довжини коренів на висоту стебла сходів сосни звичайної

Тут десять точок A_1, A_2, ..., $A1_0$, зображених на рис. 13.2 мають відповідно абсциси X_1 X_2, ..., X_{10}, та ординати $У_1$ $У_2$, ..., $У_{10}$.

$$\left.\begin{aligned} aN + b\sum x &= \sum y \\ a\sum x + b\sum x^2 &= \sum xy \end{aligned}\right\}$$

Для знаходження коефіцієнтів a та b необхідно мати конкретні значення N, ΣX, ΣX^2, Σy, Σxy. При обчисленні показників кореляції між висотою (довжиною) стебел x і довжиною коріння y, зазначені сумі отримаємо з таблиці 13.3, побудованої на основі таблиці 13.2.

Таблиця 13.3

Вихідні дані для обчислення показників кореляції довжини коріння та висоти сіянців сосни звичайної

№ п/п	1	2	3	4	5	6	7	8	9	10	Σ
Довжина коріння (x)	4	4	5	5	5	6	6	6	7	7	55
Висота сіянців (y)	3,0	3,1	3,5	3,5	4,1	3,5	4	5	5	5,3	40
x^2	16	16	25	25	25	36	36	36	49	49	313
y^2	9	9,61	12,25	12,25	16,81	12,25	16	25	25	28,09	227

Тут маємо: N=10; Σx=55; Σx^2=313: Σy=40; Σxy=227.

Отже, нормальні рівняння у конкретному вигляді будуть:

$$10a+55b=40$$

$$55a+313b=227$$

Поділивши всі члени на коефіцієнти при Ь, рівні 55 і 313, отримаємо:

$$0,1757a+b = 0,7252$$

$$0,1818a+b = 0,7273$$

Віднімаючи рівняння маємо: 0,0061a = 0,0021, звідки a = 0,344. Підставляючи a в рівняння, отримаємо b= 0,665.

Рівняння регресії буде таким:

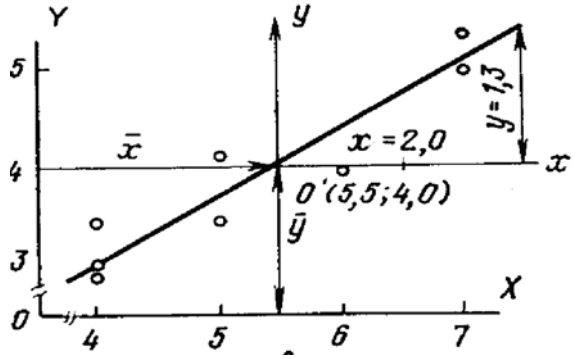


Рис. 13.3 Регресія довжини коріння на довжину стебел сходів сосни звичайної

Завдяки зазначеній замінній, маємо:

$$\sum x = \sum(x+\bar{x}) = \sum x + \sum \bar{x} = \sum x + N\bar{x}$$

Але так як із формули середньої величини

$$\bar{x} = \sum x_i / N, \quad N\bar{x} = \sum x_i$$

то зрозуміло, що в рівності Σx=0. Аналогічно й Σy = 0, якщо подібні дії провести для змінної y:

$$\sum x^2=\sum(x+\bar{x})^2=\sum(x^2+2x\bar{x}+\bar{x}^2)=\sum x^2+2x\sum\bar{x}+N\bar{x}^2=\sum x^2+N\bar{x}^2$$

$$\sum xy=\sum(x+\bar{x})(y+\bar{y})=\sum(xy+\bar{x}y+\bar{y}x+\bar{x}\bar{y})=$$
$$=\sum xy+\bar{x}\sum y+\bar{y}\sum x+N\bar{x}\bar{y})=\sum xy+N\bar{x}\bar{y}$$

Підставивши вирази в рівняння отримаємо:

$$\left.\begin{aligned} aN+b(\sum x+N\bar{x})&=N\bar{y}\\ aN\bar{x}+b(\sum x^2+N\bar{x}^2)&=(\sum xy+N\bar{x}\bar{y})\end{aligned}\right\}$$

Помноживши на x і віднімаючи 227, отримаємо однонормальне рівняння:

$$b\sum x^2=\sum xy$$

з якого випливає:

$$b=(\sum xy)/\sum x^2$$

Величина b називається коефіцієнтом регресії. Вона показує, на скільки одиниць застосованого заходу змінюється y при зміні x на одиницю її міри.

Зважаючи на загальне рівняння лінійної регресії y=a+bx, маємо в окремому випадку при X = $\bar{x}$:

$$\bar{y}=a+b\bar{x}$$

З цього рівняння отримаємо вираз для a = $\bar{y}-b\bar{x}$

Цей спосіб знаходження коефіцієнтів рівняння називається способом визначень. Алгоритм його наступний.

Нехай маємо вихідне рівняння y=y+bx.

Нормальне рівняння тут наступне:

$$b\sum x^2=\sum xy,\quad b=(\sum xy)/\sum x^2.$$

При рівнянні y=a+bx, де: x - варіанти в початкових одиницях вимірювання, нормальні рівняння буде мати вигляд:

$$\left.\begin{aligned} aN+b\sum x&=\sum y\\ a\sum x+b\sum x^2&=\sum xy\end{aligned}\right\}$$

Визначник:

$$D = N \sum x^2 - (\sum x)^2$$

$$a = (\sum y \sum x^2 - \sum x \sum xy) / D,$$

$$b = (N \sum xy - \sum x \sum y) / D,$$

Застосуємо метод визначників для знаходження коефіцієнтів a, b у регресії довжини стебел сосни звичайної на довжину коренів. Вихідні дані поміщені у таблиці 13.1. Вони такі: $N=10;\ \bar{x}=5,5;\ \bar{y}=4,0;\ \sum x^2=10,50;\ \sum y^2=6,26;\ \sum xy=7,0.$

Вихідне рівняння y = y+bx, нормальне рівняння буде $b \sum x^2 = \sum xy$ звідки $b=(\sum xy) / (\sum x^2) = 7,0 / 10,50 = 0,667$ $\hat{y}= 4,0+0,667\ x.$ Зауважимо, що як змінна тут беруть участь відхилення варіант x від середньої $\bar{x}$. Якщо потрібно знайти вираз регресії з варіантами x і y, слід поставити у вихідне рівняння замість x його значення $(x-\bar{x})$.

Отримаємо: $\hat{y}=\bar{y}+b(x - \bar{x})$

13.4 Теоретичні основи використання покрокової регресії (STEPWISE REGRESSION) в лісовому господарстві.

Покрокова регресія (stepwise regression) - один із ключових методів у багатофакторному статистичному аналізі, що широко застосовується в лісовому господарстві для моделювання, прогнозування та інтерпретації взаємозв'язків між різноманітними екологічними та інвентаризаційними показниками. Вона базується на принципі автоматизованого добору змінних для побудови оптимальної регресійної моделі. Це особливо важливо в лісових дослідженнях, де кількість потенційних незалежних змінних може бути значною — вікові характеристики насаджень, типи ґрунтів, кліматичні чинники, індекси продуктивності, просторові координати та інші екологічні параметри. Метою цього методу є створення простої, але ефективної моделі, яка пояснює змінну відгуку (наприклад, запас деревини або біомаса) з найменшою кількістю незалежних змінних при збереженні високої точності прогнозу.

Покрокова регресія поєднує два підходи: forward selection (пряме включення) і backward elimination (зворотне виключення). У першому випадку аналіз починається з порожньої моделі, що включає лише константу, після чого поступово додаються змінні, які найбільше покращують модель за обраними критеріями — коефіцієнтом детермінації (R^2), інформаційним критерієм Акаіке (AIC), Байєсовим

інформаційним критерієм (BIC) або p-значенням у t-тесті. У другому випадку, навпаки, аналіз починається з повної моделі, яка включає всі можливі змінні, і далі відбувається послідовне вилучення тих змінних, які є статистично несуттєвими. Комбінований підхід, який поєднує обидва методи, дозволяє як додавати, так і видаляти змінні на кожному кроці, забезпечуючи гнучке налаштування моделі.

Математично модель описується рівнянням множинної лінійної регресії:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

де: Y – залежна змінна (у даному прикладі – запас деревини, $м^3$/га), X_i - незалежні змінні (вік насадження, діаметр стовбурів дерев, тип грунту, індекс вологості), β_0, β_1, β_2, β_p – коефіцієнт регресії, ε – випадкова похибка.

Кожне включення чи виключення змінної супроводжується статистичною перевіркою її значущості через t-тест для окремих коефіцієнтів або F-тест для оцінки зміни залишкової дисперсії між моделями. Формула для F-критерію має вигляд:

$$F = \frac{(RSS_0 - RSS_1)/(p_1 - p_0)}{RSS_1/(n - p_1 - 1)},$$

де: RSS_0 та RSS_1 – залишкова сума квадратів до й після введення нової змінної, p_0, p_1 – кількість параметрів у відповідних моделях, n – кількість спостережень.

Особливо ефективною покрокова регресія є тоді, коли дослідник не має попередніх припущень щодо важливості предикторів (x_n). Наприклад, при оцінці запасів деревини у змішаних насадженнях Полісся України можна мати десятки факторів: географічні координати, клас бонітету, глибина ґрунту, середня висота, тип дренажу, склад насаджень, вологість повітря, рівень освітлення тощо.

У такому випадку покрокова регресія дозволяє автоматично виявити ті змінні, які мають найбільший вплив на цільовий показник, і побудувати модель, яка легко інтерпретується та узагальнюється. Наприклад, модель може показати, що лише три з десяти змінних суттєво впливають на запас: середній вік, діаметр і тип ґрунту, тоді як інші (наприклад, глибина підґрунтя або річна сума опадів) - статистично незначущі.

Критерії добору змінних у покроковій регресії визначаються заздалегідь. Найчастіше використовують p-value (при $p < 0.05$ змінна вважається значущою), скоригований R^2 (adjusted R^2), що враховує кількість предикторів, або інформаційні критерії. Adjusted R^2 обчислюється як:

$$R^2_{adj} = 1 - \left(\frac{(1-R^2)(n-1)}{n-p-1} \right),$$

де: n – кількість спостережень, p – кількість незалежних змінних. Цей показник особливо важливий під час проведення лісоінвентраизаційних робіт, де спостережень може бути не багато, а кількість потенційних зміних досить велика.

Крім цього, в практиці лісового господарства застосовують AIC:

$$AIC = 2k - 2\ln(L)$$

де: k – кількість параметрів, L – об'єктивність моделі.

Менше значення AIC вказує на кращу модель.

У лісогосподарській практиці покрокова регресія застосовується не лише для прогнозу біомаси чи запасу деревини, а й для моделювання приросту, визначення зони ризику висихання лісів, оцінки ефективності меліоративних заходів, створення регіональних нормативів рубок та оцінки просторової структури насаджень. Наприклад, за даними спостережень у соснових насадженнях Лісостепу дослідник може встановити, що приріст найтісніше пов'язаний із класом бонітету, глибиною залягання ґрунтових вод та інтенсивністю освітлення. Покрокова регресія дозволяє виділити ці змінні з-поміж багатьох інших і створити обґрунтовану аналітичну модель для прийняття лісівничих рішень.

З технічної точки зору покрокова регресія реалізується в більшості статистичних пакетів. У середовищі R використовується функція STEPAIC із пакета MASS. У Python можна скористатися модулем mlxtend, а також реалізувати алгоритм вручну через порівняння моделей у statsmodels. У SPSS, SAS, JMP існують графічні інтерфейси, які дозволяють обрати напрямок аналізу (forward, backward, both) і вказати критерії зупинки.

Приклад типового завдання в лісовому господарстві

Типове завдання: Побудувати модель для прогнозу запасу деревини в соснових деревостанах зони Центрального Полісся України.

Залежна змінна: Y - запас деревини, (м³/га)

Незалежні змінні:

X_1 - середній вік насадження, (роки)

X_2 - середній діаметр дерев, (см)

X_3 - тип ґрунту, (категоріальна змінна)

X_4 - висота над рівнем моря, (м)

X_5 - кількість опадів, (мм/рік)

X_6 - щільність насадження, (дерев/га)

Покрокова регресія є ефективним інструментом математичного моделювання в лісовому господарстві, який дозволяє ***об'єктивно відібрати найбільш впливові фактори*** з великого масиву потенційних змінних.

Зауважимо, що до складу регресійного рівняння під час обробки регресійної множинної залежності не рекомендується до її складу включати більше 3-4 предикторів (Xn). Тому кількість обрахувань коефіцієнта множинної регресії може становити значну кількість, яка залежить від кількості предикторів.

Завдяки автоматизованому алгоритму включення та виключення предикторів, метод забезпечує побудову статистично обґрунтованої, простої за структурою, але інформативної моделі. Це особливо важливо для аналізу складних природних систем, де чинники мають різну природу ранжування покрокової регресії дослідник отримує не лише регресійне рівняння, а й глибше розуміння структури впливів, що визначають цільову змінну (наприклад, запас деревини чи біомасу). Метод дозволяє уникати надмірного ускладнення моделей, зменшує ризик мультиколінеарності та сприяє більш точному прогнозуванню, що робить його важливим інструментом у прийнятті управлінських рішень у лісовому секторі.

Попри свою зручність, покрокова регресія має обмеження. Вона не гарантує виявлення глобально найкращої моделі, може включати змінні через випадковий шум або виключати ті, що мають складну нелінійну взаємодію з іншими предикторами. Також є ризик мультиколінеарності, коли змінні сильно корелюють між собою.
Проте, при правильному налаштуванні і в комбінації з іншими аналітичними підходами (наприклад, аналізом залишків, перевіркою допущень, або використанням компонентного аналізу) покрокова регресія лишається цінним інструментом для формалізованої побудови моделей у лісовому менеджменті, охороні природи, плануванні лісокористування та оцінці впливу кліматичних змін на лісові екосистеми.

13.5 Обчислення значень залежної ознаки на основі рівнянь регресій в лісовому господарстві.

Рівняння регресії дає можливість знайти значення у, які називають обчисленими чи вирівняними (іноді - найбільш ймовірними значеннями).

Наприклад з сіянцями сосни, застосовуючи рівняння y=0,332+0,667x, для x: 4, 5, 6, 7 см, отримаємо у відповідно рівні 3,0; 3,7; 4,3 та 5,0 см; або найточніше: 3,000; 3,667; 4,334; 5,001 см. Обчислені значення слід розуміти як вирівняні (середні) за допомогою регресії величини у, які найбільш близькі до істинних значень цієї величини за даних х, якби справжні значення були знайдені за великою кількістю спостережень. На цій підставі обґрунтовано вважати у найімовірнішими значеннями величини. Найчастіше величини, які обчислені за рівняннями регресії, є загальною закономірністю для деякої сукупності. Наприклад, розробляючи таблиці ходу росту для основних лісоутворюючих

порід України, були виведені різні рівняння, що описують зв'язку таксаційних показників:

$$H=f(A);\ D=f(A);\ G=f(H);\ F=f(H)$$

Так, для березових деревостанів залежність величини суми площ перерізів дерев на висоті 1,3 м (Σq) від середньої висоти (H) описана рівнянням:

$$\sum q = 5{,}151 + 1{,}3208 * H - 0{,}00842 * H^2 - 0.0000979 * H^3$$

де: H - висота деревостану в межах від 5 до 32м. Для обчислення середніх видових чисел деревостану (F) використовують рівняння:

$$F = 0.398 + \frac{0.9845}{H}$$

Розвязавши це рівняння, тобто підставляючи послідовно значення H, рівні 5, 6, 7, ..., 32м, отримали такі величини Σq (таблиця 13.4).

Таблиця 13.4

Величини Σq з і F для березових деревостанів, пробної площі №2 Перганського ПНДВ, обчисленні за рівняннями регресії

H, м	ΣG, м^2	F
8	15Д	0,521
10	17,4	0,497
12	19,6	0,481
14	21,7	1,469
16	23,7	0,460
18	25,6	0,453
20	27,4	0,448
22	29,1	0,443
24	30,6	0,440
26	32,1	0,437
28	33,4	0,434
30	34,6	0,432

Наведені величини є середніми для сукупності всіх березових деревостоїв не лише Перганського ПНДВ, а й всього Поліського природного заповідника. Для окремих насаджень, відхилення можуть досягати 4 - 5%, але вже для сукупності в 5 - 10 виділів (ділянок однорідного березового деревостану) відхилення не виходять за межі 1 - 1,5%. Тому при таксації такого лісового фонду, де однорідних виділів зазвичай не менше, загальна помилка визначення за відсутності систематичних відхилень становитиме незначну величину 0,1 - 0,3%, а видового числа ще менше. Подібним чином в лісовому господарстві використовується більшість даних, отриманих за рівняннями регресії.

РОЗДІЛ 14. ОЦІНКА РЕГРЕСІЙНИХ РІВНЯНЬ

14.1 Помилки регресійних рівнянь.
14.2 Оцінки коефіцієнтів рівнянь регресії.
14.3 Залишкова дисперсія та її аналіз.
14.4 Взаємокорелюючі аргументи. Вибір аргументів у рівнянні регресії при їх взаємній кореляції у лісовому господарстві.

14.1 Помилки регресійних рівнянь.

Встановити наявність зв'язку між залежною та незалежною (незалежними) змінними звісно важливо, але недостатньо. В наш час дослідник озброєний сучасним комп'ютером, що має потужне програмне забезпечення (наприклад, систему «MathCAD»), яку ми й використовували при проведенні математично-статистичної обробки отриманих результатів з пробних площ дендрохронологічних досліджень, за кілька хвилин отримуючи коефіцієнти заданого рівняння. Сьогодні для дослідження на перший план виходить завдання оцінки та інтерпретації отриманих рівнянь. Викликано це тим, що всі регресійні рівняння дають деяке наближення до того закону чи закономірності, що існує в природі. Рівняння регресії характеризує цю закономірність із деякими помилками. Вони викликані в основному недоліками в експериментальному матеріал, а іноді і невисокою якістю моделі. Може виявитися, що вибірка не охоплює всі особливості досліджуваного явища, бувають помилки вимірів, на величину вихідних даних впливають випадкові причини і т. д. Але бувають помилки й іншого роду. Ми не завжди знаємо причинно-наслідковий зв'язок явищ, що вивчаються. Іноді неправильно підбираємо форму зв'язку, що виражається конкретним рівнянням. Наприклад, німецький вчений проф. Продан ще в 60-ті роки минулого століття зазначив, що рівняння параболи другого порядки $y = a+bx+cx^2$, яке часто використовували для опису ходу росте деревостанів по висоті, є занадто «жорстким», воно занижує значення висот в молодому віці та завищує їх для стиглих та перестиглих деревостанів. Тому дуже важливо встановити, наскільки обчислена лінія регресії достовірна, тобто знайти основну помилку регресійного рівняння.

14.2 Оцінки коефіцієнтів рівнянь регресії.

Подібно до коефіцієнта кореляції та ряду інших статистичних показників, коефіцієнт регресії завжди визначається на основі вибіркової сукупності. Отже, він є вибірковим показником, саме конкретне рівняння регресії також можна назвати вибірковим. Рівняння справжньої лінії регресії виглядатиме наступним чином:
$y = a + bx$.

Параметри a і b вибіркового рівняння регресії служать для оцінки справжніх значень a і b, тобто їх значень в генеральній

сукупності. Нульовою гіпотезою відсутность зв'язку є те, що коефіцієнт регресії (b) не відрізняється від нуля. Для того, щоб мати право відкинути нульову гіпотезу, необхідно встановити достатню достовірність цього коефіцієнта σ_b по відношенню σ_b до t: критерія Стьюдента, що може бути зроблено шляхом зіставлення b з його помилкою:

$$\sigma_{b/t} = b_p / \sigma_b)$$

Про достовірність b можна судити за величиною помилки σ_b, а також оцінити ступінь близькості b до β. Оскільки у визначенні лінії регресії беруть участь два параметри a і b, слід окремо розглянути, як можуть вони варіювати у вибіркових сукупностях, що взяті з однієї й тієї ж генеральної сукупності. Теоретична лінія регресії зазвичай розташована під великим або меншим кутом по відношенню до осі абсцис. Цей кут визначається величиною (коефіцієнтом) b. У геометричному сенсі b є тангенс кута між лінією регресії та віссю абсцис (або ординат – якщо розглядати другу лінію регресії) за відсутності регресії b=0. Тоді лінія регресії y по x повинна йти горизонтально по відношенню до осі абсцис, а лінія регресії x по y - вертикально. Місце їх перетину відповідає середнім значенням обох ознак. Таким чином, кожна лінія регресії обов'язково пройде через точку (рис. 14.1), координати якої x та y.

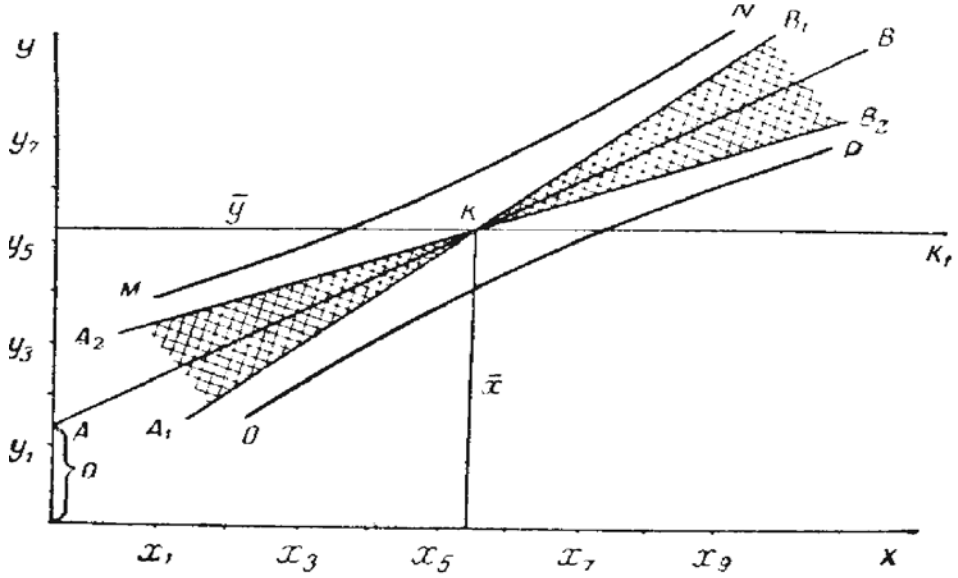


Рис. 14.1 Довіряючі межі для вибіркової лінії регресії AB.

Це важливо для розуміння можливого коливання лінії регресії. Так як b має помилку σ_b, то, очевидно значення вибіркових b можуть перебувати у межах, визначених цією помилкою. Це означає, що кут нахилу лінії регресії може бути або більшим, або меншим.

Як показано рис. 14.1, лінія регресії AB перетинає точку K і має кут нахилу по відношенню до горизонталі. Лінія регресії вміщена

всередині пари кутів, утворених перетином ліній A_1B_1 та A_2B_2. Якщо кути B_1K та B_2K побудовані по верхній і нижній межі b з врахуванням лише однієї помилки, то ймовірність знаходження справжньої лінії регресії в цих межах дорівнює 0,68. Однак рівняння регресії має ще вільний член a. Він визначає величину відрізка, що відсікається на осі у лінії регресії AB. Величина a і має свої межі коливання, тому лінія регресії в тому ж значенні b, тобто при тому ж куті нахилу її до осі абсцис, може проходити або трохи нижче лінії AB, або вище. Так як треба враховувати обидва параметри рівняння регресії, то встановлення довіряючих меж для лінії регресії не так просто. Загалом можна вважати, що межі довіряючого інтервалу являють собою криві лінії типу гіпербол (лінії MN та OP на рис. 14.1). Це означає, що в міру віддалення від середньої точки (x, y) вони розширюються. Крайні точки, якими будується лінія регресії, мають більшу помилку. Однак при проведенні спеціальних досліджень можна досягти достатньо великих n на всіх частинах інтервалу змін x (або відповідно y) та прийняти, що дисперсія окремих значень y (або x) буде приблизно однаковою всіх частинах інтервалу. У такому разі можна застосовувати порівняно прості методи для оцінки достовірності коефіцієнта та лінії регресії. Основою визначення можливої варіації лінії регресії є сума квадратів відхилень фактичних значень y_i від обчислених теоретично $y_{i\ теор}$, за тими ж значеннями ряду x_i. Формула визначення основної помилки регресії y по x, тобто y-x (якщо регресія x по y, то записується x-y) наступна:

$$\sigma^2_{y\cdot x} = \frac{\Sigma(y_i - \hat{y}_i)^2}{N-2}$$

$$\sigma_{y\cdot x} = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{N-2}}$$

Тут N-2 – число ступенів свободи. В даному випадку число ступенів свободи визначається як N-2 тому, що при обрахунку відхилень, використовуються дві величини y_i та $У_{i\ теоретичне}$ а не одна, як це робилося при аналізі варіаційного ряду, коли число ступенів свободи визначалося як N-1. Пояснимо викладене прикладом обчислення основної помилки регресійного рівняння прямої, що описує залежність зміни висоти від діаметра у молодих соснових деревостанах. Згідно даних щодо висоти та діаметру стовбурів сосни звичайної в соснових деревостанах, які розглядались вище в таблиці 12.1, отримуємо значення регресії:

$$\sigma^2_{y\cdot x} = \frac{0{,}89\text{м}}{9} = 0{,}099\text{м}$$

$$\sigma_{y\cdot x} = \sqrt{0{,}099}\text{м} = 0{,}314\text{м}$$

Тут величина σ_{yx} має таке ж значення, як і σ в варіаційному ряді. У межах однієї σ_{yx} відхилення розподіляються вгору і вниз від лінії регресії у 68% випадків. У 95% вони лежать в межах 2 σ_{yx} а 99,7% випадків відхилення від теоретичної лінії регресії становлять величину 3 σ_{yx}, тобто в нашому прикладі це 0,942 м.

Межі вимірювань наведено в таблиці 14.1.

Таблиця 14.1

Обчислення статистик розподілу та зв'язку за вихідними даними таблиці 12.1

№ п/п	x_i	y_i	$K_1=x_i-\bar{x}$	$K_2=y_i-\bar{y}$	K_1^2	K_2^2	K_1K_2
1	3,9	4,8	-2,97	2,55	8,82	6,50	7,57
2	5,0	6,0	-1,87	1,35	3.50	1,82	2,52
3	5,5	6,4	-1,37	0,95	1,88	0,90	1,30
4	6.0	6,1	-0,87	1,25	0,76	1,56	1,09
5	6,5	7Д	-0,37	0,25	0,14	0,06	0,09
6	6,8	7,7	-0,07	0,35	0,05	0,12	0,02
7	7,4	7,6	0,53	0,25	0,28	0,06	0,13
8	7,7	8,2	0,83	0,85	0,69	0,72	0,71
9	8,2	8,2	1,33	0,85	1,77	0,72	1,13
10	8,9	9,1	2,03	1,75	4,12	3,06	3,55
11	9,7	9,7	2,83	2,35	8,01	5,52	6,65
Σ	75,1	80,9	0,03	0,05	30,02	21,04	24,76
Середнє	6,87	7,35	-	-	-	-	-

$$\sigma_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = 1{,}65$$

$$\sigma_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n}} = 1{,}38$$

$$r = \frac{\sum K_1K_2}{\sqrt{K_1^2K_2^2}} = \frac{24{,}76}{\sqrt{30{,}02\cdot 21{,}04}} = \frac{24{,}76}{25{,}13} = 0{,}985$$

$$R_p = 0{,}985 \cdot \frac{1{,}38}{1{,}65} = 0{,}824$$

Тоді:

Тепер знайдемо помилку коефіцієнта регресії:

$$\sigma_R = \frac{\sigma_{y \cdot x}}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{0{,}099}{\sqrt{30{,}02}} = \frac{0{,}099}{5{,}48} = 0{,}02$$

Найбільша величина помилки, говорить про достовірність лінії регресії. Це виражається в формулі:

$$t = \frac{R_p}{\sigma_R} = \frac{0{,}824}{0{,}02} = 41{,}2$$

За спеціальними таблицями t-критеріz Стьюдента (додаток Е) знаходимо, що критична величина t: при числі ступенів свободи 9 дорівнює 5,04 за p=0,001, тобто обчислене значення t: перевищує p при 0,001. Тому нульова гіпотеза про те, що коефіцієнт регресії в генеральній сукупності (β) дорівнює 0, має бути відкинута. Обчислене рівняння є достовірним.

14.3 Залишкова дисперсія та її аналіз.

Для коректного вирішення питання про адекватність прийнятої моделі, що описує деяку закономірність недостатньо знати її основну помилку та визначити значущість коефіцієнтів рівняння. Дуже велике значення має аналіз залишкової дисперсії. Щоправда, найчастіше цей аналіз біологи і лісівники не роблять, так як він досить важкий. Просто передбачається, що залишкові величини які виходять за межі рівняння регресії, тобто ті значення $_{xi}$, які викликані випадковими причинами (їх зазвичай називають просто залишками) розподілені нормально і не впливають на результат.

Для названого аналізу розглянемо три величини, саме:

$\sigma^2_y, \sigma^2_{yx}, \sigma^2_0$ де $\sigma^2_y = \sum (y_i - \bar{y})$, тут x – загальне середнє змінної y, σ_y^2 – сума квадратів відхилень y_i від середньої.

Суми квадратів відхилень знаходяться у співвідноенні:

$$\sigma^2_y = \sigma^2_{y \cdot x} + \sigma^2_0$$

Цей вираз представляє розкладання загальної суми квадратів відхилень на дві складові: обумовлену регресією σ^2_{yx} і залишкову, що пояснюється іншими причинами, ніж залежність y від x. Очевидно, що чим наша модель краще описує досліджувану залежність, тим меншою має бути залишкова сума квадратів σ_0^2 і тим ближче до одиниці коефіцієнт детермінації:

$$r^2 = \sigma^2_{y \cdot x} / \sigma^2_y$$

При цьому в основу оцінки рівняння регресії може бути покладений аналіз дисперсій з урахуванням того, що сумі σ_y^2 відповідає N -1 ступінь свободи, σ_{yx}^2 - має один ступінь свободи, оскільки є єдиною функцією від y_i та σ_0^2 має N -2 ступенів свободи, тому що відповідно (N-1) = (N-2) +1.

Сума квадратів відхилень, поділена на число ступенів свободи, що дає відповідну оцінку дисперсії називають середнім квадратом відхилень.

Позначимо оцінку дисперсії, обумовлену регресією, через σ^2_{yx}.

Якщо застосована модель справедлива, то $\sigma^2_{xy.} = \sigma^2$ в іншому випадку $\sigma^2_{xy.} > \sigma^2$ Оцінка залишкової дисперсії $\sigma^2 = \sigma^2_0/(N-2)$. Якщо в генеральної сукупності зв'язок між величинами x і y лінійний, то σ^2_0 є незміщеною оцінкою дисперсії σ^2 помилок ε_i або, що те ж саме дисперсії величини y_i, позначеної через σ_y^2. Якщо зв'язок не лінійний, тобто модель обрана не вірно, то останнє твердження не правильне. Вихідна таблиця для аналізу лінійного рівняння регресії має вигляд (таблиця 14.2).

Таблиця 14.2

Вихідні дані для аналізу дисперсій регресійного рівняння

Джерело	Сума квадратів	Кількість ступенів свободи	Оцінка дисперсії
регресії	σ^2_{xy}	1	-
залишкова	σ^2_0	N-2	$\bar{\sigma}^2_0 = \sigma^2_0/(N-2)$
загальна	σ_y^2	N-1	-

Якщо розрахункове значення F перевищує табличне F_{1-a} [1;(N-2)], то гіпотезу $H_0 = 0$ відхиляють із рівнем значущості a табличну величину F беруть із додатка Ж. Припустивши, що лінійна модель дійсна, можна вивести формули основних помилок коефіцієнтів βo і β_1 та рівняння регресії в цілому, а також побудувати довіряючі інтервали для дійсних β_0 та β_1.

Звідси виходить, що для збільшення надійності вибіркового коефіцієнта регресії, слід брати якомога більший інтервал Xmax – Xmin. Довіряючі інтервали для β_1 мають вигляд:

$$b_1 - t_{1-\alpha/2}\,(N-2)\,m_{b1} \le \beta_1 \le b_1 + t_{1-\alpha/2}\,(N-2)\,m_{b1}$$

де: $t_{1-\alpha/2}\,(N-2)$ - табличні значення t для двостороннього рівня значущості α та при числу ступенів свободи k=N-2.

Для перевірки гіпотези $H_0:\beta_1 = 0^*$ проти альтернативі: $H_0:\beta_1 \neq \beta^*$ досить переконатися, що β^* знаходиться в довіряючому інтервалі, якщо це не так, то альтернативну гіпотезу відкидають як хибну.

В такому випадку, довіряючий інтервал розраховують за формулою:

$$b_0 - t_{1-\alpha/2}\,(N-2)\,m_{b1} \leq \beta_0 \leq b_0 + t_{1-\alpha/2}\,(N-2),$$

а перевірку гіпотези $H_0:\beta_1 = 0^*$ проти $H_0:\beta_1 \neq \beta^*$ проводять аналогічно коефіцієнту регресії b_o.

Раніше нами розглядався коефіцієнт детермінації r^2, що обчислюється (якщо лінійна модель правильна) як відношення суми квадратів, обумовленої регресією, до загальної суми квадратів, і вимірює ту частина мінливості залежної змінної у, яка визначається залежністю у від х. В прикладі $r^2 = 148{,}31/162{,}31 = 0{,}913$, тобто на 91% мінливість залежить від зміни х. Відповідність моделі вихідним даним можна оцінити за графіком, завдавши вихідні точки та обчислене рівняння регресії. Для лінійного рівняння з однією змінною такий шлях цілком прийнятний, але якщо модель складна (багато змінних, складний вид залежності), то для висновку про точність, потрібні об'єктивні статистичні критерії. Зазвичай з цією метою застосовують F-критерій. Слід підкреслити, що у всіх випадках, коли є можливість перевірити точність, ця процедура потрібна. Якщо модель дієва, то залишкова сума квадратів $\sigma^2_{\text{залишкове}}$ пояснюється лише дисперсією помилок ε_1 σ^2; в іншому випадку $\sigma^2_{\text{залишкове}}$ додатково включає суму квадратів, породжену не справедливим розподілом або функіональністю моделі, тобто суму квадратів відстаней між обчисленим та істинним рівнянням регресії.

Залишки можна досліджувати за допомогою спеціальних критеріїв. Однак цілком достатні результати дає графічний аналіз, який використовує загальний графік залишків (для висновків про нормальність ε_i), графіки залежності залишків від x_i, та y_i. Якщо модель справедлива, то точки ε_i повинні розташовуватися в лінії, що лежить паралельно осі абсцис на графіку залежності ε_i від x_i та y_i.

Певна суб'єктивність висновків про дієвість представлених статистичних моделей зрозуміло залишається. В даний час повноцінні ліцензійні програми матзабезпечення для ПК включають аналіз залишків. Тому цю процедуру треба обов'язково застосовувати під час проведення регресійного аналізу результатів наукових досліджень.

14.4 Взаємокорелюючі аргументи. Вибір аргументів у рівнянні регресії при їх взаємній кореляції у лісовому господарстві.

Якщо ми визначаємо функцію від кількох аргументів, то треба знати, як впливає кожен аргумент, його значущість, які ми розібрали вище. Але залишається питання про необхідну кількості аргументів. На перший погляд може здатися, що чим більше аргументів, тим

точніше обчислення функції, оскільки враховується більша кількість факторів, які впливають на результати статистичного аналізу.

Наприклад, якщо ми вивчаємо залежність висоти дерева від лісорослинних умов, то можемо взяти як аргумент механічний склад грунту. Якщо ж ці аргументи доповнимо відсотком гумусу в генетичних горизонтах за результатами грунтових обстежень за допомогою шурфів, то прогноз функції буде точніше. Можемо додати умови зволоження (за едафічною сіткою від 0 до 5), скажімо записати глибину рівня залягання ґрунтових вод, що дасть ще більш точні показники. Але нескінченно додавати аргументи не можна. По-перше, це просто складно реалізується, особливо у стадії експерименту. Головне ж – аргументи можуть бути взаємно корельовані. В нашому випадку, якщо у нас глибина залягання грунових вод на рівні - 1,0 м, то немає сенсу брати відсоток вологи в кореневмісному шарі. Ці два аргументи взаємозалежні. Таке ж явище ми спостерігаємо в раніше наведеному прикладі визначення коефіцієнта форми q_2. Раніше використовували вираз: $q_2 = f(H, D)$, але пізніше (після досліджень Ф. П. Моісеєнко) зрозуміли, що достатньо буде прийняти $q_2 = f(H)$. Справа тут у тому, що аргументи в рівнянні регресії не повинні корелювати один з одним. Так рівень залягання грунтових вод та механічний склад грунту якщо й корелюють, то досить слабко. Тому застосування їх обох цілком виправдано. Але вже зволоженість кореневмісного шару прямо залежить від рівня залягання грунтових вод. Взаємно корелюють діаметр та висота дерева. Застосування множинного регресійного аналізу передбачає наявність некорельованих аргументів. У разі, якщо аргументи регресії корельовані, що часто буває в лісівничих дослідженнях, то коефіцієнти, що стоять при них (a, b, ...) визначаються в залежності один від одного, і теж корельйовані.

Кореляція між коефіцієнтами регресії спотворює оцінку індивідуальної зміни кожного аргумента величини функції. Це означає, що обчислені коефіцієнти a, b, ..., не визначені суворо, а можуть змінюватися взаємно: збільшується a, тоді зменшується b і т.д.

Взаємозв'язок між аргументами знаходиться через внутрішню межу визначеності $d_{bH}=1-(1-r_{ij})$, що лежить у межах від 0 до 1.

Тут і j - коефіцієнт кореляції між аргументами і та j. При збільшенні d_{bH} зростає невизначеність щодо коефіцієнтів регресії. Межа для d_{bH}, де інтерпретація окремих факторів ще можлива, визначають з умов завдання та суті досліджуваного явища. При $d_{bH>}0,5$, коефіцієнти регресії спотворені і не дуже сильно піддаються інтерпретації. Вибір конкретних аргументів у рівнянні регресії визначається з умов завдання. Насамперед робиться логічний аналіз та вибираються ті аргументи, які краще відповідають біологічним та лісівничим особливостям. Серед вибраних аргументів, якщо вони взаємокорельовані, залишають саме ті, де великі коефіцієнти кореляції $x - y_0$. Два аргументи, кореляція між якими більше ніж 0,5 не повинні бути в одному рівнянні. Без шкоди для результату обчислень, один із них опускається.

РОЗДІЛ 15. ВИБІР РІВНЯНЬ РЕГРЕСІЇ, ЇХ ВЕРИФІКАЦІЯ І ПРАКТИЧНЕ ЗАСТОСУВАННЯ У ЛІСОВОМУ ГОСПОДАРСТВІ

15.1 Принципи та основні критерії вибору регресійної моделі.
15.2 Основні види закономірностей у лісівництві, лісовій таксації та інших лісівничих компонентах, що виражаються за допомогою регресійних моделей.
15.3 Верифікація регресійних моделей.
15.4 Застосування регресійних моделей в лісовому господарстві.

15.1 Принципи та основні критерії вибору регресійної моделі.

Попередньо ми детально описали як обчислити коефіцієнти регресійного рівняння, та як оцінити це рівняння. Але для проведення аналізу біологічних та лісівничих явищ цього недостатньо. Нам потрібно вибрати вид зв'язку, визначитися з конкретною його моделлю для того, щоб вона відображала сутність явища, що вивчається та описувала закономірності його зміни. Найкраще модель буде обрана тоді, коли вона не просто згладить експериментальні дані, але й відображатиме суть досліджуваного явища. В цьому випадку, з певними застереженнями, цю модель можна використовувати для екстраполяції дослідних даних. Останнє є найбільш складною та неоднозначною частиною щодо вибору та використання моделей. Якщо модель обрана не зовсім правильно і не відображає всієї суті того, що досліджується, то хоча на деякому відрізку кривої згладжування може бути задовільним, але для екстраполяції така модель не підходить. Наприклад, якщо ми описуємо зростання дерева або деревостану у висоту, то модель, виходячи з основних законів росту деревостану, повинна відповідати наступним умовам:

- крива зростання має починатися на початку координатних осей, так як зростання починається від насіння, тобто від нульового значення висоти;
- в початковий період життя (у різних деревних порід цей період неоднаковий), зростання йде повільно;
- поступово відбувається прискорення росту, досягаючи певного максимуму та приросту за висотою. Останній спостерігається у різному віці залежно від деревиної породи та лісорослинних умов;
- поступове уповільнення приросту за висотою продовжується до старшого віку, який теж буває різним, аналогічно до попередньої умови;
- у найстаршому віці зростання у висоту практично припиняється (по діаметру дерево ще росте), і крона дерев набуває своєрідну ущільнену форму;

• наприкінці життя дерево (деревостан) гине і починається новий цикл.

Виходячи з описаних умов, крива зростання повинна відповідати наступним умовам:

■ починатися на початку координат;

■ мати не менше 2 (іноді 3) точки згину: початок інтенсивного зростання, перехід до уповільнених темпів приросту, припинення приросту.

В практиці зазвичай використовують останню умову, так як при інтенсивному веденні лісового господарства до природної стиглості та розпаду, деревостани можуть дожити тільки в умовах повної заповідності у заповідниках, заказниках або національних парках;

■ абсолютне максимальне значення висоти, яке дає модель, не має перевищувати максимально досяжну висоту, до якої виростають дерева певної деревної породи в певних лісорослинних умовах.

Оскільки для моделювання приросту дерев (древостанів) у висоту часто використовують експериментальний матеріал (пробні площі), де вік деревостану коливається від 40-60 до 70-90 років, то цей відрізок може бути описаний різними кривими, навіть параболою 2 порядку - $y=a_0+_{a1X}+a_{2X}^2$. Але ця крива через її «жорсткість» зовсім непридатна для екстраполяції отриманих наукових даних про прирости деревостанів. Тому, застосувавши криву, що не відповідає суті явища, яке вивчається, ми не можемо зробити висновки про дійсну траєкторію зміни деякої ознаки під впливом різних чинників. Так, у наведеному прикладі не можна сказати, яка буде висота деревостану за межами досліджуваного віку - в 10 чи 100 років. Крива може тут піти далеко в бік, навіть донизу у віці 100 років, просто спотворивши суть явища. Тому в лісовій таксації з цією метою з XIX століття застосовують досить гнучкі моделі. Найбільш відомі та добре себе зарекомендували такі:

Фунція Вебера:

$$H_a = H_{\max}(1 - \frac{1}{1.0p^c})$$

Формула Я. А. Юдицького (НАУ (НУБіП України, м. Київ):

$$H_a = b_1 \Phi\left[b_2 (A - b_3)\right] + b_4$$

де: H_a – висота у віці «а»;

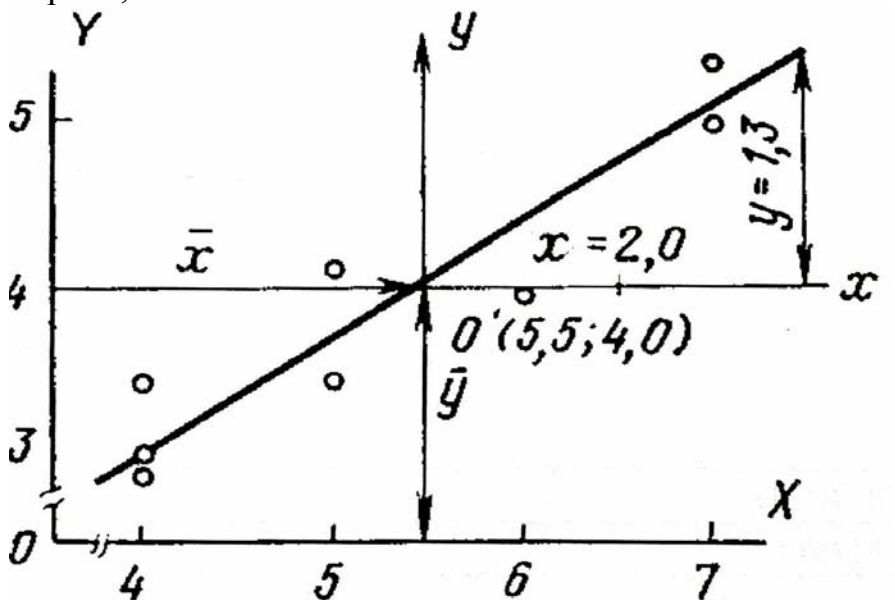
H_{max} - максимальна висота даної породи в конкретних лісорослинних умовах;

Ф – функція Маркова.

Інші позначення характеризують параметри рівнянь. Дещо інша тенденція спостерігається в моделі росту дерев за діаметром. Пояснюється це тим, що у товщину дерева ростуть до їх відмирання, хоча темпи приросту в пристигаючом та стиглому віці зповільнюються. Тут залежною ознакою є діаметр (незалежною – вік) – має позитивний приріст практично протягом всього періоду життя, тоді як приріст у висоту в стиглому деревостані припиняється.

На основі знань біології можна передбачити, що криві, які відображають зростання у висоту, в зв'язку з віком та діаметром $H=f(A, D)$, будуть відрізнятись один від одної. Це пов'язано з тим, що зміна віку характеризується рівними приростами протягом всього життя. Приріст діаметра є неоднаковим. У перші роки життя він невеликий, потім досягає максимуму в молоді роки, після чого прогресивно зменшується до періоду відмирання дерева. Тут ми характеризуємо середню поведінку зростання, не звертаючи уваги на річні чи короткоперіодичні коливання приросту висоти або діаметра, які можливі в зв'язку з впливом різних факторів середовища, наприклад, посухи, але не біологічних законів. З цього далеко не повного переліку моделей видно, що вони поступово удосконалюються, зберігаючи основні вимоги до моделей росту. Існують певні правила, прийоми та критерії, що в сукупності визначають вибір моделей. У більшості випадків, вибір рівняння регресії робляють на основі аналізу, використовуючи професійні знання, як це робили вище при розгляді прикладу зростання дерев за висотою та діаметром. Наразі накопичені знання про конкретні рівняння регресії для опису найважливіших біологічних процесів, подібні вище розглянутим. Проте слід зазначити, що з вивченням багатьох явищ виникають великі труднощі у виборі відповідного рівняння регресії. Навіть встановлення загальної її форми (прямолінійна вона або криволінійна), на основі професійних знань часто не може бути зроблено. Статистичні методи у таких випадках дають основу для прийняття рішень про форму регресії та вибір рівняння. Наприклад, візьмемо раніше розглянуту залежність довжини коренів сіянців сосни звичайної та висоти її сходів. Розміщення точок на графіку часто вказує форму кривої. Якщо точки розташувалися, як на рис. 15.1, то є підстави прийняти зв'язок за лінійний. Теоретичний аналіз сутності явища не дає аргументації, що суперечить цьому висновку. Справді, для сходів сосни звичайної немає обмежувальних факторів до розвитку обох досліджуваних ознак – довжини (висоти) стебла та довжини коренів. Щодо регресії довжини стебел і коріння сходів або молодих рослин можна сказати, що вона лінійна. Якщо точки на графіку розташувалися так, що вказують на згин узагальнюючої їх кривої, існують підстави перевірити гіпотезу про лінійність регресії, тобто розрахувати та оцінити достовірність міри криволінійності.

Довжина кореня, см



Довжина стебла, см

Рис. 15.1 Регресія довжини коріння на довжину стебел сходів сосни звичайної.

Часто при дослідженнях зв'язку між регресійними ознаками, їх аналіз рухається за кореляційним, або здійснюється разом з ним. Тоді певні статистичні висновки про лінійність регресії роблять на основі t-критерію Стьюдента, але краще на основі F-критерію Фішера. Зазначимо, що самі по собі критерії не дають вичерпної відповіді про вибір рівняння, а лише про форму регресії: прямолінійна вона чи криволінійна. Якщо отримано дані, які свідчать, що зв'язок прямолінійний, цього достатньо щоб перейти до наступного кроку регресійного аналізу. Посилання на криволінійний характер регресії зобов'язує вести подальший пошук функції серед багатьох функцій цього виду. У такому випадку слід випробувати найбільш прості та доступні досліднику функції. Зупиняють вибір на функції, що дає краще наближення до дослідних даних, або має найменше середнє квадратичне відхилення обчислених даних (σ_{yx}) або (σ'_{yx}). Тут часто допомагають згадані альбоми, де вміщено графіки різних функцій. В лісовому господарстві у багатьох випадках задовільну апроксимацію дослідних даних отримують на основі парабол 2-ї, 3-ї та більш високих ступенів, оцінюючи точність кожної з регресій. При малому числі груп (класів) залежної змінної, можна отримати параболу з числом коефіцієнтів, рівним числу груп, яка проходить через всі точки, що характеризують групові середні. Проте цінність регресії у цьому випадку знижується. Крива не виражає в такому разі закономірності зв'язку, а відображає випадковості вибіркових спостережень. Застосовувати таку криву можна тільки для апроксимації та знаходження проміжних значень у конкретному випадку, не виходячи за межі вимірювань. Наприклад,

якщо ми виміряли на зрубаному дереві діаметри в 6 точках, то для отримання проміжних діаметрів, що лежать між вимірами, доцільно використовувати рівняння виду цілих поліномів 5-6 ступеня. Але вони описуватимуть твірну криву тільки цього дерева. Криволінійний характер залежності між змінними іноді вдається замінити на прямолінійний, шляхом перетворення х або у шляхом логарифмування, що часто дає суттєві уточнення виразу зв'язку. Логарифмічні параболи внаслідок розтягнутості осей взагалі більш гнучкі.

15.2 Основні види закономірностей у лісівництві, лісовій таксації та інших лісівничих компонентах, що виражаються за допомогою регресійних моделей.

В лісовому господарстві, як видно з вище викладеного, застосовуються різні моделі. Вони вибираються, виходячи з умов проведеного експерименту і з сутності досліджуваного явища. Якщо ми вивчаємо приріст чого-небудь: (дерева, насадження, гілки, тварини і т.д.), то повинні визначитися, яким умовам відповідатиме наша модель. Для опису приросту у висоту та по діаметру, такі підходи викладені вище. Якщо ж ми вивчаємо інші показники приросту дерева або деревостану, то маємо враховувати вже їх особливості. Наприклад, вивчаючи зміну запасу (М) деревостану, ми повинні пам'ятати такі умови:

• крива починається на початку координат;

• крива повинна відображати повільний приріст у наймолодшому віці;

• необхідно, щоб знайшло відображення уповільнення збільшення запасу деревостану після певного віку (у сосни звичайної після 60-70 років, у берези повислої після 40-50 років), що буде видно з наявності точки згину;

• у найстаршому віці йде процес розпаду деревостану, з'являється багато сухостою, і величина запасу зменшується. При цьому він не може бути від’ємним - граничний випадок за повного розпаду М=0.

Таким чином, модель зміни запасу насадження матиме приблизно такий вид, як крива на рис. 15.2.

Якщо ж ми будемо досліджувати приріст запасу деревостану (Z_M), то він повинен відповідати таким умовам:

■ початок кривої починається з точку 0;

■ повільне зростання на початку життєвого циклу;

■ різке збільшення в період інтенсивного приросту (сосна звичайна в 20-50 років(;

■ досягнення максимуму приросту та поступове його зниження;

■ у перестійних деревостанах, коли переважають процеси розпаду, поточний приріст може бути від’ємним за рахунок відмирання дерев;

■ для окремого дерева приріст може дорівнювати 0, але від'ємні величини виключаються.

Графічно зміна приросту запасу деревостану (М) та фактичного деревостану (Z) показано рис. 15.2.

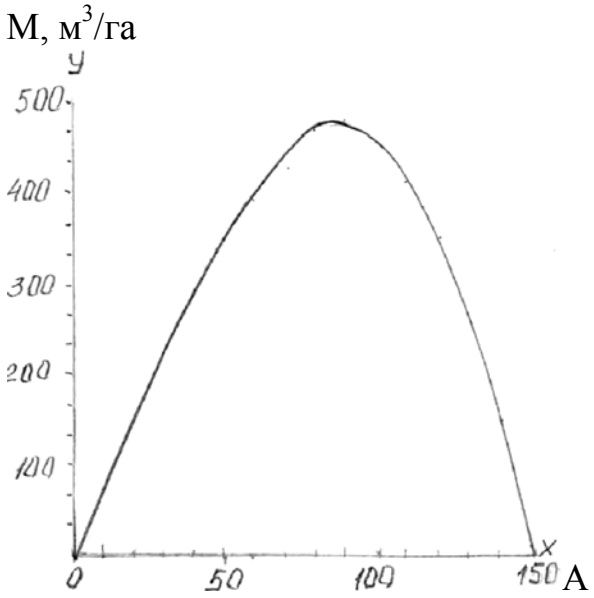


Рис. 15.2 Хід росту по фактичному запасу деревостану сосни звичайної I класу бонітету та його поточна зміна із збільшенням віку в умовах пробної площі №2 Перганського ПНДВ

Дослідження інших закономірностей в лісовому господарстві наводять для використання інших кривих. Вибір форми залежності у всіх випадках йде від простого до складного. Моделі, що відображають причинно-наслідкові взаємозв'язки та взаємодії в системах (або моделі зв'язку) - основний тип моделей, що застосовуються в практиці та дослідженнях з лісового господарства. Нижче ми розглянемо моделі зв'язку двох класів: емпіричні та структурні або функціональні.

В якості математичної форми емпіричних моделей зв'язку здебільшого використовують регресійні рівняння, а рідше – інтерполяційні багаточлени. В першому випадку, для знаходження коефіцієнтів рівнянь застосовують різні модифікації методу найменших квадратів, що дозволяють просто та досить надійно оцінити статистичним шляхом модель, що розробляється. Методом регресійного аналізу отримано практично всі найбільш змістовні біометричні закономірності в лісовому господарстві.

У той же час, метод найменших квадратів має суттєві недоліки суто пізнавального плану: по-перше, він не враховує природної сутності досліджуваного явища і допускає довільність у виборі конкретних типів рівнянь. По-друге, цей метод передбачає детермінований характер досліджуваного процесу. Тому останній час все більше уваги привертають імовірнісні моделі, що використовують методи випадкових функцій.

Основний засіб об'єктивної розробки моделей емпіричних зв'язків - використання методів біометрії, заснованих на законах математичної статистики. Це дозволяє ухвалити обґрунтовані рішення, оцінити структуру, робочий діапазон та надійність моделей. Однак повною мірою, методи статистики можна застосовувати лише до матеріалу, отриманого із дотриманням статистичних передумов. У всіх інших випадках, застосування щодо трудомістких методів статистичного аналізу взаємозв'язків невиправдано. Те саме стосується і матеріалу, з якого довільно виключено значну частину тільки на тій підставі, що вона не входить в деякі уявлення експериментатора, бо прості графічні методи це дозволяють робити. Якщо для розробки моделі зв'язку інформацію зібрана, то планування експерименту експериментатором дозволяє значно покращити результати, так як краще витратити частину часу та коштів для попередньої оцінки ситуації, вибору незалежних змінних та проведення їх аналізу. Головне тут, як і в багатьох інших випадках, застосування математичних методів, - це точне формулювання завдання та кінцевих цілей, виходячи з сутності досліджуваного явища. Тому слід застосовувати такі терміни як:

<u>адекватність моделі</u> - відповідність вихідним даним, що підтверджені статистичними критеріями;

<u>коректність</u> - її прийнятність (з погляду дослідника, практика), та відповідність змодельованому процесу або системі.

Ці терміни є звичайними у математиці. Статистичні методи можуть підтвердити високу ймовірність адекватності моделі, але особливості інформації можуть призвести до результатів, неприйнятних з погляду сутності явища, інакше кажучи, - коректна модель є у певному сенсі і найкраща.

Особливо зростає небезпека помилок при малих вибірках.

Моделі, що пов'язують понад дві змінні, називають багатовимірними.

До них відносяться множинні регресійні рівняння, багатовимірні явища, випадкові процеси. За формою, моделі зв'язку можуть бути в табличному, графічному чи аналітичному вигляді, тобто у вигляді рівнянь. Регресійні рівняння бувають лінійні та нелінійні, причому цей термін може відноситися як до коефіцієнтів рівняння, так і до незалежної змінної. Наприклад, рівняння $y=a+bx+cx^2+dx^3$ є лінійним за коефіцієнтами a, b, c, d і не лінійними по x, а множина рівняння $y = ax_1^b+cx_2^{dx_3}$ не лінійне як за коефіцієнтами, так і за змінними x_1, x_2, x_3. Ми розглянемо лінійні рівняння щодо коефіцієнтів, так як моделі такого роду цілком достатні для моделювання зв'язків в лісових дослідженнях, а теорія нелінійного (за коефіцієнтами) оцінювання дуже складна. ***Багато нелінійних моделей можна привести до лінійного вигляду, наприклад, логарифмування. Такі моделі називають внутрішньо лінійними.*** Моделі можуть бути представлені не тільки рівняннями,

але й в графічній та табличній формі. Табличне та графічне уявлення закономірностей зв'язку традиційно властиві лісовому господарству та широко поширені, особливо для даних, отриманих без належного статистичного обґрунтування. Так, лісотаксаційні таблиці (об'ємні, сортиментні, таблиці ходу росту та ін.), розроблені до 60-х років XX століття, представлені переважно у вигляді числових масивів, отриманих графічним вирівнюванням дослідних матеріалів. Недолікі такого роду моделей відомі: суб'єктивізм кінцевих результатів, неможливість статистичної оцінки відповідності моделі досліджуваному явищу, незадовільність форми для обробки інформації при допомозі ПК. У той же час моделі, виражені в табличному вигляді, де величини отримані шляхом графічного вирівнювання, розроблені дуже кваліфікованими та відповідальними фахівцями на основі великого (навіть величезного) експериментального матеріалу, часто виявлялися адекватними і коректними та знаходили (знаходять і сьогодні) широке застосування в практиці. До таких моделей можна віднести таблиці об’ємів ствбурів А. Крюденера та об'ємні таблиці М. М. Орлова (1867-1932), Д. І. Товстоліс, В. К. Захарова, Б. А. Шустова, А. В. Тюріна, сортиментні таблиці Ф. П. Моїсеєнко, таблиці ходу росту А. В. Тюріна та інші. Ці табличні дані зазвичай точніше результатів, отриманих шляхом використання математичних моделей, що мають недостатнє лісівниче обґрунтування. В принципі будь-який числовий чи графічний матеріал можна виразити як формулу. Проте досвід моделювання чисельних масивів, вирівняних графічно свідчить, що досить точне їхнє відображення вимагає використання громіздких та різнорідних математичних виразів, що створює певні незручності у використанні їх в системах автоматизованої обробки даних В даний час можливості комп'ютерів дозволяють користуватися величезними масивами інформації, вираженої в табличній формі, не вдаючись до її спрощення у вигляді рівнянь. При обробці лісівничої інформації часто трапляються випадки, коли графічні методи вирівнювання найдоцільніші. Це буває тоді, коли способи збору вихідних даних невідомі чи статистично неможливі, а збір інформації (метод збору, обсяг) не узгоджений з кінцевою метою роботи. В силу цього застосування статистичних методів може дати результати, що суперечать суті досліджуваного явища. Проілюструємо друге положення прикладом. Нехай у нас є певний масив даних, який характеризує зміну запасу деревостану за віком. Відомо, що запас деревостану залежить від породи дерев, лісорослинних умов, віку та повноти. При зборі експериментального матеріалу, найбільш складно витримати однаковість по повноті. Якщо ми хочемо визначити хід росту за запасом нормальних деревостанів (з повнотою 1,0), але в дослідному матеріалі є пробні площі з повнотою від 0,6 до 1,0, то математичне вирівнювання з використанням всього наявного матеріалу дасть спотворену картину, що видно на рис.15.3.

В даному випадку встановлено, що використання всього масиву пробних площ, які представляють різнорідний матеріал (за повнотою) веде до систематичних помилок: в цьому випадку правильно попередньо збудувати графік, визначити головну закономірність (хід росту за повноти 1,0), навіть зробити графічне вирівнювання, а потім розробляти математичну (регресійну) модель.

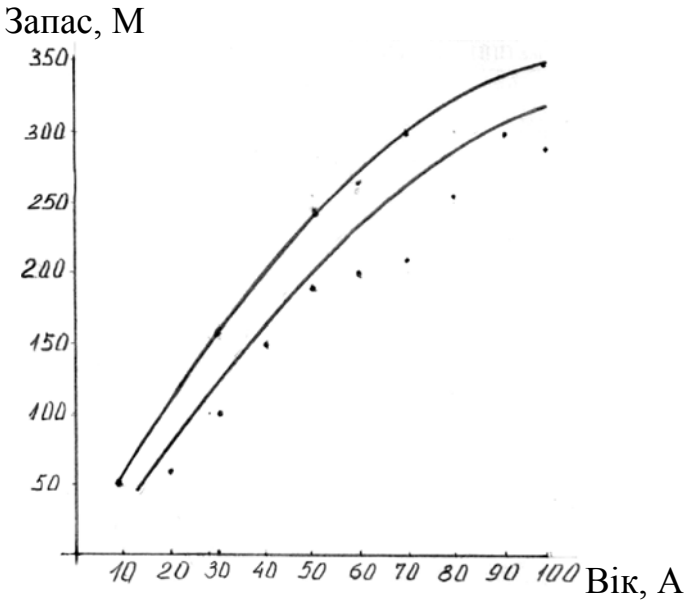


Рис. 15.3 Модель зміни запасу (М) деревостану від віку (А) при неправильно організованих вихідних даних для соснових деревостанів II класу бонітету пробної площі №3 в лісорослинних умовах B_{2-3} Копищанського ПНДВ

1 - лінія, побудована за даними з різноповнотних деревостанах;
2 - графічне вирівнювання за даними деревостанів з повнотою 1,0.

Подібний приклад ми наводимо на рис. 15.4. На ньому показано 15 випадковим чином виміряних висот дерев у різновіковому насадженні сосни звичайної, на пробній площі №2 Селезівського ПНДВ. Припустимо, що потрібно встановити залежність висоти деревостану від віку. Застосування стандартних статистичних методів дає залежність, показану штриховою лінією: починаючи з деякого віку, висота дерев поступово зменшується, тобто модель адекватна, але не коректна. Такий результат отримано через недостатність вихідних даних. У різновіковому деревостані сосни звичайної дерева, що мають однаковий вік, але які відрізняються зростанням в ярусах, мають різні висоти. Якщо немає можливості зібрати додатковий матеріал, то графічне вирівнювання - єдина можливість розв'язання цієї задачі. Надійність такого вирівнювання не можна оцінити, оскільки в основі його лежить єдина передумова: висота дерев із віком повинна збільшуватись, але не може зменшуватись.

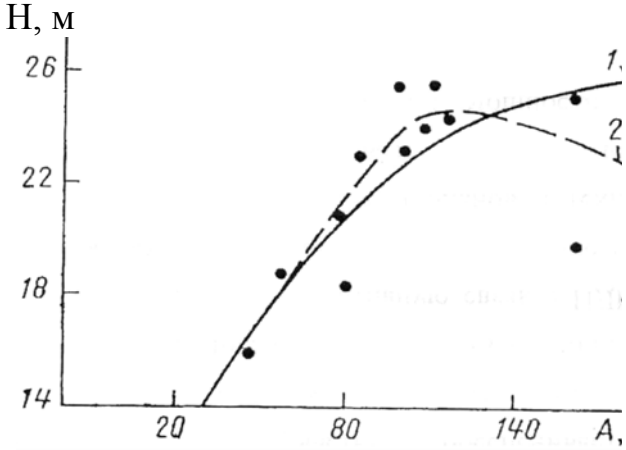


Рис. 15.4 Вирівнювання не коректно організованих даних:
1 – графічне; 2 – аналітичне.

Іноді виникає необхідність - графічну модель надати в аналітичному вигляді. Для цього застосовують два практичні прийоми: знімають із графіка значення залежної змінної та застосовують один із статистичних методів обчислення рівняння регресії або, якщо конкретний вид рівняння відомий, то коефіцієнти рівнянь доцільно обчислювати розв'язком системи рівнянь. Статистична оцінка моделей в обох випадках може дати лише відповідність графіка отриманого рівняння. Як приклад, візьмемо дані із рис. 15.5.

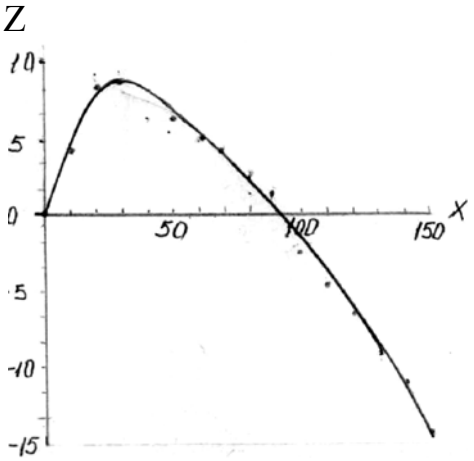


Рис. 15.5 Зміна приросту наявного деревостану у різновіковому насадженні сосни звичайної на пробній площі №2 Селезівського ПНДВ

Тут потрібно підібрати коефіцієнти рівняння, що відображає залежність висоти від віку. Виходячи з графіка, форма зв'язку тут може бути у вигляді параболи другого ступеня.

$$y = b_0 + b_1 x + b_2 x^2$$

де: x – вік, y – висота.

Для знаходження коефіцієнтів b, $Ь_1$, b_2 візьмемо три точки, наприклад, зі значеннями x, що дорівнюють 20, 80 та 140 рокам. За графіком знаходимо y_1=12,3; y_2 =21,1; y_3=25,2. Оскільки рівняння наведене вище справедливе для будь-якої точки, що лежить на кривій, то підставивши значення x і y в це рівняння, розв'язуємо систему із трьох рівнянь для знаходження коефіцієнтів. Змінну x для спрощення розрахунків можна масштабувати в 20 разів і записати:

$$x'_1 = x_1/20 = 1, \qquad x'_2 = x_2/20 = 4, \qquad x'_3 = x_3/20 = 7$$

$$\left.\begin{aligned} 12{,}3 &= b_0 + b_1 + b_2; \\ 21{,}1 &= b_0 + 4b_1 + 16b_2; \\ 25{,}2 &= b_0 + 7b_1 + 49b_2. \end{aligned}\right\}$$

Вирішуючи систему, знаходимо b_0=8,32, b_1=4,239, b_2=-0,2611; повернувшись до вихідної змінної, остаточно отримаємо

$y = 8{,}32+4{,}239x/20\ 0{,}2611(x/20)^2 = 8, +0{,}212-0{,}000652x^2$.

Легко переконатися, що суцільна лінія малюнку 15.4 є графічне зображення отриманого рівняння. Для цього обчислимо для деякого значення x, наприклад 80 років і отримаємо:

y=8,32+0,212*80+0,000652*802=8,32+16,96-4,17=21,11 м,

тобто визначений результат 21,1 м висоти в 80 років. Аналогічно розраховуємо значення для будь-якого віку на наявному інтервалі спостережень. Наведений приклад є окремий випадок так званих інтерполяційних багаточленів. При численних наближеннях інтерполяційний багаточлен ступеня n однозначно визначається n+1 його значенням. В прикладі для знаходження коефіцієнтів многочлена другого ступеня, ми взяли 3 складові. Застосування інтерполяційних багаточленів такого типу зручно, якщо модель, що використовується, повинна передбачати поведінку функції лише в точках (x_i, y_i,) - вузлах інтерполяції, оскільки оцінка поведінки многочлена між вузлами ускладнена. Якщо статистичний аналіз (чи інші міркування) дозволяє обмежитися лінійним (щодо незалежної змінної) рівнянням, то вибір моделі зв'язку однозначний. Але якщо зв'язок нелінійний, то процедури вибору конкретного аналітичного виразу як рівняння регресії не існує: одна і та ж залежність y від x може бути відображена багатьма формулами. Тут потрібно, з одного боку - знання загальних властивостей модельованої залежності, з іншого – математичних властивостей виразів, що використовуються.

Як приклад розглянемо дослідний матеріал, що отриманий при проведенні дендрохроноіндикаційних досліджень низки пробних площ в умовах Перганського, Копищанського, Селезівського ПНДВ Поліського природного заповідника за період з 2022 по 2025 роки. Для прикладу нам потрібно знайти залежність коефіцієнта варіації ($\upsilon_м$) запасу деревостану, визначеного за матеріалами вимірів на кругових пробних площах, від величини цих кругових пробних площ (b). Результати показані на рис. 15.6.

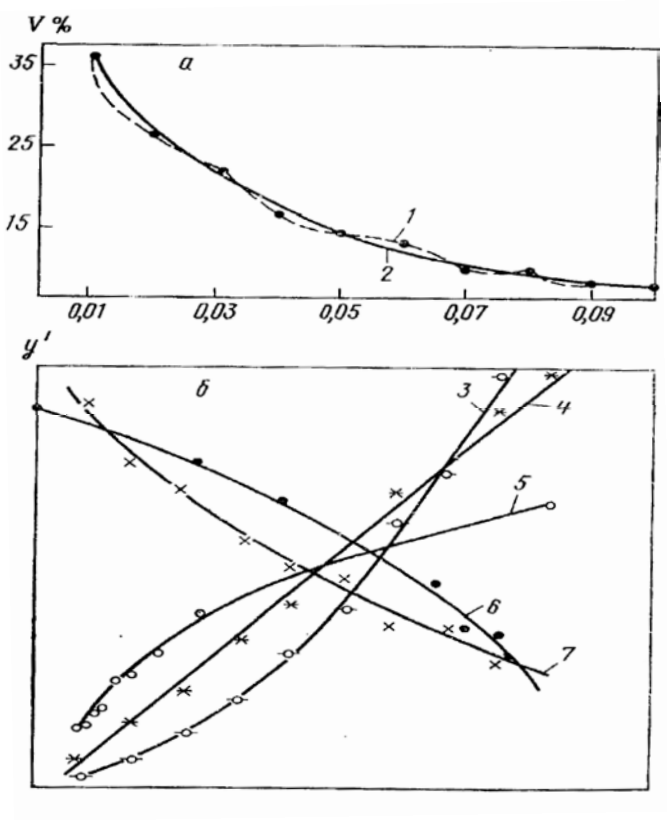


Рис. 15.6:

а) модель $\upsilon_м = f(l)$ 1- вихідні дані; 2 - вирівнююча крива гіперболічного виду: $y = x(a + \frac{1}{bx})$**;**

б) залежність коефіцієнта мінливості запасу від величини кругових пробних площ та перетворення її до лінійного вигляду.

Можна підібрати таке рівняння регресії, яке пройде через всі точки (так, інтерполяційний багаточлен n-1 ступеня пройде точно через n точок), але безглуздість подібного підходу очевидна. Наприклад, нам потрібно отримати закономірність зміни показників форми стовбура (q_2) залежно від висоти дерева в цілому, тобто врахувати зміну умовних середніх (а не кожного конкретного значення), які є випадковими величинами, схильні до мінливості і містять помилки вимірювань. Аналіз ситуації дозволяє припустити, що зменшення q_2 зі збільшенням висоти, має бути представлене монотонно спадаючою кривою. Отже, модель, що вибирається не повинна утримувати всередині свого робочого інтервалу особливих точок мінімуму, максимуму, точок згину.

Важливе значення має простота моделі та кількість коефіцієнтів, що підлягають визначенню. Тому вирази, з яких вибирають рівняння регресії, мають бути обмежені простими функціями. З них особливе місце посідають ті, які шляхом алгебраїчних перетворень можуть бути приведені до лінійного вигляду. Це дозволяє застосувати відносно прості, але глибоко розроблені методи статистичної оцінки До них насамперед належать

показові, статичні, логарифмічні та гіперболічні функції. В таблиці 15.1 наведено типи рівнянь, що найчастіше використовують у лісвничих дослідженнях. Ці моделі з однією незалежною змінною та з двома коефіцієнтами, що приводяться до лінійного простого вигляду. У таблиці показаний вид перетворення та критерії, що дозволяють оцінити прийнятність цього виразу.

Таблиця 15.1

Деякі елементарні функції, які найчастіше використовуються в лісовому господарстві, однієї змінної та перетворенням їх до лінійного вигляду

Рівняння	Тип перетворення	Рівняння прямої	Рівняння застосовується, якщо постійній величині дорівнює співвідношення
$y=ax^b$	логарифмування	$\lg y=\lg a+b\lg x$	$\Delta(\lg x)/\Delta(\lg y)$
$y=ae^{bx}$	логарифмування	$\lg y=\lg a+(b\lg e)x$	$\Delta x/\Delta(\lg y)$
$y=a+bx^n$ n - відоме	заміна $x'=x^n$ $y'=y$	$y'=a+bx'$	$\overline{\Delta(x^n)/\Delta y}$
$y=a+b/x$	заміна $x'=1/x$ $y'=y$	$y'=a+bx'$	$\Delta(1/x)/\Delta\lg$
$1/y=a+bx$	заміна $x'=x,$ $y'=1/y$	$y'=a+bx$	$\Delta x/\Delta(1/y)$
$x/y=a+bx$ $y=\frac{x}{a+bx}$	заміна $x'=x,$ $y'=x/y$	$y'=a+bx'$	$\Delta x/\Delta(x/y)$

Завдання також можна вирішити шляхом попереднього перетворення вихідних даних та нанесення їх на графік у новій системі координат (точки повинні розташовуватися вздовж прямої лінії). Перший шлях набагато кращий, оскільки легко реалізується в програмах автоматичного вибору моделі на комп'ютері, другий - наочніший. В лісовому господарстві крім рівнянь, наведених у таблиці 15.1 дуже широко використовуються рівняння виду цілих поліномів, переважно 2-4 ступеня:

$$y=a_0+a_1x+a_2x^2+a_3x^3+\ldots$$

До лінійного вигляду вони можуть бути наведені шляхом логарифмування. Цими рівняннями добре виражаються залежності

H=f(D). При цьому тут небажане використання параболи 2 порядку, яка занижує початкові та завищує кінцеві дані.

А. Г. Мошкальов (1926-1992) використовував для зв'язку H = f(D) поліноми 3-го ступеня. В. Ф. Багінський іноді використовував параболу 4 порядку. Для обчислення коефіцієнтів в рівняннях виду цілих поліномів для ПК є хороше матзабезпечення, де дається оцінка рівняння. З рівнянь, рекомендованих К. Є. Нікітіним та А. З. Швиденко, при проведенні досліджень в лісовому господарстві найбільш вживаними є (таблиця 15.1)

:

$$\Delta(\lg x)/\ \Delta(\lg y)$$

$$\Delta(1/x)/\ \Delta\lg$$

$$\Delta x/\ \Delta(1/y)$$

$$\Delta x/\ \Delta(x/y)$$

Аналіз основних моделей та виділення найбільш адекватних, що використовуються для опису динаміки деревостанів, їх будови, зв'язків між запасами насаджень і класами бонітету, типів лісу та опис взаємозв'язків між іншими таксаційними показниками, виконав О. О. Атрощенко в своїй монографії «Моделювання зростання лісу та лісогосподарських процесів», якау наведено у списку літератури.

15.3 Верифікація регресійних моделей.

Верифікація моделі - це перевірка, підтвердження того, що модель вірна (або не вірна), і відповідає законам та закономірностям, чинним у генеральній сукупності.

Верифікація проводиться у кілька етапів.

• Перевіряється ступінь згладжування (вирівнювання) дослідних даних, тобто ступінь збігу теоретичних та експериментальних значень функції за заданих аргументів. Для цього використовують методи, розглянуті раніше: основну помилку рівняння регресії, залишкову дисперсію, відношення загальної та залишкової дисперсій, аналіз залишків.

• Перевіряють значущість коефіцієнтів рівняння регресії за t-критерію, де t - має бути 2 або 3, залежно від прийнятого рівня достовірності.

• Встановлюють ступінь кореляції аргументів та приймають рішення про їх усунення або використання.

Після виконання перерахованих процедур приходять до висновку, що рівняння підібрано правильно і добре описує закономірність за результатами досліду. Якщо це не так, то підбирають іншу модель. Але під час постановки досліду можуть бути помилки. Кількість первинного матеріалу може виявитися недостатнім для відображення всіх особливостей досліджуваного явища ув генеральній сукупності. Тому потрібна додаткова перевірка моделі, яка описує ту чи іншу закономірність, але на новому матеріалі, зібраному незалежно від вже раніше закладеного досліду. Його кількість має бути не меншою, ніж використано для побудови моделі. При цьому збір матеріалу для перевірки повинен здійснюватися відповідно до правил планування експерименту, - правилам організації вибіркових спостережень. Інакше вибірка буде нерепрезентативною. Тут не можна використовувати типові місця вимірів, типові ділянки і т.д., так як можуть виникнути зміщені оцінки, тобто ті, що мають систематичну помилку. Наприклад, якщо ми, визначаючи вихід сіянців у розпліднику, взяли більш врожайну ділянку як типову, то показники будуть завищені, і навпаки. Коли немає систематичної помилки, то точність можна розрахувати (і планувати), але систематичну помилку виправити майже неможливо, так як вона піддається статистичній інтерпретації. Тому при плануванні експерименту треба суворо витримувати статистичний метод.

В лісовому господарстві зазвичай досліджують досить неоднорідні сукупності. Наприклад, вивчаючи ступінь приживлювання лісових культур, ми можемо виявити на ділянці місця, де культури прижилися добре або де вони взагалі загинули, скажімо у разі вимочок на пониженнях. Оцінюючи ділянки деревостанів, пройдені рубками догляду, зустрічаємося з нерівномірністю вибірки дерев по всій площі. Вивчаючи змішаний деревостан бачимо, що дерева, скажімо, сосни звичайної та берези повислої можуть розташовуватися відносно рівномірно або біогрупами, але зустрічаються й відносно великі куртини (до 0,05 га., та до 0,10 га, й більше). Під час проведення досліджень (обстежень), причому не лише в наукових, а й практичних цілях, потрібно збирати різну інформацію. Тому планування експерименту щодо подібних об'єктів має будуватися на застосуванні вибіркових методів, відкидаючи закладку дослідних ділянок у типових місцях.

Сказане підтвердимо прикладом, коли потрібно провести дослідження в сукупності різних об'єктів (таблиця 15.2).

Таблиця 15.2

Інформація, необхідна щодо проведення досліджень різних об'єктів лісового господарства

№ п/п	Назва сукупносте (об'єктів)	Необхідна інформація	Метод збору
1.	середній склад деревостанів в умовах лісогосподарської філії чи природно-заповідного фонду	склад у виділі	маршрутний вибірковий облік або аналіз таксаційних матеріалів
2.	поновлення під наметом лісу	середня кількість та вік поновлення на 1 га.	закладання облікових площ
3.	стан 1-2 річних культу сосни звичайної	% приживлюваності	закладання облікових площ
4.	середній запас стиглих соснових деревостанів	запас, $м^3$/га	закладання пробних площ

Щоб отримати статистично значиму інформацію, наприклад для другої із названих сукупностей, тобто при врахуванні природного поновлення, треба провести спостереження на великій кількості облікових площ чи проаналізувати багато виділів. Чисельність поновлення від виділу до виділу, і від площі до площі всередині виділу буде сильно варіювати. Коефіцієнт варіації υ може досягати 100% й вище. Зрозуміло, що в цих умовах жодна з закладених пробних площ за принципом типової вибірки, не дадуть надійної інформації.

Так, для обліку з точністю 10% треба закласти 100 майданчиків:

$$N = \frac{V^2}{P^2} = \frac{100^2}{10^2} = 100 .$$

За 5% точності (з достовірністю 0,68) вже потрібно:

$$N = \frac{100^2}{5^2} = \frac{10000}{25} = 400 \text{ шт.}$$

Тому в практиці зазвичай обмежуються точністю 10-15% при достовірності 0,68. Аналогічна картина буде для першого прикладу. Склад соснових деревостанів від виділу до виділу змінюватиметься. Зрозуміло, що типова вибірка тут нічого не дає, так як типовість, тобто середній склад, ще належить визначити. Для обстеження лісових культур з метою встановлення їх приживлювання, в

лісовому господарстві існують спеціальні правила, які вимагають підрахувати кількість сіянців на певній площі. При цьому передбачається вибірковим шляхом, найчастіше методом систематичної вибірки, закладати облікові площі чи смуги (ряди). Для знаходження середнього запасу в деревостанах, які намічаються в головне користування, необхідно закласти пробні площі в природно - стиглих сосняках. Але варіація тут висока через різницю в умовах зростання та повноті. Тому і в подібних ситуаціях планування експерименту базується на методах, розроблених в математичній статистиці.

Непридатність типового способу відбору вибірки в описаних випадках полягає не в тому, що спосіб недостатньо точний, а в тому, що він не вільний від суб'єктивізму. Незважаючи на цей недолік, спосіб типової вибірки доводиться нерідко застосовувати в оцінці (таксації) насаджень. Наприклад, при окомірній таксації середній діаметр дерев деревостану визначають як середню величину з 3-4 окомірно взятих середніх по товщині дерев. При детальному обстеженні ділянки лісу на ураженість шкідниками, доводиться закладати пробну площу в середніх (типових) умовах. Іноді так само вирішують і завдання оцінки поновлення. За неминучістю як спосіб, що вимагає найменшого обсягу спостережень, метод типової вибірки знаходить широке застосування при вирішення багатьох завдань в лісовому господарстві. Слід зазначити, що інформацію, придатну для статистичної оцінки досліду з певною точністю саме цим способом одержати неможливо. Можна обчислити середню величину ознаки з врахуванням такого матеріалу. Але помилку цієї середньої не слід визначати, оскільки вона неправомірна, бо випадкова мінливість ознаки (середнє квадратичне відхилення) досліду не вимірювалася. В цьому випадку доведеться для оцінки досліду обмежитися типовою вибіркою, тобто тільки отриманою середньою величиною ознаки.

Для статистичного тлумачення результатів досліду розглянутих сукупностей (відновлення на вирубці, сосняки Селезівського ПНДВ, культури сосни звичайної), з певною точністю потрібно закладати пробні площі по одному із способів випадкового відбору, що розглядаються нижче. При цьому досліднику доведеться вирішити два основні питання:

1) визначити достатню кількість спостережень;

2) правильно відібрати одиниці для спостережень.

При вирішенні першого питання можна скористатися формулою:

$$N = \frac{V^2}{P^2},$$

де: V – коефіцієнт варіації; P – показник заданої точності.

Вище показано, що з метою отримання результату з точністю 5% для оцінки поновлення сосни звичайної, де коефіцієнт варіації (υ) нами встановлено у 100%, потрібно було б закласти 400 пробних площ. При цьому наш висновок про те, що отримана вибіркова середня відрізнятиметься не більше ніж на 5% від генеральної середньої дається з ймовірністю 0,68. Це означає, що в 32 випадках зі 100, ця закономірність може й не підтвердитися.

Якщо прийняти рівень безпомилкового висновку 0,05, тобто робити висновок з ймовірністю 0,95, яку слід вважати достатньою, то у формулу для визначення числа спостережень потрібно ввести множник t; (t – критерій). При ймовірності 0,95 t-критерій приблизно дорівнює 2 (1,96). Тоді формула для числа спостережень буде мати вигляд:

$$N = \frac{V^2 \cdot t^2}{P^2}$$

Для нашого випадку, число спостережень стало б:

$$N = (2^2(100^2) / 5^2 = 1600.$$

В інших випадках, коли υ за пробною вибіркою характеризувалося б меншим числом, можна було б планувати меншу кількість спостережень.

З наведеного прикладу видно, що якщо варіювання значень ознаки в сукупності велике, та отримане N за вище наведеною формулою практично недосяжне. В таких обставинах дослідник має піти на звуження об'єкта досліджень або іноді задовольнятися точністю досліду в 10 та навіть 15%. Правильніше буде перше рішення. Можна взяти, наприклад у досліді не всі дерева сосни звичайної в межах виділу, а лише якогось типу лісу чи лісорослинних умов, бажано найпоширеніших. В межах таких об'єктів, варіювання буде значно меншим. Взагалі найбільш оптимально дотримуватися наступного принципу: брати під час закладання пробних площ більш обмежені сукупності.

15.4 Застосування регресійних моделей в лісовому господарстві.

В лісівничих дослідженнях регресійні моделі – необхідний інструмент дослідника для вирішення статистичних завдань. При цьому найважливішим етапом, який треба суворо витримувати і в практичній роботі, є планування експерименту в суворій відповідності до законів біометрії.

Вирішення питання щодо планування експерименту полягає в правильному відборі чи розміщенні одиниць спостереження. Сучасна статистична теорія рекомендує при цьому ряд методів.

Зазвичай застосовують такі способи: простий випадковий відбір, або випадкове безповторне вибіркове спостереження, випадкове пошарове вибіркове спостереження, систематичне вибіркове спостереження та субвибіркове спостереження або двостадійне спостереження.

Простий випадковий відбір є найбільш поширеним та статистично розробленим методом. Його організують за допомогою будь-якого механізму, що забезпечує рівну можливість для будь-якої одиниці потрапити у вибірку.

Зазвичай для вибору одиниць використовують таблицю випадкових чисел. Але застосувати цей спосіб в чистому вигляді при дослідженнях в лісовому господарстві часто дуже важко, інколи й неможливо. Наприклад, закладка облікових майданчиків в межах лісокористувача або об’єкта природно-заповідного фонду вимагатиме великої кількості часу на переїзди та пошук потрібного об'єкта.

Найприйнятніша і найчастіше застосовується систематична вибірка. За точністю та обґрунтованістю, вона практично не поступається випадковій, будучи своєрідним варіантом останньої.

Систематична вибірка повністю визначається вибором першого її члена. Вибирають для обміру або спостереження, припустимо, кожен десятий член, наприклад 10, 20, 30 і т.д. дерево за переліком або 10, 20, 30-й і т.д. ряд культур.

Закладка облікових майданчиків через певну відстань одна від одної, представляє також систематичну чи механічну вибірку. *Перевага такої вибірки* – легкість її отримання та рівномірність розподілу по всьому об'єкту. *Її недоліком* можуть бути випадки, коли сукупність має періодичну мінливість, і якщо інтервал між одиницями, що відбираються збігається з довжиною хвилі цієї зміни (або кратний їй), то отримаємо вибірку зі зсувом, тобто із систематичною помилкою.

Допустимо, що багаторядний агрегат, який застосовували при посіві будь-якої сільськогосподарської культури, мав один із шести або іншого числа несправний захват площі посіву. Якщо у наступному обліку номери облікових рядів с.-г. культур збігаються з рядами, зробленими з несправним захопленням, то вибірка матиме систематичну помилку. Це завжди слід враховувати під час планування досліду.

В умовах рівномірної мінливості ознаки, можна застосовувати систематичну вибірку і без суттєвої похибки обробляти як випадкову. Після отримання додатково матеріалу, зібраного з використанням статистичного методу, можна рекомендувати її використання в широкій лісогосподарській практиці. Остаточний

висновок про вірність моделі та її широке застосування може бути лише після перевірки практично, тобто модель має бути випробувана у виробничих умовах. Тільки тоді можна врахувати всі можливі особливості закономірностей, що діють в генеральної сукупності. Якщо модель підібрана правильно та добре перевірена, то при її використанні в практиці не виявляється суттєвих недоліків. Але невелика корекція все одно може мати місце.

Наприклад, сортиментні таблиці Ф. П. Моїсеєнко були неодноразово перевірені в практиці. Але їх новий, вдосконалений варіант пройшов додаткову перевірку, і лише після внесення невеликих уточнень було прийнято та видано. Натомість із 2007 року, ці таблиці застосовуються у практиці без зауважень. Це позитивний приклад, хоча є такі, коли модель доводилося змінювати. Наприклад, початкову модель для спрощеного відведення лісосік, розроблену В. Ф. Багінським через рік після дослідної перевірки автор замінив більш досконалу.

Регресійні моделі – основа для складання майже всіх нормативів в лісовому господарстві і вони повинні ретельно відбиратись та перевірятися, щоб господарство в лісі велося на належному науковому рівні. Всі нормативи для лісового господарства, де зустрічаються кількісні величини, розроблені за допомогою описаних методів. Так, «Правила закладання пробних площ», «Проведення відводів рубок головного користування», «Правила проведення відводів вибіркових санітарних рубок» базуються на закономірностях зміни приросту деревостану (z) залежно від повноти (П). Це регресійний зв'язок z =*f*(А, П, Б), де А і Б - вік та клас бонітету. Вихід сортиментів (С) в товарних таблицях визначається за його зв'язком з середніми діаметром (D), висотою (Н) та класом товарності (К), тобто С=*f*(D, Н, К). Вартість деревини (Cm) залежить від деревної породи (Р), розміру сортименту (Кр) та його сорту (S), тобто. Cm = *f*(Р, Кр, S). Цей показник враховують під час розрахунків ефективності проведених лісогосподарських заходів. Стиглість лісу та вік рубки знаходять, використовуючи зв'язки виходу сортиментів, їх вартості та інших від віку. Ці приклади можна продовжити. Все вище наведене свідчить про те, що в лісовому господарстві регресійні методи застосовуються цілком й повністю.

РОЗДІЛ 16. ВИМІРЮВАННЯ ЗВ'ЯЗКУ МІЖ ЯКІСНИМИ ОЗНАКАМИ

16.1 Особливості статистичного аналізу якісних ознак.

В лісовому господарстві доводиться мати справу не лише з об'єктами, що відрізняються кількісно, а й якісно. Наприклад, рослини та тварини одного виду варіюють за забарвленням. Іноді замість числівникових вимірів величини (ваги, розміру) насіння, їх можна розділити за такими якісним категоріям, як колір; листя відрізняють за формою (еліпсоподібні, круглі) і т.д. У цьому випадку, звісно можна зробити відцифровування якісних ознак, але це несе певну умовність. Тому краще скористатися спеціальними методами, розробленими для таких випадків. Такі методи розроблено достатньо повно та добре викладені К. Є. Нікітіним, А. З. Швиденко, П. Ф. Рокицьким, Н. І. Сваловим та іншими авторами, матеріалом яких ми скористаємося при вивяенні даного питання.

Якісні ознаки в конкретній задачі розглядають як постійні або мінливі. Будемо називати якісну ознаку мінливою, якщо вона може приймати різні стани чи градації x_i. Наприклад, різні інтенсивності забарвлення хвої, насіння, кори та інші, багатство умов зростання деревостану - бори, субори, сугруди, груди; категорії захисності лісів та ін. Для мінливих ознак доцільно вводити поняття, аналогічні статистикам розподілу та зв'язку випадкової величини - середнє, медіану, моду, кореляції та ін. Найпростіший випадок - коли в досліді відзначають лише наявність чи відсутність деякої ознаки «А», тобто «А» має дві градації: «А» та не «А». Тоді, позначивши одну з них 1, а другий 0, вивчення статистичних задач, пов'язаних з даною якісною ознакою, можна звести до закономірностей біномного розподілу. Загалом до цього прийому вдаються і в складніших ситуаціях: різні якісні градації ставлять у відповідність кількісним ознакам. Наприклад, деяка кількість жолудів може бути розподілена за градаціями інтенсивності блиску (блискучі, тьмяні, матові), деревостани розрізняються за лісорослинними умовами, де визначено їх площі чи запаси тощо. Тоді, позначивши значення кількісної ознаки через y_i, можна записати якісну ознаку аналогічно ряду розподілу випадкової величини:

$$\begin{matrix} x_1 & x_2 & \dots & x_k \\ y_1 & y_2 & \dots & y_k \end{matrix}$$

Приклади розподілу якісних ознак для різних об'єктів показано у таблиці 16 1.

Таблиця 16.1

Якісні ознаки для різних сукупностей

Назва сукупностей	Градація	Кількість	
		шт	%
насіння сосни звичайної в партії насіння	блискучі	-	75
	тьмяні	-	20
	матові	-	5
	всього:	-	100
мишевидні гризуни	чорні	-	85
	сірі	-	12
	інші	-	3
	всього:	-	85
розподіл соснових деревостанів по групам за лісорослинними умовами	бори	43260	65,0
	субори	20880	31,4
	сугруди	1820	2,7
	груди	580	0,9
	всього:	66540	100

За термінологією Клауса Джині, якісну ознаку називають впорядкованою, якщо її градації утворюють природну послідовність, в іншому випадку вона є невпорядкованою.

Впорядковані ознаки поділяють на прямолінійні та циклічні. *Для прямолінійних ознак* природним чином можна виділити дві крайні градації, інакше кажучи, упорядкувати ознаку за градаціями від меншого до більшого чи навпаки. Як приклад, наведемо інтенсивність забарвлення насіння сосни звичайної або мишовидних гризунів, а також збільшення грунтового багатства лісорослинних умов що показано в таблиці 16.1.

Об'єкти, що мають якість прямолінійного типу, можна ранжувати, тобто зіставити кожному їх число натурального ряду. Якщо два об'єкти (градації) мають ранги 2 та 5, значить між ними знаходиться 3 інші об'єкти або градації. Ранжування вихідних даних - широко поширений прийом при обробці не тільки якісних, а й кількісних ознак. Можна ранжувати області України за відсотком лісистості, мисливські угіддя за кількістю диких тварин, підприємства лісового господарства за кількістю працюючих чи об’ємом деревини, що заготовлюється і т.д.

Ранги бувають нескладними, коли градації об'єктів дозволяють здійснити ранжування повністю, і ***зв'язковими*** - коли деяка частина об'єктів займає однакове місце. Наприклад: 4 об'єкти можна

рівноправно розмістити після об'єкта із рангом 2. Тоді їм приписують ранг:

$$\frac{1}{4}(3+4+5+6) = 4\frac{1}{2}$$

Для циклічних ознак дві крайні градації не можна виділити на об'єктивній основі. Приклади циклічних ознак - пори року (весна, літо, осінь, зима), орієнтування схилів у гірській місцевості по напрямах світу і т. д. Приклади невпорядкованих якісних ознак - типи лісу деякого регіону (за будь-якою з існуючих класифікацій), методи окулірування при щепленнях і т.д.

16.2 Основні методи аналізу якісних ознак.

Для обчислення якісних ознак виходять із умови, що різні градації цих ознак кількісно невимірні чи вимірні, але виміри з якихось причин проведено не були. Однак для подальшого аналізу допускаються такі припущення:

• є можливість встановлення між градаціями деякої міри розходження;

• різниця між сусідніми градаціями однакова.

Такі ознаки називають рівнопроміжними. Якщо виконується перша умова, то ознака може бути зроблена рівнопроміжною введенням відсутніх значень. Для якісного рівнопроміжного ряду типу (таблиця 16.1) середнього значення, вводиться за аналогією з кількісними ознаками: градації X_i ставиться у відповідність деяка кількість, а постійної різниці між X_i та X_{i+1} - деяке число h. Тоді замість ряду (таблиця 16.1) маємо:

$$\begin{array}{cccccc} a, & a+h, & \ldots, & a+kh, \\ y_1, & y_2, & \ldots, & y_k, \end{array}$$

Досліджуючи приведений вище ряд, можна отримати деякі чисельні значення щодо якісної ознаки X. Так, середню арифметичну отримаємо з наступного рівняння:

$$\tilde{X}_{\text{кат}} = -\frac{\sum_{i=0}^{k}(a+ih)y_i}{\sum_{i=1}^{k} y_i}$$

Але формальне перенесення операцій, введених для випадкових величин, виправдане лише в тому випадку, якщо обчисленим

величинам можна приписати реальний зміст, і вони пояснюють будь-які обставини, пов'язані з розв'язуваним завданням.

Наприклад, візьмемо розподіл соснових деревостанів деякого (і - того) філії «Коростенське ЛГ», Столичного офісу ДП «Ліси України». Загальна площа соснових деревостанів в філії за результатами останнього лісовпорядкування 2020 року становить 66540 га. За класами бонітету вони розподілені в такий спосіб (таблиця 16.2).

Таблиця 16.2

Розподіл соснових деревостанів за класами бонітету

Клас бонітету	I^a	I	II	III	IV	V	V^a	Всього:
площа, га	720	2070	41130	17480	2920	1810	410	66540
площа, %	1,1	3,1	61,8	26,3	4,4	2,7	0,6	100

Клас бонітету тут виступає як якісна ознака, замінюючи показник лісорослинних умов. Клас бонітету виражають римськими цифрами, а крайні бонітети мають індекси «а» та «б». Для того, щоб знайти середній клас бонітету, перетворимо ряд в таблиці 16.2 на рівнопроміжний, зашифрувавши римські цифри арабськими (таблиця 16.3).

Таблиця 16.3

Розподіл соснових деревостанів за класами бонітету під час заміни римських цифр арабськими в рівнопроміжному ряду

Клас бонітету	1	2	3	4	5	6	7	Всього:
площа, га	720	2070	41130	17480	2920	1810	410	66540
площа, %	1,1	3,1	61,8	26,3	4,4	2,7	0,6	100

За даними таблиці 16.3 звичайним шляхом обчислимо умовне значення класу бонітету:

$$\tilde{X}_{\text{к а}} = -\frac{\sum_{i=0}^{k}(a + ih)y_i}{\sum_{i=1}^{k} y_i}$$

При цьому результат має бути однаковим, в межах точності округлення, якщо використовуємо показники площ в га чи відсотках.

Таблиця 16.4

Розподіл площ та середніх запасів ($м^3$/га) за класами бонітету для соснових деревостанів

Клас бонітету	I^a	I	II	III	IV	V	V^a	Всього:
шифр класа бонітета	1	2	3	4	5	6	7	-
площа, га (П)	720	2070	41130	20400	-	-	2220	66540
запас, $м^3$ (М)	220	180	150	110	-	-	40	-

Наведений в таблиці ряд не є рівнопроміжним, тому що відсутні значення між III та V^a класів бонітету.

Поширеним методом аналізу якісних ознак є метод чотирьох полів. Враховуючи позначення, ступінь сполученості, що існує між якісними ознаками чи альтернативами визначається за формулою Д. Юла.

$$r = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

В даному випадку, приклади зручно брати з галузі генетики. Тому проведемо аналіз якісних ознак на улюбленому об'єкті генетиків - мушці дрозофілі. Від схрещування сірого самця, що мав довгі крила, із чорною самкою, у якої вони були рудиментарними, все потомство виявилося однаковим. А при схрещуванні особин із першого покоління між собою, у другому поколінні гібридів були отримані такі категорії мух (таблиця 16.5).

Таблиця 16.5

Розподіл ознак у потомстві мушки дрозофіли

Ознака	Кількість, шт	%
сірі, довгокрилі з придатками крил	20	41,7
чорні довгокрилі	20	8,3
чорні, довгокрилі з придаткми крил	100	41,7
всього:	210	100

Тепер зробимо аналіз наведених якісних ознак для встановлення між ними кореляційних залежностей. Розв'язання цього завдання зводиться до обчислення коефіцієнта кореляції за методом чотирьох полів, тобто за формулою Юла. Позначимо через х чорне

забарвлення тіла, а через у – зачатки крил. Потім розмістимо дані експерименту в чотирипільної кореляційної таблиці (16.6).

Таблиця 16.6

Розподіл розщеплених ознак у потомства дрозофіли

Х/У	Крила без придатків, y_1	Крила з придатками, y_2	Всього:
не чорне тіло, x_1	a=100	b=20	a+b=120
чорне тіло, x_2	c=20	d=100	c+d=120
всього:	a+c=120	b+d=120	a+c+b+d=240

Користуючись цією таблицею, підставляємо у формулу Юла потрібні дані і знаходимо величину коефіцієнта кореляції між зазначеними ознаками:

$$r = \frac{100 \cdot 100 - 20 \cdot 20}{\sqrt{120 \cdot 120 \cdot 120 \cdot 120}} = \frac{9600}{14400} = +0,67$$

Оцінка достовірності коефіцієнта кореляції, обчисленого за способом чотирьох полів, що нічим суттєво не відрізняється від звичайної оцінки значущості цього показника обчислимо за формулою:

$$m_r = \frac{1 - r}{\sqrt{N}}$$

Звідси:

$$r = \sqrt{\frac{106,7}{240}} = \sqrt{0,45} = +0,67.$$

Отримуємо той самий результат, що при розрахунку за формулою Юла.

Коефіцієнт асоціації.

Англійський вчений Джозеф Юл запропонував для оцінки тісноти сполучення між двома парами альтернатив підносити значення, що стоять в чисельнику, не до кореня квадратному з суми даних по стовбцям та рядкам чотирьохпільної таблиці, а до суми двох добутків ad+bc. отриманий таким чином показник, що позначається К, отримав назву коефіцієнта асоціації Юла. Його формула така:

$$K = \frac{ad - bc}{ad + bc}$$

Для нашого прикладу цей показник складає:

$$K = \frac{100 \cdot 100 - 20 \cdot 20}{100 \cdot 100 + 20 \cdot 20} = +0{,}92$$

Коефіцієнт асоціації завжди більший за коефіцієнт кореляції. Коефіцієнт асоціації свідчить про наявність паралелізму між числовими значеннями ознак, без урахування їхньої варіабельності, а отже і не дає точного уявлення про існуючий між ними зв'язок. У цьому полягає причина теого, що коефіцієнт асоціації не отримав широкого застосування в практиці.

Коефіцієнт контингенції.

Існує ще один оригінальний показник взаємної спряженості, так званий коефіцієнтом контингенції, що позначається С. Він обчислюється за формулою Пірсона:

$$C = \sqrt{\frac{\Phi^2}{1 + \Phi^2}}$$

Де величина $1+\Phi^2 = \left(\frac{p_{xy}^2}{p_x} : p_y \right)$

В цьому виразі P_{xy} – частоти, що вміщено в клітинках варіаційної решітки; p_y – кількість випадків в кожному горизонтальному ряду кореляційної таблиці.

Покажемо обчислення коефіцієнта контингенції на тому ж прикладі розщеплення ознак в гібридному потомстві дрозофіли. Хід обчислення представлено в таблиці 16.7.

З цієї таблиці випливає, що $1+\Phi^2=1{,}444$, звідки $\Phi^2 = 0{,}444$. Підставляючи знайдені значення у формулу отримаємо:

$$C = \sqrt{\frac{0{,}444}{1{,}444}} = \sqrt{0{,}307} = +0{,}55$$

Коефіцієнт контингенції менший за коефіцієнт кореляції, і по різниці цих показників можна робити висновки про прямолінійну або криволінійну залежність між ознаками. Перевага коефіцієнта контингенції перед коефіцієнтом асоціації полягає в тому, що він дозволяє вимірювати сполученість не тільки між двома, а й великою кількістю корельованих ознак.

Таблиця 16.7

Обчислення коефіцієнта контингенції

x_i/y_i		Крила без придатків, y_1	Крила з придатками, y_2	P_x	$\sum\left(\frac{P^2_{xy}}{P_x}:P_y\right)$
1	P_{xy}	100	20	120	86,67/120=0,722
	P^2_{xy}	10000	400	-	
	$P^2_{xy}: P_{xy}$	83,33	3,33	86,67	
2	P_{xy}	20	100	120	0,722
	P^2_{xy}	400	10000	-	
	$P^2_{xy}: P_{xy}$	3,33	83,33	86,67	
	P_y	120	120	240	1,444

В лісогосподарських дослідженнях для визначення тісноти зв'язку між якісними ознаками застосовують методи, що ґрунтуються на кореляції рангів. Таку кореляцію з обчисленням критерію Спірмена ми описали в розділі 10, коли досліджували кореляцію між кількісними ознаками (x) та (y). Цей коефіцієнт можна використовувати і для оцінки рангової кореляції якісних ознак. Для встановлення кореляції якісних ознак існує також ранговий коефіцієнт кореляції Кендела. Рангові коефіцієнти кореляції, що описують якісні ознаки змінюються як і звичайні коефіцієнти кореляції від -1 до +1 у тому самому значенні, тобто як показники прямого та зворотнього зв'язку. Вони змінюють знак на протилежний, якщо одну з послідовностей рангів замінити на сполучену, тобто на місце першого елемента поставити останній, за ним передостанній і т.д. Коефіцієнт рангової кореляції Кендела визначається за формулою:

$$\tau = \frac{2\sum x_j n \cdot y_i n}{N(N-1)}$$

При заміні суми $\Sigma x_j n * y_i n$ на S отримуємо:

$$r = \frac{2S}{N(N-1)}$$

Вище зазначені формули застосовуються в послідовності x_1, x_2, …, x_n та y_1, y_2, …y_n. Кожній парі рангів(x_i-x_j, y_i-y_j) ставлять у відповідність певну функцію. Остання дорівнює наступним величинам:

x_{in} або y_{in} $\begin{cases} +1 \text{ при } x_i,\ y_i < x_j, y_i \\ 0 \text{ при } x_i,\ y_i = x_j, y_i \\ -1 \text{ при } x_i > x_j, y_j\ . \end{cases}$

Обчислення, як правило, виконують на комп'ютері за спеціальною програмі, що входить до сертифікованого матзабезпечення. Для простих випадків, тобто при невеликій величині N, можна провести ручний розрахунок, скориставшись формулою:

$$\tau = \frac{4P}{N(N-1)} - 1$$

де: P - кількість пар, що утворюють пряму послідовність, мають прямий порядок (j<i). Основна помилка τ визначається за формулою:

$$m_\tau = \sqrt{\frac{2(2N+5)}{9N(N-1)}}$$

Як приклад, розглянемо, чи є зв'язок між світлолюбивістю деревної породи та її продуктивністю. Візьмемо 6 деревних порід, що ростуть за однакових лісорослинних умов – $C_{2.}$. Перший ряд (x_i) характеризує ранжування за світлолюбивістю, другий – за продуктивностю (запас на 1 га).

Деревні породи	Сосна звичайна	Ялина звичайна	Липа серцелиста	Дуб звичайний	Береза повисла	Осика
x_i	2	6	1	4	3	5
y_i	3	2	1	5	6	4

Перепишемо тепер ряд x в порядку зростання рангів:

Деревні породи	Липа серцелиста	Сосна звичайна	Береза повисла	Дуб звичайний	Осика	Ялина звичайна
x_i	1	2	3	4	5	6
y_i	1	3	6	5	4	2

Тепер підрахуємо, скільки пар в ряду утворюють прямий порядок. Для липи серцелистої таких пар 5 (сосна звичайна, береза повисла, дуб звичайний, осика, ялина звичайна), для породи сосна звичайна, таких пар 3 (береза повисла, дуб звичайний, осика), відповідно для берези повислої – 0, дуба звичайного – 0, осики – 0, ялини звичайної – 0.

Тоді P=5+3+0+0+0+0=8. За формулою визначимо основну помилку:

$$\tau = \frac{4 \cdot 8}{6(6-1)} - 1 = \frac{32}{30} - 1 = 1{,}07 - 1 = 0{,}07$$

$$m_\tau = \sqrt{\frac{2(2 \cdot 6 + 5)}{9 \cdot 6(6-1)}} = \sqrt{\frac{34}{270}} = \sqrt{0{,}015} = \pm 0{,}12$$

Відношення τ до її основної помилки дорівнює:

$$t = \frac{0{,}07}{0{,}12} = 0{,}58$$

Так як m_i<3, й навіть <2 то шукана залежність недостовірна. Достовірність коефіцієнта рангової кореляції Кендела t можна встановити і за таблицею критичних значень t-критерія Стьюдента, наведеної в додатку Е.

16.3 Метод індексів та його значення в лісовому господарстві.

Поряд з методами кореляції та регресійного аналізу в практичній роботі лісівників застосовується метод індексів. Сутність його полягає в тому, що величина однієї ознаки, зазвичай менша, виражається у частках одиниці або у відсотках від величини іншої ознаки:

$$I = \frac{Y}{X} \cdot 100\%$$

де: I – індекс, виражений у відсотках;
Y та X – величини відповідних ознак.

Метод індексів давно знайшов широке застосування у сфері морфології. Наприклад, ним користуються, оцінюючи статури тварин в мисливському господарстві і навіть людини. Перевага цього методу полягає в його простоті та загальнодоступності, а також у тому, що він дозволяє більш повно, ніж окремо взяті виміри, характеризувати особливості статури та їхню динаміку в онтогенезі. Тому найбільше прикладів описано саме біологами. Наведемо окремі з них, які описав Г. Ф. Лакін. Зробимо аналіз рухової активності мишевидних гризунів різних видів (таблиця 16.8).

Таблиця 16.8

Аналіз рухової активності мише видних гризунів різних видів

Вид гризуна	Вага тіла, г	Вага мозку, г	Ваговий індекс мозку, г	Рухова активність тварини
миша жовтогорла	24,2	726,0	21,8	рухома, здійснює далекі переходи
миша польова	23,5	569,9	13,6	менше рухома, ніж миша жовтогорла
полівка руда	21,3	541,3	13,8	рухома, не тримається однієї нірки
полівка сіра	23,8	462,0	8,9	мало рухома, тримається біля нірки

Посилення рухової активності тварин призводить очевидно до того, що більш швидко збільшується ваги мозку порівняно з вагою тіла. Тому в тварин, що ведуть активніший спосіб життя, виявляються і вищі вагові показники мозку.

Щоб встановити філогенетичне положення виду, характеризувати висоту його прогресивного розвитку, М. Ф. Нікітенко використав інший індекс, ***названий ним коефіцієнтом теленцефалізації, що представляє вагу переднього мозку, виражену в відсотках до загальної ваги головного мозку.***

В таблиці 16.9 показано значення цього індексу деяких видів ссавців та в людини.

Таблиця 16.9

Значення індексу в деяких ссавців та людини (за даними М. Ф Нікітенко)

Види	Вага, г		Коефіцієнт теленцефалізації, %
	головного мозку	переднього мозку	
їжак	2,85	1,36	47,8
кріт	0,85	0,44	51,2
кролик	22,05	11,31	51,3
білка	5,89	3,13	53,3
бобер річковий	41,40	26,60	64,3
собака	71,10	48,49	68,2
вовк	218,80	153,20	70,3
дикий кабан	173,50	123,35	71,1
дельфін	612,00	464,00	74,6
шимпанзе	93,20	74,83	80,3
людина	1482,70	1260,3-1289,9	85,0-87,0

Видно, що в міру руху вгору сходинками еволюції, індекс поступово збільшується: у комахоїдних він дорівнює 47%, а у приматів сягає 80%, тобто виявляє певну закономірність в філогенетичному ряді тварин.

Недоліки методу індексів.

За більш ніж сторічну історію цього методу було запропоновано чимало різних індексів як зоологами, так й антропологами. Однак незважаючи на простоту, загальнодоступність та широку популярність методу, він має дуже суттєві недоліки. Справа в тому, що індекси – величини відносні. Вони не враховують жодної варіабельності ознак, ні ступінь сполученості між ними, як це властиво іншим біометричним показникам – коефіцієнтам регресії та кореляції. Відомо також, що біологічні ознаки не тільки варіюють, але виявляють неоднакову швидкість розвитку та зростання (гетеродинамію), що не може не позначитися на величині індексів. Постійність індексів зберігається лише за відносно однакових швидкостях розвитку частин тіла в онтогенезі, тобто у випадках гомодинамії, коли відносини середніх арифметичних ознак X і У дорівнюють коефіцієнтам регресії, тобто за умови, що:

$$\frac{\overline{X}_x}{\overline{X}_y} = R_{x/y}$$

та:

$$\frac{\overline{X}_y}{\overline{X}_x} = R_{y/x}$$

Але такої рівність у співвідносній мінливості різних частин організму, зазвичай немає. Індекси та коефіцієнти регресії зазвичай повністю збігаються один з одним. Ілюстрацією цього може бути таблиця 16.10, у якій наводяться порівняльні дані визначення нормальної ваги людини за висотою її тіла на основі використання індексу Брока.

З таблиці 16.10 видно, що оцінка ваги тіла за його висотою виходячи з індексу Брока лише приблизно збігається з величинами, обчисленими за рівнянням регресії, та й то лише за середнього зростання росту людини в 160 – 165 см.

Антропологи вважають, що метод індексів абсолютно непридатний для оцінки фізичного розвитку організму, і повинен бути залишений як пройдений етап в розвитку науки. Водночас зоологи та багато лікарів продовжують користуватися індексами в своїй практиці, застосовуючи їх там, де можна обмежитися

наближеними даними, та де не потрібно досить точних показників фізичного розвитку організму. В даний час цей недолік в науці минулого століття повністю подолано; стали застосовуватися і досконаліші методи дослідження біологічних явищ, і це також цілком закономірне явище.

Таблиця 16.10

Порівняльні дані визначення нормальної ваги людини за висотою її тіла (індекс Брока)

Висота тіла (зріст), см	Вага тіла, кг		
	по індексу Брока	за рівнянням регресії	
		чоловіки	жінки
150	50	54	51,5
155	55	57	55,0
160	60	60	58,5
165	65	63,5	62,0
170	70	66,5	65,5
175	75	70	69,0
180	80	73	72,5
185	85	76	76,0

Безперечно, що найбільш точні методи для оцінки фізичного розвитку людини та тварин - це методи регресії та кореляції. Тільки завдяки цим методам оцінка фізичного розвитку людини була поставлена на наукову основу - вона використовується в даний час за допомогою номограм за трьома шкалами вимірювань – висотою (довжиною) тіла, обхвату грудей та живої ваги людини. Однак, мабуть, всілякий метод має властиві йому позитивні риси та недоліки, й різні за своєю конструкцією біометричні показники не виключають, а доповнюють один одного. Тому в практичній діяльності лісівникам доводиться вибирати то одні, то інші методи, відповідно до цілого ряду згідно певних обставин, від яких залежить цей вибір. Вся річ у тому, щоб розуміти конструкцію даного показника, бачити його позитивні сторони та недоліки. В умілих руках, метод індексів може надати лісівникам гарну послугу. Індекси однаково характеризують і прямолінійні, а також криволінійні співвідношення, що існують між біологічними ознаками, чого не можна сказати про інші показники статистичного зв'язку. Тому там, де цей метод застосовуємо, він в перспективі дає задовільні результати, його можна використовувати хоча б як допоміжний засіб в комплексній оцінці різних ознак в біології, сільському та лісовому господарстві.

16.4 Застосування статистичних методів дослідження якісних ознак в лісовому господарстві.

Методи дослідження якісних ознак в лісовому господарстві застосовуються досить широко, хоча за обсягом використання ці способи поступаються кількісним аналізам явищ. Застосування методів дослідження якісних ознак значно розширилося протягом останніх 10-15 років. Лісівники, як і вчені інших напрямів досліджень, вивчають лісових звірів та птахів, де необхідно виділяти ряд якісних ознак. Якісні ознаки застосовують для оцінки доброякісності лісового насіння, стиглості лісових ягід і т.д. Тут можна назвати роботи О. З. Швиденко, В. Б. Гедих, Г Г. Гончаренко, В. Є. Падутова, П. Г. Вакулюка та інших. Для вивчення лісових звірів та птахів, де необхідно виділити ряд якісних ознак (забарвлення, наявність або відсутність рогів і т.д.), для аналізу лісового насіння (колір, блиск та ін), форми листя (округла, еліпсоподібна та ін.) використовують різні якісні показники рангової кореляції Кендала, метод індексів і т.д. Знаходять застосування та інші статистичні методи оцінки якісних ознак.

Особливе місце ув дослідженнях з лісового господарства займає метод індексів. Його широко застосовують лісівники в своїй практиці, знаходячи його дуже зручним та часто незамінним. Найбільші застосування методу індексів знаходить при зіставленні величин, що мають різні одиниці виміру. Так, якщо потрібно врахувати біомасу живого грунтового покриву, скажімо чорниці в сукупності, із запасом деревостану, то для обчислення показників комплексної продуктивності різних типів лісу в перерахунку на 1 га, метод індексів є оптимальним.

Колли В. Л. Мешковій необхідно було визначити вік еколого-економічної стиглості деревостанів наших лісів, потрібно було поєднати економічні показники, що виражалися величиною найвищої рентабельності в певному віці, та екологічні, - що виражаються в тоннах зв'язаного CO_2 на 1 га, знайти вік оптимального поєднання цих показників, що оцінюються за різними критеріями, виявилося можливим застосувати лише метод індексів.

Вік деревостанів, в якому сума індексів, що виражають економічні та екологічні показники була максимальною, характеризував еколого-економічну стиглість.

Метод індексів широко використовується в лісовій таксації. Так В. В.Загреєв (1932-1993) розробив спеціальний метод дослідження динаміки деревостанів та складання таблиць ходу росту за допомогою методу індексів. Він встановив, що серед великої кількості кривих ходу росту нормальних насаджень за будь-яким таксаційним показником, завжди існує багато подібних. На підставі аналізу понад 400 таблиць ходу росту нормальних насаджень різних природно-кліматичних зон України та низки закордонних країн, цей автор виявив можливість систематизації, класифікації та

стандартизації таких таблиць. Для соснових деревостанів ним розроблено таблиці лісорослинних умов, що є середніми значенями і систематизовані з певною градацією відносні (індексні) ряди ходу росту за окремими таксаційними показниками. Наприклад, для характеристики з точністю ±3% великої кількості наявних таблиць ходу рост, встановлено, що соснові деревостани мають велику різноманітність ліній ходу росту у висоту. Тому виявилося достатнім мати одну таблицю, що включає лише 13 типових рядів. Такі індексні таблиці (графіки) служать для порівняльної якісної оцінки, групування таблиць за рівнем подібності та відмінності в характері (типі кривих) ходу росту.

На рис. 16.1 для прикладу наведено графік типових ліній ходу росту соснових деревостанів за сумою площ перерізів. Типові графіки та таблиці мають загальне значення, оскільки замінюють всі існуючі в природі лінії ходу росту насаджень по окремим таксаційним показникам. Типізація ходу росту насаджень дозволяє навести всі таблиці ходу росту нормальних насаджень певну систему, усунути існуючу плутанину в їх практичному застосуванні, оцінити існуючі таблиці ходу росту, виявити загальні закономірності, географічні відмінності та особливості зростання лісових деревостанів в певних лісорослинних умовах.

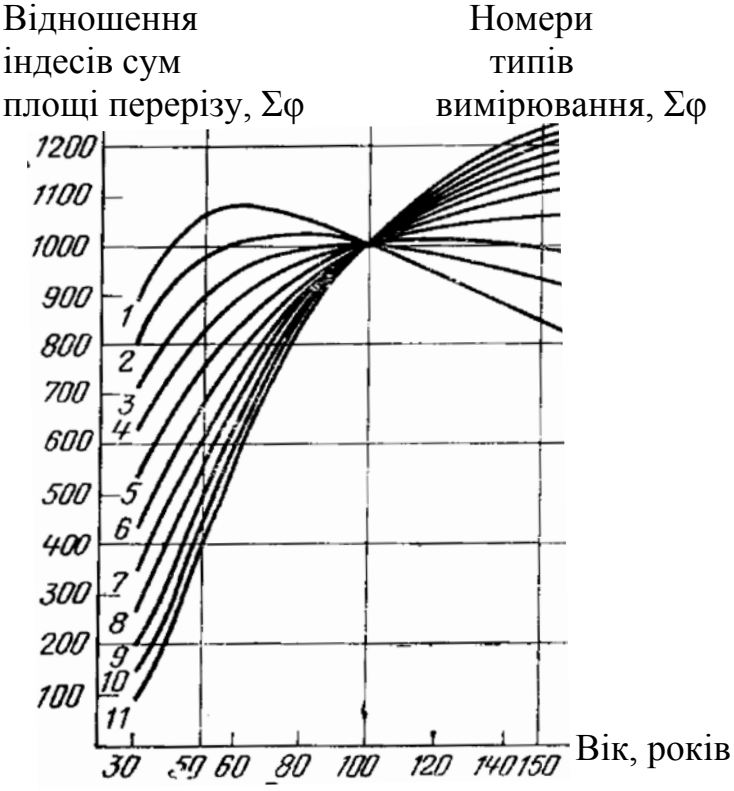


Рис. 16.1 Графік типових ліній ходу росту соснових деревостанів за сумою площ перерізів.

Але метод індексів має і недоліки, які виявились і в даному випадку. Типові лінії (рис.16.1) мають достовірність 0,68, тобто в

32% випадків точність типових ліній виходить за встановлені межі. Тому типові лінії правомірно застосувати лише для великої кількості досліджуваних об'єктів. В нашому випадку цей метод ми використовували в дослідженнях деревостанів в умовах пробних площ Перганського, Селезівського, Копищанського ПНДВ Поліського природного заповідника. В той же час метод індексів вимагає дуже ретельного та високопрофесійного попереднього аналізу досліджуваного явища. Без дотримання цієї умови, як і у випадках, що було описано вище, він призводить до помилкових висновків. Прикладом некоректного використання методу індексів можуть бути дослідження, проведені в умовах водно-болотних масивів 22-24 кварталів урочища Висока Піч ДП «Зарічанське ЛГ» в 2012-2016 роках щодо вивчення впливу локальної гідротехнічної меліорації в умовах бувшої розформованої та занедбаної ракетно-операціної бази в/ч 99931 «Висока Піч», на приріст сосни звичайної прилеглих лісових масивів. На той момент дослідники використовували відносний (індексний) показник $K_i = Z_a/M_a$, де: Za - поточний приріст в $м^3$ у віці а; Ма - запас в цьому ж віці. Але через те, що застосовані індекси Кі мають збіжність рядів, то через 10 років після спостереження відмінності штучно нівелюються, хоча насправді суттєва різниця зберігалась. Отримана величина виявилась непридатною для аналізу цього явища, так як спотворювала його сутність і призвела до помилкових висновків. Запобігти негативним наслідкам цих помилкових висновків стало можливим лише завдяки коректному аналізу роботи при її подальшому рецензуванні та перевірці для подальшого практичного використання.

З вище наведених прикладів випливає висновок, що проходить червоною лінією через все вивчення біометрії: перш ніж приступати до оперування з цифрами, треба вивчити суть явища, зрозуміти його основні закони та закономірності на загальному (професійному) рівні. Для цього потрібен попередній аналіз, а потім коректно слід змоделювати явище чи процес за допомогою прийомів біометрії.

РОЗДІЛ 17. ДИСПЕРСІЙНИЙ АНАЛІЗ

17.1 Поняття про дисперсійний аналіз.

При проведенні досліджень часто треба визначити, наскільки суттєвий вплив одного чи кількох факторів на кінцевий їх результат. Не завжди тут можна застосувати регресійний аналіз з отриманням різного виду моделей. Найчастіше такі проблеми виникають в генетиці, лісовій селекції, лісовідновленні, лісової ентомології та інших галузях лісівничої науки. В цьому випадку необхідно проводити статистичний аналіз результатів спостережень, що залежать від різних одночасно впливаючих факторів, робити вибір цих чинників та оцінку сили їх впливу. Основою вирішення цих питань є вивчення рівноцінності і однорідності дисперсій досліджуваної ознаки та розкладання її на складові, що породжені дією аналізованих факторів.

Деякі з факторів, що змінюються в експерименті чи спостереженні (наприклад, порода дерев, типи лісу, колір жолудів тощо), можуть бути якісними, інші кількісними. Кількісними величинами зазвичай виражають параметри дерев та деревостанів: висота, діаметр, запас деревостану тощо. У той самий час деякі з цих показників можуть виражатися і якісними ознаками, наприклад, дерева великі, дрібні або дерева першої, другої, третьої величини і т. д. Залежно від співвідношення між кількісними та якісними факторами застосовують один із трьох досить близьких за підходами та математичному апарату методів аналізу: регресійний, дисперсійний та коваріаційний. В регресійному аналізі підхід кількісний, в дисперсійному - всі фактори розглядають як якісні; в коваріаційному аналізі - частину факторів вивчають як кількісні, іншу частину – як якісні.

Сказане не означає, що в дисперсійному аналізі факторами не можуть бути кількісні змінні - застосування дисперсійного аналізу до кількісних факторів при обробці лісівничої інформації трапляється повсякденно. Однак висновок про вплив факторів роблять на якісній основі: перевіряють гіпотезу про вплив даного фактора за обраного рівня значущості. ***Якщо вивчають вплив одного змінного фактора, то дисперсійний аналіз називають однофакторним, двох - відповідно двофакторним і т.д.***

Для характеристики основних типів моделей, що зустрічаються в дисперсійному аналізі, розглянемо два реальні приклади.

В першому прикладі припустимо, що у вегетаційних дослідах випробовували вплив двох рівнів радіоактивного забруднення ґрунту Цезієм-137 (Cs^{137}) на ріст та розвиток сіянців сосни звичайної, ялини звичайної, дуба звичайного та берези повислої. Методами дисперсійного аналізу потрібно оцінити, чи впливає порода і рівень забруднення Цезієм - 137 на інтенсивність росту. У другому випадку із сукупності стиглих соснових одновікових деревостанів природного заповідника «Древлянський» відібрано випадковим чином 50 насаджень, для яких визначені середня висота та середній коефіцієнт форми q_2. Потрібно встановити, чи впливає середня висота деревостану на величину q_2.

При подібності в спільній постановці питання, ситуація у прикладах істотно різниться. В першому прикладі фактори є фіксовані, незмінні величини. Вихідні дані для аналізу можна тут розглядати як вибірку з незліченної множини дослідів над конкретною деревною породою з показниками щільності радіоактивного забруднення. В другому - досліджуваний фактор (середня висота) сам є випадковою величиною, а обрані для дослідження насадження можуть розглядатися як випадкова вибірка з усіх соснових деревостанів як даного об'єкта природно-заповідного фонду, так й всієї Житомирської області, ліси якої потрапили в зону радіоактивного забруднення після аварії на Чорнобильській АЕС.

Аналогічно наведеним прикладам, можливі три основні моделі дисперсійного аналізу: з фіксованими (постійними) факторами, з випадковими та зі змішаними, тобто коли частина факторів постійна, а інша – випадкова. Загальна теорія розроблена для пергої моделі. Для двох інших, в складних завданнях не завжди можна знайти прийнятну теоретичну схему. Щоправда, в більшості практичних випадків для всіх трьох моделей можна застосовувати однакові обчислювальні схеми та оцінки. До того ж в багатьох завданнях, що виникають в лісовому господарстві, можна обмежитися першою моделлю. В прикладі по оцінки впливу висоти на q_2, доцільно використовувати моделі з постійними факторами: спочатку зафіксувати деякі значення висот, а потім за ними вибирати насадження.

Сутність дисперсійного аналізу.

Дисперсійний, або варіантний, аналіз являє собою в даний час самостійний та дуже важливий розділ біологічної статистики. Сутність його полягає у визначенні ролі окремих факторів в мінливості тієї чи іншої ознаки.

Справа в тому, що вплив тих чи інших факторів на ознаку, що вивчається (або суми ознак) ніколи не може бути виділено в чистому вигляді. Хоча під час проведення досліджень і намагаються зберегти умови максимально однорідними, все ж таки різні досліди дають дещо неоднакові результати. Пояснюється це тим, що на них

впливають численні випадкові обставини та інші фактори, що змінюються від досліду до досліду і не піддаються контролю. Тимпаче, великий вплив таких додаткових неконтрольованих факторів при проведенні аналізу здійснюється не в експериментальних умовах, а безпосередньо в лісі.

Тому виникає завдання розкладання загальної мінливості ознаки на складові: з одного боку, які визначаються конкретними, дослідними факторами, а з іншого — викликані випадковими, неконтрольованими причинами. Дисперсійний аналіз дозволяє оцінювати значущість впливу окремих чинників, і навіть їх відносну роль загальної мінливості.

Методи дисперсійного аналізу були розроблені англійським математиком і біологом Р. Фішером і застосовувалися спочатку головним чином для аналізу результатів дослідів в рослинництві та тваринництві. Для різних схем дослідів були розроблені відповідні схеми дисперсійного аналізу. Однак надалі виявилася повна можливість використання дисперсійного аналізу як для вивчення біологічного матеріалу, взятого з природи, так і будь-яких експериментальних даних, в тому числі й в лісовому господарстві. Тому опис дисперсійного аналізу наводиться у всіх підручниках з біометрії. Найбільш повними є роботи К. Є. Нікітіна та А. З. Швиденко, П. Ф. Рокицького, Н. М. Свалова, на які ми посилаємося у викладі цього розділу.

Загальні причини. Уявімо, що ми аналізуємо відхилення деякої випадкової величини (або кількох випадкових величин) від середньої арифметичної. Досліджуваним об'єктом нехай буде сукупність дерев сосни звичайної на закладеній пробній площі 1,0 га. При цьому вважаємо, що відхилення від середнього значення ($\overline{X}$) до певної міри пов'язано з дією на дану величину якогось певного фактора, наприклад, географічного, тобто його вплив може бути виражено в належності до певних лісорослинних умов. Тоді:

$$x - \overline{X} = A + C$$

де: $\overline{X}$ – середня арифметична популяції;

x – конкретне значення змінної (варіанта);

A - частка відхилення змінної, пов'язана з впливом даного конкретного фактора;

C - залишкова частина відхилення, не зрозуміла впливом даного фактора. Це суміш усіх неконтрольованих та невизначених факторів, інакше кажучи, є результатом випадкових відхилень.

Дуже важливо, що у фактичному відхиленні варіанти (змінної) від середньої фігурують 2 компоненти:

а) та частина відхилення, яка залежить саме від цього чинника, тобто за нашою символікою – A;

б) залишкова частина, яка не залежить від даного фактора – С.

В такому разі можна порівняти значення А та С.

При достовірному впливі фактора, що вивчається, значення А буде в достатній мірі перевищувати значення С. За ступенем перевищення А над С можна зробити висновок про те, наскільки достовірний вплив чинника А.

Наведену загальну схему, що відноситься до окремого відхилення, можна перенести на варіацію багатьох варіантів, тобто виразити ступінь варіації дисперсій через величину:

$$\sigma_0^2 = \sigma_A^2 + \sigma_C^2$$

Тобто загальна дисперсія дорівнює сумі 2-х дисперсій, а саме визначається варіацією фактора А, та дисперсії, що визначається іншими, неконтрольованими (випадковими) причинами – С.

Більш складний випадок - відхилення змінної х від середньої арифметичної популяції $\overline{X}$ під впливом 2-х причин: впливу факторів А та В. Наприклад, фактором А для наших деревостанів сосни звичайної може бути географічний район, тобто вплив території, а чинником В — клас зростання.

Тоді:

$$x - \overline{X} = A + B + AB + C$$

Тут: А - частка відхилення, пов'язана із впливом фактора А;

В - частка відхилення, пов'язана з впливом фактора В;

АВ - частка відхилення, пов'язана з впливом не окремих факторів А та В, а їх взаємодії;

С - залишкова, випадкова частина відхилення.

У значеннях дисперсій, загальна дисперсія σ_0^2 може бути представлена як:

$$\sigma_0^2 = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_e^2$$

За великої кількості факторів, схему можна ускладнювати й надалі. Так, при 3 факторах:

$$x - \overline{X} = A + B + C + AB + BC + AC + ABC + C$$

А, В, С-головні фактори;

АВ, ВС та АС - взаємодія першого порядку;

АВС-взаємодія другого порядку.

Аналогічно можна виразити мінливість в дисперсіях (σ^2) та середньоквадратичних відхиленнях - σ.

Неважко помітити, що сказане вище безпосередньо пов'язане з викладеним в розділі, коли ми описували регресію.

Градації факторів та їх характер. *Зазвичай кожен досліджуваний в експерименті фактор А має не одне, а кілька*

значень, які називають градаціями або рівнями фактора А. В межах кожного рівня, окремі змінні (варіанти) набувають різних значень, тобто спостерігається випадкова варіація. Те саме стосується і в більш складних випадків, коли в загальній мінливості бере участь кілька факторів, кожен з яких може мати окремі рівні. Проводячи дисперсійний аналіз впливу різних факторів слід мати на увазі різний характер рівнів факторів. В одних випадках, ці рівні практично точно встановлені. Наприклад, вивчаючи вплив сезонів року, виділяють окремо зиму, весну, літо, осінь. Зовнішні умови цих сезонів року суворо фіксовано. З іншого боку, можуть бути такі фактори, рівні яких не є точно фіксованими або які мають взагалі всі можливі випадкові градації. Так, наприклад, серед факторів, що впливають на плодоношення дуба звичайного є вік насадження, його повнота, умови погоди (особливо пізні весняні заморозки) та багато іншого. Але кожній з цих ознак властива своя варіація. ***Такі фактори називають випадковими, розуміючи під цим лише те, що випадковими можуть бути різні рівні.*** Втім, треба мати на увазі, що випадкові рівні деяких із них, також можна зробити фіксованими. Звідси випливає, що можливі дуже різні схеми чи моделі дисперсійного аналізу. Вони можуть відрізнятися за кількістю аналізованих факторів (одно-, дво-, трифакторні і т. д.), за характером градацій в середині факторів: з фіксованими факторами, з випадковими чи змішані схеми.

Є так звані ієрархічні моделі, які широко використовуються в лісовому господарстві. В цьому випадку рівні одного фактора не розташовуються випадково серед рівнів інших факторів, але пов'язані з ними ієрархічно. Так в лісівництві, вивчаючи ліси ми виділяємо намет, види, підвиди деревних порід. Але в межах деяких географічних районів дерева навіть одного виду можуть мати різний приріст залежно від кількості тепла, опадів та грунтів. Детальні схеми будуть розглянуті нижче.

За наявності єдиних загальних принципів, конкретні методи дисперсійного аналізу залежатимуть від того, з якою схемою розташування матеріалу доводиться мати справу. Таким чином весь матеріал, що вивчається, може бути розбитий на ряд груп, що різняться як за окремими чинниками, так і їх градаціям. Вивчення методами дисперсійного аналізу варіації в середині цих груп, між групами та нарешті варіації всього матеріалу в цілому, дає можливість встановити, чи впливають ці фактори на мінливість чи ні і які з них мають більшу вагу загальної мінливості.

Нульова гіпотеза. Як і в інших випадках статистичного аналізу, при дисперсійний аналіз слід виходити з спочатку прийнятої нульової гіпотези, а саме: що даний фактор А (або В, або С і т.д.) не впливає. Якщо правильна нульова гіпотеза, σ_A^2 повинна дорівнювати нулю (то це відноситься й до σ_B^2; σ_c^2 і т.д.), тобто вся варіація зводиться тільки до випадкової. Для того щоб відкинути нульову

гіпотезу, потрібно довести, що σ_A^2 достовірна (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, як це зазвичай роблять по відношенню до будь-якого статистичного показника, тобто шляхом поділу його на його помилку:

$$\sigma_A = \frac{A}{m_A}$$

де: А вказує на рівень достовірності.

Найпростіша схема варіювання за відмінності по одному фактору. Для того, щоб розуміти сенс розрахунків при дисперсійному аналізі, дуже важливо з самого початку ясно уявляти можливу варіацію у тих групах, куди розбивається фактичний матеріал. Розберемо найпростішу схему, коли аналізується вплив лише одного фактора, що може приймати різні межі, чи кількісні рівні:

1, 2, ….., i, ……. , a

Окремі спостереження (варіанти) розбиваються на групи згідно цим градаціям фактора, що вивчається в досліді або при спостереженнях в природі. Важливо, що фактор, який вивчається, тільки один, наприклад: вид добрива, що вноситься в розсаднику, або приналежність до різних деревних видів, або вплив радіоактивного забруднення, або роль способів обробітку ґрунту і т. д. За наявності двох або кількох факторів потрібно більш складні схеми.

Розподіл варіант при відмінності по одному фактору представлено в таблиці 17.1.

Таблиця 17.1

Схема варіювання за відмінності группо одному фактору

Групи по одному фактору	Окремі варіанти (спостереження) x_{ij}							Суми по групам	Середнє по групам $\bar{x}_i$
1	X_{11}	X_{12}	X_{13}	…	j	…	n	$\sum x_1 = T_1$	$\bar{x}_1$
2	X_{21}	X_{22}	X_{23}	:	X_{2j}	:	X_{2n}	$\sum x_2 = T_2$	$\bar{x}_2$
:	:	:	:	:	:	:	:	:	:
i	X_{i1}	X_{i2}	X_{i3}	:	X_{ij}	:	X_{in}	$\sum x_i = T_i$	$\bar{x}_i$
a	X_{a1}	X_{a2}	X_{a3}	:	X_{aj}	:	X_{an}	$\sum x_a = T_a$	$\bar{x}_a$
								$\sum xij = T$	$\bar{x}$

Число спостережень (варіант) в кожній групі n, але не обов'язково рівне числу в групах. При нерівному числі можна виходити із середнього числа n_i.

$$N = an\ (=an_i)$$

Зазвичай різні рівні прийнято позначати літерою i, а окремі варіанти (спостереження) - літерою j. Тому кожен варіант, незалежно від того, де він знаходиться, можна позначати в загальному вигляді як Xij. В межах кожного рівня (групи), окремі варіанти набувають випадкових значень: суми варіант за кожною групою (у графі «суми за групами») позначені літерами T_1 T_2,..., T_i..., T_a. Загалом їх можна позначити Ti. Загальна сума всіх варіант Σxji=T. В останній графі наведено середні по групам: x_1, x_2, …, x_i , …, x_a. Загалом групові середні можна позначити через X_i. Загальну ж середню для всіх варіант всіх груп через $\overline{X}$.

Різне варіювання варіант та його характеристика. Після введення всіх цих позначень можна приступити до аналізу варіювання даних, представлених у таблиці 17.1. Можна виділити 3 типи, або напрямки, варіювання:

а) загальне варіювання всіх варіантів (x_{ij}), незалежно від того, в якій групі вони знаходяться навколо загальної середньої x;

б) варіювання групових середніх x_i, або інакше, - середніх кожного рівня даного досліджуваного фактора, навколо загальної середньої x;

в) варіювання варіант X_{ij} всередині кожної групи навколо кожної групової середньої X_i.

Для характеристики цих варіювань під час проведення дисперсійного аналізу використовуються вже відомі з вище описаних розділів цього посібника величин:

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 обчислити суми квадратів. Слово «відхилення» для стислості відкидатимемо. В загальному вигляді вони будуть такими:

1. Загальна сума квадратів:

$$\sum_{ij} (x_{ij} - \overline{x})^2$$

Значок біля знака суми означає, що підсумовування проводиться за всіма варіантами всіх груп.

2. Сума квадратів для групових середніх:

$$\sum_{i} n_i (\overline{x}_i - \overline{x})^2$$

Щоб ця величина була того ж порядку, що і перша, введено множник n_i, тобто середнє число варіант в кожній групі. Якщо кількість варіантів всіх групах однакове, то просто n.

3. Сума квадратів відхилень варіант від групових середніх всередині кожної групи, інакше кажучи, для випадкової варіації усередині груп:

$$\sum_i \left[\sum_j (x_{ij} - \bar{x}_i)^2 \right]$$

Два знаки сум вказують, що підсумовування проводиться двічі: всередині кожної групи, тобто по окремих j (від 1 до n), а потім по всіх рівнях i – від 1 до a.

Ступені свободи. Щоб вирахувати середні квадрати (дисперсії), треба розділити кожну суму квадратів на відповідні числа степенів свободи, які будуть наступними:

1) для загальної дисперсії^

$$df=N - 1 \quad (N=an);$$

2) для дисперсії групових середніх:

$$df= a - 1$$

3) для випадкового варіювання варіант всередині групи:

$$df = (n - 1)\, a = na - a = N - a.$$

Неважко помітити, що сума чисел ступенів свободи для групових середніх і для варіації в всередині груп повинні дорівнювати числу ступенів свободи для загальної дисперсії:

$$(N - a) + (a - 1) = N - 1$$

Загальна схема дисперсійного аналізу при одному факторі.

Загальна схема дисперсійного аналізу наведена в таблиці 17.2. З таблиці 17.2 видно, що загальна варіація розкладається на 2 компонента: один з них - це варіація групових середніх (по градаціям фактора А) навколо загальної середньої x; інший - варіація окремих варіант всередині групи. Останню варіацію можна вважати як випадкову в тому сенсі, що вона створюється багатьма неконтрольованими факторами, крім враховуючого фактора А. При розподілі сум квадратів, зазначених як ss, на число ступенів свободи, отримуємо середні квадрати (дисперсії) - ms, що визначають сумарну варіацію її компонента.

Таблиця 17.2

Схема дисперсійного аналізу (аналізу дисперсії) при одному факторі

Джерело варіювання	Сума квадратів ss	Число ступенів свободи df	Середній квадрат ms
загальне (всі варіанти)	$\sum_{ij}(x_{ij}-\bar{x})^2$	N-1	$\frac{1}{N-1}\sum_{ij}(x_{ij}-\bar{x})^2$
групові середні (фактор А)	$\sum_{i} n_i(\bar{x}_i-\bar{x})^2$	a-1	$\frac{1}{a-1}\sum_{i} n_i(\bar{x}_i-\bar{x})^2$
варіанти в середині груп (випадкові відхилення)	$\sum_{i}\left[\sum_{j}(x_{ij}-\bar{x}_i)^2\right]$	N-a	$\frac{1}{N-a}\sum_{i}\left[\sum_{j}(x_{ij}-\bar{x}_i)^2\right]$

Надалі ми побачимо, що весь цей аналіз знадобиться для того, щоб порівняти два середні квадрати — другий і третій, користуючись критерієм:

$$F=\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$$

Робочі формули обчислення сум квадратів. Обчислення сум квадратів відхилень безпосередньо за вихідними даними цілком можливо, але потребує багато праці. Його рідко використовують навіть при комп'ютерній обробці. Тому краще скористатися робочими формулами, заснованими на одній із формул для суми квадратів відхилень, а саме тієї, де сума квадратів відхилень обчислюється за значеннями варіант:

$$\sum x_i^2-\frac{(\sum x_i)^2}{n}$$

Другий член є хіба що поправкою до першого. В літературі він позначається літерою С, тобто:

$$C=\left(\sum x_i\right)^2/n\,.$$

Якщо далі використовувати наведені вище позначення ΣX_i для кожної групи (рівня фактора А) через T_i (T_1, T_2, . . T_a), суми всіх варіант - Т, число спостережень в кожній групі позначати n_i, загальне число варіант - N, то робочі формули виглядатимуть досить просто:

1) загальна сума квадратів:

$$\sum_{ij} x_{ij}^2 - \frac{T^2}{N}$$

2) сума квадратів для групових середніх:

$$\sum_{i} \frac{T_i^2}{n_i} - \frac{T^2}{N}$$

3) сума квадратів для варіант всередині групи, тобто для випадкових відхилень:

$$\sum_{ij} x_{ij}^2 - \sum_{i} \frac{T_i^2}{n_i}$$

Практично зовсім не обов'язково обчислювати всі 3 суми квадратів, достатньо обчислити лише 2, наприклад, першу та другу. Третя може бути отримана шляхом віднімання другої з першої.

При розподілі сум квадратів на числа ступенів свободи отримують середні квадрати (варіанси). Таким чином, робочі формули для них будуть наступними:

1) для загального варіювання:

$$ms = \sigma^2 = \frac{1}{N-1}\left(\sum_{ij} x_{ij}^2 - \frac{T^2}{N}\right)$$

2) для групових середніх:

$$ms = \sigma^2 = \frac{1}{a-1}\left(\sum_{i} \frac{T_i^2}{n_i} - \frac{T^2}{N}\right)$$

3) для випадкових відхилень:

$$ms = \sigma^2 = \frac{1}{N-a}\left(\sum_{ij} x_{ij}^2 - \sum_{i} \frac{T_i^2}{n_i}\right)$$

Нижче на прикладах однофакторної та двофакторної моделі, розглянуто основні методи дисперсійного аналізу та ілюструючи їх схеми обчислення.

17.2 Однофакторний дисперсійний аналіз.

Для розуміння суті однофакторного дисперсійного аналізу припустимо, що вивчається залежність випадкової величини (x) від мінливого фактора A, градації (або рівні) якого позначені A_i. Тоді кожне спостереження можна позначити через Xij, де i - рівень фактора A, j –номер спостереження. Вихідні дані зручно подати у вигляді таблиці 17.3.

Таблиця 17.3

Вихідні дані для однофакторного дисперсійного аналізу

Рівні фактора A_i	Результати вимірювань	Середнє по факторам
A_1	$x_{11}\ x_{12}\ \ldots\ x_{1j}\ \ldots x_{1m1}$	$\tilde{X}_1$
A_2	$x_{21}\ x_{22}\ \ldots\ x_{2j}\ \ldots x_{2m1}$	$\tilde{X}_2$
…	. . . . .	. . .
A_i	$x_{i1}\ x_{i2}\ \ldots\ x_{ij}\ \ldots x_{im1}$	$\tilde{X}_i$
…	. . . . .	. . .
A_k	$x_{k1}\ x_{k2}\ \ldots\ x_{kj}\ \ldots x_{km1}$	$\tilde{X}_k$

В таблиці 17.3 до рядків (або груп) - за кількістю рівнів фактора A, в кожній групі m_i спостережень (кількість спостережень неоднакове); при рівному числу спостережень підхід не змінюється, але наведені нижче формули дещо спрощуються, тому що всі m_i рівні m . Для підготовки даних до аналізу утворюємо суми квадратів, де знаком S_2 позначені дисперсії:

$$s_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{m_i}(x_{ij} - \tilde{X})^2$$

- загальну суму квадратів всіх спостережень від загального середнього $\tilde{X}$

$$s_M^2 = \sum_{i=1}^{k} m_i(\tilde{X}_i - \tilde{X})^2$$

- суму квадратів відхилення групових середніх $\tilde{X}$ т а від загального середньо $\tilde{X}$ визначеної через число спостережень по групах:

$$s_b^2 = \sum_{i=1}^{k}\sum_{j=1}^{m_i}(x_{ij} - \tilde{X})^2$$

- суму квадратів відхилень всередині групи (від групових середніх).

Прості перетворення дозволяють розкласти першу суму на дві інші, що аналогічно:

$$s_0^2 = s_M^2 + s_b^2$$

Ця вже відома нам формула є основою дисперсійного аналізу. Розглянемо оцінки дисперсій, пов'язаних із запровадженими сумами.

Сума S_0^2 пов'язана з оцінкою загальної дисперсії ознаки, що вивчається, якщо її поділити на число ступенів свободи $n-1=k\sum_i (m_i - 1)$.

По сумі S_M^2 можна оцінити дисперсію між рівнями факторів A_i - міжгрупову дисперсію Число ступенів свободи k - 1. Нарешті S_b^2, дозволяє оцінити дисперсію всередині груп (або залишкову). Оскільки оцінка дисперсії кожної з груп пов'язана з m_i- 1 ступенем свободи, то загальна кількість ступенів свободи

$$k\sum (m_i - 1) = N-k,$$

де: N - число спостережень.

Подальший аналіз залежить від типу моделі. Для моделі з фіксованими факторами, відповідь на основне питання дисперсійного аналізу зводиться до перевірки гіпотези $H_0 : \overline{X}_1 = \overline{X}_2 = \ldots = \overline{X}_k$; тобто, що всі групові середні не залежать від впливу фактора А. Тоді, якщо вірна H_0, міжгрупова дисперсія (в генеральній сукупності) повинна дорівнювати внутрішньогруповій, тобто сформульована гіпотеза може бути замінена еквівалентною $H_0 : \sigma_M^2 = \sigma_b^2$

Припустимо, що X_{ij} – незалежні спостереження над випадковою величиною X, розподіленою нормально, але із середнім X і дисперсією σ^2. Тоді відношення

$$F(k, n-k) = \frac{s_M^2 / (k-1)}{s_b^2 / (n-k)}$$

використовують як статистичної характеристики критерію, якщо обчислене значення F менше табличного при рівні значущості а, то гіпотезу про відсутність впливу фактора А не відхиляють. Якщо ж фактори випадкові, то перевірка гіпотези про рівність групових середніх виявляє невеликий інтерес (рівні фактора А самі випадкові величини) і перевіряють гіпотезу про те, що міжгрупова дисперсія в генеральній 2 сукупності дорівнює нулю: $H_0 : \sigma_M^2 = 0.$

В статистичній теорії показано, що як статистичну характеристику критерія, можна застосовувати величину $F(k, n-k)$. Надалі ми не обговорюватимемо відмінності між моделями з випадковими та фіксованими факторами. Повна схема однофакторного дисперсійного аналізу наведена в таблиці 17.4.

Таблиця 17.4

Повна схема однофакторного дисперсійного аналізу

Тип дисперсії	Число ступенів свободи	Сума квадратів	Оцінка дисперсії	Гіпотеза при факторах		Статистичний критерій
				фіксованих	випадкових	
міжфакторна	k-1	$s_M^2 = \sum_{i=1}^{k} m_i(\tilde{X}_i - \tilde{X})^2$	$\sigma_M^2/(k-1)$	$\bar{X}_1 = \bar{X}_2 = \ldots = \bar{X}_k$,	$\sigma_M^2=0$	$\frac{s_M^2/(k-1)}{s_b^2/(n-k)}$
всередині факторів	$k\sum(m_i-1) = n-k$	$s_b^2 = \sum_i \sum_j (x_{ij} - \tilde{X}_i)^2$	$\sigma_b^2/(n-k)$	-	-	-
загальна	n-1	$s_o^2 = \sum_i \sum_j (x_{ij} - \tilde{X})^2$	-	-	-	-

Вплив чинника можна оцінити іншим шляхом. Тоді кореляційне відношення $\eta^2_{yx} = \sigma^2_M / \sigma^2_o$ та гіпотеза про рівність групових середніх в генеральній сукупності зводиться до виду: $H_0 : \eta^2 = 0$ при альтернативній $H_a : \eta^2 > 0.$ Розмір η^2 пов'язані з F простим співвідношенням і підпорядковується β-розподілу зі ступенями свободи k=1. Є таблиці критичних значень кореляційного відношення η^2 при a = 0,05 (додаток Т); якщо фактичне значення η^2 більше табличного, то гіпотеза про рівність групових середніх відхиляється. Використання η^2 як оцінку впливу іноді досить зручно. Для розрахунку потужності F-критерію (якщо H_0 не відхилена), необхідно припустити, що вірна альтернативна гіпотеза $H_a : \eta^2 > 0.$. Тоді F-критерій підпорядковується нецентральному F-розподілу, що залежить від числа ступенів свободи k_a і k_b та параметра нецентральності:

$$\delta^2 = n\eta^2$$

Зазвичай під час розрахунку потужності, тобто при роботі з нецентральним F-розподілом, користуються графіками, побудованими аналогічно графіку F-розподілу. На рисунку 17.1 як приклад наведено графік потужності F-критерію при a = 0,05 для числа ступенів свободи чисельника k_i =7; при k_1=1 можна скористатися графіком потужності одностороннього t-критерія. Для інших, графіки потужності наведені в спеціальних виданнях.

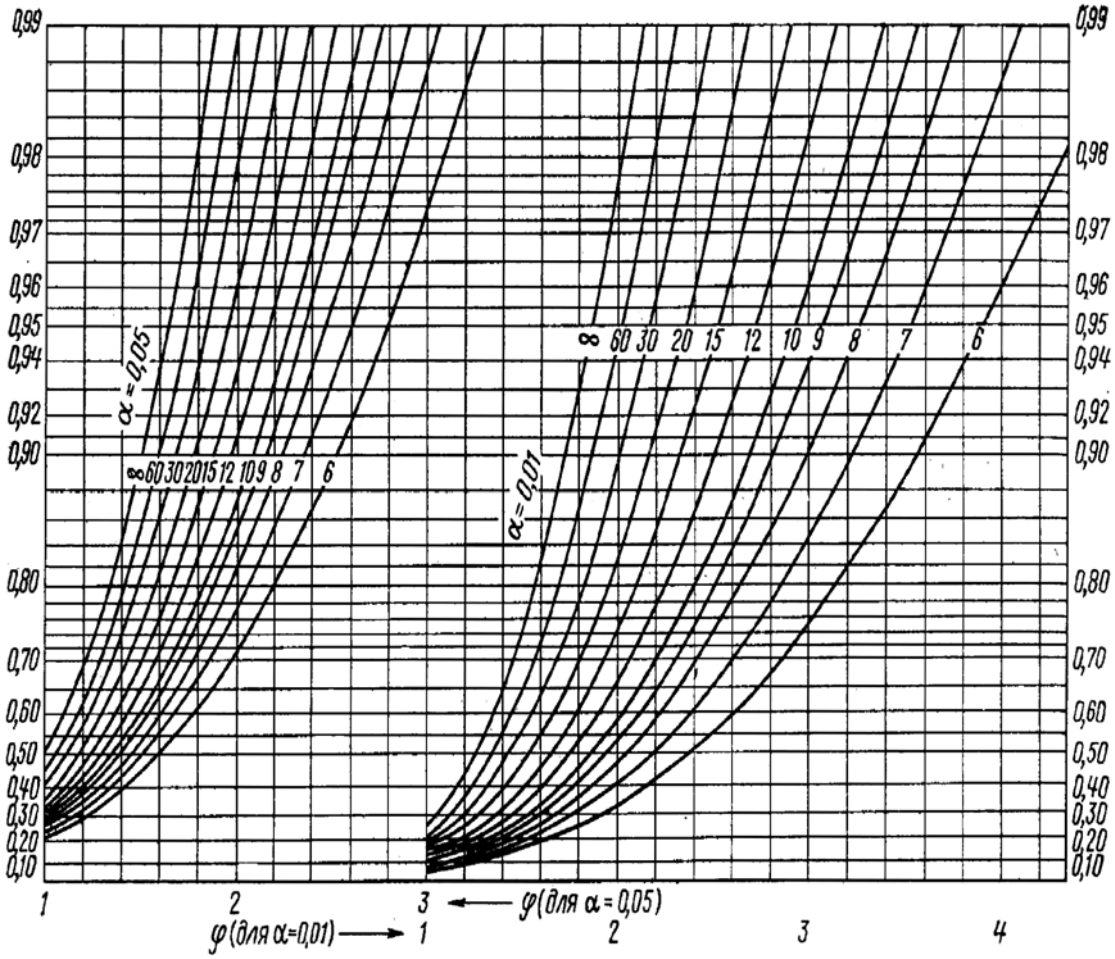


Рис. 17.1 Графік функції потужності F-критерію для k=1 (за К. Є. Нікітіном та А. З. Швиденко)

Як статистична характеристика на рис. 17.1 використана величина:

$$\varphi = \sqrt{n\eta_a^2 / k_1}$$

де: η_a^2 – альтернативне кореляційне відношення.

При розрахунках потужності F-критерія слід мати на увазі, що альтернативна гіпотеза $H_a : \eta^2 > 0$ складна й для розрахунків необхідно вибрати деяке конкретне значення $\eta|_a^2$, у якому тіснота зв'язку в генеральній сукупності має в межах розв'язуваного завдання практичне значення, тобто перевірити H_0 проти простої

$$H_a : \eta^2 = \eta_a^2.$$

Техніка розрахунку потужності F-критерію - в таблиці 17.4. Зауважимо, що за допомогою φ визначають η^2, яке може бути ще визнано суттєвим за даними $k_м$, k_b, та $1 - \beta$.

$$\eta^2 = \varphi^2 k_a / n$$

а також кількість спостережень, що забезпечує визнання суттєвості наявності зв'язку з даними η^2 та при заданих а й β.

Зазвичай в дисперсійному аналізі реальних завдань мало відкинути гіпотезу. Але якщо визнається, що досліджуваний фактор впливає на ознаку, яка вивчається, то для з'ясування співвідношення між груповими середніми можна застосувати t-критерій Стьюдента для попарного порівняння x_i.

Певний інтерес представляє метод множинного порівняння, що дозволяє оцінити порівняння будь-яких поєднань групових середніх. Припустимо, наприклад, що в таблиці 17.4 вихідні дані представляють результати вимірювань для складання деяких лісотаксаційних таблиць та необхідно вибрати доцільне поєднання рівнів, тобто вирішити питання про складання двох окремих таблиць для A_1+A_2 та $A_з+...+A_к$ або трьох таблиць $A_1 + A_2 + A_3$, $A_4 + A_5 + . . , + A_к$ і т.д. Застосовують два методи множини порівняння: S-метод Шеффе і Т-метод Тьюкі. Розглянемо застосування першого з них (аналіз однофакторний). Нехай нас цікавить множинне порівняння:

$$\varphi = c_1 \bar{x}_1 + c_2 \bar{x}_2 + \ldots c_k \bar{x}_k, \quad \sum c_i = 0$$

Для якого не зміщеною оцінкою є:

$$\hat{\varphi} = c_1 \tilde{X}_1 + c_2 \tilde{X}_2 + \ldots c_k \tilde{X}_k,$$

Для статистики $\hat{\varphi}$ оцінку дисперсії обчислюють за формулою:

$$\hat{\sigma}_\varphi^2 = \frac{s_b^2}{k \sum (m_i - 1)} \sum_i \frac{c_i^2}{m_i^2}$$

$$\hat{\sigma}_\varphi^2 = \frac{s_b^2}{n - k} \sum_i \frac{c_i^2}{m_i^2}$$

де: σ_b^2 – сума квадратів внутрішньо групових відхилень.

Тоді для порівняння φ довірчий інтервал при довіряючій ймовірності 1 - а має вигляд:

$$\hat{\varphi} - s\,\hat{\sigma}_{\varphi} \le \varphi \le \hat{\varphi} + s\,\hat{\sigma}_{\varphi}$$

де постійна s визначається за формулою:

$$s^2 = (k-1)\,F_a(k-1;\ n-k)$$

де: $F_a(k-1;\ n-k)$ - квантиль F – розподілу з числом ступенів свободи (k-1), (n-k).

Як приклад, проведемо дисперсійний аналіз для визначення впливу середньої висоти (H) на видове число (f). Вихідні дані для розрахунку показано в таблиці 17.5.

Таблиця 17.5

Вихідні дані для дисперсійного аналізу впливу середньої висоти на видове число

H_i, м	Видові числа f_{ij}*1000	$\sum f_{ij}$	n_i	$\tilde{f}_i$
1	2	3	4	5
22	455, 436, 466, 467, 446, 483	2753	6	458,8
24	467, 446, 502, 448, 429	2292	5	458,4
26	465,466,417, 510, 480	2238	5	467,6
28	502, 489, 442, 530, 467, 501	2931	6	488,5
30	452, 467, 456, 433, 467, 456	2731	6	455,2
32	503,483,458, 451,469	2364	5	472,8
34	446, 427, 430	1303	3	434,3
36	468, 434, 407, 370	1679	4	419,8
		Σ218391	Σ40	460

Тут вивчається вплив середньої висоти деревостану на величину середнього видового числа умовно одновікових стиглих соснових насаджень. При розрахунках на комп'ютерах суми s^2_0, $s^2_м$ і $s^2_в$ зручніше обчислювати за формулами:

$$s_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{m_i} x_{ij}^2 - n\tilde{X}^2$$

$$s_M^2 = \sum_{i=1}^{k} m_i \tilde{X}_i^2 - n\tilde{X}^2$$

$$s_B^2 = \sum_{i=1}^{k}\sum_{j=1}^{m_i} x_{ij}^2 - \sum_{i=1}^{k} m_i \tilde{X}_i^2$$

а замість вихідних даних, використовувати їх відхилення від деякого початкового значення наприклад, від загального середнього X, що спрощує розрахунки.

Групові середні наведені в колонці 5. Число груп k=8, загальне число спостережень n = 40. Загальне середнє $\tilde{f}$=460. Перейдемо до відхилень від середнього (таблиця 17.6) та обчислимо показники, необхідні для застосування формул.

З таблиці 17.5 s_0^2 = 36195, s_M^2 = 14587, s^2_B = 36195 - 14587 = 21608. Результати обчислень запишемо до таблиці 17.6 враховуючи, що число ступенів свободи для групової дисперсії до k-1=8-1=7, для загальної N -1=40-1=39, а внутрішньогрупової N -k =40-8=32.

Таблиця 17.6

Обчислення сум квадратів в однофакторному дисперсійному аналізі

H_i, м	$f_{ij} - \tilde{f} = x_{ij}$	$\sum x_{ij}$	m_i	$\tilde{X}_i$	$\sum x_{ij}^2$	$\sum m_i \tilde{x}$
22	-5, -24, +6, +7,-14,+23	7	6	-1,2	1411	8,6
24	+7,-14, +42,-12,-31	-8	5	-1,6	3114	12,8
26	+5, +6, -43, +50, +20	+38	5	+7,6	4810	288,8
28	+42,+29,-18,+70, +7,+41	+171	6	+28,5	9559	4873,5
30	8, +7, -4, -4, -27,+7	-29	6	-4,8	923	138,2
32	+43, +23, -2, -9, +9	+64	5	+12,8	2544	819,2
34	14, -33, -30,	-77	3	-25,7	2185	1981,5
36	+8, -26, -53,-90	-161	4	-40,2	11649	6464,2
		-9	40	-0,23	36195	14587

Таблиця 17.7

Підсумки однофакторного дисперсійного аналізу

Тип дисперсії	Сума квадратів	Число ступенів свободи	Оцінка дисперсії
міжгрупова	14587	7	2084
внутрішньогрупова	21608	32	675
загальна	36195	39	928

Статистична характеристика ($F_{розрахункове}$), що отримана, дорівнює $F_{розрах}$ = 2084/675 = 3,09. При a=0,05 табличне значення F взяте з додатку Ж, при v=7 та N - k=32 буде дорівнюватиме 2,3. Оскільки $F_{розрах} > F_{табл.}$, то гіпотезу про відсутності впливу висоти на

середнє видове число деревостану відхиляють: середні f у генеральній сукупності в повному обсязі не всі рівні між собою, а залежать від значення середньої висоти. Для розрахунку методом множинного порівняння припустимо, що необхідно з'ясувати, з яких значень висот можна скласти єдині таблиці, що використовують середні видові числа. Випробуємо, наприклад, можливість об'єднання висот 22, 24, 26 в одну групу, решти - в другу. Функція порівняння має вигляд:

$$\hat{\varphi} = \frac{1}{3}(458{,}8+458{,}4+467{,}6) - \frac{1}{5}(488{,}5+455{,}2+472{,}8+434{,}3+419{,}8) = 7{,}5$$

А оцінка дисперсії за рівняннм:

$$\hat{\sigma}_{\varphi}^2 = \frac{s_b^2}{k\sum(m_i - 1)}\sum_i \frac{c_i^2}{m_i^2}$$

$$\hat{\sigma}_{\varphi} = \left\{675\left[\frac{1}{9}\left(\frac{1}{36}+\frac{1}{25}+\frac{1}{25}\right)+\frac{1}{25}\left(\frac{1}{36}+\frac{1}{36}+\frac{1}{25}+\frac{1}{9}+\frac{1}{16}\right)\right]\right\}^{1/2} = 3{,}85$$

Постійну s знаходимо за формулою:

$$s^2=(k\text{-}1)\,F_a(k\text{-}1;\ n\text{-}k)$$

$$s=[(8\text{-}1)\,F_{0,05}\,(7,32)]^{1/2} = (7\cdot 2{,}3)^{1/2} \approx 4$$

тобто для ймовірності 0,95 маємо довірчий інтервал 7,5–4*3,85≤φ≤7,5+4*3,85 або -7,9≤φ≤22,9.

Так як довіряючий інтервал містить нуль, немає підстав об'єднувати матеріал у зазначені групи в залежності від значення висоти. В цьому прикладі змінено постановку завдання: замість однофакторного застосовано двофакторний аналіз (включено додатково середній діаметр), після чого вдалося задовільним чином згрупувати матеріал. Далі використовуємо кореляційне відношення та розрахунок потужності критерію. Перевіримо H_0 трохи інакше. Обчислимо кореляційне відношення η^2, що дорівнює відношенню міжгрупової дисперсії до суми квадратів (таблиця 17.7): η^2 = 14587/36195 = 0,403. Гіпотеза H0:η^2 = 0 рівнозначна гіпотезі H_0: $X_1=X_2\ldots.. = X_к$. Перевіримо H_0 η^2 = 0 при альтернативній Ha:η^2>0.

З додатку Т знаходимо критичні значення η^2 = 0,387 при k_1 = 7, k_2 = 32, тобто гіпотезу про відсутність впливу висоти на видове число за a =0,05 відхиляємо.

17.3 Двофакторний дисперсійний аналіз.

Вище вже зазначалося, що за участі у спільній варіації 2 факторів А та В, аналіз ускладнюється наявністю взаємодії між цими факторами. Тому загальна сума квадратів за двофакторної схеми

розкладається на 4 компоненти: а) варіація під впливом фактора А; б) варіація під впливом фактора В; в) варіація під спільним впливом А й В, тобто взаємодії А та В; г) випадкові відхилення.

Крім того, треба пам'ятати, що за двофакторною схемою кожен рівень одного фактора повинен поєднуватися з будь-яким рівнем другого фактора. Так, якщо вивчаються якісь дані за 3 роки про тварин із 3 різних місць проживання, то необхідно щоб по кожному місцю були дані за всі три роки. Якщо цього немає, потрібно застосовувати іншу схему аналізу. Розподіл варіант при варіюванні за 2 факторами показано в таблиці 17.8. У графах «варіант» вміщено варіанти, у графах «показники» - показники Т та х. Символом r позначається кількість груп (рівнів) за фактором А, тобто кількість горизонтальних рядів (1, 2, 3, ..., 1, ..., r); с –кількість груп (рівнів) за чинником В, тобто кількість вертикальних стовбців, або колонок (1,2,3,...,j, ..., с); n - число спостережень в кожній клітині таблиці. В даному випадку n=3, але не обов'язково, щоб воно було однаковим у всіх клітинах. Все ж таки для простоти розрахунків вигідніше останнє, тоді nrb = N тобто загальному числу всіх спостережень.

Так як варіювання груп за фактором А і фактором В завжди порівнюють з випадковими відхиленнями варіант в межах кожної групи, як мірилом випадкової варіації (σ_e^2), то остання має бути виміряна на достатньому матеріалі. Це означає, що у кожній клітці треба мати щонайменше 2-а спостереження, а ще краще, якщо їх буде більше. Кожний варіанта (спостереження) може бути позначене у загальному вигляді як k_e, тобто як е спостереження в ряду і та вертикальному стовпці j. Конкретна ж варіанта х має 3 значки. Перший позначає номер групи за фактором А, тобто номер горизонтального рядка, другий - номер групи по фактору В, тобто номер вертикального стовбця, третій - номер у цій клітці.

В кожній клітині дано зведені показники: сума варіант клітинки (Т..) та середня арифметична їхня (х..). Значки при них вказують номери горизонтальними рядки та вертикального стовпця. Загалом показники для кожної клітинки T_{ij} та X_{ij}. Показники для горизонтальних рядків, тобто для градацій фактора А, дані праворуч у вертикальних стовпцях: T_1 T_2, . . ., Ті, . . . , T_r і відповідно x_1, x_2. й т.д. В загальному вигляді їх позначатимемо T_i та x_i або просто Ті Хі.

Для вертикальних стовпців (градацій за фактором В) показники представлені у нижній частині таблиці 17.8. Це суми T_1. $T_{.2}$, . . ., T_j, . . . , Тс та середні x_1 x_2, . . ., x_j, . . . , x_c У загальному вигляді вони будуть позначатися як T_j, і x_j.

Загальна сума всіх варіантів всіх клітинок позначається Т, а загальна середня арифметична - х.

Таблиця 17.8

Схема варіювання ознак при різниці груп за 2-ма факторами

Групи по фактору А	Групи по фактору В та окремі спостереження x_{ijk} всередині них												Сума по групам факторів А, T_i	Середнє по групам факторві А, x_i
	1		2		3		…	i		…	c			
	вар.	пок.	вар.	пок.	вар.	пок.	…	вар.	пок.	…	вар.	пок.		
1	x_{111} x_{112} x_{113}	T_{11} $\bar{x}_{11}$	x_{121} x_{122} x_{123}	T_{12} $\bar{x}_{12}$	x_{131} x_{132} x_{133}	T_{13} $\bar{x}_{13}$		x_{1j1} x_{1j2} x_{1j3}	T_{1j} $\bar{x}_{1j}$		x_{1c1} x_{1c2} x_{1c3}	T_{1c} $\bar{x}_{1c}$	$T_{1.}$	$\bar{x}_{1.}$
2	x_{211} x_{212} x_{213}	T_{21} $\bar{x}_{21}$	x_{221} x_{222} x_{223}	T_{22} $\bar{x}_{22}$	x_{231} x_{232} x_{233}	T_{23} $\bar{x}_{23}$		x_{2j1} x_{2j2} x_{2j3}	T_{2j} $\bar{x}_{2j}$		x_{2c1} x_{2c2} x_{2c3}	T_{2c} $\bar{x}_{2c}$	$T_{2.}$	$\bar{x}_{2.}$
3	x_{311} x_{312} x_{313}	T_{31} $\bar{x}_{31}$	x_{321} x_{322} x_{323}	T_{32} $\bar{x}_{32}$	x_{331} x_{332} x_{333}	T_{33} $\bar{x}_{33}$		x_{3j1} x_{3j2} x_{3j3}	T_{3j} $\bar{x}_{3j}$		x_{3c1} x_{3c2} x_{3c3}	T_{3c} $\bar{x}_{3c}$	$T_{3.}$	$\bar{x}_{3.}$
...														
i	x_{i11} x_{i12} x_{i13}	T_{i1} $\bar{x}_{i1}$	x_{i21} x_{i22} x_{i23}	T_{i2} $\bar{x}_{i2}$	x_{i31} x_{i32} x_{i33}	T_{i3} $\bar{x}_{i3}$		x_{ij1} x_{ij2} x_{ij3}	T_{ij} $\bar{x}_{ij}$		x_{ic1} x_{ic2} x_{ic3}	T_{ic} $\bar{x}_{ic}$	$T_{i.}$	$\bar{x}_{i.}$
…														
r	x_{r11} x_{r12} x_{r13}	T_{r1} $\bar{x}_{r1}$	x_{r21} x_{r22} x_{r23}	T_{r2} $\bar{x}_{r2}$	x_{r31} x_{r32} x_{r33}	T_{r3} $\bar{x}_{r3}$		x_{rj1} x_{rj2} x_{rj3}	T_{rj} $\bar{x}_{rj}$		x_{rc1} x_{rc2} x_{rc3}	T_{rc} $\bar{x}_{rc}$	$T_{r.}$	$\bar{x}_{r.}$
Суми по групам фактора В Tj		$T_{.1}$		$T_{.2}$		$T_{.3}$			$T_{.j}$			$T_{.c}$	T	—
Середнє по групам фактора В x_j		$\bar{x}_{.1}$		$\bar{x}_{.2}$		$\bar{x}_{.3}$			$\bar{x}_{.j}$			$\bar{x}_{.c}$	—	$\bar{x}$

Обчислення сум квадратів та середніх квадратів. Після введення всіх цих позначеннях можна перейти до побудови загальних формул сум квадратів, необхідних для проведення дисперсійного аналізу при 2 –х факторах. Вони будуть такими:

Загальна сума квадратів:

$$\sum_{ijk} (x_{ijk} - \bar{x})^2$$

тобто проста сума квадратів відхилень всіх спостережень від загальної середньої арифметичної.

Сума квадратів для варіювання за фактором А:

$$nc \sum_{i} (\bar{x}_i - \bar{x})^2$$

тобто помножена на nc сума квадратів відхилень всіх значень від загальної середньої арифметичної.

Сума квадратів для варіювання за фактором В:

$$nr \sum_{j} (\bar{x}_j - \bar{x})^2$$

тобто помножена на nr сума квадратів відхилень всіх значень xj від загальної середньої арифметичної.

Сума квадратів для взаємодії А та В:

$$n \sum_{ij} \left[\bar{x}_{ij} - (\bar{x}_i - \bar{x}) - (\bar{x}_j - \bar{x}) - \bar{x} \right]^2 = n \sum_{ij} \left[\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x} \right]^2$$

Зрештою, сума квадратів для випадкових відхилень:

$$\sum_{ijk} (x_{ijk} - \bar{x}_{ij})^2$$

тобто сума квадратів відхилень варіантів від середніх окремих клітинок таблиці.

Числа ступенів свободи df такі: для загального варіювання rcn-1, для варіювання за фактором А r-1, для варіювання за фактором В c-1, для взаємодії А та В (r-1)(c-1), для випадкових відхилень rc (n-1).

Загальна схема розкладання варіації при двофакторній дисперсійній схемі аналізу наведена в табл. 17.9. У ній дано і загальні формули для середніх факторів, що отримують шляхом поділу сум квадратів на число ступенів свободи.

Робочі формули при двофакторному аналізі. Для спрощення розрахунків краще застосувати робочі формули для сум квадратів, а саме:

Для загального варіювання:

$$\sum_{ijk} x_{ijk}^2 - \frac{T^2}{nrc}$$

Для варіювання по фактору А:

$$\frac{1}{nc}\sum_{i} T_i^2 - \frac{T^2}{nrc}$$

Для варіювання по фактору В:

$$\frac{1}{nr}\sum_{j} T_j^2 - \frac{T^2}{nrc}$$

Для варіювання, що характеризує взаємодію А та В:

$$\frac{1}{n}\sum_{ij} T_{ij}^2 - \frac{1}{nc}\sum T_i^2 - \frac{1}{nr}\sum T_j^2 + \frac{T^2}{nrc}$$

Для варіювання випадкових відхилень (всередині всіх груп):

$$\sum_{ijk} x_{ijk}^2 - \frac{1}{n}\sum_{ij} T_{ij}^2$$

В цих формулах:

n - число варіант у кожній клітині;

c - число вертикальних стовпців, тобт. груп за фактором В;

r - число горизонтальних рядків, тобто груп за фактором А.

Величина $\sum x_{ijk}^2$ означає фігуруючу раніше суму квадратів всіх варіант. Далі доводиться мати справу з різними сумами варіант:

T_{ij} - сума варіант по окремих клітинках (як рядів, так і стовпців);

T_i; - сума варіант для і-рядів, тобто рядів за рівнями (групами) фактора А;

T_j - сума варіант для j-стовпців, тобто колонок за рівнями (групами) фактора В;

Т - загальна сума всіх варіантів.

Застосування цих формул для сум квадратів дає можливість користуватися не середніми, що є в таблиці 18., а лише сумами варіант $\sum x_{ijk}^2$ Тому в схемі варіювання таблиці 17.8 можна не записувати середні окремі клітинки, вводячи всі величин одного порядку. Число ступенів свободи для всіх сум квадратів наведено в таблиці 17.9.

Таблиця 17.9

Схема дисперсійного аналізу при 2 факторах досліджень

Джерело варіювання	Сума квадратів ss	Число ступенів свободи df	Середній квадрат ms
загальне	$\sum_{ijk}(\bar{x}_{ijk} - \bar{x})^2$	rcn-1	$\frac{1}{rcn-1}\sum_{ijk}(\bar{x}_{ijk} - \bar{x})^2$
фактор А (групові середні по фактору А)	$nc\sum_{i}(\bar{x}_i - \bar{x})$	r-1	$\frac{nc}{r-1}\sum_{i}(\bar{x}_i - \bar{x})^2$
фактор В (групові середні по фактору В)	$nr\sum_{j}(\bar{x}_j - \bar{x})^2$	c-1	$\frac{nr}{c-1}\sum_{j}(\bar{x}_j - \bar{x})^2$
взаємодія А та В	$n\sum_{ij}(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$	(r-1)(c-1)	$\frac{n}{(r-1)(c-1)}\sum_{ij}(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$
випадкове відхилення	$\sum_{ijk}(\bar{x}_{ijk} - \bar{x}_{ij})^2$	rc(n-1)	$\frac{1}{rc(n-1)}\cdot\sum_{ijk}(\bar{x}_{ijk} - \bar{x}_{ij})^2$

Тому можна записати наступні робочі формули для середніх квадратів, що отримують шляхом поділу сум квадратів на відповідні числа ступенів свободи:

$$ms = \frac{1}{rcn-1}\left(\sum_{ijk} x_{ijk}^2 - \frac{T^2}{nrc}\right)$$

для варіювання по фактору А:

$$ms = \frac{1}{r-1}\left(\frac{1}{nc}\sum_{i} T_i^2 - \frac{T^2}{nrc}\right)$$

для варіюванню по фактору В:

$$ms= \frac{1}{c-1}\left(\frac{1}{nr}\sum_{j} T_j^2 - \frac{T^2}{nrc}\right)$$

для варіювання А та В:

$$ms= \frac{1}{(r-1)(c-1)}\left(\frac{1}{n}\sum_{ij} T_{ij}^2 - \frac{1}{nc}\sum T_i^2 - \frac{1}{nr}\sum T_j^2 + \frac{T^2}{nrc}\right)$$

для випадкового відхилення:

$$ms= \frac{1}{rc(n-1)}\left(\sum_{ijk} x_{ijk}^2 - \frac{1}{n}\sum_{ij} T_{ij}^2\right)$$

17.4 Багатофакторний дисперсійний аналіз. Використання дисперсійного аналізу в лісовому господарстві.

Дисперсійний аналіз за трифакторної схеми. При структурі матеріалу, що відрізняється за 3 факторами, застосовується принципово та ж схема аналізу, як і за відмінностями по за 2 – м факторами, але вона складніша і тому потребує великої уваги під час розрахунків.

Загальна сума квадратів розкладається на 8 компонентів:

1. Ефект фактора А.
2. Ефект фактора В.
3. Ефект фактора С.
4. Взаємодія А та В.
5. Взаємодія А та С.
6. Взаємодія В та С.
7. Взаємодія А, В та С разом (взаємодія другого порядку).
8. Випадкові відхилення.

Кожна окрема варіанта позначається 4 значками, саме x_{ijkl}.

Відповідні середні: x – середня арифметична всіх спостережень;

x_i, x_j і x_k - середні для рівнів за фактором А, фактором В і по фактору С З окремо: x_{ij}, x_{ik} і x_{jk} - середні для всіх рівнів по 2 факторам без урахування третього; x_{ijk} – середні всіх клітинок решітки. Щоб не сплутати літери, можна позначити кількість груп факторів А, В та С однією літерою г зі значками 1, 2, 3. Тоді загальна схема аналізу може бути подана, як у таблиці 17.10. Середній квадрат, як завжди, отримують поділом суми квадратів на кількість ступенів свободи, тому для економії місця його можна не включати до таблиці. Загальна схема аналізу по суті та сама, яка була викладена вище для дисперсійного аналізу за 2-х факторним дослідом. Зокрема, такий же розрахунок сум квадратів та ступенів

свободи для взаємодії з двох факторів. Поряд з урахуванням взаємодії А та В додається облік взаємодій А і С та В і С. Новим є облік взаємодії всіх 3 факторів А, В та С. Користування квадратами відхилень різних середніх від загальної середньої у разі аналізу за 3 факторами ще більше ускладнило б техніку розрахунків, тому і тут для підрахунку сум квадратів доцільно користуватися робочими формулами, у яких фігурують квадрати варіант та суми варіант за групами.

Таблиця 17.10

Схема дисперсійного аналізу при 3 факторному досліді

Джерело варіювання	ss	df
Загальне	$\sum_{ijkl} (x_{ijkl}-\bar{x})^2$	$nr_1r_2r_3-1$
Фактор А	$nr_2r_3 \sum_{i} (\bar{x}_i-\bar{x})^2$	r_2-1
Фактор И	$nr_1r_3 \sum_{j} (\bar{x}_j-\bar{x})^2$	r_2-1
Фактор С	$nr_1r_2 \sum_{k} (\bar{x}_k-\bar{x})^2$	r_3-1
Взаємодія А та В	$nr_3 \sum_{ij} (\bar{x}_{ij}-\bar{x}_i-\bar{x}_j+\bar{x})^2$	$(r_1-1)(r_2-1)$
Взаємодія В та С	$nr_1 \sum_{jk} (\bar{x}_{jk}-\bar{x}_j-\bar{x}_k+\bar{x})^2$	$(r_2-1)(r_3-1)$
Взаємодія А та С	$nr_2 \sum_{ik} (\bar{x}_{ik}-\bar{x}_i-\bar{x}_k+\bar{x})^2$	$(r_1-1)(r_3-1)$
Взаємодія А, В та С	$n\sum_{k} (\bar{x}_{ijk}-\bar{x}_{ij}-\bar{x}_{ik}-\bar{x}_{jk}+\bar{x}_i+\bar{x}_j+\bar{x}_k-\bar{x})^2$	$(r_1-1)(r_2-1)(r_3-1)$
Випадкове відхилення	$\sum_{ijkl} (x_{ijkl}-\bar{x}_{ijk})^2$	$r_1r_2r_3(n-1)$

Вони будуть такими:
загальна мінливість:

$$\sum_{ijkl} x_{ijkl}^2 - \frac{T^2}{nr_1r_2r_3}$$

Ефект А:

$$\frac{1}{nr_2r_3}\sum_{i} T_i^2 - \frac{T^2}{nr_1r_2r_3}$$

Ефект В:

$$\frac{1}{nr_1r_3}\sum_j T_j^2 - \frac{T^2}{nr_1r_2r_3}$$

Ефект C:

$$\frac{1}{nr_1r_2}\sum_k T_k^2 - \frac{T^2}{nr_1r_2r_3}$$

Взаємодія A та B:

$$\frac{1}{nr_3}\sum_{ij} T_{ij}^2 - \frac{1}{nr_2r_3}\sum_i T_i^2 - \frac{1}{nr_1r_3}\sum_i T_j^2 + \frac{T^2}{nr_1r_2r_3}$$

Взаємодія B та C:

$$\frac{1}{nr_1}\sum_{jk} T_{jk}^2 - \frac{1}{nr_1r_3}\sum_j T_j^2 - \frac{1}{nr_1r_2}\sum_k T_k^2 + \frac{T^2}{nr_1r_2r_3}$$

Взаємодія A та C:

$$\frac{1}{nr_2}\sum_{ik} T_{ik}^2 - \frac{1}{nr_2r_3}\sum_i T_i^2 - \frac{1}{nr_1r_2}\sum_k T_k^2 + \frac{T^2}{nr_1r_2r_3}$$

Взаємодія A. B та C:

$$\frac{1}{n}\sum_{ikj} T_{ijk}^2 - \frac{1}{nr_3}\sum_{ij} T_{ij}^2 - \frac{1}{nr_1}\sum_{jk} T_{jk}^2 - \frac{1}{nr_2}\sum_{ik} T_{ik}^2 +$$

$$+\frac{1}{nr_2r_3}\sum_i T_i^2 + \frac{1}{nr_1r_3}\sum_j T_j^2 - \frac{1}{nr_1r_2}\sum_k T_k^2 - \frac{T^2}{nr_1r_2r_3}$$

Випадкове відхилення:

$$\sum_{ijkl} x_{ijkl}^2 - \frac{1}{n}\sum_{ijk} T_{ijk}^2$$

У всіх цих формулах поправка одна й та сама - $\frac{T^2}{nr_1r_2r_3}$, тобто квадрат суми всіх варіантів, поділений на загальну їх кількість. При обчисленні першої частини робочих формул важливо не сплутати, які саме суми треба зводити в квадрат. Щоб не захаращувати текст,

обмежимося тільки цими формулами для суми квадратів в літерній символіці без остаточних формул для середніх квадратів та без наведення конкретних прикладів. Параметри, що оцінюються середніми квадратами, при 3 факторах принципово не відрізняються від зазначених вище для випадку двофакторного дисперсійного аналізу. Вони наведені у таблиці 17.11 для випадку, коли рівні з усіх факторів будуть випадковими. Якщо ж відмінності між рівнями за одним із факторів є не випадковими, а фіксованими (змішана схема), відповідний компонент варіації треба позначати не σ^2, а будь-яким іншим значком, наприклад x^2, як вказувалося при розборі однофакторної схеми, або просто К зі значком, що позначає даний фактор А, В, С і т.д.

Таблиця 17.11

Параметри, що оцінюються,
при трифакторному дисперсійному аналізі

Джерело варіювання	ms	Визначальні параметри
Фактор А	1	$\sigma_e^2+ n\sigma_{ABC}^2+nr_3\sigma_{AB}^2+nr_2\sigma_{AC}^2+nr_2r_3\sigma_A^2$
Фактор И	2	$\sigma_e^2+ n\sigma_{ABC}^2+nr_3\sigma_{AB}^2+nr_1\sigma_{BC}^2+nr_1r_3\sigma_B^2$
Фактор С	3	$\sigma_e^2+ n\sigma_{ABC}^2+nr_2\sigma_{AC}^2+nr_1\sigma_{BC}^2+nr_1r_2\sigma_C^2$
Взаємодія А та В	4	$\sigma_e^2+ n\sigma_{ABC}^2+nr_3\sigma_{AB}^2$
Взаємодія В та С	5	$\sigma_e^2+n\sigma_{ABC}^2+nr_1\sigma_{BC}^2$
Взаємодія А та С	6	$\sigma_e^2+ n\sigma_{ABC}^2+nr_2\sigma_{AC}^2$
Взаємодія А, В та С	7	$\sigma_e^2+n\sigma_{ABC}^2$
Випадкове відхилення	8	σ_e^2

Розбір параметрів, що оцінюються різними середніми квадратами показує, що в даному випадку оцінити за допомогою критерію F достовірність впливу окремих факторів та їх взаємодії значно складніше, ніж у попередніх випадках дисперсійного аналізу. Тому ми обмежимося вище описаним.

Ієрархічна схема дисперсійного аналізу.

Всі попередні схеми були факторними. В них передбачалося, що рівні одного фактора поєднуються з будь-якими рівнями всіх інших факторів. Таким чином створюються групи варіант, на які діють будь-які поєднання всіх досліджуваних факторів. Звичайні факторні схеми найчастіше застосовуються в дослідах, плани яких будується експериментатором заздалегідь. Очевидно, що такий план має передбачати наявність всіх поєднань градацій різних факторів. Однак при аналізі матеріалу, взятого з природи або з лісового

господарства, звичайні факторні схеми можуть бути нездійсненними, тому що всередині градацій (груп) фактора А можливі різні, що відрізняються одина від одної градації (групи) факторів В, С і т.д. Так, наприклад, при вивченні даних про приріст соснових деревостанів, що відбуваються в різних лісорослинних умовах, виявиться певний зв'язок між групами та факторами, що впливають на них (рис. 17.2).

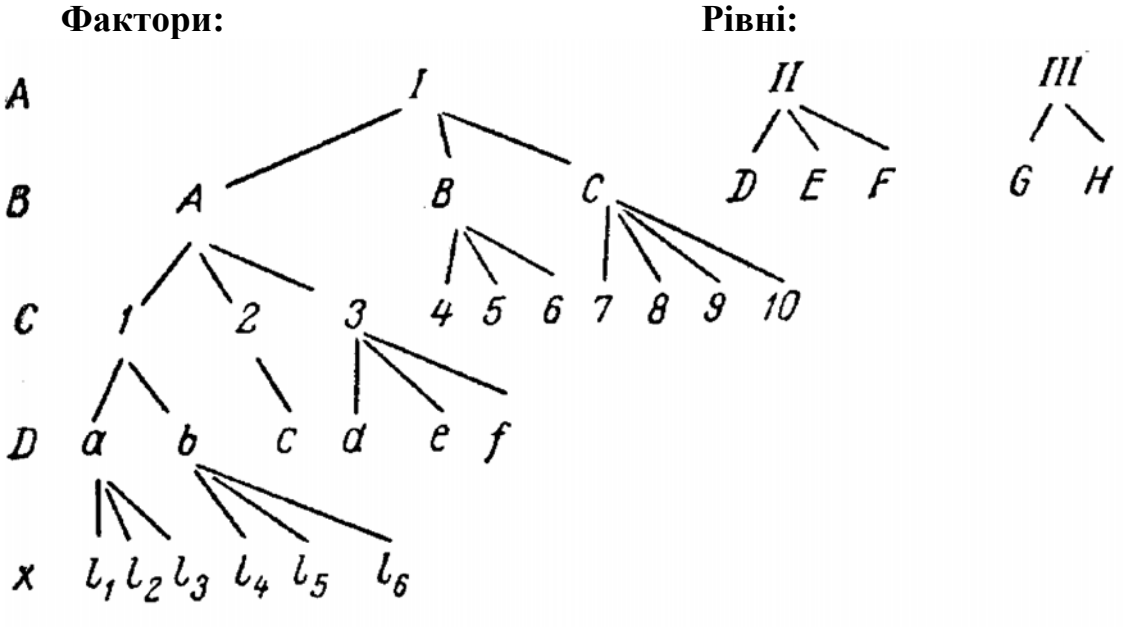


Рис. 17.2 Схема ієрархічних зв'язків між факторами та їх рівнями (за П.Ф. Рокицьким).

Рівні нижчого порядку розташовуються лише всередині певних рівнів вищого порядку.
Фактори: А – деревна порода; В – лісорослинні умови;
С – ранній та пізній приріст; D – пізній приріст; х – варіанти.

Варіюючі за окремими приростами насадження сосни звичайної залежать від 4 факторів: деревної породи, лісорослинних умов, раннього та пізнього приросту, варіантів в розрізі закладеної пробної площі, але зв'язок між ними здійснюється по ієрархічних сходинках - від більш загальних факторів до більш конкретних, або від факторів вищого порядку до факторів нижчого порядку. Тому такі схеми, чи моделі, отримали назву ієрархічних.

Так само може бути згрупований ботанічний матеріал за факторами: види, підвиди, екотипи, місця проживання (поширення), вибірки, окремі екземпляри. Для ієрархічних схем характерна відсутність вільних поєднань між градаціями факторів А, В, С і т.д., у цьому їхня відмінність від розглянутих вище факторних схем. Ієрархічні сходинки можуть бути коротшими або довшими в залежності від кількості факторів, що враховуються.

Застосування дисперсійного аналізу в лісовому господарстві. В лісовому господарстві при проведенні досліджень дисперсійний аналіз застосовується дуже широко. Він є основою при доказі подібності та відмінностей у потомства під час проведення робіт із селекції лісових деревних видів. Ним користуються для визначення ефективності проведених лісогосподарських заходів, зокрема при застосуванні добрив, засобів захисту, навіть під час проведення рубок догляду, хоча в цьому випадку найчастіше використовують регресійний аналіз. Обчислення в даний час проводять тільки на комп'ютерах, за сертифікованими стандартними програм, які є у всіх системах математичного забезпечення для ПК, зокрема «Статистика», «Пакет аналіза» для ліцензованої версії Microsoft Excel.

СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ:

1. Айвазян С. А. Прикладна статистика дослідження лісів. (2000). Київ. Наукова думка 487 с.
2. Алексєєв А. С. Математичні моделі та методи в лісовому господарстві. Навчальний посібник. (1988). Львів. ЛТА. 88 с.
3. Антанайтіс В. В. Закономірності лісової таксації. (2003). Каунас. ЛітСХА. 127 с.
4. Антанайтіс В. В. Сучасний напрямок лісовпорядкування. (2002). Харків. Лісова промисловість. 280 с.
5. Араб-Огли Е. А. Робоча книга з прогнозування. (1982). Харків. Наукова думка. 430 с.
6. Атрощенко О. А. Моделювання зростання лісу та лісогосподарських процесів. (2004). Харків. 249 с.
7. Багінський В. Ф. Про метод дослідження факторів впливу на зростання деревостану. Ботаніка. Дослідження. (2002). Полтава. Наука та техніка. 171 с.
8. Багінський В. Ф. Лісокористування в Україні. (2000). Чернігів. Либідь. 367 с.
9. Багінський В. Ф. Системний аналіз в лісовому господарстві. Навчальний посібник.(2009). Харків. ХНАУ. 168 с.
10. Бейлі Н. С. Статистичні методи в біології. (2004). Київ. Наукова думка. 480 с.
11. Ван-дер-Ванден. Математична статистика. (2006). Талін. Наука. 424 с.
12. Воїнов Н. Т. Математичні моделі об'ємних та сортиментних таблиць. (2004). Київ. Урожай. 134 с.
13. Горошко М. П., Миклуш С. І., Хомюк П. Г. Біометрія: Навчальний посібник. (2004). Львів: Камула 236 с.
14. Горошко М. П., Миклуш С.І., Хомюк П.Г. Практикум з лісової біометрії. (1999). Львів. УкрДЛТУ. 1999. 108 с.
15. Гнєденко Б.В. Курс теорії ймовірностей. (2003). Черкаси. Фітматгіз. 406 с.
16. Грейг-Сміт П. Кількісна екологія рослин. (2002). Тарту. Мир. 349 с.
17. Гурський Є. І. Теорія ймовірностей з елементами математичної статистики. (2003). Київ. Вища школа. 224 с.
18. Дворецький М. Л. Практичний посібник з варіаційної статистики. (2006). Йошкар-Ола. Знання. 99 с.
19. Доспехов Б. А. Методика польового досліду: (з основами статистической обработки результатов досліджень). (1979). Харків. Колос. 416 с.
20. Дукарський О. М. Статистичний аналіз та обробка спостережень на ЕОМ. (2010). Херсон. Наукова думк. 244 с.

21. Заграїв В. В. Географічні закономірності зростання та продуктивності деревостанів. (2006). Харків. Лісова промисловість. 240 с.

22. Зайцев Г. М. Математична статистика в експериментальній ботаніці. (2000). Київ. Наука. 424 с.

23. Зайцев Г. М. Математичний аналіз біологічних даних (2003). Київ. Наука. 184 с.

24. Звонкова Т. В., Ксімова Н. С. Географічне прогнозування і охорона природи. (1990). Вінниця. Кальварія. 176 с.

25. Здорік Н. Г. Статистика для лісових спеціалістів. (2002). Херсон. Кльварія. 225 с.

26. Кендал Д. Дж. Статистичні висновки та зв'язки (1999). Київ. Наука. 320 с.

27. Кравченко Г. Л. Варіаційна статистика (2004). Харків. ХНАУ. 86 с.

28. Крамер Г. К. Математичні методи статистики (1993). Чернігів. Світ. 248 с.

29. Кузмічов В. В. Закономірності зростання деревостанів. (2006). Київ. Наука. 160 с.

30. Кудрявцев Л. Д. Математичний аналіз. (1973). Видання 2-ге, доповнене. Харків. Вища щкола. 470 с.

31. Калінін М. І. Біометрія: Підручник для студентів ЗВН біологічних і екологічних напрямків. (2000). Миколаїв. Вид-во МФ НаУКМА. 204 с.

32. Лакін Г. Ф. Біометрія. Навчальний посібник. (1990). Київ. Вища школа. 352 с.

33. Levchenko V. B., Shulga I. V., Ivanyk I. D., Romanyuk A. A., Rusetskaya N. M. Innovative forest and biologikalmetods of entomological monitoring of trumpet pest in the conditions of the Pergan nature cjnservation research department of Poliska nature reserve. DOI 10.26886/2520-7474.1(51)2022.1. Paradigm of knowledge № 1(51), 2022. P. 5-29.

34. Levchenko V. B., Shulga I. V., Fuchilo Y. D., Karpovych M. S., Romanyuk A. A., Belska O. V. Forest pathological monitoring of pine stands in the conditions of the Pergans scientific and research nature protection department Polissky nature reserve. Innovative Solutions In Modern Science № 3(55), 2022. DOI 10.26886/2414-634X.3(55)2022.2.. P. 18-62.

35. Levchenko V. B., Shulga I. V., Fuchilo Y. D., Hurzhii R. V., Romanyuk A. A., Belska O. V. Fall of Pine phytomass after large scale forest fires in the conditions nature protection scientific research departments Polisky nature reserve. Paradigm of knowledge № 1(59), 2024. DOI 10.26886/2520-7474.1(59)2024.1. S. 5 – 32.

36. Леман Е. Ф. Перевірка статистичних гіпотез. (2004). Вінниця. Мир. 498 с.

37. Лосицький К. Б. Еталонні ліси. (1990). Харків. Лісова промисловість. 191 с.

38. Мармоза А. Т. Практикум з теорії статистики. Навчальний посібник. (2007). 3 – тє видання, доповнене. Київ. Ельга, Ніка-Центр. 348 с.

39. Митропольський А. К. Техніка статистичних обчислень. (2004). Харків. Наукова думка. 479 с.

40. Моісеєнко Ф .П. Про закономірності в зростанні, будові та товарності насаджень. (2003). Київ. НАУ. 78 с.

41. Моісеєнко Ф. П. Таблиці для сортиментного обліку лісу на пні. (2001). Київ. НАУ. 328 с.

42. Мошкальов А. Г. Таксація товарної структури деревостанів. (1994). Харків. Лісова промисловість. 157 с.

43. Неверов А. В. Стиглість лісу в системі сталого природокористування. (2002). Харків. ХНАУ. 207 с.

44. Нікітін К. Є. Модрина в Україні. (1993). Київ. Урожай. 331 с.

45. Нікітін К. Є. Методи та техніка обробки лісівничої інформації. (2002). Київ. НАУ. 270 с.

46. Орлов М. М. Лісовиробництво. (2004). Львів. Рута. 247 с.

47. Плохінський Н.А. Біометрія. (2002). Київ. НАУ. 267 с.

48. Рокіцький П. Ф. Біологічна статистика. (2004). Київ. Вища школа. 326 с.

49. Свалов Н. М. Варіаційна статистика. Навчальний посібник. (1996). Харків. Лісова промисловість. 177 с.

50. Свалов Н. М. Моделювання продуктивності деревостанів та теорія лісокористування. (2002). Харків. Лісова промисловість. 216 с.

51. Суслов Н. П. Загальна теорія статистики. (2008). Вінниця. Статистика. 391 с.

52. Третьяков Н. В. Закон єдності у будові лісових насаджень. (2017). Тернопіль. Зання. 113 с.

53. Трулль О. А. Математична статистика в лісовому господарстві. (2016). Киїів.Наукова думка. 233 с.

54. Тюрін А. В. Нормальна продуктивність сосни звичайної, берези повислої, осики звичайної та ялини звичайної. (1993). Черкаси. Промінь. 189 с.

55. Тюрін А. В. Основи варіаційної статистики в лісівництві. (2004). Рівне. Райдуга. 94 с.

56. Устінов М. В. Математичні методи в лісовому господарстві. (20150. Харків. Наукова думка. 98 с.

57. Штоф В. А. Моделювання та філософія. (2004). Київ. Кальварія. 202 с.

58. Юркевич І.Д. Географія, типологія та районування лісової рослинності. (2018). Київ. Наукова думка. 248 с.

59. Чепур С. С. Біометрія: Методичний посібник. (2015).Ужгород. Видавництво УжНУ «Говерла». 40 с.

ДОДАТКИ
(МАТЕМАТИЧНО-СТАТИСТИЧІ ТАБЛИЦІ)

Додаток А

Випадкові числа по А. К. Митропольському

1	2	3	4	5	6	7	8	9	10
1534	7106	2836	7873	5574	7545	7590	5574	1202	7712
6128	8993	4102	2551	0303	2358	6427	7067	9325	2454
6047	8566	8644	9343	9297	6751	3500	8754	2913	1258
0806	5201	5705	7355	1448	9562	7514	9205	0402	2427
9915	8274	4525	5695	5752	9630	7172	6988	0227	4264
2882	7158	4341	3463	1178	5786	1173	0670	0820	5067
9213	1223	4388	9760	6691	6861	8214	8813	0611	3131
8410	9836	3899	3683	1253	1683	6988	9978	8026	6751
9974	2362	2103	4326	3825	9079	6187	2721	1489	4216
3402	8162	8226	0782	3364	7871	4500	5598	9421	3816
8188	6596	1492	2139	8823	6878	0613	7161	0241	3834
3825	7020	1124	7483	9155	4919	3209	5959	2364	2555
9801	8788	6338	5899	3309	0807	0968	0539	4205	8257
5603	1251	6352	6467	0231	3556	2569	9446	4174	9219
0714	3757	0378	8266	8864	1374	6687	1221	0678	3714
4617	5652	7627	0372	8151	3668	1994	4402	2124	0016
6789	6279	7306	1856	7028	9043	7161	7526	6923	6393
8705	4978	8621	1790	4433	6298	0854	9127	3445	1111
3840	1086	0774	9241	9297	4239	1739	7734	0119	2436
7662	3939	2965	3273	0551	1645	8477	1877	5327	8629
7639	2868	4391	2950	7122	7325	9727	0080	7467	7947
3237	3203	4246	7329	7936	0065	4146	0866	4916	8648
3917	6271	1721	5469	1914	8653	0387	2756	6073	8984
9138	9395	6005	6423	7977	1873	7103	4267	9316	7206
8358	5896	6286	9242	5040	8509	2941	3913	3028	1563
1030	5094	1745	2975	2018	7340	6547	0207	5587	0300
6606	6305	1564	6668	7822	7142	6564	1659	5369	1659
4533	8841	4922	9365	1361	6692	1633	6764	0747	3881
4258	2012	0992	0106	1542	4760	0392	4057	0092	5203
5224	5128	8949	7928	7267	0116	1476	2009	1772	3860
6872	7492	7962	1867	7437	1526	3516	9129	4153	8084
8638	8407	7198	0956	0950	7753	5144	3914	5596	6104
9958	7172	5822	4224	6701	7559	4985	4856	4461	6147
0265	3086	2996	0699	3584	9702	1665	0446	9107	6437
8987	5441	7878	9404	0487	2939	3805	9172	7887	5197
5552	3529	9627	9362	6298	6021	0024	9520	9154	0643

9383	6640	7394	9592	9903	7699	8939	9972	1257	0994
9903	4059	0332	9109	0182	6721	9163	9008	2542	4461
6530	5070	7589	6928	6014	1832	9307	5107	1354	9257
8679	8953	8310	2060	6277	1773	7979	6741	6033	3588
5765	4987	1639	3512	9843	5286	3786	2384	4919	5611
7198	2447	6716	0291	5585	1106	5330	0504	6346	3679
2385	0605	2678	1399	2371	7968	1212	9569	8650	5841
0732	8732	8660	5836	9065	4603	0029	8042	0159	0345
1642	6094	3795	2600	4532	9740	0376	4384	9203	5387
4514	1956	7212	0687	7632	2106	0846	7055	4106	9157
8744	5580	8038	9087	7222	0424	0028	4511	3191	9846
3729	6225	5397	6790	2157	3414	6509	5204	4779	5641
8858	3147	8410	2783	1290	9796	8873	7585	7185	4726
3522	5601	6197	6051	3470	8283	5702	0103	8726	5282
11	12	13	14	15	16	17	18	19	20
5489	5583	3156	0835	1988	3912	0938	7460	0869	4420
3522	0935	7877	5665	7020	9255	7379	7124	7878	5544
7555	7579	2550	2487	9477	0864	2349	1012	8250	2633
5759	3554	5080	9074	7001	6249	3224	6368	9102	2672
6303	6895	3371	3196	7231	2918	7380	0438	7547	2644
7351	5634	5323	2623	7803	8374	2191	0464	0696	9529
7068	7803	8832	5119	6350	0120	5026	3684	5657	0304
3613	1428	1796	8447	0503	5654	3524	7336	9536	1944
5143	4534	2105	0368	7890	2473	4240	8652	9435	1422
9815	5144	7649	8638	6137	8070	5345	4865	2456	5708
5780	1277	6316	1013	2867	9938	3930	3203	5696	1769
1187	0951	5991	5246	5700	5564	7352	0891	6249	6568
4184	2179	4554	9083	2254	2435	2965	5154	1209	7069
2916	2972	9885	0275	0144	8034	8122	3213	7666	0230
5524	1341	9860	6565	6981	9842	0171	2284	2707	3008
0146	5291	2354	5694	0377	5336	6460	9585	3415	2358
4920	2826	5238	5402	7937	1993	4332	2327	6875	5230
7978	1947	6380	3425	7267	7285	1130	7722	0164	8573
7453	0653	3645	7497	5969	8682	4191	2976	0361	9334
1473	6938	4899	5348	1641	3652	0852	5296	4538	4456
8162	8797	8000	4707	1880	9660	8446	1883	9768	0881
5645	4219	0807	3301	4279	4168	4305	9937	3120	5547
2042	1192	1175	8851	6432	4635	5757	6656	1660	5389
5470	7702	6958	9080	5925	8519	0127	9233	2452	7341
4045	1730	6005	1704	0345	3275	4738	4862	2556	8333

Продовження додатку А

5880	1257	6163	4439	7276	6353	6912	0731	9033	5294
9083	4260	5277	4998	4298	5204	3965	4028	8936	5148
1762	8713	1189	1090	8989	7273	3213	1938	9321	4820
2023	2589	1740	0424	8924	0005	1969	1636	7237	1227
7965	3855	4765	0703	1678	0841	7543	0308	9732	1289
7690	0480	8098	9629	4819	4219	7241	5128	3853	1921
9292	0426	9573	4903	5916	6576	8368	3270	6641	0033
0867	1656	7016	4220	2533	6345	8227	1904	5138	2537
0505	2127	8255	5276	2233	3956	4118	8199	6380	6340
6295	9795	1112	5761	2575	6837	3336	9322	7403	8345
6323	2615	3410	3365	1117	2417	3176	2434	5240	5455
8672	8536	2966	5773	5412	8114	0930	4697	6919	4569
1422	5507	7596	0670	3013	1351	3886	3268	9469	2584
2653	1472	5113	5735	1469	9545	9331	5303	9914	6394
0438	4376	3328	8649	8327	0110	4549	7955	5275	2890
2851	2157	0047	7085	1129	0460	6821	8373	2572	8962
7962	2753	3077	8718	7418	8004	3125	3706	8822	1494
3837	4098	0220	1217	4732	0150	1637	1097	1040	7372
8542	4126	9274	2251	0607	4301	8730	7690	6235	3477
0139	0765	8039	9484	2577	7859	1976	0623	1418	6685
6687	1943	4307	0579	8171	8224	8641	7034	3595	3875
8242	5582	5872	3197	4919	2792	5991	4058	9769	1918
2851	2157	0047	7085	1129	0460	6821	8373	2572	8962
6859	9606	0522	4993	0345	8958	1289	8825	6941	7685
6590	1932	6043	3623	1973	4112	1795	8465	2110	8045
3482	0478	0221	6738	7323	5643	4767	0106	2372	9862

Додаток Б

N	V		N	V		N	V	
	5%	1%		5%	1%		5%	1%
5	0,96	0,99	13	0,41	0,52	22	0,32	0,41
6	0,81	0,92	14	0,40	0,50	23	0,31	0,41
7	0,69	0,80	15	0,38	0,49	24	0,31	0,40
8	0,61	0,74	16	0,37	0,47	25	0,30	0,39
9	0,55	0,68	17	0,36	0,46	26	0,30	0,39
10	0,51	0,64	18	0,35	0,45	27	0,30	0,39
11	0,48	0,60	19	0,34	0,44	28	0,29	0,38
12	0,45	0,57	20	0,33	0,43	29	0,29	0,37

Додаток В

N	V		N	V		N	V	
	5%	1%		5%	1%		5%	1%
5	1,92	1,97	21	2,80	3,11	80	3,33	3,70
6	2,07	2,16	22	2,82	3,13	90	3,37	3,74
7	2,18	2,31	23	2,84	3,16	100	3,40	3,77
8	2,27	2,43	24	2,86	3,18	120	3,46	3,83
9	2,35	2,53	25	2,88	3,20	150	3,53	3,90
10	2,41	2,62	26	2,90	3,22	200	3,61	3,98
11	2,47	2,69	27	2,91	3,24	300	3,73	4,09
12	2,52	2,75	28	2,93	3,26	400	3,80	4,17
13	2,56	2,81	29	2,94	3,28	500	3,87	4,24
14	2,60	2,86	30	2,96	3,29	600	3,92	4,28
15	2,64	2,90	35	3,02	3,36	700	3,96	4,32
16	2,67	3,94	40	3,08	3,42	800	3,99	4,35
17	2,70	2,98	45	3,12	3,48	900	4,02	4,35
18	2,73	3,02	50	3,16	3,52	1000	4,05	4,41
19	2,75	3,05	60	3,22	3,58	1500	4,14	4,50
20	2,78	3,08	70	3,28	3,64	2000	4,21	4,56

Додаток Г

P	p									
	0,10	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	0,01
0,75	33	40	51	67	91	132	206	367	827	3308
0,80	41	50	64	83	114	164	256	456	1026	4105
0,85	51	63	80	105	143	207	323	575	1295	5180
0,90	67	83	105	138	187	270	422	751	1690	6763
0,91	71	88	112	146	199	287	449	798	1796	7185
0,92	76	94	119	156	212	306	478	851	1915	7662
0,93	82	101	128	167	227	328	512	911	2051	8207
0,94	88	109	138	180	245	353	552	982	2210	8843
0,95	96	118	150	195	266	384	600	1067	2400	9603
0,96	105	130	164	215	292	421	659	1171	2636	10544
0,965	111	137	173	226	308	444	694	1234	2778	11112
0,970	117	145	183	240	327	470	735	1308	2943	11773
0,975	125	155	196	256	348	502	784	1395	3139	12559
0,980	135	167	211	276	375	541	845	1503	3382	13529
0,985	147	182	231	301	410	591	924	1643	3697	14791
0,990	165	204	259	338	460	663	1036	1843	4146	16587
0,991	170	210	266	348	473	682	1066	1895	4264	17057

0,992	175	217	274	358	488	703	1098	1953	4395	14583
0,993	181	224	284	371	505	727	1136	2020	4545	18182
0,994	188	233	294	385	524	755	1179	2097	4718	18875
0,995	196	243	307	402	547	787	1231	2188	4924	19698
0,996	207	255	323	422	575	828	1294	2301	5177	20709
0,997	220	271	344	449	611	880	1376	2446	5504	22018
0,998	238	294	373	487	663	954	1492	2652	5968	23873
0,999	270	334	422	552	751	1082	1691	3007	6767	27069

Додаток Д

Значення t – критерія Стюдента

V	p												V
	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,01	0,001	
1	0,128	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,700	63,650	636,600	
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	9,925	31,590	
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	5,841	12,940	
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	4,604	8,610	
5	0,132	0,267	0,408	5,559	0,727	0,920	1,156	1,476	2,015	2,571	4,032	6,859	1
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,707	5,959	2
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	3,499	5,405	3
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	3,355	5,041	4
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	3,250	4,781	5
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	3,169	4,587	6
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	3,106	4,437	7
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	3,055	4,318	8
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	3,012	4,221	9
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,977	4,140	10
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,947	4,073	11
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,921	4,015	12
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,898	3,965	13
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,878	3,922	14
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,861	3,883	15
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,845	3,850	16
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,831	3,819	21
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,819	3,792	22
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,807	3,767	23
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,797	3,745	24
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,787	3,725	25
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,779	3,707	26
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,771	3,690	27
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,763	3,674	28
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,756	3,659	29
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,750	3,646	30
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,704	3,551	40
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,660	3,460	60
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,617	3,373	120
>120	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,576	3,291	>120

Додаток Е

Значення F - критерію Фішера (верхній ряд – 99,9%, середній – 99%, нижній – 95%)

v2	v2																							v2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	
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
	46,3	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	
10	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	
	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	
11	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	
	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	
12	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	
	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	
13	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	
14	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	2,5	2,4	2,4	2,3	2,3	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	5,4	5,3	5,2	5,1	5,0	3,1	3,1	3,0	3,0	
	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	3,4	2.3	3,3	3,2	3,1	2,2	2,2	2,1	2,1	

v2	v2																								v2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	
15	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	
	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	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	
	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	
17	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	
	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	
18	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	
	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	
19	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	
	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	
20	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	
	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	
21	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	3,5	3,6	3,5	3,4	3,3	
	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	
22	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	
	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	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,8	1,8	
23	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	
	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	
24	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	

Продовження додатку Е

v2	v2																								v2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	
28	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	
29	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	
30	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	
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	

Продовження додатку Е

v2	v2																								v2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	>500	
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	
	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	55
	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	
	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	60
	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	
	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	65
	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	
	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	70
	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	
80	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	
	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	
100	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	
	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	
125	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	
	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	
150	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	
	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	
200	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	
	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	
400	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	
	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	
1000	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	
	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	
>1000	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	
	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	

Додаток Є

$\frac{R}{N_1+N_2+1}$	0	1	2	3	4	5	6	7	8	9
0,00	- ∞	-3,09	-2,88	-2,75	-2,65	-2,58	-2,51	-2,46	-2,41	-2,37
0,01	-2,33	-2,29	-2,26	-2,23	-2,20	-2,17	-2,14	-2,12	-2,10	-2,07
0,02	-2,05	-2,03	-2,01	-2,00	-1,98	-1,96	-1,94	-1,93	-1,91	-1,90
0,03	-1,88	-1,87	-1,85	-1,84	-1,83	-1,81	-1,80	-1,79	-1,77	-1,76
0,04	-1,75	-1,74	-1,73	-1,72	-1,71	-1,70	-1,68	-1,67	-1,66	-1,65
0,05	-1,64	-1,64	-1,63	-1,62	-1,61	-1,60	-1,59	-1,58	-1,57	-1,57
0,06	-1,55	-1,55	-1,54	-1,53	-1,52	-1,51	-1,51	-1,50	-1,49	-1,48
0,07	-1,48	-1,47	-1,46	-1,45	-1,45	-1,44	-1,43	-1,43	-1,42	-1,41
0,08	-1,41	-1,40	-1,39	-1,39	-1,38	-1,37	-1,37	-1,36	-1,35	-1,35
0,09	-1,34	-1,33	-1,33	-1,32	-1,32	-1,31	-1,30	-1,30	-1,29	-1,29
0,10	-1,28	-1,28	-1,27	-1,26	-1,26	-1,25	-1,25	-1,24	-1,24	-1,23
0,11	-1,23	-1,22	-1,22	-1,21	-1,21	-1,20	-1,20	-1,19	-1,19	-1,18
0,12	-1,18	-1,17	-1,17	-1,16	-1,16	-1,15	-1,15	-1,14	-1,14	-1,13
0,13	-1,13	-1,12	-1,12	-1,11	-1,11	-1,10	-1,10	-1,09	-1,09	-1,09
0,14	-1,08	-1,08	-1,07	-1,07	-1,06	-1,06	-1,05	-1,05	-1,05	-1,04
0,15	-1,04	-1,03	-1,03	-1,02	-1,02	-1,02	-1,01	-1,01	-1,00	-1,00
0,16	-0,99	-0,99	-0,99	-0,98	-0,98	-0,97	-0,97	-0,97	-0,96	-0,96
0,17	-0,95	0,95	-0,95	-0,94	-0,94	-0,93	-0,93	-0,93	-0,92	-0,92
0,18	-0,92	-0,91	-0,91	-0,90	-0,90	-0,90	-0,89	-0,89	-0,89	-0,88
0,19	-0,88	-0,87	-0,87	-0,87	-0,86	-0,86	-0,86	-0,85	-0,85	-0,85
0,20	-0,84	-0,84	-0,83	-0,83	-0,83	-0,82	-0,82	-0,82	-0,81	-0,81
0,21	-0,81	-0,80	-0,80	-0,80	-0,79	-0,79	-0,79	-0,78	-0,78	-0,78
0,22	-0,77	-0,77	-0,77	-0,76	-0,76	-0,76	-0,75	-0,75	-0,75	-0,74
0,23	-0,74	-0,74	-0,73	-0,73	-0,73	-0,72	-0,72	-0,72	-0,71	-0,71
0,24	-0,71	-0,70	-0,70	-0,70	-0,69	-0,69	-0,69	0,68	-0,68	-0,68
0,25	-0,67	-0,67	-0,67	-0,67	-0,66	-0,66	-0,66	-0,65	-0,65	-0,65
0,26	-0,64	-0,64	-0,64	-0,63	-0,63	-0,63	-0,63	-0,62	-0,62	-0,62
0,27	-0,61	-0,61	-0,61	-0,60	-0,60	-0,60	-0,60	-0,59	-0,59	-0,59
0,28	-0,58	-0,58	-0,58	-0,57	-0,57	-0,57	-0,57	-0,56	-0,56	-0,56
0,29	-0,55	-0,55	-0,55	-0,54	-0,54	-0,54	-0,54	-0,53	-0,53	-0,53
0,30	-0,53	-0,52	-0,52	-0,52	-0,51	-0,51	-0,51	-0,50	-0,50	-0,50
0,31	-0,50	-0,49	-0,49	-0,49	-0,48	-0,48	-0,48	-0,47	-0,47	-0,47
0,32	-0,47	-0,46	-0,46	-0,46	-0,46	-0,45	-0,45	-0,45	-0,45	-0,44
0,33	-0,44	-0,44	-0,43	-0,43	-0,43	-0,43	-0,43	-0,42	-0,42	-0,42
0,34	-0,41	-0,41	-0,41	-0,40	-0,40	-0,40	-0,40	-0,39	-0,39	-0,39
0,35	-0,39	-0,38	-0,38	-0,38	-0,37	-0,37	-0,37	-0,37	-0,36	-0,36
0,36	-0,36	-0,36	-0,35	-0,35	-0,35	-0,35	-0,34	-0,34	-0,34	-0,33
0,37	-0,33	-0,33	-0,33	-0,32	-0,32	-0,32	-0,32	-0,31	-0,31	-0,31
0,38	-0,31	-0,30	-0,30	-0,30	-0,30	-0,29	-0,29	-0,29	-0,28	-0,28
0,39	-0,28	-0,28	-0,27	-0,27	-0,27	-0,27	-0,26	-0,26	-0,26	-0,26
0,40	-0,25	-0,25	-0,25	-0,25	-0,24	-0,24	-0,24	-0,24	-0,23	-0,23
0,41	-0,23	-0,23	-0,22	-0,22	-0,22	-0,21	-0,21	-0,21	-0,21	-0,20
0,42	-0,20	-0,20	-0,20	-0,19	-0,19	-0,19	-0,19	-0,18	-0,18	-0,18
0,43	-0,18	-0,17	-0,17	-0,17	-0,17	-0,16	-0,16	-0,16	-0,16	-0,15
0,44	-0,15	-0,15	-0,15	-0,14	-0,14	-0,14	-0,14	-0,13	-0,13	-0,13
0,45	-0,13	-0,12	-0,12	-0,12	-0,12	-0,11	-0,11	-0,11	-0,11	-0,10
0,46	-0,10	-0,10	-0,10	-0,09	-0,09	-0,09	-0,09	-0,08	-0,08	-0,08
0,47	-0,08	-0,07	-0,07	-0,07	-0,07	-0,06	-0,06	-0,06	-0,06	-0,05
0,48	-0,05	-0,05	-0,05	-0,04	-0,04	-0,04	-0,04	-0,03	-0,03	-0,03
0,49	-0,03	-0,02	-0,02	-0,02	-0,02	-0,01	-0,01	-0,01	-0,01	-0,00
0,50	0,00	0,00	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,02

Продовження додатку Є

$\frac{R}{N_1+N_2+1}$	0	1	2	3	4	5	6	7	8	9
0,51	0,03	0,03	0,03	0,03	0,04	0,04	0,04	0,04	0,05	0,05
0,52	0,05	0,05	0,06	0,06	0,06	0,06	0,07	0,07	0,07	0,07
0,53	0,08	0,08	0,08	0,08	0,09	0,09	0,09	0,09	0,10	0,10
0,54	0,10	0,10	0,11	0,11	0,11	0,11	0,12	0,12	0,12	0,12
0,55	0,13	0,13	0,13	0,13	0,14	0,14	0,14	0,14	0,15	0,15
0,56	0,15	0,15	0,16	0,16	0,16	0,16	0,17	0,17	0,17	0,17
0,57	0,18	0,18	0,18	0,18	0,19	0,19	0,19	0,19	0,20	0,20
0,58	0,20	0,20	0,21	0,21	0,21	0,21	0,22	0,22	0,22	0,23
0,59	0,23	0,23	0,23	0,24	0,24	0,24	0,24	0,25	0,25	0,25
0,60	0,25	0,26	0,26	0,26	0,26	0,26	0,27	0,27	0,27	0,28
0,61	0,28	0,28	0,28	0,29	0,29	0,29	0,30	0,30	0,30	0,30
0,62	0,31	0,31	0,31	0,31	0,32	0,32	0,32	0,32	0,33	0,33
0,63	0,33	0,33	0,34	0,34	0,34	0,35	0,35	0,35	0,35	0,36
0,64	0,36	0,36	0,36	0,37	0,37	0,37	0,37	0,38	0,38	0,38
0,65	0,39	0,39	0,39	0,39	0,40	0,40	0,40	0,40	0,41	0,41
0,66	0,41	0,42	0,42	0,42	0,42	0,43	0,43	0,43	0,43	0,44
0,67	0,44	0,44	0,45	0,45	0,45	0,46	0,46	0,46	0,46	0,46
0,68	0,47	0,47	0,47	0,48	0,48	0,48	0,48	0,49	0,49	0,49
0,69	0,50	0,50	0,50	0,50	0,51	0,51	0,51	0,52	0,52	0,52
0,70	0,52	0,53	0,53	0,53	0,54	0,54	0,54	0,54	0,55	0,55
0,71	0,55	0,56	0,56	0,56	0,57	0,57	0,57	0,57	0,58	0,58
0,72	0,58	0,59	0,59	0,59	0,59	0,60	0,60	0,60	0,61	0,61
0,73	0,61	0,62	0,62	0,62	0,63	0,63	0,63	0,63	0,64	0,64
0,74	0,64	0,65	0,65	0,65	0,66	0,66	0,66	0,67	0,67	0,67
0,75	0,67	0,68	0,68	0,68	0,69	0,69	0,69	0,70	0,70	0,70
0,76	0,71	0,71	0,71	0,72	0,72	0,72	0,73	0,73	0,73	0,74
0,77	0,74	0,74	0,75	0,75	0,75	0,76	0,76	0,76	0,77	0,77
0,78	0,77	0,78	0,78	0,78	0,79	0,79	0,79	0,80	0,80	0,80
0,79	0,81	0,81	0,81	0,82	0,82	0,82	0,83	0,83	0,83	0,84
0,80	0,84	0,85	0,85	0,85	0,86	0,86	0,86	0,87	0,87	0,87
0,81	0,88	0,88	0,89	0,89	0,89	0,90	0,90	0,90	0,91	0,91
0,82	0,92	0,92	0,92	0,93	0,93	0,93	0,94	0,94	0,95	0,95
0,83	0,95	0,96	0,96	0,97	0,97	0,97	0,98	0,98	0,99	0,99
0,84	0,99	1,00	1,00	1,01	1,01	1,02	1,02	1,02	1,03	1,03
0,85	1,04	1,04	1,05	1,05	1,05	1,06	1,06	1,07	1,07	1,08
0,86	1,08	1,09	1,09	1,09	1,10	1,10	1,11	1,11	1,12	1,12
0,87	1,13	1,13	1,14	1,14	1,15	1,15	1,16	1,16	1,17	1,17
0,88	1,18	1,18	1,19	1,19	1,20	1,20	1,21	1,21	1,22	1,22
0,89	1,23	1,23	1,24	1,24	1,25	1,25	1,26	1,26	1,27	1,28
0,90	1,28	1,29	1,29	1,30	1,30	1,31	1,32	1,32	1,33	1,33
0,91	1,34	1,35	1,35	1,36	1,37	1,37	1,38	1,39	1,39	1,40
0,92	1,41	1,41	1,42	1,43	1,43	1,44	1,45	1,45	1,46	1,47
0,93	1,48	1,48	1,49	1,50	1,51	1,51	1,52	1,53	1,54	1,55
0,94	1,55	1,56	1,57	1,58	1,59	1,60	1,61	1,62	1,63	1,64
0,95	1,64	1,65	1,66	1,67	1,68	1,70	1,71	1,72	1,73	1,74
0,96	1,75	1,76	1,77	1,79	1,80	1,81	1,83	1,84	1,85	1,87
0,97	1,88	1,90	1,91	1,93	1,94	1,96	1,98	2,00	2,01	2,03
0,98	2,05	2,07	2,10	2,12	2,14	2,17	2,20	2,23	2,26	2,29
0,99	2,33	2,37	2,41	2,46	2,51	2,58	2,65	2,75	2,88	3,09

Додаток Ж

Значення Х-критерія Ван-дер-Вардена

V	N_1-N_2=0 або 1		N_1-N_2=2 або 3		N_1-N_2=3 або 4	
	5%	1%	5%	1%	5%	1%
8	2,40	-	2,30	-	-	-
9	2,48	-	2,40	-	-	-
10	2,60	3,20	2,49	3,10	2,30	-
11	2,72	3,40	2,58	3,40	2,40	-
12	2,86	3,60	2,79	3,58	2,68	3,40
13	2,96	3,71	2,91	3,64	2,78	3,50
14	3,11	3,94	3,06	3,88	3,00	3,76
15	3,24	4,07	3,19	4,05	3,06	3,88
16	3,39	4,26	3,36	4,25	3,28	4,12
17	3,49	4,44	3,44	4,37	3,36	4,23
18	3,63	4,60	3,60	4,58	3,53	4,50
19	3,73	4,77	3,69	4,71	3,61	4,62
20	3,86	4,94	3,84	4,92	3,78	4,85
21	3,96	5,10	3,92	5,05	3,85	4,96
22	4,08	5,26	4,06	5,24	4,01	5,17
23	4,18	5,40	4,15	5,36	4,08	5,27
24	4,29	5,55	4,27	5,53	4,23	5,48
25	4,39	5,68	4,36	5,65	4,30	5,58
26	4,50	5,83	4,48	5,81	4,44	5,76
27	4,59	5,95	4,56	5,92	4,51	5,85
28	4,68	6,09	4,68	6,07	4,64	6,03
29	4,78	6,22	4,76	6,19	4,72	6,13
30	4,88	6,35	4,87	6,34	4,84	6,30
31	4,97	6,47	4,95	6,44	4,91	6,39
32	5,07	6,60	5,06	6,58	5,03	6,55
33	5,15	6,71	5,13	6,69	5,10	6,64
34	5,25	6,84	5,24	6,82	5,21	6,79
35	5,33	6,95	5,31	6,92	5,28	6,88
36	5,42	7,06	5,41	7,05	5,38	7,02
37	5,50	7,17	5,48	7,15	5,45	7,11
38	5,59	7,28	5,58	7,27	5,55	7,25
39	5,67	7,39	5,65	7,37	5,62	7,33
40	5,75	7,50	5,74	7,49	5,72	7,47
41	5,83	7,62	5,81	7,60	5,79	7,56
42	5,91	7,72	5,90	7,71	5,88	7,69
43	5,99	7,82	5,97	7,81	5,95	7,77
44	6,04	7,93	6,06	7,92	6,04	7,90
45	6,14	8,02	6,12	8,01	6,10	7,98
46	6,21	8,13	6,21	8,12	6,19	8,10
47	6,29	8,22	6,27	8,21	6,25	8,18
48	6,36	8,32	6,35	8,31	6,34	8,29
49	6,43	8,41	6,42	8,40	6,39	8,37
50	6,50	8,51	6,51	8,50	6,48	8,48

Додаток З

Значення Т-критерія Уайта

N_1 (велика вибірка	N_2 (менша вибірка)													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Рівень значущості 5%														
5		6	11	17										
6		7	12	18	26									
7		7	13	20	27	36								
8	3	8	14	21	29	38	49							
9	3	8	15	22	31	40	51	63						
10	3	9	15	23	32	42	53	65	78					
11	4	9	16	24	34	44	55	68	81	96				
12	4	10	17	26	35	46	58	71	85	99	115			
13	4	10	18	27	37	48	60	73	88	103	119	137		
14	4	11	19	28	38	50	63	76	91	106	123	141	160	
15	4	11	20	29	40	52	65	79	94	110	127	145	164	185
16	4	12	21	31	42	54	67	82	97	114	131	150	169	
17	5	12	21	32	43	56	70	84	100	117	135	154		
18	5	13	22	33	45	58	72	87	103	121	139			
19	5	13	23	34	46	60	74	90	107	124				
20	5	14	24	35	48	62	77	93	110					
21	6	14	25	37	50	64	79	95						
22	6	15	26	38	51	66	82							
23	6	15	27	39	53	68								
24	6	16	28	40	55									
25	6	16	28	42										
26	7	17	29											
27	7	17												
Рівень значущості 1%														
5				15										
6			10	16	23									
7			10	17	24	32								
8			11	17	25	34	43							
9		6	11	18	26	35	45	56						
10		6	12	19	27	37	47	58	71					
11		6	12	20	28	38	49	61	74	87				
12		7	13	21	30	40	51	63	76	90	106			
13		7	14	22	31	41	53	65	79	93	109	125		
14		7	14	22	32	43	54	67	81	96	112	129	147	
15		8	15	23	33	44	56	70	84	99	115	133	151	171
16		8	15	24	34	46	58	72	86	102	119	137	155	
17		8	16	25	36	47	60	74	89	105	122	140		
18		8	16	26	37	49	62	76	92	108	125			
19	3	9	17	27	38	50	64	78	94	111				
20	3	9	18	28	39	52	66	81	97					
21	3	9	18	29	40	53	68	83						
22	3	10	19	29	42	55	70							
23	3	10	19	30	43	57								
24	3	10	20	31	44									
25	3	11	20	32										
26	3	11	21											
27	4	11	21											

Додаток И

Значення Z – критерія

N	Рівень значущості		N	Рівень значущості		N	Рівень значущості		N	Рівень значущості	
	5%	1%		5%	1%		5%	1%		5%	1%
N	Уровень значимости		N	Уровень значимости		N	Уровень значимости		N	Уровень значимости	
	5%	1%		5%	1%		5%	1%		5%	1%
6	6	-	30	21	23	54	35	37	78	49	51
7	7	-	31	22	24	55	36	38	79	49	52
8	8	8	32	23	24	56	36	39	80	50	52
9	8	9	33	23	25	57	37	39	81	50	53
10	9	10	34	24	25	58	37	40	82	51	54
11	10	11	35	24	26	59	38	40	83	51	54
12	10	11	36	25	27	60	39	41	84	52	55
13	11	12	37	25	27	61	39	41	85	53	55
14	12	13	38	26	28	62	40	42	86	53	56
15	12	13	39	27	28	63	40	43	87	54	56
16	13	14	40	27	29	64	41	43	88	54	57
17	13	15	41	28	30	65	41	44	89	55	58
18	14	15	42	28	30	66	42	44	90	55	58
19	15	16	43	29	31	67	42	45	91	56	59
20	15	17	44	29	31	68	43	46	92	56	59
21	16	17	45	30	32	69	44	46	93	57	60
22	17	18	46	31	33	70	44	47	94	57	60
23	17	19	47	31	33	71	45	47	95	58	61
24	18	19	48	32	34	72	45	48	96	59	62
25	18	20	49	32	34	73	46	48	97	59	62
26	19	20	50	33	35	74	46	49	98	60	63
27	20	21	51	33	36	75	47	50	99	60	63
28	20	22	52	34	36	76	48	50	100	61	64
29	21	22	53	35	37	77	48	51	-	-	-

Додаток І

Значення W – критерія Вілконса

N	Рівень значущості		N	Рівень значущості	
	5%	1%		5%	5%
6	1	-	16	31	21
7	3	-	17	36	24
8	5	1	18	41	29
9	7	3	19	47	33
10	9	4	20	53	39
11	12	6	21	60	44
12	15	8	22	67	50
13	18	11	23	74	56
14	22	14	24	82	62
15	26	17	25	90	69

Додаток Ї

Критичне значення коефіцієнта ексцеса

N	Рівень значущості		
	10%	5%	1%
11	0,890	0,907	0,936
16	0,873	0,888	0,914
21	0,863	0,877	0,900
26	0,857	0,869	0,890
31	0,851	0,863	0,883
36	0,847	0,858	0,877
41	0,844	0,854	0,872
46	0,841	0,851	0,868
51	0,839	0,848	0,865
61	0,835	0,843	0,859
71	0,832	0,840	0,855
81	0,830	0,838	0,852
91	0,828	0,835	0,848
101	0,826	0,834	0,846
201	0,818	0,823	0,832
301	0,814	0,818	0,826
401	0,812	0,816	0,822
501	0,810	0,814	0,820

ЗМІСТ

Навчальне видання

ЛЕВЧЕНКО Валерій Борисович, кандидат с.-г. наук, доцент;

ШУЛЬГА Ігор Володимирович, кандидат с.-г. наук, доцент;

ТРОФИМЕНКО Петро Іванович, доктор с.-г. наук, доцент;

РОМАНЮК Алла Андріївна, спеціаліст вищої категорії, викладач-методист, Заслужений працівник освіти України;

МАЧУЛЬСЬКИЙ Григорій Миколайович, кандидат с.-г. наук, доцент.

БІОМЕТРІЯ В ЛІСОВОМУ ГОСПОДАРСТВІ З ОСНОВАМИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

Навчальний посібник для студентів спеціальності 205 «Лісове господарство» освітньо-професійного ступеня «фаховий молодший бакалавр», початкового рівня (короткого циклу) вищої освіти «молодший бакалавр», першого рівня вищої освіти «бакалавр», другого рівня вищої освіти «магістр». За ред. кандидата с.-г. наук, доцента В. Б. Левченко

Надруковано з оригінал-макета авторів

Підписано до друку 08.01.26. Формат 60х90/16. Папір офсетний.
Гарнітура Times New Roman. Друк офсетний.
Ум. друк. арк. 18.7. Обл. вид. арк. 16.8. Наклад 300. Зам. 02/26.

Видавництво Житомирського державного університету імені Івана Франка
м. Житомир, вул. Велика Бердичівська, 40
Свідоцтво про державну реєстрацію:
серія ЖТ №10 від 07.12.04 р.
електронна пошта (E-mail): zu@zu.edu.ua