57
Стат

Ю.І. ПРИЛУЦЬКИЙ,
О.В. ІЛЬЧЕНКО, О.В. ЦИМБАЛЮК,
С.О. КОСТЕРІН

СТАТИСТИЧНІ МЕТОДИ В БІОЛОГІЇ

Підручник для студентів вищих навчальних закладів

КИЇВ
НАУКОВА ДУМКА
2017

УДК 519.2

Статистичні методи в біології: підручник / Ю.І. Прилуцький, О.В. Ільченко, О.В. Цимбалюк, С.О. Костерін. — К.: Наук. думка, 2017. — 216 с. ISBN 978-966-00-1629-3

Підручник написано відповідно до навчальної програми дисципліни "Статистичні методи в біології" для підготовки бакалаврів з біології в Науково-навчальному центрі "Інститут біології та медицини" Київського національного університету імені Тараса Шевченка. У виданні систематизовано лекційний матеріал з таких важливих розділів вищої математики, як теорія ймовірностей і математична статистика, які є основою обробки результатів вибіркових досліджень у рамках комплексного аналізу різноманітних біологічних явищ і процесів, наведено теорію похибок, без якої неможливо оцінити точність отриманих експериментальних даних, велику кількість типових прикладів із різних розділів біології, які пояснюють теоретичний матеріал, а також завдання як для виконання студентами під час аудиторних занять, так і для самостійної роботи. Основну увагу зосереджено на чіткому роз'ясненні суті математичних понять, методів і формул, навчанні студентів вмілому їх застосуванню під час розв'язування завдань різного типу складності. Детально розглянуто деякі алгоритми прикладних пакетів програм "Excel", "Origin", "Statistica", які знадобляться для проведення статистичних розрахунків.

Для студентів біологічних спеціальностей вищих навчальних закладів, які вивчають дисципліну "Теорія ймовірностей та математична статистика".

Рецензенти:
член-кореспондент НАН України,
доктор біологічних наук, професор *Д.М. Говорун*,
доктор фізико-математичних наук, професор *О.І. Клесов*

Рекомендовано до друку вченою радою
Інституту біохімії ім. О.В. Палладіна НАН України
(протокол № 10 від 15.07.2016 р.)

Видання здійснено за кошти Цільової комплексної програми «Створення та розвиток науково-видавничого комплексу НАН України»

Науково-видавничий відділ медико-біологічної, хімічної та геологічної літератури

Редактор *Н.А. Серебрякова*

ISBN 978-966-00-1629-3

ПЕРЕДМОВА

Одна з особливостей природничих наук на нинішньому етапі їх розвитку полягає в міждисциплінарності (або й трансдисциплінарності). Справді, найцікавіші фундаментальні та прикладні питання сучасного Природознавства локалізовані саме "на перехресті" різних наук і наукових напрямів, інакше чим можна пояснити виникнення і бурхливий розвиток в останні десятиліття таких комплексних дисциплін, як фізико-хімічна й математична біологія, біофізична хімія, фізична біохімія, хімічна біофізика, фізика живого, системна біологія, нанобіотехнологія, комп'ютерна біологія тощо. Адже ПРИРОДА зовсім "не знає", що ми, відповідно до нашої освіти, наукових уподобань, націленості на вирішення конкретних проблем, диференціюємо її на Біологію, Фізику, Хімію, Математику; вона — ЄДИНА! Так, до найцікавіших "перехресних" проблем сучасної фізико-хімічної біології належать комплексні питання структурної геноміки, протеїн-протеїнової взаємодії, ензиматичного каталізу, внутрішньоклітинної сигналізації, транспорту речовин крізь біологічні мембрани, ліганд-рецепторної взаємодії, поширення нервового імпульсу і м'язового скорочення, біосенсорики, популяційної динаміки тощо. До розв'язання цих питань учені активно залучають сучасні уявлення, експериментальні й теоретичні методи і методології, які притаманні фізиці та фізичній хімії.

Натомість, як добре відомо, на сьогодні дедалі більшого значення у комплексних міждисциплінарних біологічних дослідженнях (зокрема, у галузі біохімії, молекулярної біології, біофізики) набуває використання математичних методів (я міг би тут згадати відомі вислови видатних особистостей на кшталт: "Велику книгу природи написано математичними символами" (*Галілей*) чи "В кожній природничій науці вміщено стільки істини, скільки в ній є математики" (*Кант*)), а також комп'ютерних технологій. Може навіть йтися про так звану обчислювальну біологію, яка широко використовує досягнення інформатики, обчислювальної техніки, прикладної математики й математичної статистики для вирішення вищенаведених фундаментальних біологічних проблем, а також намагається моделювати природні біологічні процеси за допомогою методів аналітичної і прикладної математики. Сучасну математичну біологію (математичне біомоделювання, біометрію) ми сприймаємо як комплексну міждисциплінарну галузь наукових досліджень, її цілком можна розглядати як підрозділ обчислювальної біології.

У контексті вищезазначеного особливу роль відіграє методологія статистичного об'єктивного аналізу експериментальних результатів. Математична статистика є розділом математики, в якому на основі експериментальних даних вивчаються ймовірнісні закономірності масових явищ. Основними завданнями математичної статистики є статистична перевірка гіпотез, оцінювання розподілу статистичних ймовірностей та його параметрів, вивчення статистичної залежності, визначення основних числових характеристик випадкових вибірок. Математична статистика ґрунтується, зокрема, на використанні методів теорії кореляції, теорії ймовірностей та теорії похибок. Ці методи розширюють можливості наукового передбачення, а також прийняття найефективнішого рішення у випадку багатьох наукових завдань і проблем (щодо наукового передбачення, мабуть доцільно згадати вислів проф. С. Хокінга: "Будь-яка теорія є доброю, якщо вона задовольняє щонайменше дві вимоги: точно описує великий клас спостережень на основі моделі, яка містить лише кілька елементів; дає змогу робити точні передбачення щодо результатів майбутніх спостережень").

Окремим випадком математичної статистики, аплікованої до аналізу результатів біологічних досліджень, є біологічна статистика (біометрія). Біологічну статистику широко застосовують при вивченні різних теоретичних і практичних питань біології, а також медицини, рослинництва і тваринництва, суміжних дисциплін.

Проте абсолютно очевидно, що опанування знаннями з математичної біології, які вкрай необхідні для коректного (я б навіть сказав грамотного) опису і тлумачення експериментальних даних у різних галузях біологічних знань, майбутні дослідники мають розпочинати ще під час навчання у вищих закладах освіти.

Сподіваємось, що підручник "Статистичні методи в біології" буде корисним не лише для студентів, аспірантів і пошукачів, а й для досвідчених учених — фахівців у різних галузях сучасної біологічної науки (біохімії, молекулярної біології, біофізики, фізіології, мікробіології, біотехнології тощо), які працюють у науково-дослідних інститутах НАН України.

ВСТУП

За формою прояву причинних зв'язків закони природи і суспільства поділяють на дві категорії: *детерміновані* та *статистичні.* Наприклад, на основі законів небесної механіки за відомим на сьогодні положенням планет Сонячної системи можна практично однозначно передбачити їх положення в довільний, наперед заданий момент часу. Це приклад детермінованого закону.

Для біологічних об'єктів характерним є те, що вони в переважній більшості створюють однорідні групи, а саме популяції, види, породи, сорти тощо, які достатньо чітко відрізняються між собою. Властивості саме таких груп цікавлять дослідників при вивченні окремих особин. При цьому параметри конкретних особин різні, тому постає запитання, як знайти значення параметра для всієї групи. Задача ускладнюється тим, що, як правило, всіх представників групи обстежити неможливо. Подібна ситуація складається, зокрема, у медико-біологічних випробуваннях під час вивчення результатів впливу лікарського засобу на певну категорію пацієнтів. Задачі такого типу трапляються досить часто. Їх об'єднує наявність комплексу чинників впливу, які не можна описати точно. Такий вплив називають *випадковим* і вивчають за допомогою статистичних методів, які ґрунтуються на теорії ймовірностей.

Теорія ймовірностей вивчає властивості масових випадкових подій, здатних неодноразово повторюватись при відновленні певного комплексу умов. Основна властивість будь-якої випадкової події незалежно від природи – міра або ймовірність її здійснення. На теорію ймовірностей спирається математична статистика, завдання якої полягає в тому, щоб за обмеженою кількістю обстежених об'єктів (вибіркою) визначити з певним ступенем вірогідності характеристики, притаманні генеральній сукупності, тобто всім об'єктам, що вивчаються.

Пропонований підручник охоплює всі програмні питання теорії ймовірностей і математичної статистики та приклади їх застосування для розв'язання задач біології, в тому числі із залученням комп'ютерних пакетів прикладних програм. При його написанні автори спирались на багаторічний досвід викладання курсу "Теорія ймовірностей і математична статистика" студентам різних факультетів (біологічного, механіко-математичного, геологічного, соціології та психології, географічного) Київського національного університету імені Тараса Шевченка.

РОЗДІЛ I

ОСНОВИ ТЕОРІЇ ЙМОВІРНОСТЕЙ

Теорія ймовірностей — математична наука, яка вивчає властивості масових випадкових подій, що неодноразово повторюються при відновленні певного комплексу умов. Основна властивість будь-якої випадкової події незалежно від її природи — міра або ймовірність її здійснення. На теорію ймовірностей спирається математична статистика, завдання якої полягає в тому, щоб за обмеженими даними (вибіркою) відновити з певним ступенем вірогідності характеристики, притаманні генеральній сукупності, тобто всьому можливому набору даних, які описують явище, що вивчається.

Теорія ймовірностей зародилася у XVI ст. і пов'язана з появою азартних ігор та визначенням шансів гравця на виграш. У перекладі з французької слово "азарт" означає "випадок". Схеми азартних ігор були першими математичними моделями для дослідження й перевірки законів теорії ймовірностей. Завдяки працям таких видатних математиків, як Дж. Кардано, П. Ферма, Б. Паскаль, Х. Гюйгенс, Я. Бернуллі, А. Муавр, П.С. Лаплас, С.Ф. Гаусс, С.Д. Пуассон, Т. Байєс, П.Л. Чебишов, О.М. Ляпунов, А.А. Марков, А.М. Колмогоров та інших теорія ймовірностей стала потужним інструментом дослідження складних процесів і явищ у різних галузях науки, в тому числі й біології. Особливістю методів, які використовує теорія ймовірностей, є те, що вони розглядають явище загалом, вивчають результати дії усіх причинних зв'язків, які неможливо відстежити окремо.

§ 1. ОСНОВНІ ПОНЯТТЯ ТЕОРІЇ ЙМОВІРНОСТЕЙ

1.1. Частота випадкових подій

Багато явищ, що відбуваються у світі, випадкові, тобто допускають різні наслідки. Так, виграш партії у шаховому матчі, отримання студентом відмінної оцінки на іспиті, випуск певного числа стандартних (якісних) вакцин за зміну залежать від різних чинників, що мають неоднаковий вплив і не можуть бути точно передбачені заздалегідь. Разом з тим деякі випадкові події за неодноразового повторення виявляють статистичну регулярність. Вона полягає в тому, що відношення числа ξ появ деякого фіксованого результату до загального числа n проведених експериментів (ξ/n), у кожному з яких цей результат міг відбутися або не відбутися, коливається навколо певної величини. Це відношення називають ***відносною частотою*** результату експерименту. В цьому випадку кажуть, що відносна частота випадкової події виявляє ***стійкість***. Наприклад, якщо підкидати монету багато разів, то відносна частота випадання "герба" буде близькою до 1/2. Так, англійський математик К. Пірсон підкинув монету $n = 24\,000$ разів, при цьому "герб" випав $\xi = 12\,012$ разів; відносна частота появи "герба" $\frac{\xi}{n} = \frac{12\,012}{24\,000}$ відрізняється від 1/2 усього на 0,0005.

Отже, за неодноразового підкидання монети виявляється деяка математична закономірність. Зрозуміло, що для виявлення математичних закономірностей інших випадкових явищ достатньо побудувати їх математичну модель, яка відображатиме статистичну регулярність. *Теорія ймовірностей є розділом математики, що вивчає математичні закономірності й моделі випадкових явищ, які виявляють стійкість відносних частот.*

Розглянемо первинні поняття теорії ймовірностей.

1.2. Випадковий експеримент. Випадкові події

Під ***випадковим експериментом*** розуміють деяку дію, яка здатна неодноразово повторюватися при відновленні певного комплексу умов, результат якої неможливо точно передбачити до її здійснення. Підкидання монети або грального кубика, шаховий турнір, час очікування автобуса на зупинці, стрільба з рушниці по мішені, перевірка на придатність партії медпрепаратів — усе це випадкові експерименти.

Випадковою подією називають результат випадкового експерименту. При проведенні експерименту випадкова подія може відбутися, а може й ні. Випадання "герба" за одноразового підкидання монети, випадання "шістки" при підкиданні грального кубика, прибуття автобуса через 1 хв після появи пасажира на зупинці, влучення в центр мішені при стрільбі з рушниці, стандартність виробу з деякої партії, яку перевіряють, є випадковими подіями.

Випадкові події називають ***рівноможливими***, якщо немає об'єктивних передумов, щоб віддати перевагу настанню однієї події відносно іншої. Наприклад, випадання "герба" або "числа" при підкиданні монети, випадання "одиниці", "двійки" і т. д. аж до "шістки" при підкиданні грального кубика є рівноможливими подіями, якщо монета і кубик симетричні за формою.

1.3. Елементарні події. Простір елементарних подій

Випадкову подію називають ***елементарною*** (***елементарним наслідком***), якщо її неможливо розкласти на простіші випадкові події. Наприклад, випадання "двійки" при підкиданні грального кубика є елементарною подією, а випадання непарної цифри — ні, оскільки ця подія може бути розкладена на простіші — випадання "одиниці", "трійки" або "п'ятірки".

Множину всіх елементарних подій деякого випадкового експерименту назвемо ***простором елементарних подій*** або ***ймовірнісним простором*** цього експерименту. Простір елементарних подій позначатимемо літерою Ω, елементарні події — літерою ω, $\omega \in \Omega$.

Будь-яку ***неелементарну випадкову подію*** можна інтерпретувати як множину всіх тих елементарних подій, на які вона може бути розкладена. При цьому кажуть, що кожна з таких елементарних подій сприяє появі зазначеної неелементарної випадкової події. Надалі випадкові події позначатимемо літерами A, B, C і т. д. Отже, кожна випадкова подія A є підмножиною свого простору елементарних подій: $A \subseteq \Omega$.

Приклад 1. У закритій клітці знаходиться 100 кролів, з яких 50 — альбіноси. Навмання обирають одного кроля. Описати простір елементарних подій Ω. Описати подію A — обраний кріль виявився альбіносом.

Розв'язання. Простір елементарних подій Ω складається з усіх кролів у клітці і містить 100 елементів. Подія A складається з усіх кролів-альбіносів і має 50 елементів.

Підкреслимо, що простір елементарних подій Ω може складатись зі скінченного, нескінченного зліченного або нескінченного незліченного числа елементарних подій. Так, простір Ω, який відповідає випадковому експерименту "відповідь студента на екзаменаційний квиток", складається з 4 елементарних подій:

ω_1 = {*отримана оцінка* "5"}; ω_2 = {*отримана оцінка* "4"};

ω_3 = {*отримана оцінка* "3"}; ω_4 = {*отримана оцінка* "2"}.

Простір Ω, який відповідає випадковому експерименту "підкидання монети до першого випадання герба" є нескінченним, але зліченним. Він складається з таких елементарних подій:

ω_1 = {герб випав при першому підкиданні};

ω_2 = {герб випав при другому підкиданні};

....

ω_n = {*герб випав при n-му підкиданні*}.

Простір Ω, який відповідає випадковому експерименту "визначення часу очікування автобуса на зупинці", є нескінченним і незліченним. Якщо інтервал між автобусами не більш як Т хвилин, то $\Omega = \{\omega : \omega \in [0, T]\}$.

Неелементарна випадкова подія відбувається, якщо відбувається хоча б один елементарний наслідок, який входить у цю подію. Так, подія A = {*При підкиданні грального кубика випала непарна цифра*} відбувається, якщо, наприклад, випадає цифра 3.

§ 2. ВИПАДКОВІ ПОДІЇ ТА ОПЕРАЦІЇ НАД НИМИ

При подальшому вивченні випадкових подій необхідно спиратися на наведені нижче визначення.

Визначення. Випадкову подію A називають ***достовірною***, якщо в результаті випадкового експерименту вона обов'язково відбувається.

Наприклад, подія Ω є достовірною.

Визначення. Випадкову подію A називають ***неможливою***, якщо в результаті випадкового експерименту вона не може відбутися.

Приклади неможливих подій: випадання цифри "7" при підкиданні шестигранного грального кубика; одночасне випадання "одиниці" й "двійки". Неможлива подія є порожньою множиною елементарних подій, її позначають $\varnothing$.

Визначення. Кажуть, що подія B *випливає* з події A, якщо A є підмножиною B: $A \subseteq B$.

Визначення. Подію C називають *перетином* або *перерізом* подій A і B (позначають $C = A \cap B$), якщо C складається з тих і тільки тих елементарних подій, які входять і в A, і в B одночасно.

Перетин двох подій відбувається тоді і тільки тоді, коли відбуваються обидві ці події.

Визначення. Подію C називають *об'єднанням* подій A і B (позначають $C = A \cup B$), якщо C складається з тих і тільки тих елементарних подій, які входять хоча б в одну з подій A або B.

Об'єднання двох подій відбувається тоді і тільки тоді, коли відбувається хоча б одна з цих подій.

Аналогічно визначають перетин і об'єднання будь-якого скінченного числа подій.

Визначення. Подію C називають *різницею* подій A і B (позначають $C = A \setminus B$), якщо C складається з тих і тільки тих елементарних подій, які входять в A, але не входять у B.

Різниця двох подій $A \setminus B$ відбувається тоді і тільки тоді, коли відбувається подія A, але не відбувається подія B.

Визначення. Подію $\overline{A}$ називають *протилежною* події A, якщо $\overline{A} = \Omega \setminus A$.

Подія $\overline{A}$ відбувається тоді і тільки тоді, коли не відбувається подія A.

Визначення. Події A і B називають *несумісними*, якщо $A \cap B = \varnothing$.

Несумісні події не можуть відбутися в результаті одного й того самого випадкового експерименту.

Визначення. Події A і B називають *сумісними*, якщо $A \cap B \neq \varnothing$.

Сумісні події не суперечать одна одній у результаті одного й того самого випадкового експерименту.

Проілюструємо операції над подіями за допомогою діаграм Ейлера—Венна (рис. 1; заштриховано подію, позначену над діаграмою).

Приклад 2. У першій клітці знаходиться 100 білих щурів, у другій — 30 білих і 70 чорних. Для проведення досліджень навмання обирають по одному щуру з першої та другої кліток. Вибір білого щура з першої клітки є *достовірною подією*, а вибір чорного щура — *неможливою подією*. Водночас вибір білого щура з другої клітки є *випадковою подією*.

Приклад 3. У клітці знаходяться здорові та інфіковані вірусом тварини. Навмання обирають одну тварину. Нехай A_1 — подія, яка полягає в тому, що обрана тварина здорова; подія A_2 — обрана тварина інфікована вірусом. Події A_1 та A_2 — *протилежні*.

Приклад 4. Події A_1 та A_2, які розглядали в прикладі 3 — *несумісні*.

Приклад 5. Для вивчення ефективності зв'язування молекул α і β з деяким білковим комплексом були розглянуті події A_1 та A_2. Подія A_1 полягає в тому, що молекула α активує комплекс.

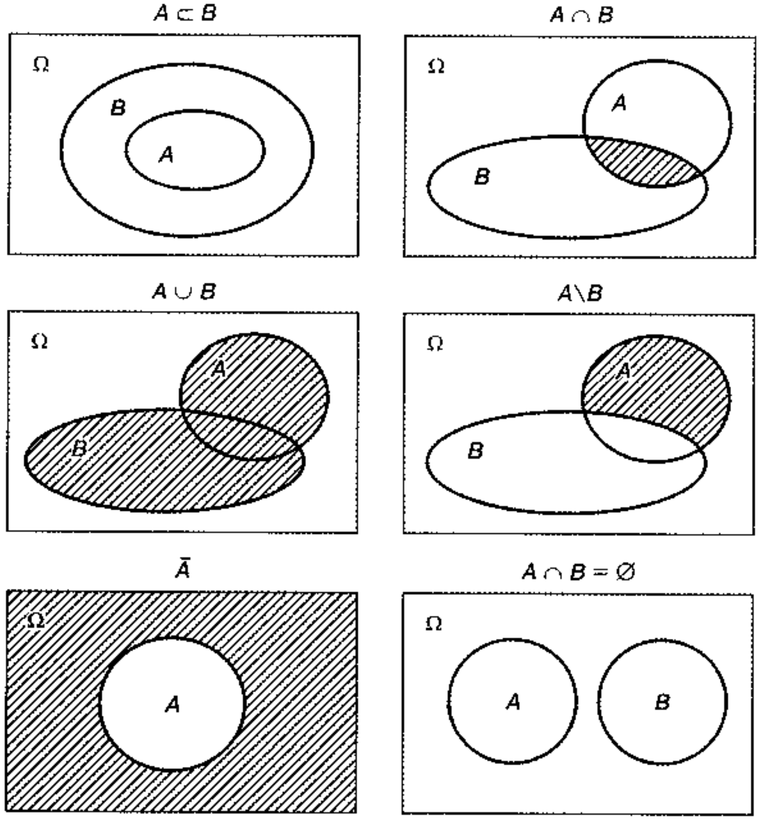


Рис. 1. Діаграми Ейлера—Венна

Подія A_2 полягає в тому, що молекула β активує комплекс. Події A_1 та A_2 *сумісні*.

Приклад 6. Відомо, що досліджуваний білок має три сайти зв'язування з препаратом. Розглядають три події: A_1 — активується перший сайт, A_2 — активується другий сайт, A_3 — активується третій сайт. Нехай подія B полягає в тому, що препарат зв'язується з першим або другим сайтом білка; подія C — препарат зв'язується з будь-яким сайтом білка. Для подій B і C справедливі такі зображення: $B = A_1 \cup A_2$; $C = A_1 \cup A_2 \cup A_3$.

Приклад 7. З метою вивчення залежності захворюваності на рак легенів від тютюнопаління обрали піддослідну групу, серед учасників якої можуть траплятися курці — подія A і хворі на рак — подія B. Нехай C_1 — навмання обрана особа палить і

хворіє на рак легенів; C_2 — не палить і хворіє на рак легенів; C_3 — палить і не хворіє на рак легенів; C_4 — не палить і не хворіє на рак легенів; C_5 — або хворіє на рак легенів, або палить. Тоді справедливі записи: $C_1 = A \cap B$; $C_2 = \overline{A} \cap B$; $C_3 = A \cap \overline{B}$; $C_4 = \overline{A} \cap \overline{B}$; $C_5 = C_2 \cup C_3 = (\overline{A} \cap B) \cup (A \cap \overline{B})$.

Приклад 8. Для виконання експериментального вимірювання група студентів може використовувати три прилади. Розглянемо такі події: A_1 — перший прилад вийде з ладу; A_2 — другий прилад вийде з ладу; A_3 — третій прилад вийде з ладу. Тоді події $\overline{A}_1$, $\overline{A}_2$, $\overline{A}_3$ полягають у тому, що жоден із приладів окремо не вийде з ладу; події $B_1 = A_1 \cap \overline{A}_2 \cap \overline{A}_3$, $B_2 = \overline{A}_1 \cap A_2 \cap \overline{A}_3$ і $B_3 = \overline{A}_1 \cap \overline{A}_2 \cap A_3$ полягають у тому, що з трьох приладів вийде з ладу лише перший, другий і третій прилади відповідно. Подія $F = B_1 \cup B_2 \cup B_3$ означає, що упродовж експерименту вийде з ладу лише один прилад. Подію G, яка полягає в тому, що під час експерименту вийде з ладу хоча б один прилад (тобто не менш як один прилад), можна зобразити так: $G = B_1 \cup B_2 \cup B_3 \cup C_1 \cup C_2 \cup C_3 \cup D$, де $C_1 = \overline{A}_1 \cap A_2 \cap A_3$, $C_2 = A_1 \cap \overline{A}_2 \cap A_3$, $C_3 = A_1 \cap A_2 \cap \overline{A}_3$, $D = A_1 \cap \cap A_2 \cap A_3$. Подія $H = \overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3$ означає, що жоден із трьох приладів не вийде з ладу.

В силу теоретико-множинної інтерпретації подій, операції над подіями мають такі самі властивості, що й операції над множинами. Для будь-яких подій $A, B, C \subseteq \Omega$ справедливі такі властивості:

1) $A \cap \Omega = A$; $A \cap \varnothing = \varnothing$; $A \cup \Omega = \Omega$; $A \cup \varnothing = A$; $A \cap \overline{A} = \varnothing$; $A \cup \overline{A} = \Omega$; $A \setminus B = A \cap \overline{B}$; $\varnothing \subseteq A \subseteq \Omega$ для будь-якої множини A;

2) $A \cap B = B \cap A$; $A \cup B = B \cup A$ (переставна властивість перетину й об'єднання);

3) $A \cap (B \cap C) = (A \cap B) \cap C$; $A \cup (B \cup C) = (A \cup B) \cup C$ (сполучна властивість перетину й об'єднання);

4) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (розподільна властивість перетину відносно об'єднання);

5) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (розподільна властивість об'єднання відносно перетину);

6) $\overline{A \cup B} = \overline{A} \cap \overline{B}, \overline{A \cap B} = \overline{A} \cup \overline{B}$ (формули Де Моргана).

§ 3. КЛАСИЧНЕ ВИЗНАЧЕННЯ ЙМОВІРНОСТІ ПОДІЇ

Як уже зазначалось, за неодноразового повторення випадкового експерименту, проведеного за однакових умов, відносна частота ξ/n появи деякої випадкової події може лише незначно коливатися відносно деякої сталої величини. Виходячи з цього, в теорії ймовірностей прийнято за *аксіому* той факт, що у такої випадкової події A існує числова характеристика $P(A)$, яку називають *ймовірністю* цієї *події*, біля якої і коливається частота появи події A.

Образно кажучи, ймовірність події — це міра об'єктивної впевненості в тому, що ця подія відбудеться. Розглянемо класичну схему.

Надалі для довільної події $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_m}\}$ через $|A|$ позначатимемо кількість елементарних подій, які входять в A, тобто $|A| = m$.

Визначення. Нехай простір $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ усіх елементарних подій деякого випадкового експерименту є *скінченним* і складається з N рівноможливих елементарних подій. Подія $A \subseteq \Omega$, $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_K}\}$ складається з K елементарних подій $(K \le N)$. Тоді ймовірність події A — це число

$$P(A) = \frac{|A|}{|\Omega|} = \frac{K}{N} \tag{1}$$

відношення числа елементарних подій, які сприяють події A, до загального числа елементарних подій.

Так, ймовірність випадання "герба" за одноразового підкидання монети дорівнює 1/2, а ймовірність випадання "одиниці" за одноразового підкидання грального кубика дорівнює 1/6.

Властивості ймовірності події

1. Для будь-якої події $A \subseteq \Omega$ $0 \le P(A) \le 1$.

Ця властивість випливає з того, що $0 \le |A| \le |\Omega|$. Зокрема, ймовірність достовірної події дорівнює 1, тобто $P(\Omega) = 1$, оскільки $|\Omega| = K = N$; ймовірність неможливої події дорівнює 0, тобто $P(\varnothing) = 0$, оскільки $|\varnothing| = 0$.

2. Теорема додавання ймовірностей для несумісних подій.

Якщо події A і B несумісні, то

$$P(A \cup B) = P(A) + P(B). \tag{2}$$

Справді, якщо $A \cap B = \varnothing$, то $|A \cup B| = |A| + |B|$, звідки і випливає властивість 2.

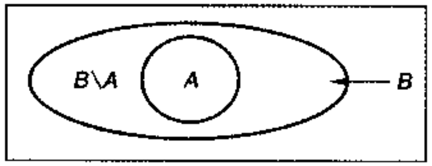


Рис. 2

Наслідок. Якщо події $A_1, A_2, \ldots, A_n$ попарно несумісні ($A_i \cap A_j = \varnothing$ за $i \ne j$), то ймовірність їх об'єднання дорівнює сумі ймовірностей цих подій:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

3. Ймовірність протилежної події $\overline{A}$ дорівнює

$$P(\overline{A}) = 1 - P(A). \tag{3}$$

Доведення. Оскільки $A \cup \overline{A} = \Omega$, тоді, згідно з властивістю 1,

$$P(A \cup \overline{A}) = P(\Omega) = 1.$$

Водночас $A \cap \overline{A} = \varnothing$, тоді, згідно з властивістю 2,

$$P(A \cup \overline{A}) = P(A) + P(\overline{A}).$$

З цих формул випливає: $P(A) + P(\overline{A}) = 1$, що і треба було довести.

4. Якщо подія B випливає з події A, то $P(A) \le P(B)$.

Доведення. Оскільки $A \subseteq B$, то B можна представити, як зображено на рис. 2: $B = A \cup (B \setminus A) = A \cup (B \cap \overline{A})$, де $A \cap (B \cap \overline{A}) = \varnothing$. Тому $P(B) = P(A \cup (B \cap \overline{A})) = P(A) + P(B \cap \overline{A}) \ge P(A)$.

5. Теорема додавання ймовірностей для довільних подій.

Для будь-яких подій A і B, що належать одному простору елементарних подій Ω,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{4}$$

Доведення. Подію $A \cup B$ зображено на рис. 3: $A \cup B = A \cup (B \setminus A) = A \cup (B \cap \overline{A})$, причому $A \cap (B \cap \overline{A}) = \varnothing$. Тому

$$P(A \cup B) = P(A) + P(B \cap \overline{A}).$$

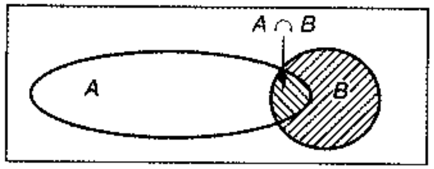


Рис. 3

Для B маємо: $B = (B \cap A) \cup (B \cap \overline{A})$, $(B \cap A) \cap (B \cap \overline{A}) = \varnothing$. Тому $P(B) = P(B \cap A) + P(B \cap \overline{A})$ і $P(B \cap \overline{A}) = P(B) - P(B \cap \overline{A})$.

Врахувавши ці вирази, отримаємо формулу (4).

Для більшого числа подій теорема додавання ймовірностей має такий вигляд: нехай $A_1, A_2, \ldots, A_n$ — довільні події з одного й того ж простору елементарних подій Ω. Тоді

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) +$$

$$+ \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right).$$

Приклад 9. У закритій клітці знаходиться 70 білих, 20 сірих і 10 чорних кролів. Яка ймовірність того, що навмання взятий з клітки кріль матиме сірий колір?

Розв'язання. Простір елементарних подій Ω складається з усіх кролів у клітці і містить 100 елементів. Подія A складається з усіх кролів сірого кольору і має 20 елементів. Тоді

$$P(A) = \frac{|A|}{|\Omega|} = \frac{K}{N} = \frac{20}{100} = \frac{1}{5}.$$

Приклад 10. За статистикою, 68 % чоловіків, які досягли 60-річного віку, доживають і до 70 років. Яка ймовірність того, що 60-річний чоловік не доживе до 70 років?

Розв'язання. Позначимо через A подію, яка полягає в тому, що 60-річний чоловік досягне свого 70-річчя. Тоді протилежна подія $\overline{A}$ — 60-річний чоловік не доживе до 70 років. Згідно з формулою (3), маємо: $P(\overline{A}) = 1 - P(A) = 1 - 0,68 = 0,32$. Отже, 32 % 60-річних чоловіків не проживуть наступні 10 років.

3.1. Геометричні ймовірності

Звернемо увагу на той випадок, коли Ω є деякою областю n-вимірного координатного простору (у практичних застосуваннях $n = 1, 2, 3$). Розглянемо випадок $n = 2$. Нехай Ω — деяка квадровна область на площині. Позначимо через F систему квадровних областей $A \subseteq \Omega$. Для будь-якої області $A \subseteq F$ покладемо, за визначенням,

$$P(A) = \frac{S(A)}{S(\Omega)},$$

де $S(\cdot)$ — площа відповідної області.

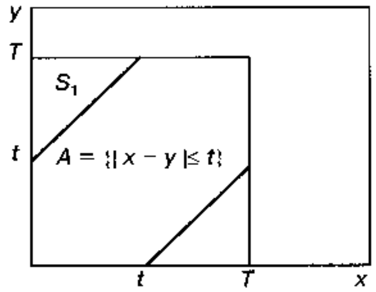


Рис. 4

Нескладно перевірити, що визначена так ймовірність задовольняє властивості 1—5 (див. § 3). Її називають *геометричною ймовірністю.*

Приклад 11 (задача про зустріч). Дві людини домовилися про зустріч. Кожна з них може прийти на місце зустрічі в інтервалі часу між 0 і T, чекає t хвилин і потім іде. Яка ймовірність того, що вони зустрінуться?

Розв'язання. Нехай x — момент приходу однієї людини, y — момент приходу іншої. Тоді "сумісний момент" їхнього приходу (x, y) можна інтерпретувати як точку на координатній площині з відповідними координатами. Такі точки заповнюють квадрат зі стороною T, тобто $\Omega = [0, T] \times [0, T]$. Дві людини зустрінуться (відбудеться подія A), якщо $|x - y| \le t$. Область $A = \{(x, y) \in |x - y| \le t\}$ зустрічі людей зображена на рис. 4.

Оскільки $S(\Omega) = T^2$ і $S(A) = T^2 - 2S_1 = T^2 - 2 \cdot \frac{1}{2}(T - t)^2 = T^2 - (T - t)^2$, шукана ймовірність $P(A) = \frac{S(A)}{S(\Omega)} = \frac{T^2 - (T - t)^2}{T^2}$.

§ 4. УМОВНІ ЙМОВІРНОСТІ ПОДІЙ

На практиці для визначення ймовірності подій застосовують переважно непрямі методи, які дають можливість за відомими ймовірностями одних подій знайти ймовірності інших. Для цього розглядають додаткові ймовірнісні конструкції, такі як *незалежні події, умовні ймовірності, формула повної ймовірності, формула Байєса.*

Перш ніж сформулювати визначення умовної ймовірності події, розглянемо таку схему. Нехай випадковий експеримент проводиться у два етапи. На першому етапі може відбутися або не відбутися деяка подія A і настання або ненастання події A *впливає* на другий етап експерименту. Тоді ймовірність деякої події B, яка може відбутися або не відбутися на другому етапі експерименту, обчислена *за умови*, що подія A *відбулася* (умовна ймовірність події B за умови A, $P(B/A)$) відрізнятиметься від звичайної (*безумовної*) ймовірності події B, $P(B)$.

Приклад 12. У класі навчається 25 учнів. Із них 6 учнів захоплюються тільки біологією, 5 учнів — тільки математикою, інші — і біологією, і математикою. Визначити ймовірність того, що учень, обраний навмання: а) захоплюється біологією; б) захоплюється біологією за умови, що він захоплюється і математикою.

Розв'язання. а) Нехай подія B полягає в тому, що учень, обраний навмання, захоплюється біологією. Оскільки всього учнів 25, із них 20 захоплюються біологією, тоді $P(B) = \frac{20}{25} = 0,8$.

б) Тепер випадковий експеримент проведемо в два етапи: спочатку з'ясуємо, чи захоплюється учень математикою, потім — біологією. На першому етапі відбувається подія A — учень, обраний навмання, захоплюється математикою. Ймовірність події B обчислюється за умови, що відбулася подія A. Оскільки учнів, які захоплюються математикою — 19, із них 14 — захоплюються також біологією, то $P(B/A) = \frac{14}{19} \approx 0,736$. Зауважимо, що у цьому випадку $P(B/A) = \frac{14}{25} : \frac{19}{25} = \frac{P(A \cap B)}{P(A)}$.

Наведений приклад пояснює визначення умовної ймовірності події.

Визначення. Нехай ймовірність події A не дорівнює нулю. ***Умовною ймовірністю*** події B за умови, що відбулася подія A, називають число $P(B/A) = \frac{P(A \cap B)}{P(A)}$.

Властивості умовної ймовірності

1. $0 \le P(B/A) \le 1$.

2. Якщо події B і C несумісні, то $P(B \cup C/A) = P(B/A) + P(C/A)$.

З наведеного визначення випливає також ***формула множення ймовірностей***: для будь-яких подій A і B з одного й того ж простору елементарних подій

$$P(A \cap B) = P(B/A)P(A) = P(A/B)P(B). \qquad (5)$$

Приклад 13. Відомо, що в групі з 15 тварин 3 інфіковані вірусом. Навмання для подальших досліджень обирають двох тварин. Яка ймовірність того, що обидві тварини: *1*) хворі, *2*) здорові?

Розв'язання. Нехай події A і B полягають у тому, що перша і друга тварини інфіковані. Тоді у випадку 1) за формулою (5) отримаємо: $P(A \cap B) = P(B/A)P(A) = \frac{2}{14} \cdot \frac{3}{15} = \frac{1}{35}$. Аналогічно у випадку 2) для подій A і B, які полягають у тому, що перша і друга тварини здорові: $P(A \cap B) = P(B/A)P(A) = \frac{11}{14} \cdot \frac{12}{15} = \frac{22}{35}$.

Формулу множення ймовірностей можна узагальнити для будь-якої кількості скінченного числа подій.

Нехай $A_1, A_2, \ldots, A_n$ — довільні події, задані на одному й тому ж просторі елементарних подій Ω. Тоді

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2)\ldots P(A_n/\bigcap_{i=1}^{n} A_i). \qquad (6)$$

Справді, відповідно до формули (5), отримаємо

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_n/A_1 \cap A_2 \cap \ldots \cap A_{n-1})\, P(A_1 \cap \ldots \cap A_{n-1}) =$$

$$= P(A_n/A_1 \cap A_2 \cap \ldots \cap A_{n-1})\, P(A_{n-1}/A_1 \cap A_2 \cap \ldots \cap A_{n-2}) \times$$

$$\times P(A_1 \cap A_2 \cap \ldots \cap A_{n-2}) = \ldots =$$

$$= P(A_n/A_1 \cap \ldots \cap A_{n-1})\, P(A_{n-1}/A_1 \cap \ldots \cap A_{n-2})\ldots P(A_1).$$

Цей вираз відрізняється від правої частини залежності (6) лише порядком співмножників.

Приклад 14. У папці знаходяться 12 заявок на постачання лабораторного біоматеріалу, з них 5 заявок — від вітчизняних виробників, 7 — від іноземних. Обирають навманная 4 заявки. Яка ймовірність того, що всі вони — від іноземних підприємств?

Розв'язання. Нехай подія A_i полягає в тому, що i-та обрана заявка — від іноземних підприємств, $i = 1, 2, 3, 4$. Потрібно обчислити $P(A_1 \cap A_2 \cap A_3 \cap A_4)$. Згідно з формулою (6),

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)\, P(A_2/A_1)\, P(A_3/A_1 \cap A_2) \times$$

$$\times P(A_4/A_1 \cap A_2 \cap A_3) = \frac{7}{12} \cdot \frac{6}{11} \cdot \frac{5}{10} \cdot \frac{4}{9} = \frac{7}{99}.$$

4.1. Незалежні події та їх властивості

Вивчивши поняття умовної ймовірності, варто звернути особливу увагу на незалежні події.

Визначення. Події A і B, задані на одному й тому самому просторі елементарних подій Ω, називають незалежними, якщо

$$P(A \cap B) = P(A)P(B). \qquad (7)$$

Порівнявши формули (5) і (7), для незалежних подій A і B отримаємо $P(B/A) = P(B)$, $P(A/B) = P(A)$, тобто ці події не впливають одна на одну в ймовірнісному сенсі.

Лема. Нехай події A і B — незалежні. Тоді події A і $\overline{B}$, $\overline{A}$ і B, $\overline{A}$ і $\overline{B}$ також незалежні.

Доведення. Справді, $A = (A \cap \overline{B}) \cup (A \cap B)$, де $(A \cap \overline{B}) \cap (A \cap B) = \varnothing$. Тому $P(A) = P(A \cap \overline{B}) + P(A \cap B)$, тобто $P(A \cap \overline{B}) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\overline{B})$, звідки, за визначенням, випливає незалежність подій A і $\overline{B}$. Незалежність інших пар подій доводиться аналогічно.

Приклад 15. Ймовірність інфікуватися вірусом для щура становить 0,5, а для кроля — 0,6. Для дослідження дії вірусу обрано дві тварини. Знайти ймовірність інфікування вірусом обох тварин.

Розв'язання. Нехай подія A — обраний щур інфікований вірусом, подія B — обраний кріль інфікований вірусом. Ці події незалежні. Тому маємо: $P(A \cap B) = P(A)P(B) = 0,5 \cdot 0,6 = 0,3$.

Приклад 16. Відомо, що ймовірність вилікувати хвору тварину вакциною A_1 дорівнює 0,9, вакциною A_2 — 0,8, вакциною A_3 — 0,7. Вакцини діють незалежно. Знайти ймовірність того, що застосування хоча б однієї з цих вакцин дасть позитивний лікувальний ефект.

Розв'язання. Позначимо через C подію, ймовірність якої потрібно обчислити, через B_1, B_2 і B_3 — події, які полягають у позитивному лікувальному ефекті від дії вакцин відповідно A_1, A_2 та A_3. За умовою задачі, ці події незалежні. Отже, події $\overline{B}_1$, $\overline{B}_2$ і $\overline{B}_3$ — також незалежні. З урахуванням формул (3) і (7) отримаємо

$$P(C) = 1 - P(\overline{B}_1 \cap \overline{B_2} \cap \overline{B}_3) = 1 - P(\overline{B}_1)P(\overline{B_2})P(\overline{B}_3) =$$

$$= 1 - (1 - 0,9) \cdot (1 - 0,8) \cdot (1 - 0,7) = 1 - 0,1 \cdot 0,2 \cdot 0,3 = 0,994.$$

Приклад 17. Відомо, що піддослідна тварина може бути незалежно інфікована вірусом α з імовірністю 0,7 і вірусом β — з імовірністю 0,8. Яка ймовірність того, що піддослідна тварина інфікована вірусом α або вірусом β?

Розв'язання. Розглянемо події A — інфікування вірусом α і B — інфікування вірусом β. За умовою задачі, ці події незалежні. Тоді, згідно з формулами (4) і (7),

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) =$$

$$= 0,7 + 0,8 - 0,7 \cdot 0,8 = 0,94.$$

Приклад 18. Три програмісти незалежно один від одного складають окремі частини спільної програми для ПК. Ймовірність того, що перший програміст припуститься помилки у своїй частині програми, дорівнює $P_1 = 0,15$; для другого і третього програмістів ці ймовірності відповідно дорівнюють $P_2 = 0,2$ і $P_3 = 0,08$. Знайти ймовірність того, що: а) тільки один програміст припуститься помилки; б) хоча б один програміст припуститься помилки; в) усі три програмісти припустяться помилок.

Розв'язання. а) Нехай подія A_i полягає в тому, що i-й програміст припуститься помилки $(i = 1, 2, 3)$. Розглянемо подію $A = \{$*Тільки один програміст припуститься помилки*$\}$. Тоді $A = (A_1 \cap \overline{A}_2 \cap \overline{A}_3) \cup (\overline{A}_1 \cap A_2 \cap \overline{A}_3) \cup (\overline{A}_1 \cap \overline{A}_2 \cap A_3)$, причому події в дужках попарно несумісні. В силу незалежності подій A_1, A_2, A_3, а також леми

$$P(A) = P(A_1 \cap \overline{A}_2 \cap \overline{A}_3) + P(\overline{A}_1 \cap A_2 \cap \overline{A}_3) + P(\overline{A}_1 \cap \overline{A}_2 \cap A_3) =$$

$$= P(A_1)P(\overline{A}_2)P(\overline{A}_3) + P(\overline{A}_1)P(A_2)P(\overline{A}_3) + P(\overline{A}_1)P(\overline{A}_2)P(A_3) =$$

$$= 0,15 \cdot 0,8 \cdot 0,92 + 0,85 \cdot 0,2 \cdot 0,92 + 0,85 \cdot 0,8 \cdot 0,08 = 0,3212.$$

б) Позначимо через B подію, яка полягає в тому, що хоча б один із програмістів припуститься помилки. В цьому разі простіше знайти $P(\overline{B})$ і скористатися формулою (3), де $\overline{B} = \{$*Жоден програміст не припуститься помилки*$\}$.

$$P(\overline{B}) = P(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3) = P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3) =$$

$$= 0,85 \cdot 0,8 \cdot 0,92 = 0,6256.$$

Тому $P(B) = 1 - P(\overline{B}) = 1 - 0,6256 = 0,3744$.

в) Нехай подія C полягає в тому, що всі три програмісти припустяться помилок. Тоді $P(C) = P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2) \times$ $\times P(A_3) = 0{,}15 \cdot 0{,}2 \cdot 0{,}08 = 0{,}0024$.

4.2. Формула повної ймовірності. Формула Байєса

Формули цього розділу випливають з означення умовної ймовірності. Їх, як правило, застосовують у випадках, коли поява події залежить від кількох інших подій, які називають *гіпотезами*. Наведемо визначення відповідних понять.

Визначення. Події $H_1, H_2, ..., H_n$ утворюють ***повну групу подій***, якщо

1) вони попарно несумісні: $H_i \cap H_j = \varnothing$, коли $i \neq j$;

2) їх об'єднання утворює весь простір елементарних подій: $H_1 \cup H_2 \cup ... \cup H_n = \Omega$.

Теорема 1 (*формула повної ймовірності*). Нехай події $H_1, H_2, ..., H_n$ утворюють повну групу подій. Тоді для довільної події A справедлива формула

$$P(A) = \sum_{i=1}^{n} P(A / H_i) P(H_i) =$$

$$= P(A / H_1) P(H_1) + P(A / H_2) P(H_2) + ... + P(A / H_n) P(H_n). \quad (8)$$

Доведення. Подію A можна подати у вигляді $A = (A \cap H_1) \cup$ $\cup (A \cap H_2) \cup ... \cup (A \cap H_n)$, причому $(A \cap H_i) \cap (A \cap H_j) = \varnothing$ якщо $i \neq j$ (рис. 5). Отже, в силу наслідку з властивості ймовірності 2 (див. с. 15) маємо

$$P(A) = P(A \cap H_1) + P(A \cap H_2) + ... + P(A \cap H_n).$$

Застосувавши формулу (5), отримаємо $P(A \cap H_1) = P(A / H_1) \times$ $\times P(H_1)$, $P(A \cap H_2) = P(A / H_2) P(H_2)$, ..., $P(A \cap H_n) = P(A / H_n) \times$ $\times P(H_n)$, звідки і випливає шукана формула.

Зауваження. Події $H_1, H_2, ..., H_n$ називають *гіпотезами*, а ймовірності $P(H_1), P(H_2), ..., P(H_n)$ — апріорними ймовірностями гіпотез (лат. слово a priori буквально означає — з попереднього, у переносному значенні — без перевірки, *до досвіду*). Формулу повної ймовірності застосовують у тих випадках, коли деяка подія A може відбутися в результаті настання однієї з кількох гіпотез $H_1, H_2, ..., H_n$, і необхідно обчислити її "повну" ймовірність з урахуванням усіх гіпотез.

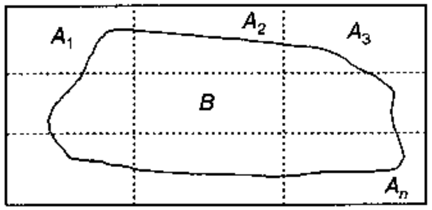


Рис. 5

Приклад 19. Лабораторія, в якій вимірюють спектри, оснащена спектрофотометрами двох виробників: 6 приладів одного виробника і 4 прилади — другого. Ймовірність виходу з ладу приладу першого і другого виробників під час проведення експерименту становить відповідно 0,05 і 0,15. Спектри вимірюють на довільно обраному приладі. Визначити ймовірність того, що під час експерименту прилад не вийде з ладу.

Розв'язання. Нехай A — подія, яка відповідає тому, що під час експерименту прилад вийде з ладу; H_1 і H_2 — події (гіпотези), які відповідають тому, що для вимірювання обрано прилади відповідно першого і другого виробників. Тоді $P(H_1) = 0{,}6$ і $P(H_2) = 0{,}4$; $P(A / H_1) = 0{,}05$ і $P(A / H_2) = 0{,}15$. Згідно з формулою повної ймовірності (8), $P(A) = 0{,}6 \cdot 0{,}05 + 0{,}4 \cdot 0{,}15 = 0{,}09$. Отже, $P(\bar{A}) =$ $= 1 - P(A) = 0{,}91$.

Приклад 20. У першій клітці знаходиться 4 чорних і 2 білих щури, у другій — 1 чорний і 3 білих щури. З першої клітки в другу пересадили одного щура. Знайти ймовірність того, що навмання взятий щур із другої клітки буде білим.

Розв'язання. Нехай A — подія, яка полягає в тому, що навмання взятий з другої клітки щур буде білим. H_1 і H_2 — події (гіпотези), які відповідають тому, що з першої клітки в другу пересадили відповідно білого і чорного щура. Тоді $P(H_1) = 1/3$ і $P(H_2) = 2/3$; $P(A / H_1) = 4/5$ і $P(A / H_2) = 3/5$. За формулою повної ймовірності (8) отримаємо $P(A) = \frac{1}{3} \cdot \frac{4}{5} + \frac{2}{3} \cdot \frac{3}{5} = \frac{2}{3}$.

Теорема 2 (*формула Байєса*). Нехай події $H_1, H_2, ..., H_n$ утворюють повну групу подій і $P(A) \neq 0$. Тоді справедливою є формула

$$P(H_i / A) = \frac{P(A / H_i) P(H_i)}{\sum_{j=1}^{n} P(A / H_j) P(H_j)}, \quad i = 1, 2, ..., n. \quad (9)$$

Доведення. За визначенням умовної ймовірності

$$P(H_i / A) = \frac{P(H_i \cap A)}{P(A)}. \qquad (9а)$$

Перетворимо чисельник за формулою (5): $P(H_i \cap A) = P(A / H_i)\, P(H_i)$; знаменник — за формулою повної ймовірності: $P(A) = \sum_{j=1}^{n} P(A / H_j)\, P(H_j)$. Підставивши ці вирази у чисельник і знаменник правої частини співвідношення (9а), отримаємо шукану формулу (9).

Зауваження. Формулу Байєса застосовують у тих випадках, коли подія B може відбутися в результаті настання однієї з кількох гіпотез, причому відомо, що подія A вже відбулася. Тоді можна "перерахувати" ймовірності гіпотез $H_1, H_2, \dots, H_n$ за умови, що подія A відбулася. Такі "перераховані" ймовірності $P(H_i/A)$, $i = 1, 2, \dots, n$ називають апостеріорними (лат. a posteriori буквально означає — з наступного, залежно від *досвіду* з практики).

Приклад 21. Дві лабораторії проаналізували один і той самий зразок біоматеріалу з метою виявлення вірусу. Ймовірність того, що в першій і другій лабораторіях результати аналізу будуть недостовірними, становить відповідно 0,05 і 0,15. При перевірці було виявлено недостовірність результату аналізу. Визначити ймовірність того, що помилка була допущена в першій лабораторії.

Розв'язання. Нехай A — подія, яка полягає в тому, що аналіз зроблено неправильно; H_1 — аналіз зроблено у першій лабораторії; H_2 — аналіз зроблено у другій лабораторії. Оскільки немає переваг щодо вибору лабораторії, то $P(H_1) = P(H_2) = 0{,}5$. Умовна ймовірність того, що перша і друга лабораторії під час аналізу допускають помилку, становить відповідно $P(A / H_1) = 0{,}05$ і $P(A / H_2) = 0{,}15$. За формулою Байєса (9) отримаємо

$$P(H_1 / A) = \frac{0{,}5 \cdot 0{,}05}{0{,}5 \cdot 0{,}05 + 0{,}5 \cdot 0{,}15} = 0{,}25.$$

Приклад 22. До діагностичного центру в однакових кількостях потрапляють пацієнти з трьох консультативних пунктів. Ймовірність того, що діагноз буде підтверджений для пацієнтів, направлених першим пунктом, становить 0,6, другим — 0,78, третім — 0,9. Яка ймовірність того, що: 1) діагноз підтвердиться

у навмання обраного пацієнта; 2) пацієнт був направлений третім консультативним пунктом?

Розв'язання. Нехай A — подія, яка полягає в тому, що діагноз підтвердиться у навмання обраного пацієнта; H_1 — діагноз поставлено у першому, H_2 — у другому, H_3 — у третьому консультативному пункті. Оскільки немає переваг щодо вибору пацієнтом того чи іншого консультативного пункту, то $P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}$. У випадку 1), згідно з формулою повної ймовірності (8), отримаємо $P(A) = \sum_{i=1}^{3} P(A / H_i) P(H_i) = \frac{1}{3}(0{,}6 + 0{,}78 + 0{,}9) = 0{,}76$, у випадку 2), згідно з формулою Байєса (9),

$$P(H_3 / A) = \frac{0{,}9 \cdot \frac{1}{3}}{0{,}76} \approx 0{,}39.$$

Приклад 23. Два ехолоти незалежно повідомляють про появу косяка риби. Ймовірності того, що при появі косяка риби спрацюють відповідно перший і другий ехолоти становлять 0,2 і 0,1. Косяк риби було виявлено. Знайти ймовірність того, що при появі косяка риби спрацює лише перший ехолот.

Розв'язання. Нехай подія A_1 — спрацював перший ехолот, подія A_2 — спрацював другий ехолот, подія B — косяк риби виявлено. Розглянемо такі гіпотези: $H_1 = A_1 \cap A_2$; $H_2 = A_1 \cap \overline{A}_2$; $H_3 = \overline{A}_1 \cap A_2$; $H_4 = \overline{A}_1 \cap \overline{A}_2$. Тоді $B = H_1 \cup H_2 \cup H_3$. Врахувавши незалежність подій A_1 та A_2, знайдемо ймовірності цих гіпотез: $P(H_1) = P(A_1 \cap A_2) = P(A_1)P(A_2) = 0{,}2 \cdot 0{,}1 = 0{,}02$. Аналогічно $P(H_2) = 0{,}18$; $P(H_3) = 0{,}08$; $P(H_4) = 0{,}72$. Далі обчислимо умовні ймовірності $P(B/H_1) = P(B/H_2) = P(B/H_3) = 1$; $P(B/H_4) = 0$. За формулою Байєса отримаємо

$$P(H_2/B) =$$

$$= \frac{P(B / H_2) P(H_2)}{P(B / H_1) P(H_1) + P(B / H_2) P(H_2) + P(B / H_3) P(H_3)} =$$

$$= \frac{0{,}18 \cdot 1}{0{,}02 \cdot 1 + 0{,}18 \cdot 1 + 0{,}08 \cdot 1} = 0{,}64.$$

§ 5. ЕЛЕМЕНТИ КОМБІНАТОРИКИ

Для підрахунку ймовірностей за формулою (1) у класичній схемі часто необхідно знати число способів, яким можна вибрати різні об'єкти із загальної їх сукупності. Задачі такого типу є предметом комбінаторики — математичної науки, що вивчає кількісні характеристики підмножин скінченної множини. Багато формул комбінаторики базується на такому простому твердженні.

Основне правило комбінаторики (*правило множення*). Нехай потрібно послідовно виконати k незалежних операцій. Якщо першу операцію можна виконати n_1 способами, другу — n_2 способами, третю — n_3 способами і т. д., k-ту — n_k способами, тоді послідовність усіх k операцій можна виконати $n_1 n_2 n_3 \ldots n_k$ способами.

Довести це твердження просто, скориставшись методом математичної індукції.

Приклад 24. У групі 15 студентів. Скількома способами можна обрати серед них старосту і профорга?

Розв'язання. Існує 15 способів вибору старости групи. Після того, як староста обраний, профоргом можна обрати будь-кого з 14 студентів, які залишилися. Отже, одному способу вибору старости відповідає 14 способів вибору профорга. Згідно з основним правилом комбінаторики, старосту і профорга можна обрати $15 \cdot 14 = 210$ способами.

Приклад 25. Для побудови біополімеру дослідник використав три різні мономери, кожен з яких складався відповідно з 7, 9 й 11 субодиниць. Скільки різних типів біополімерів, що складаються з цих мономерів, отримає дослідник?

Розв'язання. Згідно з основним правилом комбінаторики, кількість різних типів біополімерів, що складаються із зазначених мономерів, становить $N = 7 \cdot 9 \cdot 11 = 693$ одиниці.

Найважливішими поняттями комбінаторики є *розміщення, перестановки* і *сполучення.*

Розміщення без повторень. Нехай маємо деяку множину C, яка складається з n різних елементів. Розміщеннями з n елементів по k елементів ($k \le n$) називають усі впорядковані підмножини C, які складаються з k різних елементів. Розміщення відрізняються одне від одного або самими елементами, або їх порядком. Нехай, наприклад, $C = \{1,2,3,4\}$, $k = 3$; тоді всі можливі

розміщення мають вигляд:

$$\{1,2,3\}, \{1,2,4\}, \{1,3,2\}, \{1,3,4\}, \{1,4,2\}, \{1,4,3\},$$
$$\{2,1,3\}, \{2,3,4\}, \{2,3,1\}, \{2,3,4\}, \{2,4,1\}, \{2,4,3\},$$
$$\{3,1,2\}, \{3,1,4\}, \{3,2,1\}, \{3,2,4\}, \{3,4,1\}, \{3,4,2\},$$
$$\{4,1,2\}, \{4,1,3\}, \{4,2,1\}, \{4,2,3\}, \{4,3,1\}, \{4,3,2\}.$$

Теорема 1. Число розміщень з n елементів по k елементів дорівнює

$$A_n^k = n(n-1)(n-2)\ldots(n-k+1) = \frac{n!}{(n-k)!}. \tag{10}$$

Доведення. Утворимо всі можливі розміщення з n елементів по k елементів. При цьому перший елемент можна обрати n способами; після того як обраний n способами перший елемент, другий елемент можна обрати з тих, які залишились, $(n-1)$ способом, третій елемент — $(n-2)$ способами і т. д.; останній, k-й елемент, можна обрати $n-(k-1) = n-k+1$ способами. Так утворяться всі можливі розміщення. Згідно з основним правилом комбінаторики, шукане число розміщень дорівнює $n(n-1)\ldots(n-k+1)$.

Розміщення застосовують при описанні тих об'єктів, для яких важливий порядок розміщення елементів.

Приклад 26. Студенти 2-го курсу, згідно з навчальним планом, вивчають 12 дисциплін. На один день можна планувати 4 різні дисципліни. Скількома способами можна скласти розклад занять на один день?

Розв'язання. Усі можливі розклади занять на один день відрізняються один від одного як набором різних дисциплін у розкладі, так і порядком цих дисциплін. Отже, кількість таких розкладів точно дорівнює кількості розміщень з 12 по 4, а саме $A_{12}^4 = 12 \cdot 11 \cdot 10 \cdot 9 = 11\ 880$.

Приклад 27. Абонент забув 3 останні цифри телефонного номера, але пам'ятає, що всі вони різні. Яка ймовірність того, що, набираючи номер навмання, він набере його правильно?

Розв'язання. Очевидно, у телефонному номері відіграє роль порядок цифр. Загальне число способів набрати 3 різні цифри телефонного номера дорівнює $A_{10}^4 = 10 \cdot 9 \cdot 8 = 720$. Нехай подія

B полягає в тому, що номер набраний правильно. Тоді B включає одну елементарну подію, отже $P(B)=\frac{1}{720}\approx 0,00138$.

Перестановки. Нехай множина C складається з n різних елементів. Перестановками з n елементів називають усі можливі розміщення цих елементів. Нехай, наприклад, $C=\{1,2,3\}$; тоді всі перестановки матимуть вигляд

$$\{1,2,3\},\ \{1,3,2\},\ \{2,3,1\},\ \{2,1,3\},\ \{3,1,2\},\ \{3,2,1\}.$$

Теорема 2. Число P_n перестановок з n елементів дорівнює $n!$ (читається n-факторіал і означає добуток послідовних натуральних чисел від 1 до n: $n!=1\cdot 2\cdot 3\cdot\ldots\cdot n$; $0!=1$ за визначенням).

Доведення теореми випливає з визначення перестановок і теореми 1, тому що

$$P_n=A_n^n=n(n-1)(n-2)\ldots(n-n+1)=1\cdot 2\cdot 3\cdot\ldots\cdot n=n! \qquad (11)$$

Приклад 28. Відомо, що речовина, яку потрібно синтезувати в лабораторних умовах, складається з 10 різних хімічних елементів. Формула речовини залежить від послідовності сполучення елементів. Визначити кількість способів, за допомогою яких можна створити таку речовину.

Розв'язання. Загальне число способів розміщення 10 різних елементів один за одним у ряд дорівнює

$$P_{10}=10!=1\cdot 2\cdot 3\cdot\ldots\cdot 10=3\,628\,800.$$

Сполуки (комбінації) без повторень. Нехай множина C складається з n різних елементів. Сполуками (комбінаціями) з n елементів по k елементів ($k\le n$) називають усі невпорядковані підмножини C, які складаються з k елементів. Сполуки відрізняються одна від одної лише складом елементів. Нехай, наприклад, $C=\{1,2,3,4\}, k=3$; тоді всі можливі сполуки матимуть вигляд $\{1,2,3\}$, $\{1,2,4\}$, $\{1,3,4\}$, $\{2,3,4\}$.

Теорема 3. Число сполук (комбінацій) з n елементів по k елементів дорівнює

$$C_n^k=\frac{n(n-1)(n-2)\ldots(n-k+1)}{1\cdot 2\cdot\ldots\cdot k}=\frac{n!}{k!(n-k)!}. \qquad (12)$$

Доведення. Утворимо всі можливі сполуки з n елементів по k елементів. Їх число дорівнює C_n^k. Потім елементи кожної сполуки впорядкуємо, тобто переставимо всіма можливими способами. У кожній сполуці елементи можна переставити P_k способами. Після перестановки отримаємо $C_n^kP_k$ вибірок. Ці вибірки, зрозуміло, є всіма можливими розміщеннями з n елементів по k елементів, тобто $C_n^kP_k=A_n^k$, звідки

$$C_n^k=A_n^k\,/\,P_k=\frac{n(n-1)\ldots(n-k+1)}{k!}.$$

Помноживши чисельник і знаменник на $(n-k)!$, матимемо

$$C_n^k=\frac{n(n-1)(n-2)\ldots(n-k+1)(n-k)!}{k!(n-k)!}=\frac{n!}{k!(n-k)!}.$$

Сполуки застосовують при описі тих об'єктів, для яких порядок елементів не важливий, а важлива відсутність чи наявність певних елементів.

Властивості сполук (комбінацій) без повторень:

1) $C_n^1=n$; $C_n^n=C_n^0=1$; $C_n^k=C_n^{n-k}$;

2) формула бінома Ньютона $(a+b)^n=\sum_{k=0}^{n}C_n^ka^{n-k}b^k$;

3) $\sum_{k=0}^{n}C_n^k=C_n^0+C_n^1+C_n^2+\cdots+C_n^n=2^n$.

Властивість 3) є наслідком формули бінома Ньютона, якщо в ній покласти $a=b=1$.

Приклад 29. Для подарунка покупець має обрати 4 з 10 різних книг. Скількома способами він може це зробити?

Розв'язання. Оскільки порядок вибору книг не має значення, кількість наборів книг дорівнюватиме числу комбінацій з 10 по 4, тобто $C_{10}^4=\frac{10!}{4!\,6!}=\frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdot 8\cdot 9\cdot 10}{1\cdot 2\cdot 3\cdot 4\cdot 1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}=210.$

Приклад 30. Скількома способами з групи 20 біотехнологів можна обрати двох так, щоб: 1) направити їх на практику в певне фармацевтичне підприємство; 2) вони обіймали різні посади?

Розв'язання. У випадку 1) порядок відбору біотехнологів не має значення, тому кількість усіх можливих способів буде $C_{20}^2=$ $=\frac{20!}{2!\,18!}=190$. Навпаки, у випадку 2) порядок відбору біотехнологів має значення, тому $A_{20}^2=\frac{20!}{18!}=380$.

Приклад 31. У басейні плаває 10 риб, з яких 8 коропів і 2 карасі. Навмання виловлюють 5 риб. Знайти ймовірність того, що усі виловлені риби будуть коропами.

Розв'язання. Згідно з умовою задачі, немає значення, в якому порядку риби будуть виловлені, тому кількість усіх можливих елементарних подій становитиме $|\Omega| = C_{10}^5 = \frac{10!}{5!\,5!} = 252$. Нехай подія A полягає в тому, що усі виловлені риби — коропи. Тоді кількість комбінацій вилову 5 коропів з 8 дорівнює

$$C_8^5 = \frac{8!}{5!\,3!} = \frac{1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7\cdot8}{1\cdot2\cdot3\cdot4\cdot5\cdot1\cdot2\cdot3} = 56. \text{ Отже, } P(A) = \frac{C_8^5}{C_{10}^5} = \frac{2}{9}.$$

Приклад 32. У клітці знаходяться 9 кролів, 4 з яких інфіковані вірусом α_1 і 5 — вірусом α_2. З клітки навмання дістають 2 кролі. Знайти ймовірність того, що обидва кролі інфіковані різними вірусами.

Розв'язання. Усього кролів 9, тому простір елементарних подій складається з $C_9^2 = \frac{9!}{2!\,7!} = \frac{1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7\cdot8\cdot9}{1\cdot2\cdot1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7} = 36$ елементів. Нехай подія A полягає в тому, що обидва вийняті з клітки кролі інфіковані різними вірусами. Тоді, за правилом множення, $|A| = C_4^1C_5^1 = 4\cdot5 = 20$. За означенням ймовірності $P(A) = \frac{|A|}{|\Omega|} = \frac{K}{N} = \frac{20}{36} = \frac{5}{9}$.

Принцип додавання і віднімання. Припустимо, що елементи деякої скінченної множини Q можуть мати властивості A_1, $A_2, ..., A_n$. Позначимо через $k(\bigcup_{i=k}^{l} A_i)$ кількість елементів множини Q, які мають хоча б одну з властивостей $A_k, A_{k+1}, ..., A_l$, $k < l$, а через $k(\bigcap_{i=k}^{l} A)$ — кількість елементів множини Q, які мають кожну з властивостей $A_k, A_{k+1}, ..., A_l$, $k < l$. Тоді справедливою буде формула

$$k(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} k(A_i) - \sum_{1\le i<j\le n} k(A_i \cap A_j) +$$

$$+ \sum_{1\le i<j<k\le n} k(A_i \cap A_j \cap A_k) - ... + (-1)^{n+1} k(\bigcap_{i=1}^{n} A_i).$$

Правило додавання. Якщо дві властивості взаємно виключають одна одну, причому одну властивість мають q елементів, іншу — r елементів, тоді хоча б одну з властивостей мають $q + r$ елементів.

Приклад 33. У клітці знаходяться 22 щури. З них 10 тварин лінії α і 12 — лінії β. Скількома способами можна вибрати 2 щури однієї лінії?

Розв'язання. За правилом множення, 2 щури лінії α можна вибрати $10\cdot9 = 90$ способами. Аналогічно 2 щури лінії β можна вибрати $12\cdot11 = 132$ способами. Оскільки обидві лінії не мають спільних елементів (щурів), то, згідно з правилом додавання, загальне число способів вибору двох щурів однієї лінії дорівнює $90 + 132 = 222$.

Приклад 34. Скількома способами групу з 12 піддослідних тварин можна розділити на дві підгрупи, в одній з яких має бути не більше як 5, а в іншій — не більше як 9 тварин?

Розв'язання. Перша підгрупа може складатися з 3, 4 і 5 тварин, які можна вибрати $C_{12}^3 = \frac{12!}{3!\,9!} = \frac{1\cdot2\cdot3\cdot...\cdot11\cdot12}{1\cdot2\cdot3\cdot1\cdot2\cdot...\cdot8\cdot9} = 220$, $C_{12}^4 = \frac{12!}{4!\,8!} = 495$ і $C_{12}^5 = \frac{12!}{5!\,7!} = 792$ способами. Оскільки вибір першої підгрупи однозначно визначає другу підгрупу, яка може складатися з 9, 8 і 7 тварин, за правилом додавання знаходимо розділення тварин на другу підгрупу: число способів $220 + 495 + 792 = 1507$.

Розміщення з повтореннями. Нехай множина C складається з n різних елементів. Розміщеннями з повтореннями з n елементів по k елементів називають упорядковані підмножини з C, що містять k елементів, у яких елементи можуть повторюватись. Нехай, наприклад, C {1, 2, 3}, $k = 2$; тоді всі розміщення з повтореннями матимуть вигляд

$$\{1,1\}, \{1,2\}, \{2,1\}, \{2,2\}, \{1,3\}, \{3,1\}, \{2,3\}, \{3,2\}, \{3,3\}.$$

Теорема 4. Число розміщень з повтореннями з n елементів по k елементів дорівнює $\overline{A}_n^k = n^k$.

Доведення. Утворимо всі можливі розміщення з повтореннями з n елементів по k елементів. При цьому перший елемент розміщення можна обрати n способами, другий, в силу можливих повторень елементів — знову n способами і т. д. Останній, k-й елемент можна також обрати n способами. Згідно з основним правилом комбінаторики, загальне число розміщень з повтореннями дорівнює $\overline{A}_n^k = \underbrace{nn...n}_{k} = n^k$.

Приклад 35. Клавіатура ПК має 48 функціональних клавіш. Яка ймовірність того, що, набираючи клавіші наосліп, можна надрукувати слово "ймовірність"?

Розв'язання. За умовою задачі $n = 48$, $k = 11$ (оскільки у слові "ймовірність" — 11 літер). При набиранні клавіш наосліп літери можуть повторюватися, тому загальне число можливих способів набрати слово, яке містить 11 літер, дорівнює $N = \overline{A}_{48}^{11} = 48^{11}$. Нехай подія B полягає у тому, що надруковано слово "ймовірність". Тоді $|B| = K = 1$ й, отже, $P(B) = \frac{1}{48^{11}} = 48^{-11}$.

§ 6. ДИСКРЕТНІ ВИПАДКОВІ ВЕЛИЧИНИ

6.1. Поняття випадкової величини

Як відомо, математика досліджує кількісні співвідношення реального світу. Однак у реальному житті чимало кількісних параметрів об'єктів та явищ мають випадкову природу. Так, температура навколишнього середовища, кількість опадів за тиждень, ціни на товари і послуги — ці та багато інших величин залежать від різного збігу обставин, які неможливо точно описати й можна вважати випадковими. Отже, ми дійшли до поняття випадкової величини.

Визначення. ***Випадковою величиною*** називають функцію, яка залежить від елементів ймовірнісного простору Ω і набуває дійсних значень.

Випадкова величина $\xi = \xi(\omega)$ є функцією елементарної події $\omega \in \Omega$ і може набувати різних числових значень.

Розглядають лише такі випадкові величини, які можна "міряти", тобто ті, для яких множина $\{\omega : a \le \xi(\omega) < b\}$, a, $b \in R$ становить подію і, як наслідок, для яких визначена ймовірність $P\{\omega : a \le \xi(\omega) < b\}$.

6.2. Функція розподілу випадкової величини

Визначення. Функцію $F(x) = F_\xi(x) = P\{\xi < x\}$, $-\infty < x < +\infty$ називають *функцією розподілу* випадкової величини ξ.

Функція розподілу випадкової величини ξ визначає ймовірність того, що ξ набула значень, менших за x. Знаючи функцію розподілу випадкової величини ξ, можна знайти ймовірність попадання цієї випадкової величини в інтервал $[a, b)$. Справді, оскільки події $\{\xi < a\}$ і $\{a \le \xi < b\}$ несумісні, то $P\{\xi < b\} = P\{\xi < a\} + P\{a \le \xi < b\}$. Отже,

$$P\{a \le \xi < b\} = P\{\xi < b\} - P\{\xi < a\} = F_\xi(b) - F_\xi(a). \quad (13)$$

Властивості функції розподілу

1. $0 \le F_\xi(x) \le 1$ для всіх значень $x \in R$.

Ця властивість випливає з визначення функції розподілу, оскільки $0 \le P\{\omega : \xi(\omega) < x\} \le 1$, $x \in R$.

2. Функція розподілу $F_\xi(x)$ монотонно неспадна по x, тобто якщо $x_1 \le x_2$, то $F_\xi(x_1) \le F_\xi(x_2)$.

Для обгрунтування цієї нерівності достатньо зауважити, що $F_\xi(x_2) - F_\xi(x_1) = P\{x_1 \le \xi < x_2\} \ge 0$.

3. $F_\xi(-\infty) = 0$, $F_\xi(+\infty) = 1$.

Нехай x_n — довільна монотонно спадна послідовність дійсних чисел $\ldots < x_n < \ldots < x_2 < x_1$ і нехай $x_n \to -\infty$ якщо $n \to \infty$. Розглянемо події $A_n = \{\xi < x_n\}$, $n = 1,2,\ldots$. Тоді $\bigcap_{n=1}^{\infty} A_n = \varnothing$ і $\lim_{n\to\infty} P(A_n) = 0$, тобто $\lim_{n\to\infty} P\{\xi < x_n\} = 0$. Оскільки x_n — довільна послідовність, а $F_\xi(x)$ монотонна по x, то $\lim_{x\to-\infty} F_\xi(x) = F_\xi(-\infty) = 0$. Першу рівність доведено, друга доводиться аналогічно.

4. Функція розподілу $F_\xi(x)$ неперервна зліва, тобто $\lim_{\substack{y\to x \\ y<x}} F_\xi(y) = F_\xi(x)$ для будь-якого $x \in R$.

Справді, нехай y_n — довільна монотонно зростаюча послідовність дійсних чисел, тобто $y_1 < y_2 < \ldots < y_n < x$ і $y_n \to x$, якщо $n \to \infty$. Введемо позначення

$$A_n = \{ \xi < y_n \},\ A_x = \{ \xi < x \},\ B_n = A_x \setminus A_n = \{ y_n \le \xi < x \}.$$

Тоді $\bigcap\limits_{n=1}^{\infty} B_n = \varnothing$ і $\lim\limits_{n\to\infty} P(B_n) = 0$, тобто $\lim\limits_{n\to\infty} P\{ y_n \le \xi < x\} = 0$. Однак $P\{ y_n \le \xi < x\} = F_\xi(x) - F_\xi(y_n)$ і $\lim\limits_{n\to\infty}(F_\xi(x) - F_\xi(y_n)) = 0$, або $F_\xi(x) = \lim\limits_{n\to\infty} F_\xi(y_n)$, звідки, в силу довільного вибору послідовності y_n, випливає справедливість твердження.

Приклад 36. Випадкова величина ξ набуває значень на $[a,b]$ з функцією розподілу $F_\xi(x) = x^2 - 2x + 1$. Знайти величини a, b і $P\{1{,}5 \le \xi\}$.

Розв'язання. Врахувавши властивості функції розподілу і явний вигляд $F_\xi(x)$, знайдемо: $a = 1$, $b = 2$. Тепер $P\{1{,}5 \le \xi\} = P\{1{,}5 \le \xi \le 2\} = F_\xi(2) - F_\xi(1{,}5) = 1 - 0{,}25 = 0{,}75$.

За характером побудови випадкові величини поділяють на два класи: дискретні й неперервні.

6.3. Закон розподілу дискретної випадкової величини

Визначення 1. *Випадкову величину* ξ називають *дискретною*, якщо множина її значень є скінченною або зліченною множиною, тобто такою, яку можна записати у вигляді $\{x_1, x_2, \ldots, x_n, \ldots\}$.

Постає природне запитання: як можна описати випадкову величину ξ? Оскільки ми не знаємо точної залежності $\xi(\omega)$, а лише значення, яких набуває $\xi(\omega)$, то природно розглянути події A_i, які полягають у тому, що випадкова величина набуває значень x_i: $A_i = \{\omega : \xi(\omega) = x_i\}$, $i = 1, 2, \ldots$. Ймовірність подій A_i можна обчислити. Ці ймовірності у парі зі значеннями x_i утворюють повне описання випадкової величини ξ.

Визначення 2. *Законом розподілу* дискретної випадкової величини ξ називають сукупність усіх значень ξ, розглянуту разом із ймовірностями $p_i = P\{\xi = x_i\}$ цих значень. Закон розподілу зручно записувати у вигляді таблиці

ξ	x_1	x_2	...	x_n	...
p	p_1	p_2	...	p_n	...

Властивості ймовірностей p_i, $i = 1, 2, \ldots$

1. $0 \le p_i \le 1$, $i = 1, 2, \ldots$

2. $\sum\limits_{i=1}^{\infty} p_i = 1$ (зазначена сума складається зі скінченного числа доданків, якщо випадкова величина ξ набуває скінченне число різних значень).

Доведення. Перша властивість випливає з того, що $p_i = P \times \{\xi = x_i\}$, і очевидної нерівності $0 \le P\{\xi = x_i\} \le 1$.

Для доведення другої властивості зазначимо, що події A_i утворять повну групу, тобто вони попарно несумісні, і хоча б одна з них завжди відбудеться. Тому

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} P(A_i) = P(\bigcup_{i=1}^{\infty} A_i) = P(\Omega) = 1.$$

Приклад 37. Ймовірність того, що в родині народиться хлопчик становить 0,55. Відомо, що в родині є троє дітей. Нехай ξ — кількість хлопчиків у родині. Знайти закон розподілу ξ.

Розв'язання. Нехай випадкова величина ξ дорівнює кількості хлопчиків, що народилися. Тоді ξ може набувати значень $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$. Нехай подія B_i полягає в тому, що i-ю дитиною в родині є хлопчик, $i = 1, 2, 3$. Тоді з незалежності подій B_1, B_2 і B_3 отримаємо

$$p_1 = P\{\xi = 0\} = P(\bar{B}_1)P(\bar{B}_2)P(\bar{B}_3) = 0{,}45 \cdot 0{,}45 \cdot 0{,}45 = 0{,}091125;$$

$$p_2 = P\{\xi = 1\} =$$

$$= P\{(B_1 \cap \bar{B}_2 \cap \bar{B}_3) \cup (\bar{B}_1 \cap B_2 \cap \bar{B}_3) \cup (\bar{B}_1 \cap \bar{B}_2 \cap B_3)\} =$$

$$= P(B_1)P(\bar{B}_2)P(\bar{B}_3) + P(\bar{B}_1)P(B_2)P(\bar{B}_3) + P(\bar{B}_1)P(\bar{B}_2)P(B_3) =$$

$$= 0{,}55 \cdot 0{,}45 \cdot 0{,}45 + 0{,}45 \cdot 0{,}55 \cdot 0{,}45 + 0{,}45 \cdot 0{,}45 \cdot 0{,}55 = 0{,}334125;$$

$$p_3 = P\{\xi = 2\} =$$

$$= P\{(B_1 \cap B_2 \cap \bar{B}_3) \cup (B_1 \cap \bar{B}_2 \cap B_3) \cup (\bar{B}_1 \cap B_2 \cap B_3)\} =$$

$$= 0{,}55 \cdot 0{,}55 \cdot 0{,}45 + 0{,}55 \cdot 0{,}45 \cdot 0{,}55 + 0{,}45 \cdot 0{,}55 \cdot 0{,}55 = 0{,}408375;$$

$$p_4 = P\{\xi = 3\} = P\{B_1 \cap B_2 \cap B_3\} = P(B_1)P(B_2)P(B_3) =$$

$$0{,}55 \cdot 0{,}55 \cdot 0{,}55 = 0{,}166375.$$

Закон розподілу ξ має такий вигляд:

ξ	0	1	2	3
p	0,091125	0,334125	0,408375	0,166375

6.4. Функція розподілу дискретної випадкової величини

З'ясуємо, який вигляд має функція розподілу дискретної випадкової величини (рис. 6).

Нехай дискретна випадкова величина ξ набуває скінченне число різних значень і закон її розподілу має такий вигляд:

ξ	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

Тоді

$$F_\xi(x) = \begin{cases} 0, & \text{якщо} \quad x \le x_1, \\ p_1, & \text{якщо} \quad x_1 < x \le x_2, \\ p_1 + p_2, & \text{якщо} \quad x_2 < x \le x_3, \\ \dots\dots\dots & \dots\dots\dots \\ p_1 + p_2 + \dots + p_{n-1}, & \text{якщо} \quad x_{n-1} < x \le x_n, \\ 1, & \text{якщо} \quad x_n < x. \end{cases}$$

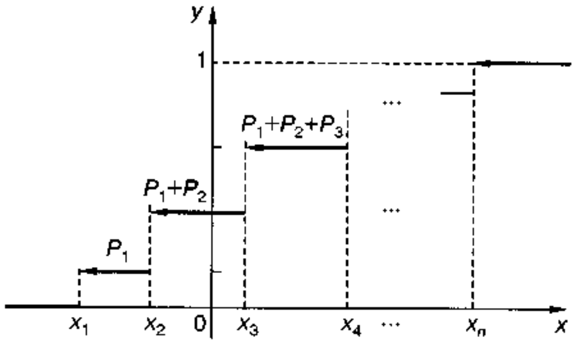


Рис. 6. Графік функції розподілу дискретної випадкової величини

6.5. Види дискретних розподілів

Серед випадкових величин є такі, що найчастіше трапляються на практиці. Це величини, які мають біноміальний розподіл, геометричний розподіл і розподіл Пуассона.

Схема Бернуллі. Біноміальний розподіл

Чимало реальних ситуацій можна звести до серії з n незалежних випадкових експериментів, кожен з яких проводять за однакових умов і який має два результати — успіх або невдача, причому ймовірність успіху в кожному з експериментів фіксована і дорівнює p, а ймовірність невдачі відповідно $q = 1 - p$. Таку схему випробувань називають *схемою Бернуллі*. Наприклад, послідовні підкидання монети (успіх — випадання герба, невдача — випадання решки), шаховий матч (успіх — виграш, невдача — нічия або програш), серія пострілів по мішені (успіх — влучення, невдача — промах), ефективність дії лікарського засобу (успіх — лікарський засіб подіяв, невдача — лікарський засіб не дав лікувального ефекту).

Розглянемо випадкову величину ξ_n, яка дорівнює числу успіхів у n випробуваннях за схемою Бернуллі. Величина ξ_n набуває значень $0, 1, 2, \dots, n$. Отже, вона дискретна. Обчислимо ймовірність $P_n(k) = P\{\xi_n = k\}$. З цією метою зафіксуємо ті номери k випробувань, які привели до успіху. В силу незалежності експериментів ймовірність появи k фіксованих успіхів і $n - k$ невдач становить $p^k q^{n-k}$. Вибрати номери k успішних випробувань із загального числа n експериментів можна у C_n^k спосіб. Тому

$$P_n(k) = C_n^k p^k q^{n-k}, k = 0,\ 1,\ 2, \dots, n. \tag{14}$$

Отриманий закон розподілу називають *біноміальним*, тому що ймовірності $P_n(k)$ є доданками у розкладі виразу $(p + q)^n$ за формулою бінома Ньютона. Таблиця біноміального розподілу має такий вигляд:

ξ	0	1	...	k	...	n
p	q^n	npq^{n-1}	...	$C_n^k p^k q^{n-k}$	...	p^n

Властивості біноміального розподілу

1. Ймовірність появи успіхів у кількості меншій, ніж m, $0 \le \le m \le n$, у серії з n випробувань знаходять за формулою

$$P_n(0 \le k < m) = \sum_{k=0}^{m-1} P_n(k) = \sum_{k=0}^{m-1} C_n^k p^k q^{n-k}. \qquad (15)$$

2. Ймовірність появи успіхів у кількості не меншій за m, $0 \le \le m \le n$, у серії з n випробувань знаходять за формулою

$$P_n(m \le k \le n) = \sum_{k=m}^{n} P_n(k) = 1 - \sum_{k=0}^{m-1} P_n(k) = 1 - \sum_{k=0}^{m-1} C_n^k p^k q^{n-k}. \qquad (16)$$

3. Ймовірність появи в хоча б одного успіху в серії з n випробувань знаходять за формулою

$$P_n(1 \le k \le n) = 1 - P_n(0) = 1 - q^n. \qquad (17)$$

Приклад 38. Два рівносильних шахісти грають у шахи. Що ймовірніше: виграти дві партії з чотирьох чи три партії з шести (нічиї до уваги не беруться)?

Розв'язання. Оскільки грають рівносильні шахісти, то ймовірність виграшу $p = 1/2$; отже, ймовірність програшу $q = 1 - p = = 1/2$. В усіх партіях ймовірність виграшу стала і не має значення, в якій послідовності будуть виграні партії. Отже, можна застосувати біноміальний розподіл.

Знайдемо ймовірність того, що будуть виграні дві партії з чотирьох:

$$P_4(2) = C_4^2 p^2 q^2 = \frac{4 \cdot 3}{1 \cdot 2}\left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{6}{16}.$$

Знайдемо ймовірність того, що будуть виграні три партії з шести:

$$P_6(3) = C_6^3 p^3 q^3 = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3}\left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^3 = \frac{5}{16}.$$

Оскільки $P_4(2) > P_6(3)$, то ймовірніше виграти дві партії з чотирьох, ніж три партії з шести.

Приклад 39. Яка ймовірність народження $0, 1, 2, 3, 4$ і 5 хлопчиків серед п'ятьох немовлят?

Розв'язання. Позначимо через p ймовірність того, що новонароджена дитина — хлопчик. Тоді ймовірність народження дівчинки — $q = 1 - p$. Вважатимемо, що ймовірність народження

хлопчиків і дівчаток однакова, тобто $p = q = 0,5$. Згідно з формулою (14), ймовірність того, що серед $n = 5$ новонароджених не буде жодного хлопчика дорівнює $P_5(0) = P_5(5) = C_5^0 p^0 q^5 = = (0,5)^5 = 0,03$. Аналогічно $P_5(1) = P_5(4) = C_5^1 p^1 q^4 = 0,16$; $P_5(2) = P_5(3) = = C_5^2 p^2 q^3 = 0,31$.

Приклад 40. Ймовірність успіху хоча б один раз у п'яти незалежних випробуваннях за схемою Бернуллі дорівнює 0,9. Яка ймовірність успіху в одному випробуванні, якщо при кожному випробуванні вона однакова?

Розв'язання. Згідно з формулою (17), маємо $P_5(1 \le k \le 5) = = 1 - (1 - p)^5 = 0,9$. Отже, ймовірність успіху p розрахуємо за рівнянням $1 - (1 - p)^5 = 0,9$, звідки $p = 0,37$.

Геометричний розподіл

Розглянемо послідовність незалежних випробувань, у кожному з яких можливі лише два результати — успіх або невдача з ймовірностями відповідно p і $q = 1 - p$. Приклади: число підкидань монети до першого випадання герба; число шахових партій, зіграних до першого виграшу одного з партнерів; число пострілів до першого влучення в мішень. Розглянемо випадкову величину η, що дорівнює числу експериментів, проведених до настання першого успіху. Очевидно, що η — дискретна випадкова величина, яка набуває значень $0, 1, 2, \ldots, k, \ldots$, причому подія $\{\omega : \eta(\omega) = k\}$ полягає в тому, що перші k випробувань призвели до невдачі, а $(k+1)$ випробування виявилося успішним. У силу незалежності випробувань $P\{\eta = k\} = \overbrace{q \ldots q}^{k} p = q^k p$, $k = 0, 1, 2, \ldots$. Отриманий закон розподілу називають *геометричним*, тому що ймовірності $q^k p$ є послідовними членами геометричної прогресії з першим членом p і знаменником q. Таблиця цього закону розподілу має такий вигляд:

η	0	1	2	...	k	...
p	p	qp	q^2p	...	q^kp	...

Найімовірнішим значенням є $\eta = 0$.

Приклад 41. Після відповіді студента на питання екзаменаційного квитка екзаменатор задає йому додаткові питання. Викладач припиняє задавати додаткові питання, як тільки студент дає неправильну відповідь. Ймовірність того, що студент правильно відповість на всі додаткові питання, дорівнює 0,9. Нехай η — число додаткових питань, які викладач задасть студенту до першої помилкової відповіді. Складіть закон розподілу дискретної випадкової величини η (а). Знайдіть найймовірніше значення η (б).

Розв'язання. а) Випадкова величина η набуває значень $0, 1, \dots, k, \dots$. Знайдемо ймовірності того, що величина η набуває цих значень. Якщо $\eta = 0$, то це означає, що студент дав помилкову відповідь на перше питання. Отже, $P\{\eta = 0\} = 1 - 0,9 = 0,1$. Випадкова величина $\eta = 1$ тоді і тільки тоді, коли студент відповів правильно на перше питання і помилково — на друге. Маємо $P\{\eta = 1\} = 0,9(1-0,9) = 0,9 \cdot 0,1 = 0,09$. Аналогічно $P\{\eta = 2\} = 0,9 \cdot 0,9(1-0,9) = 0,081$.

Закон розподілу шуканої випадкової величини η має такий вигляд:

η	0	1	2	...	k	...
p	0,1	$0,9 \cdot 0,1$	$0,9^2 \cdot 0,1$	...	$0,9^k \cdot 0,1$	...

б) Найймовірніше значення η, як випливає із закону розподілу, дорівнює 0.

Розподіл Пуассона

З практичного погляду винятково важливими є системи масового обслуговування. Це телефонні станції, на які надходять дзвінки; станції обслуговування автомобілів, на які надходять заявки на технічне обслуговування; склади і магазини, на які надходять замовлення на товари, тощо. Для перелічених систем задача формулюється у такий спосіб: скільки заявок надійде до системи упродовж інтервалу часу $[0, t]$? Очевидно, це випадкова величина, яка набуває значень $0, 1, 2, \dots$. Якщо кожне окреме надходження заявки до системи масового обслуговування розглядати як подію, то процес надходження заявок логічно назвати потоком подій. Якщо такий потік подій задовольнятиме три природні умови, то можна вивести закон розподілу кількості

подій в інтервалі часу $[0, t]$. Цей розподіл називають *розподілом Пуассона.* Надалі вважатимемо, що потік подій задовольняє такі умови:

1) *умова однорідності потоку* — число подій, що відбулись в інтервалі часу $[t, t+h]$, залежить тільки від довжини інтервалу h, але не залежить від початку t;

2) *незалежність приростів* — число подій, що настали в інтервалі часу $[t, t+h]$, не залежить від числа подій, що настали до часу t (відсутність післядії);

3) *умова ординарності потоку* — ймовірність настання однієї події в інтервалі $[0, \Delta t]$ (Δt є достатньо малим) пропорційна його довжині Δt з деяким коефіцієнтом пропорційності λ: $p_1(\Delta t) \approx \approx \lambda \Delta t$; ймовірністю настання двох або більшого числа подій у тому ж часовому інтервалі можна знехтувати: $p_2(\Delta t) + p_3(\Delta t) + \dots \approx 0$.

Нехай випадкова величина $\xi(t)$ дорівнює числу подій, що відбулись упродовж часу $[0, t]$. Якщо потік подій задовольняє умови 1)—3), то можна показати, що для деякого $\lambda > 0$

$$p_m(t) = P\{\xi(t) = m\} = \frac{(\lambda t)^m}{m!} e^{-\lambda t}, \quad m = 0, 1, 2, \dots,$$

$$P_0(t) = P\{\xi(t) = 0\} = e^{-\lambda t},$$

де $e \approx 2,71828$ — основа натурального логарифма.

Отриманий закон розподілу називають *законом Пуассона*, а параметр λ — *інтенсивністю* потоку подій. Його вводять для того, щоб розрізняти потоки, які відбуваються за різних умов. Наприклад, кількість телефонних дзвінків за одиницю часу вдень і вночі різні, відтак їм мають відповідати різні значення інтенсивності λ. Величина λ дорівнює середньому числу подій, які наступали за одиницю часу.

Приклад 42. Середнє число замовлень таксі, що надходять на диспетчерський пункт за одну хвилину, дорівнює трьом. Знайти ймовірність того, що за дві хвилини надійде: а) чотири замовлення; б) менш як чотири замовлення; в) не менш як чотири замовлення.

Розв'язання. За умовою, $\lambda = 3$, $t = 2$, $m = 4$. Скористаємось законом розподілу Пуассона $p_m(t) = P\{\xi(t) = m\} = \frac{(\lambda t)^m}{m!} e^{-\lambda t}$.

а) Шукана ймовірність того, що за дві хвилини надійде чотири замовлення таксі дорівнює $p_4(2) = P\{\xi(2)=4\} = \frac{6^4}{4!}e^{-6} \approx$ $\approx \frac{1296 \cdot 0,0025}{24} = 0,135$.

б) Подія "надійшло менш як чотири замовлення" станеться, якщо відбудеться хоча б одна з таких несумісних подій: 1) надійшло три замовлення; 2) надійшло два замовлення; 3) надійшло одне замовлення; 4) не надійшло жодного замовлення. Ці події несумісні. Отже, ймовірність того, що відбулась хоча б одна з цих подій, дорівнює сумі ймовірностей цих подій:

$$P\{\xi(2)<4\} = p_3(2)+p_2(2)+p_1(2)+p_0(2) = \frac{6^3}{3!}e^{-6} + \frac{6^2}{2!}e^{-6} +$$

$$+\frac{6^1}{1!}e^{-6} + e^{-6} \approx e^{-6}(36+18+6+1) = 0,0025 \cdot 61 = 0,1525.$$

в) Події "надійшло менш як чотири замовлення" і "надійшло не менш як чотири замовлення" протилежні, тому шукана ймовірність того, що за дві хвилини надійде не менш як чотири замовлення дорівнює

$$P\{\xi(2) \ge 4\} = 1 - P\{\xi(2)<4\} \approx 1 - 0,1525 = 0,8475.$$

Між розподілом Бернуллі і розподілом Пуассона існує важливий зв'язок, який формулює теорема Пуассона.

Теорема Пуассона. Припустимо, що у схемі Бернуллі кількість незалежних випадкових експериментів n зростає так, що $n \to \infty$, а ймовірність успіху p_n в кожному з експериментів зі збільшенням n зменшується, тобто $p_n \to 0$. Припустимо також, що $np_n \to \lambda$ при $n \to \infty$ для деякого $0 < \lambda < +\infty$. Тоді при $n \to \infty$ справедливе граничне співвідношення

$$P_n(k) \to \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0,\ 1,\ 2,\ldots, \tag{18}$$

де $P_n(k) = C_n^k p^k q^{n-k}$, $0 \le k \le n$.

Зауваження. На практиці теорему Пуассона застосовують у випадку, коли $n \ge 100$, $0 < p << 1$ і $np = \lambda$ — стала.

Приклад 43. Вакцина формує імунітет до захворювання з імовірністю 0,999. Яка ймовірність того, що імунітету не набули троє серед 2000 щеплених тварин?

Розв'язання. Оскільки $n = 2000$, $p = 1 - 0,999 = 0,001 << 1$ і $np = = \lambda = 2$, то шукану ймовірність знаходимо за формулою (18):

$$P_{2000}(k=3) = \frac{2^3}{3!}e^{-2} \approx 0,18.$$

Приклад 44. У радіобіологічній лабораторії проводять експерименти з опромінення біоматеріалу β-частинками. Ймовірність потрапляння однієї β-частинки на виділену ділянку дорівнює 10^{-3}. Яку кількість β-частинок потрібно випустити, щоб з імовірністю 0,99 більш як 3 β-частинки потрапили на виділену ділянку біоматеріалу?

Розв'язання. Нехай ξ — число β-частинок, які потрапили на виділену ділянку біоматеріалу. Ймовірність $P_n(k)$ того, що серед n випущених частинок k потрапили на виділену ділянку, має біноміальний розподіл і визначається за формулою (14). Оскільки $p = 10^{-3} << 1$, можна застосувати теорему Пуассона. Згідно з формулами (16) і (18), для достатньо великих n маємо:

$$P\{3 < \xi\} \approx P_n(4 \le k \le n) =$$

$$= 1 - \sum_{k=0}^{3} P_n(k) \approx 1 - e^{-\lambda}\left(1 + \lambda + \lambda^2/2! + \lambda^3/3!\right) = 0,99.$$

Останню рівність можна розглядати як рівняння відносно λ, звідки отримаємо $\lambda \approx 11$ і оцінку для $n = \lambda/p \approx 11 \cdot 10^3$.

6.6. Функції від випадкових величин

Нехай ξ — випадкова величина, $f(x)$ — неперервна функція. Тоді $f(\xi)$ — теж випадкова величина. Зокрема $-\xi$, $a\xi$, ξ^k — випадкові величини. Якщо ξ — дискретна величина і має закон розподілу

ξ	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

а функція $f(x)$ взаємно однозначна, то $f(\xi)$ має такий закон розподілу:

ξ	$f(x_1)$	$f(x_2)$	...	$f(x_n)$
p	p_1	p_2	...	p_n

6.7. Числові характеристики випадкових величин

Закон розподілу повністю визначає випадкову величину. Водночас він не дає можливості усвідомити структуру випадкової величини. Таку структуру можна ввести через виділення детермінованої сталої, відносно якої початкова випадкова величина перетвориться на "суто випадкову", тобто таку, що не містить жодної системної (детермінованої) складової. Для "суто випадкової" величини можна ввести показник її розпорошення. Отже, довільну випадкову величину можна описати, хоча й неоднозначно, за допомогою двох параметрів, які називають *математичним сподіванням* і *дисперсією*.

Математичне сподівання дискретної випадкової величини

Щоб з'ясувати, як визначити математичне сподівання дискретної випадкової величини, розглянемо приклад.

Приклад 45. Певний товар продають у 100 магазинах міста за ціною: 1 гривня за одиницю товару в 30 магазинах; 2 гривні за одиницю товару в 25 магазинах; 3 гривні за одиницю товару в 25 магазинах; 4 гривні за одиницю товару в 20 магазинах. Яка середня вартість товару у 100 магазинах міста? Чи варто покупцеві, який має 2,50 гривень у гаманці, заходити в перший ліпший магазин, щоб купити товар?

Розв'язання. Позначимо через m середню вартість товару у 100 магазинах міста. Тоді

$$m = \frac{1}{100}(1 \cdot 30 + 2 \cdot 25 + 3 \cdot 25 + 4 \cdot 20) = \frac{235}{100} = 2,35.$$

Маючи більше грошей, ніж величина m, покупець має більше шансів купити товар, ніж не купити його у навмання обраному магазині. Величина m має ту властивість, що ціни на товар з урахуванням кількості магазинів розміщуються відносно неї.

Розглянемо умову цієї задачі з іншого погляду. Нехай ξ — ціна товару в навмання обраному магазині. Тоді ξ — дискретна випадкова величина, яка має такий розподіл:

ξ	1	2	3	4
p	0,3	0,25	0,25	0,2

а значення m, виходячи з розподілу ξ, обчислюється за формулою

$$m = 1 \cdot 0,3 + 2 \cdot 0,25 + 3 \cdot 0,25 + 4 \cdot 0,2 = 2,35.$$

Зважаючи на властивість m, ми дійдемо до визначення математичного сподівання (середнього значення) дискретної випадкової величини.

Нехай випадкова величина ξ дискретна і набуває скінченне число значень.

Визначення. *Математичним сподіванням* (*середнім значенням*) $E\xi$ випадкової величини ξ називають суму всіх значень ξ, помножених на ймовірності цих значень:

$$E\xi = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n. \qquad (19)$$

Нехай випадкова величина ξ дискретна і набуває зліченне число значень.

Визначення. *Математичним сподіванням* ξ називають нескінченну суму

$$E\xi = \sum_{i=1}^{\infty} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n + \ldots,$$

якщо $\sum_{i=1}^{\infty} |x_i| p_i < \infty$. У протилежному випадку математичне сподівання $E\xi$ не існує.

Зауваження. Позначення математичного сподівання $E\xi$ випадкової величини ξ утворене від англійського слова expectation — сподівання. Використовують також інше позначення $M\xi$, яке наразі застаріле. Отже, рівнозначно маємо $E\xi \equiv M\xi \equiv E(\xi) \equiv M(\xi)$.

Властивості математичного сподівання. Для зручності властивості математичного сподівання розглянемо лише для дискретних випадкових величин, які набувають скінченне число значень. Для величин, що набувають зліченну множину значень, усі міркування аналогічні. Додатково потрібно лише перевіряти умову $\sum_{i=1}^{\infty} |x_i| p_i < \infty$.

1. Математичне сподівання сталої величини дорівнює цій сталій: $Ec = c$.

Справді, нехай стала величина ξ набуває значення c з імовірністю 1: $P\{\xi = c\} = 1$. Тому $Ec = c \cdot 1 = c$.

2. Сталий множник можна виносити за знак математичного сподівання: $E(c\xi) = cE(\xi)$.

Нехай ξ має розподіл

ξ	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

Тоді випадкова величина $c\xi$ матиме розподіл

$c\xi$	cx_1	cx_2	...	cx_n
p	p_1	p_2	...	p_n

Отже, $E(c\xi) = \sum_{i=1}^{n} cx_i p_i = c\sum_{i=1}^{n} x_i p_i = c\,E\xi$.

3. Якщо випадкова величина $\xi \ge 0$, то $E\xi \ge 0$.

Справді, якщо всі $x_i \ge 0$, то й $\sum_{i=1}^{n} x_i p_i \ge 0$.

4. Математичне сподівання суми двох випадкових величин дорівнює сумі їх математичних сподівань:

$$E(\xi + \eta) = E\xi + E\eta.$$

Нехай ξ — дискретна випадкова величина, яка набуває значень $x_1, x_2, ..., x_n$, а η — дискретна випадкова величина, яка набуває значень $y_1, y_2, ..., y_m$. Тоді їх сума $\xi + \eta$ — дискретна випадкова величина, яка має такий розподіл:

$\xi + \eta$	$x_1 + y_1$	$x_1 + y_2$	...	$x_n + y_m$
p	r_{11}	r_{12}	...	r_{nm}

$$r_{ij} = P(\xi = x_i, \eta = y_j),\ i = 1, 2, ..., n;\ j = 1, 2, ..., m.$$

Розглянемо такі події: $A_i = \{\xi = x_i\}$, $B_j = \{\eta = y_j\}$. Позначимо $P(A_i) = p_i$, $P(B_j) = q_j$. Тоді $E(\xi + \eta) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i + y_i)r_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i r_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} y_i r_{ij} = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} r_{ij} + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} r_{ij}$. Однак $\sum_{i=1}^{n} r_{ij} = \sum_{i=1}^{n} P(A_i \cap B_j) =$

$= P(\bigcup_{i=1}^{n}(A_i \cap B_j)) = P((\bigcup_{i=1}^{n} A_i) \cap B_j) = P(\Omega \cap B_j) = P(B_j) = q_j$. Аналогічно $\sum_{j=1}^{m} r_{ij} = p_i$. Отже,

$$E(\xi + \eta) = \sum_{i=1}^{n} x_i \sum_{i=1}^{n} r_{ij} + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} r_{ij} = \sum_{i=1}^{n} x_i p_i + \sum_{j=1}^{m} y_j q_j = E\xi + E\eta.$$

5. Якщо випадкові величини ξ і η незалежні, то математичне сподівання їх добутку дорівнює добутку їх математичних сподівань:

$$E(\xi\eta) = E\xi\, E\eta.$$

Справді, випадкова величина $\xi\eta$ має такий розподіл:

$\xi\eta$	x_1y_1	x_1y_2	...	x_ny_m
p	r_{11}	r_{12}	...	r_{nm}

де $r_{ij} = P(\xi = x_i, \eta = y_j)$, $i = 1, 2, ..., n$; $j = 1, 2, ..., m$. Із незалежності випадкових величин ξ і η (тобто $P(\xi = x_i, \eta = y_j) = P(\xi = x_i) \times$ $\times P(\eta = y_j)$) випливає, що

$$r_{ij} = P(\xi = x_i, \eta = y_j) = P(\xi = x_i)\ P(\eta = y_j) = P(A_i)\ P(B_j) = p_i q_j.$$

Маємо $E(\xi\eta) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j r_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_i q_j = \sum_{i=1}^{n} x_i p_i \sum_{j=1}^{m} y_j q_j = E\xi E\eta$.

Приклад 46. Незалежні випадкові величини ξ та η розподілені за законом

ξ	2	1
p	0,3	0,7

η	5	7	8
p	0,5	0,3	0,2

Знайти математичне сподівання випадкової величини $\xi\eta$.

Розв'язання. Спочатку знайдемо математичні сподівання для кожної з цих величин. За формулою (19) отримаємо

$$E\xi = \sum_{i=1}^{2} x_i p_i = 0,6 + 0,7 = 1,3;\quad E\eta = \sum_{i=1}^{3} x_i p_i = 2,5 + 2,1 + 1,6 = 6,2.$$

Нарешті, згідно з властивістю 5, матимемо $E(\xi\eta) = E\xi E\eta = 8,06$.

Дисперсія дискретної випадкової величини

Для опису випадкової величини недостатньо знати тільки математичне сподівання (середнє значення). Переконаємось у цьому на прикладі.

Приклад 47. Нехай ξ і η — випадкові величини

ξ	$-0,1$	$0,1$
p	$1/2$	$1/2$

η	-10	10
p	$1/2$	$1/2$

$E\xi = E\eta = 0$, але структури ξ і η різні: ξ мало відхиляється від свого середнього значення, а η значно відхиляється від 0.

Для опису відхилення випадкової величини від її середнього значення введено числову характеристику — дисперсію.

Визначення. *Дисперсією* випадкової величини ξ називають математичне сподівання квадрата відхилення ξ від свого математичного сподівання:

$$D\xi = E(\xi - E\xi)^2.$$

Перетворимо $D\xi$, скориставшись властивостями математичного сподівання:

$$D\xi = E(\xi - E\xi)^2 = E(\xi^2 - 2\xi E\xi + (E\xi)^2) = E\xi^2 - 2E\xi E\xi + (E\xi)^2 =$$
$$= E\xi^2 - (E\xi)^2. \qquad (20)$$

Якщо ξ — дискретна і набуває скінченного числа значень, то

$$D\xi = \sum_{i=1}^{n}(x_i - m)^2 p_i = \sum_{i=1}^{n} x_i^2 p_i - m^2, \text{ де } m = E\xi = \sum_{i=1}^{n} x_i p_i.$$

Якщо ξ — дискретна і набуває зліченного числа значень, то

$$D\xi = \sum_{i=1}^{\infty}(x_i - m)^2 p_i = \sum_{i=1}^{\infty} x_i^2 p_i - (\sum_{i=1}^{\infty} x_i p_i)^2$$

(за умови скінченності сум у правій частині виразу).

Властивості дисперсії. Для зручності властивості дисперсії розглянемо лише для випадкових величин, які набувають скінченного числа значень.

1. Дисперсія будь-якої випадкової величини невід'ємна. При цьому $D\xi = 0$ тоді і тільки тоді, коли ξ — стала невипадкова величина.

Справді, $D\xi = E(\xi - E\xi)^2 = \sum_{i=1}^{n}(x_i - m)^2 p_i \geq 0$. Далі, нехай $D\xi = \sum_{i=1}^{n}(x_i - m)^2 p_i = 0$. Тоді для всіх $i = 1, 2, ..., n$ $x_i = m$, тобто $\xi = m$ з імовірністю 1. Навпаки, якщо $\xi = c$, то $D\xi = E(c - c)^2 = 0$.

2. Якщо a, b — сталі, то

$$D(a\xi + b) = a^2 D\xi.$$

Справді,

$$D(a\xi + b) = E(a\xi + b - E(a\xi + b))^2 = E(a\xi + b - aE\xi - b)^2 =$$
$$= E(a\xi - a E\xi)^2 = a^2 E(\xi - E\xi)^2 = a^2 D\xi.$$

3. Якщо випадкові величини ξ і η — незалежні, то дисперсія їх суми дорівнює сумі дисперсій:

$$D(\xi + \eta) = D\xi + D\eta.$$

Справді, $E(\xi\eta) = E\xi E\eta$ і $D(\xi + \eta) = E(\xi + \eta)^2 - (E(\xi + \eta))^2 = E\xi^2 + 2E(\xi\eta) + E\eta^2 - (E\xi)^2 - 2E\xi E\eta - (E\eta)^2 = E\xi^2 - (E\xi)^2 + E\eta^2 - (E\eta)^2 = D\xi + D\eta$.

Приклад 48. Знайти дисперсію випадкової величини, яка розподілена за законом

ξ	3	5	7
p	0,2	0,5	0,3

Розв'язання. Спочатку обчислимо математичне сподівання випадкової величини за формулою (19):

$$E\xi = \sum_{i=1}^{3} x_i p_i = 0,6 + 2,5 + 2,1 = 5,2. \text{ Тоді } (E\xi)^2 = 27,04.$$

Далі

$$E\xi^2 = \sum_{i=1}^{3} x_i^2 p_i = 12,5 + 14,7 + 1,8 = 29.$$

Остаточно, згідно з формулою (20), маємо $D\xi = 29 - 27,04 = 1,96$.

Середньоквадратичне відхилення

Визначення. *Середньоквадратичним відхиленням* випадкової величини ξ називають квадратний корінь з її дисперсії:

$$\sigma(\xi) = \sqrt{D\xi}. \qquad (21)$$

Властивості середньоквадратичного відхилення

1. Якщо c — сталий множник, то: $\sigma(c\xi) = c\sigma(\xi)$.
2. Якщо ξ і η — незалежні випадкові величини, то

$$\sigma(\xi + \eta) = \sqrt{\sigma^2(\xi) + \sigma^2(\eta)}.$$

6.8. Числові характеристики біноміального, геометричного та Пуассонівського розподілів

Числові характеристики біноміального розподілу

Числові характеристики біноміального розподілу складно обчислити безпосередньо. Розглянемо для цього "індикатори успіхів" — випадкові величини $\eta_1, \eta_2, ..., \eta_n$ такого вигляду:

$$\eta_i = \begin{cases} 1, & \text{якщо в } i\text{-му експерименті настає успіх;} \\ 0, & \text{якщо в } i\text{-му експерименті настає невдача.} \end{cases}$$

Випадкові величини $\eta_1, \eta_2, ..., \eta_n$ незалежні в сукупності, однаково розподілені і $\xi_n = \eta_1 + \eta_2 + ... + \eta_n$. Тому

$$E\xi_n = E\eta_1 + E\eta_2 + ... + E\eta_n = nE\eta_1,$$
$$D\xi_n = D\eta_1 + D\eta_2 + ... + D\eta_n = nD\eta_1.$$

Далі $E\eta_1 = 1 \cdot P\{\eta_1 = 1\} + 0 \cdot P\{\eta_1 = 0\} = 1 \cdot p + 0 \cdot q = p$, $D\eta_1 =$
$= E\eta_1^2 - (E\eta_1)^2 = 1^2 \cdot P\{\eta_1 = 1\} + 0^2 \cdot P\{\eta_1 = 0\} - p^2 =$
$= p - p^2 = p(1-p) = pq.$

Отже, $E\xi_n = np$, $D\xi_n = npq$ і $\sigma(\xi_n) = \sqrt{npq}$.

Числові характеристики геометричного розподілу

За визначенням, $E\eta = \sum\limits_{k=0}^{\infty} kq^k p$. Отже,

$$E\eta = \sum_{k=0}^{\infty} kq^k p = p \sum_{k=0}^{\infty} (k+1)q^{k+1} = p \sum_{k=0}^{\infty} kq^{k+1} + p \sum_{k=0}^{\infty} q^{k+1} =$$

$$= q\sum_{k=0}^{\infty} kq^k p + pq\sum_{k=0}^{\infty} q^k = qE\eta + \frac{pq}{1-q} = qE\eta + \frac{pq}{p} = qE\eta + q,$$

звідки $E\eta = \dfrac{q}{1-q} = \dfrac{q}{p}$.

Дисперсію обчислюють аналогічно: $D\eta = \dfrac{q}{p^2}$. Нарешті, $\sigma(\eta) = \dfrac{\sqrt{q}}{p}$.

Числові характеристики розподілу Пуассона

Математичне сподівання і дисперсія закону Пуассона збігаються:

$$E\xi(t) = D\xi(t) = \lambda t.$$

Для доведення цього факту скористаємось зображенням для функції e^x, відомим із курсу математичного аналізу:

$$e^x = \sum_{m=0}^{\infty} \frac{x^m}{m!}, \quad -\infty < x < \infty.$$

Покладемо $x = \lambda t$. Отримаємо $e^{\lambda t} = \sum\limits_{m=0}^{\infty} \dfrac{(\lambda t)^m}{m!}$. Тоді за визначенням

$$E\xi(t) = \sum_{m=0}^{\infty} m\frac{(\lambda t)^m}{m!} e^{-\lambda t} = e^{-\lambda t} \sum_{m=1}^{\infty} \frac{(\lambda t)^m}{(m-1)!} = e^{-\lambda t} \times$$
$$\times \lambda t \sum_{m=1}^{\infty} \frac{(\lambda t)^{m-1}}{(m-1)!} = \lambda t\, e^{-\lambda t} \sum_{m=0}^{\infty} \frac{(\lambda t)^m}{m!} = \lambda t\, e^{-\lambda t} e^{\lambda t} = \lambda t.$$

Рівність $D\xi(t) = \lambda t$ доводиться аналогічно. Нарешті, $\sigma(\xi(t)) = \sqrt{\lambda t}$.

6.9. Коефіцієнт кореляції дискретних випадкових величин

Припустимо, що випадкова величина ξ набуває значень x_1, $x_2, ..., x_n$, а випадкова величина η — значень $y_1, y_2, ..., y_m$. Нехай також задано їх сумісний розподіл: $p_{ij} = P(\xi = x_i, \eta = y_j)$, $i = 1, 2, ..., n$; $j = 1, 2, ..., m$. Зобразимо це у вигляді таблиці:

$\eta \backslash \xi$	x_1	x_2	...	x_n
y_1	p_{11}	p_{12}	...	p_{1n}
y_2	p_{21}	p_{22}	...	p_{2n}
...	...	...	...	...
y_m	p_{m1}	p_{m2}	...	p_{mn}

Опишемо залежність між величинами ξ і η. З цією метою розглянемо коефіцієнт кореляції між ними, який і буде показником залежності величин ξ і η.

Визначення. *Коефіцієнтом кореляції* між ξ і η називають величину

$$\rho = \rho(\xi, \eta) = \frac{E[(\xi - E\xi)(\eta - E\eta)]}{\sigma(\xi)\sigma(\eta)}.$$

Для зручнішого обчислення коефіцієнта кореляції зауважимо, що $E[(\xi - E\xi)(\eta - E\eta)] = E(\xi\eta) - E\xi\, E\eta$.

Отже,

$$\rho = \rho(\xi, \eta) = \frac{E(\xi\eta) - E\xi\, E\eta}{\sigma(\xi)\sigma(\eta)}. \qquad (22)$$

З'ясуємо, як обчислити коефіцієнт кореляції $\rho(\xi,\eta)$ за даними таблиці. Для підрахунку $E\xi$, $E\eta$, $\sigma(\xi)$ і $\sigma(\eta)$ необхідно знати розподіл величин ξ і η. Позначимо через $p_i = P(\xi = x_i)$. Отримаємо

$$p_i = P(\xi = x_i) = P\{(\xi = x_i) \cap \Omega\} = P\{(\xi = x_i) \cap \bigcup_{j=1}^{m}(\eta = y_j)\} =$$

$$= P\{\bigcup_{j=1}^{m}(\xi = x_i, \eta = y_j)\} = \sum_{j=1}^{m} P(\xi = x_i, \eta = y_j) = \sum_{j=1}^{m} p_{ij}.$$

Аналогічно $q_j = P(\eta = y_j) = \sum_{i=1}^{n} p_{ij}$.

Закони розподілів випадкових величин ξ і η мають такий вигляд:

ξ	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

η	y_1	y_2	...	y_m
p	q_1	q_2	...	q_m

Підрахунок $E\xi$, $E\eta$, $\sigma(\xi)$ і $\sigma(\eta)$ виконують за визначенням. Для $M(\xi\eta)$ маємо таку формулу:

$$E(\xi\eta) = \sum_{i=1, j=1}^{n,m} x_i y_j p_{ij}.$$

Властивості коефіцієнта кореляції (без доведення)

1. Для будь-яких ξ і η: $|\rho(\xi, \eta)| \le 1$.

2. Для будь-яких ξ і η: $|\rho(\xi, \eta)| = \pm 1$ тоді і тільки тоді, коли $\xi = a\eta + b$, для деяких a і b, тобто ξ і η зв'язані лінійною залежністю; при цьому $\rho(\xi, \eta) = 1$, якщо $a > 0$ і $\rho(\xi, \eta) = -1$, якщо $a < 0$.

3. Якщо $\rho(\xi, \eta) = 0$, то випадкові величини ξ і η називають *некорельованими*. Це означає, що вони не мають статистичної залежності.

4. Якщо ξ і η незалежні, то: $\rho(\xi, \eta) = 0$.

Це випливає з того, що для незалежних випадкових величин $E(\xi\eta) = E\xi\, E\eta$.

Отже, коефіцієнт кореляції характеризує ступінь лінійної залежності між випадковими величинами.

§ 7. НЕПЕРЕРВНІ ВИПАДКОВІ ВЕЛИЧИНИ

7.1. Щільність розподілу неперервної випадкової величини

На відміну від дискретно розподілених випадкових величин, які набувають фіксованих значень із додатною ймовірністю, випадкові величини з неперервним розподілом мають нульові шанси знаходитись в окремих точках. Водночас вони з додатною ймовірністю концентруються на інтервалах ненульової довжини. Висловлюючись неформально, дискретно розподілену випадкову величину можна порівняти з краплями фарби, які вміщують в окремі точки поверхні. Натомість випадкову величину, яка має неперервну будову, розмазують по поверхні, не залишаючи крапель в окремих точках.

Для вибору математичного апарату опису неперервно розподілених випадкових величин звернемо увагу на те, що забарвлення поверхні визначається щільністю нанесення фарби і має таку властивість: кількість фарби на поверхні доволі малої площі приблизно дорівнює добутку площі цієї поверхні на значення щільності нанесення фарби в її довільній точці.

Позначимо через $\rho_\xi(x), -\infty < x < \infty$ — щільність розподілу неперервно розподіленої випадкової величини. Тоді для $\Delta = (b-a)/n$ і достатньо великих n матимемо

$$P\{a \le \xi < b\} = P\left\{\bigcup_{k=0}^{n-1}\{a + k\Delta \le \xi < a + (k+1)\Delta\}\right\} =$$

$$= \sum_{k=0}^{n-1} P\{a + k\Delta \le \xi < a + (k+1)\Delta\} \approx \sum_{k=0}^{n-1}\rho_\xi(a + k\Delta)\Delta .$$

Як відомо з курсу математичного аналізу, останню суму називають інтегральною, вона збігається до інтеграла $\int_a^b \rho_\xi(x)\,dx$ якщо $n \to \infty$ для досить широкого класу функцій $\rho_\xi(x)$. Отже, ми дійшли до співвідношення, яке описує неперервно розподілену випадкову величину ξ на інтервалі $[a,b)$:

$$P\{a \le \xi < b\} = \int_a^b \rho_\xi(x)\,dx\,, a < x < b .$$

Визначення. Випадкову величину ξ називають *неперервною*, якщо її функцію розподілу $F_\xi(x)$ можна записати у вигляді

$$F_\xi(x) = \int_{-\infty}^{x} \rho_\xi(y)dy, \tag{23}$$

де $\rho_\xi(x)$ — деяка функція, яку називають *щільністю розподілу* випадкової величини ξ.

Нехай $[a,b)$ — деякий інтервал, а ξ — неперервна випадкова величина, тоді

$$P\{a \le \xi < b\} = F_\xi(b) - F_\xi(a) = \int_a^b \rho_\xi(y)\,dy .$$

Якщо $b = a + \Delta a$, де Δa — достатньо мале число, то щільність на інтервалі $[a, a + \Delta a)$ можна вважати сталою і такою, що дорівнює $\rho_\xi(a)$, тобто

$$P\{a \le \xi < a + \Delta a\} \approx \rho_\xi(a)\Delta a .$$

Отже,

$$\rho_\xi(a) \approx \frac{P\{a \le \xi < a + \Delta a\}}{\Delta a} = \frac{F_\xi(a + \Delta a) - F_\xi(a)}{\Delta a},$$

причому рівність тим точніша, чим менше Δa. Перейшовши до границі при $\Delta a \to 0$, отримаємо точну рівність

$$\rho_\xi(a) = \lim_{\Delta a \to 0}\frac{F_\xi(a + \Delta a) - F_\xi(a)}{\Delta a} = F_\xi'(a) = \frac{dF_\xi(a)}{dx},$$

яка означає, що щільність розподілу неперервної випадкової величини є похідною від функції розподілу цієї величини.

Властивості щільності розподілу

1. Для всіх x $\rho_\xi(x) \ge 0$.

Ця властивість випливає з того, що $F_\xi(x)$ монотонно зростає по x, оскільки $\Delta F_\xi(x)/\Delta x \ge 0$ для всіх x і Δx. Перейшовши до границі при $\Delta x \to 0$, отримаємо необхідну нерівність для $\rho_\xi(x)$.

2. Умова нормування:

$$\int_{-\infty}^{+\infty} \rho_\xi(y)\,dy = 1 . \tag{24}$$

Ця властивість випливає з того, що

$$\int_{-\infty}^{+\infty} \rho_\xi(y)\,dy = F_\xi(+\infty) - F_\xi(-\infty) = 1 ,$$

відповідно до властивості 2 функції розподілу.

3. Нехай $[a,b)$ — деякий інтервал, ξ — неперервна випадкова величина, тоді

$$P\{a \le \xi < b\} = \int_a^b \rho_\xi(y)\,dy . \tag{25}$$

Приклад 49. Щільність розподілу випадкової величини ξ задана формулою $\rho(x) = \dfrac{\alpha}{1 + x^2}$, $-\infty < x < +\infty$. Знайти параметр α та функцію розподілу випадкової величини ξ.

Розв'язання. Параметр α визначимо, скориставшись умовою нормування щільності розподілу (24): $1=\int\limits_{-\infty}^{+\infty}\frac{\alpha}{1+x^2}dx=2\alpha\int\limits_{0}^{+\infty}\frac{dx}{1+x^2}=$ $=2\alpha\operatorname{arctg}x\,|_0^{+\infty}=\alpha\pi$, звідки $\alpha=1/\pi$.

Функцію розподілу обчислимо за формулою (23):

$$F(x)=\int\limits_{-\infty}^{x}\rho(y)\,dy=\int\limits_{-\infty}^{x}\frac{\alpha}{1+y^2}dy=\frac{1}{\pi}\operatorname{arctg}x\,|_{-\infty}^{x}=\frac{1}{2}+\frac{1}{\pi}\operatorname{arctg}x.$$

Приклад 50. Щільність розподілу випадкової величини ξ задана формулою

$$\rho(x)=\begin{cases}0, & -\infty<x<0;\\ (x-3)^2/9, & 0\le x\le 3;\\ 0, & 3<x<+\infty.\end{cases}$$

Знайти функцію розподілу випадкової величини ξ.

Розв'язання. Функцію розподілу знайдемо за формулою (23). Оскільки $\rho(x)=0$ при $-\infty<x<0$, тотожно отримаємо: $F(x)\equiv 0$.

Для області $0\le x\le 3$ визначимо: $F(x)=\int\limits_0^x\frac{(y-3)^2}{9}dy=\frac{(x-3)^3}{27}+1$.

Нарешті, при $3<x<+\infty$ $F(x)=1$.

Розглянемо три найпоширеніші випадкові величини, які мають неперервний розподіл.

7.2. Рівномірний розподіл

Випадкову величину ξ, яка має рівномірний розподіл на проміжку $[a,b]$, можна інтерпретувати як точку падіння деякого малого матеріального тіла, розмірами якого можна знехтувати, кинутого навмання на проміжок $[a,b]$.

Визначення. *Рівномірним* називають *розподіл*, щільність якого є сталою, відмінною від нуля на деякому проміжку $[a,b]$, і дорівнює нулю поза цим проміжком (рис. 7):

$$\rho(x)=\begin{cases}0, & x<a;\\ c, & a\le x\le b;\\ 0, & b<x.\end{cases}$$

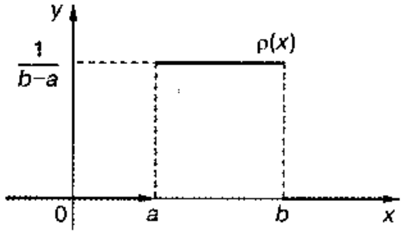


Рис. 7. Графік щільності рівномірного розподілу

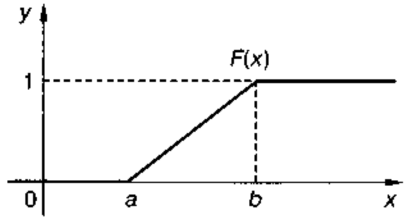


Рис. 8. Графік функції розподілу для рівномірного розподілу

Значення сталої c знайдемо з умови нормування функції щільності за формулою (24). У цьому випадку

$$1=\int\limits_{-\infty}^{+\infty}\rho(x)dx=\int\limits_a^b c\,dx=c(b-a),$$

звідки $c=\frac{1}{b-a}$. Функцію розподілу ілюструє рис. 8:

$$F(x)=\int\limits_{-\infty}^{x}\rho(y)dy=\begin{cases}0, & x\le a;\\ \frac{x-a}{b-a}, & a\le x\le b;\\ 1, & b\le x.\end{cases}$$

Нехай випадкова величина ξ рівномірно розподілена на інтервалі $[a,b]$. Тоді ймовірність попадання ξ у деякий інтервал $[c,d]\subseteq[a,b]$ дорівнює

$$P\{c\le\xi\le d\}=F(d)-F(c)=\frac{d-c}{b-a},$$

тобто пропорційна довжині інтервалу $d-c$ і не залежить від його розміщення на $[a,b]$.

7.3. Показниковий розподіл

Нехай ξ — випадкова величина, яка дорівнює проміжку часу між двома послідовними подіями у потоці подій, який має розподіл Пуассона з інтенсивністю λ. Тоді подія $\{\xi>t\}$ означає, що дві послідовні події в потоці розділені інтервалом часу більшим,

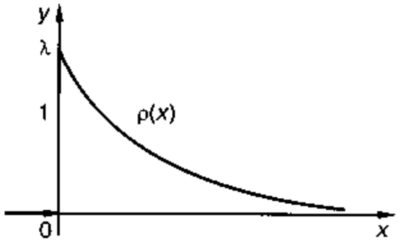


Рис. 9. Графік щільності показникового розподілу

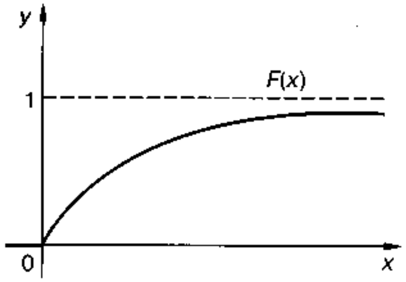


Рис. 10. Графік функції розподілу для показникового розподілу

ніж t, і згідно з властивістю потоку подій $P\{\xi > t\}$ є ймовірністю того, що в інтервалі часу $[0,t]$ не відбулось жодної події. Тому $P\{\xi > t\} = P_0(t) = e^{-\lambda t}$. Ймовірність протилежної події дорівнює $P\{\xi \le t\} = 1 - e^{-\lambda t}$, $t \ge 0$. Зрозуміло, що $P\{\xi > t\} = 0$, якщо $t < 0$.

Визначення. *Розподіл* випадкової величини ξ називають *показниковим*, якщо його щільність розподілу (рис. 9) дорівнює:

$$\rho(x) = \frac{dF}{dx}(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 \le x; \\ 0, & x < 0. \end{cases}$$

Функцію розподілу ілюструє рис. 10:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & 0 \le x; \\ 0, & x < 0. \end{cases}$$

Якщо, наприклад, $\lambda = 1$, то при $3 \le x$

$$\rho(x) \le \rho(3) = e^{-3} \approx 0,05; \quad F(x) \ge F(3) \approx 0,95,$$

тобто випадкова величина показниково розподілена з параметром $\lambda = 1$, з імовірністю, близькою до 1, набуває значень від 0 до 3, а більші значення — малоймовірні.

7.4. Нормальний (Гауссівський) розподіл

Важливість нормального, або Гауссівського, розподілу полягає в тому, що його застосовують у задачах, пов'язаних з підсумовуванням великої кількості незалежних випадкових величин, зокрема незалежних вимірювань об'єктів.

Визначення. *Нормальний (Гауссівський) розподіл* із параметрами a і σ^2 — це розподіл, щільність якого задає формула

$$\rho_{a,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

$$x \in (-\infty, \infty), \qquad (26)$$

де a — будь-яке фіксоване число; $\sigma > 0$.

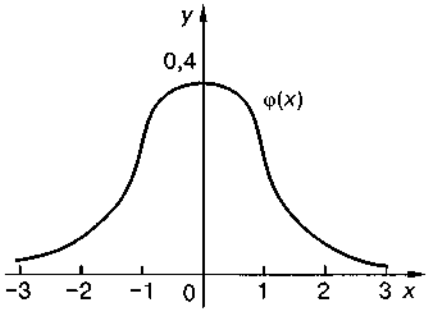


Рис. 11. Графік щільності стандартного нормального розподілу

Зауваження. Сукупність випадкових величин, які мають нормальний розподіл із параметрами a і σ^2 позначають через $N(a, \sigma^2)$. Якщо випадкова величина ξ має нормальний розподіл із параметрами a і σ^2, це записують так: $\xi \sim N(a, \sigma^2)$.

Множник $\dfrac{1}{\sigma\sqrt{2\pi}}$ введено для виконання умови нормування щільності розподілу (24).

Нормальний розподіл називають *стандартним*, якщо його параметри набувають значень $a = 0$, $\sigma = 1$. При цьому

$$\varphi(x) \equiv \rho_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in (-\infty, \infty) \text{ (рис. 11)}.$$

Властивості щільності стандартного нормального розподілу

1. $\varphi(x) > 0$ для всіх $x \in (-\infty, \infty)$.

2. $\varphi(x)$ — парна функція, тобто $\varphi(x) = \varphi(-x)$.

3. У точці $x = 0$ функція $\varphi(x)$ має локальний максимум, причому $\varphi(0) = \dfrac{1}{\sqrt{2\pi}} \approx 0,3989 \approx 0,4$.

4. При $x \in (-\infty, 0)$ функція $\varphi(x)$ монотонно зростає, при $x \in (0, \infty)$ — монотонно спадає.

5. $x = \pm 1$ — точки перегину функції $\varphi(x)$.

6. Вісь $0x$ є горизонтальною асимптотою, причому $\varphi(x) \to 0$ при $x \to \pm\infty$.

Таблиця 1. **Значення функції Лапласа $\Phi(x)$**

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,0	0,000	0,55	0,2088	1,3	0,4032	2,5	0,4938
0,01	0,0040	0,6	0,2257	1,4	0,4192	2,6	0,4953
0,05	0,0199	0,65	0,2422	1,5	0,4332	2,7	0,4965
0,1	0,0398	0,7	0,2580	1,6	0,4452	2,8	0,4974
0,15	0,0596	0,75	0,2734	1,7	0,4554	2,9	0,4981
0,2	0,0798	0,8	0,2881	1,8	0,4641	3,0	0,4986
0,25	0,0987	0,85	0,3023	1,9	0,4713	3,2	0,4993
0,3	0,1179	0,9	0,3159	2,0	0,4772	3,4	0,4997
0,35	0,1368	0,95	0,3289	2,1	0,4821	3,6	0,49984
0,4	0,1554	1,00	0,3413	2,2	0,4861	3,8	0,49992
0,45	0,1736	1,1	0,3643	2,3	0,4893	4,0	0,49996
0,5	0,1915	1,2	0,3849	2,4	0,4918	4,5	0,49997
						5,0	0,49998

7. Якщо $x \ge \pm 4$, то $\varphi(x) \le 0,0001$.

Через важливість функції $\varphi(x)$ вона табульована, тобто для неї складено таблиці значень.

Функція розподілу стандартного нормального розподілу має такий вигляд:

$$F(x) = \int_{-\infty}^{x} \varphi(y)\,dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}}\,dy.$$

Скориставшись властивостями визначених інтегралів, $F(x)$ можна подати у вигляді

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{y^2}{2}}\,dy + \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{y^2}{2}}\,dy.$$

Інтеграл $\int_{-\infty}^{0} e^{-\frac{y^2}{2}}\,dy = \int_{0}^{\infty} e^{-\frac{y^2}{2}}\,dy = \frac{\sqrt{2\pi}}{2}$ називають *інтегралом **Пуассона***, функцію

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{y^2}{2}}\,dy \quad — \qquad (27)$$

стандартним інтегралом ймовірностей, або *функцією **Лапласа***. Отже, $F(x) = \frac{1}{2} + \Phi(x)$.

Стандартний інтеграл ймовірностей не можна виразити через скінченний набір елементарних функцій, тобто він не обчислюється звичайними методами. Для обчислення функції $\Phi(x)$ складено таблицю (табл. 1), де x змінюється від 0 до 5.

Властивості стандартного інтеграла ймовірностей

1. $\Phi(0) = \frac{1}{\sqrt{2\pi}} \int_{0}^{0} e^{-\frac{y^2}{2}}\,dy = 0$, $\Phi(\infty) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{y^2}{2}}\,dy = \frac{1}{2}$.

2. Оскільки підінтегральна функція $e^{\frac{-y^2}{2}}$ є додатною, то функція $\Phi(x)$ монотонно зростає при $x > 0$ і при $x = 5$ $\Phi(5) = 0,499997 \approx 0,5$. Тому для усіх $x \ge 5$ $\Phi(x)$ вважають такою, що дорівнює 0,5.

3. $\Phi(x)$ є непарною функцією, тобто $\Phi(-x) = -\Phi(x)$.

Справді,

$$\Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{-x} e^{-\frac{y^2}{2}}\,dy = \left| \begin{array}{lll} \text{заміна} & y = -z & \\ & dy = -dz & \\ \text{при} & y = 0 & z = 0 \\ & y = -x & z = x \end{array} \right| =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{(-z)^2}{2}}(-dz) = -\frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{z^2}{2}}\,dz = -\Phi(x).$$

4. Нехай випадкова величина ξ має стандартний нормальний розподіл, $[\alpha, \beta]$ — деякий інтервал. Тоді $P\{\alpha \le \xi \le \beta\} = F(\beta) - F(\alpha) = \frac{1}{2} + \Phi(\beta) - \left(\frac{1}{2} + \Phi(\alpha)\right)$ й, отже, $P\{\alpha \le \xi \le \beta\} = \Phi(\beta) - \Phi(\alpha)$.

Вивчення загального нормального розподілу зводиться до стандартного нормального розподілу.

1. Нехай $\sigma = 1$ і $a \ne 0$. Тоді $\rho_{a,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2}}$.

Графік функції $\rho_{a,1}(x)$ буде горизонтально зміщеним відносно графіка функції $\rho_{0,1}(x)$ на a одиниць (рис. 12, 13).

Параметр a природно назвати ***центром нормального розподілу***.

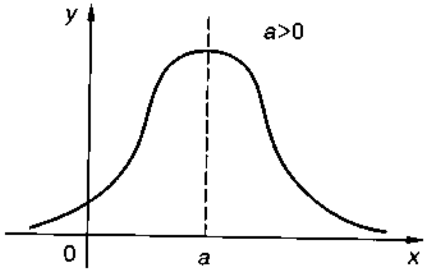


Рис. 12. Графік щільності нормального розподілу для $a > 0$

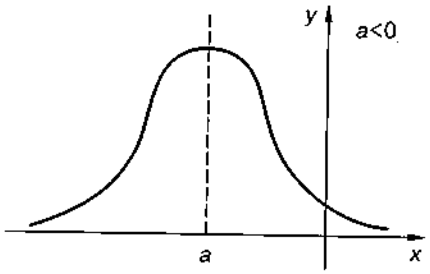


Рис. 13. Графік щільності нормального розподілу для $a < 0$

2. Якщо значення a фіксоване, то при збільшенні σ максимальне значення щільності $p_{a,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$, яке дорівнює $\frac{1}{\sigma\sqrt{2\pi}}$, буде зменшуватись. У цьому випадку графік функції $\rho_{a,\sigma}(x)$ "розповзається". Навпаки, при зменшенні σ максимальне значення функції $\rho_{a,\sigma}(x)$ збільшується і її графік "звужується" (рис. 14).

Отже, параметр σ визначає відхилення нормально розподіленої випадкової величини від свого центра.

Нехай ξ — випадкова величина, яка має нормальний розподіл з параметрами a і σ^2, $[\alpha,\beta]$ — деякий інтервал. Тоді

$$P\{\alpha \le \xi \le \beta\} = \int_{\alpha}^{\beta} \rho_{a,\sigma}(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} e^{-\frac{y^2}{2}}\,dy = \left| \begin{array}{l} \text{заміна } \dfrac{x-a}{\sigma} = y \\ dx = \sigma\,dy \end{array} \right| =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\beta-a}{\sigma}} e^{-\frac{y^2}{2}}\,dy - \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\alpha-a}{\sigma}} e^{-\frac{y^2}{2}}\,dy = \Phi\left(\frac{\beta-a}{2}\right) - \Phi\left(\frac{\alpha-a}{2}\right).$$

Остаточно отримуємо формулу

$$P\{\alpha \le \xi \le \beta\} = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right). \qquad (28)$$

Ймовірності попадання нормально розподіленої випадкової величини у деякий інтервал обчислюють за формулою (28) з використанням таблиці для стандартного інтеграла ймовірностей.

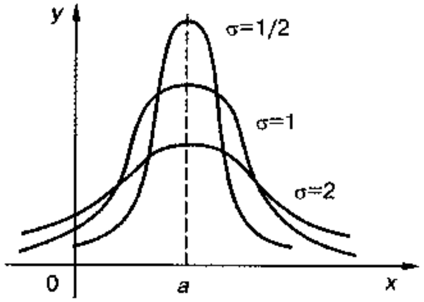


Рис. 14. Графіки нормального розподілу для різних σ

Зокрема, нехай $\alpha = a - t\sigma$, $\beta = a + t\sigma$. Тоді $P\{a - t\sigma \le \xi \le a + t\sigma\} = \Phi\left(\frac{a+t\sigma-a}{\sigma}\right) - \Phi\left(\frac{a-t\sigma-a}{\sigma}\right)$, тобто $P\{|\xi - a| \le t\sigma\} = 2\Phi(t)$.

Якщо $t = 1$, то $P\{\,|\xi - a| \le \sigma\} = 2\Phi(1) \approx 0{,}68268$.

Якщо $t = 2$, то $P\{\,|\xi - a| \le 2\sigma\} = 2\Phi(2) \approx 0{,}9545$.

Якщо $t = 3$, то $P\{\,|\xi - a| \le 3\sigma\} = 2\Phi(3) \approx 0{,}9973 \approx 1$.

Останню формулу називають *правилом "3σ"*. Згідно з цим правилом, практично достовірною є подія, яка полягає в тому, що нормально розподілена випадкова величина з параметрами a і σ^2 відхилиться від свого центра a не більш як на 3σ.

Правило "3σ" часто застосовують у практичних обчисленнях, пов'язаних із нормальним розподілом.

Наступні три розподіли часто застосовують у математичній статистиці.

χ^2-Розподіл Пірсона

Нехай $\xi_1, \xi_2, \ldots, \xi_n$ — n однаково розподілених незалежних випадкових величин, які мають стандартний нормальний розподіл: $\xi_i \sim N(0,\,1)$, $1 \le i \le n$.

Визначення. χ^2 (хі-квадрат)-*розподілом Пірсона* з n ступенями вільності називають випадкову величину

$$\chi^2 = \sum_{i=0}^{n} \xi_i^{\,2}.$$

Графіки щільності χ^2-розподілу Пірсона наведено на рис. 15.

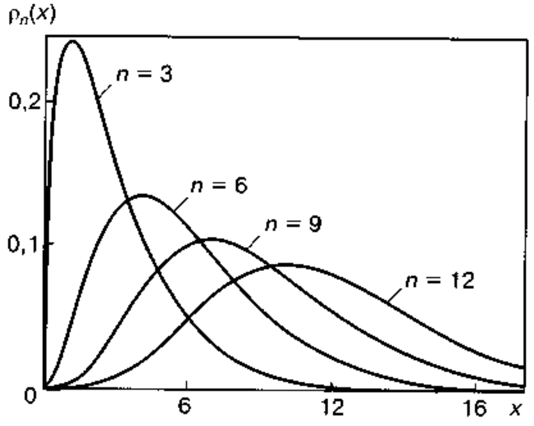


Рис. 15. Графіки щільності χ^2-розподілу Пірсона за різних n

t-Розподіл Стьюдента

Нехай $\xi_0, \xi_1, \xi_2, ..., \xi_n$ — $n+1$ однаково розподілена незалежна випадкова величина, яка має стандартний нормальний розподіл: $\xi_i \sim N(0,1)$, $0 \le i \le n$.

Визначення. *t-Розподілом Стьюдента* з n ступенями вільності називають випадкову величину

$$t = \frac{\xi_0}{\sqrt{\frac{1}{n}\sum_{k=0}^{n}\xi_i^2}}.$$

Зауваження. Стьюдент — це псевдонім Госсета. Щільність *t-розподілу Стьюдента* прямує до щільності нормального розподілу, якщо $n \to +\infty$.

Графік щільності t-розподілу Стьюдента наведено на рис. 16.

F-розподіл Фішера

Нехай $\xi_1, \xi_2, ..., \xi_m, \eta_1, \eta_2, ..., \eta_n$ — $m + n$ однаково розподілених незалежних випадкових величин, які мають стандартний нормальний розподіл: $\xi_i, \eta_i \sim N(0, 1)$, $1 \le i \le m$, $1 \le j \le n$.

Визначення. *F-розподілом Фішера* називають випадкову величину F, яку обчислюють за формулою

$$F = \frac{\frac{1}{m}\sum_{i=0}^{m}\xi_i^2}{\frac{1}{n}\sum_{j=0}^{n}\eta_j^2}.$$

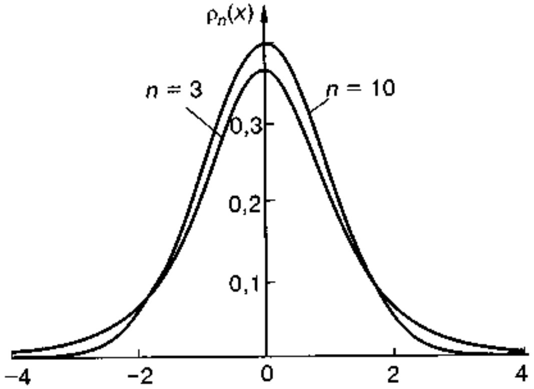


Рис. 16. Графік щільності t-розподілу Стьюдента

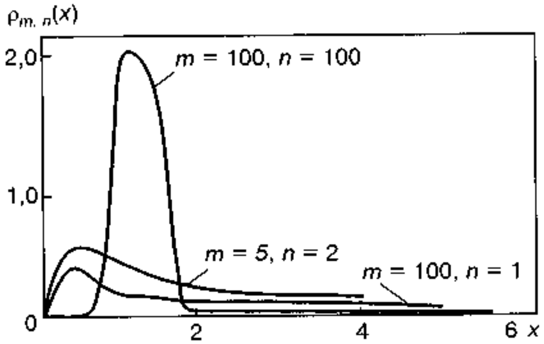


Рис. 17. Графіки щільності F-розподілу Фішера за різних m і n

Графіки щільності F-розподілу Фішера наведено на рис. 17.

7.5. Числові характеристики неперервних випадкових величин

Математичне сподівання і дисперсія виконують для неперервних випадкових величин таку саму роль, як і для дискретних.

Нехай неперервна випадкова величина ξ задана щільністю розподілу $\rho_\xi(x)$.

Визначення. *Математичним сподіванням* неперервної випадкової величини ξ називають число

$$E\xi = \int_{-\infty}^{+\infty} x\rho_\xi(x)\,dx, \tag{29}$$

якщо інтеграл $\int_{-\infty}^{+\infty}|x|\rho_\xi(x)dx < \infty$.

Визначення. *Дисперсією* неперервної випадкової величини ξ називають число

$$D\xi = E\xi^2 - (E\xi)^2 = \int_{-\infty}^{+\infty} x^2 \rho_\xi(x)\,dx - \left[\int_{-\infty}^{+\infty} x\rho_\xi(x)\,dx\right]^2. \qquad (30)$$

Визначення. *Середньоквадратичним відхиленням* неперервної випадкової величини ξ називають квадратний корінь з її дисперсії:

$$\sigma(\xi) = \sqrt{D\xi}.$$

Математичне сподівання, дисперсія і середньоквадратичне відхилення неперервної випадкової величини мають такі самі властивості, як і математичне сподівання, дисперсія і середньоквадратичне відхилення дискретної випадкової величини.

Числові характеристики рівномірного розподілу

$$E\xi = \int_{-\infty}^{+\infty} x\rho(x)dx = \int_a^b x\frac{1}{b-a}dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

$$D\xi = \int_{-\infty}^{+\infty} x^2\rho(x)dx - \left(\int_{-\infty}^{+\infty} x\rho(x)dx\right)^2 = \int_a^b x^2\frac{1}{b-a}dx - \frac{(a+b)^2}{4} =$$

$$= \left.\frac{x^3}{3(b-a)}\right|_a^b - \frac{(a+b)^2}{4} = \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4} =$$

$$= \frac{4\left(b^2+ab+a^2\right) - 3\left(a^2+2ab+b^2\right)}{12} = \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}.$$

Нарешті, $\sigma(\xi) = \dfrac{b-a}{2\sqrt{3}}$.

Числові характеристики показникового розподілу

$$E\xi = \int_{-\infty}^{+\infty} x\rho(x)\,dx = \int_0^{+\infty} xe^{-\lambda x}dx = \lambda\lim_{A\to+\infty}\int_0^A xe^{-\lambda x}dx =$$

$$= \left|\begin{array}{l} u = x, dv = e^{-\lambda x}dx \\ du = dx, v = -\frac{1}{\lambda}e^{-\lambda x}\end{array}\right| = \lambda\lim_{A\to+\infty}\left[-\frac{A}{\lambda}e^{-\lambda A} - \frac{1}{\lambda^2}e^{-\lambda A} + \frac{1}{\lambda^2}\right] = \frac{1}{\lambda},$$

тому що $\lim\limits_{A\to+\infty} Ae^{-\lambda A} = \lim\limits_{A\to+\infty} e^{-\lambda A} = 0$. Дисперсію обчислюють аналогічно: $D\xi = \dfrac{1}{\lambda^2}$.

Нарешті, $\sigma(\xi) = \dfrac{1}{\lambda}$.

7.5.3. Числові характеристики стандартного нормального розподілу

$E\xi = \displaystyle\int_{-\infty}^{\infty} x\varphi(x)\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}}dx = 0$ в силу непарності підінтегральної функції $f(x) = xe^{-\frac{x^2}{2}}$. Далі

$$D\xi = \int_{-\infty}^{\infty} x^2\rho(x)\,dx - \left(\int_{-\infty}^{\infty} x\rho(x)\,dx\right)^2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}dy =$$

$$= \frac{1}{\sqrt{2\pi}}\lim_{A\to\infty}\int_{-A}^{A} x^2e^{-\frac{x^2}{2}}dx = \frac{2}{\sqrt{2\pi}}\lim_{A\to\infty}\int_0^A x^2e^{-\frac{x^2}{2}}dx =$$

$$= \left|\begin{array}{ll} u = x & dv = xe^{-\frac{x^2}{2}}dx \\ du = dx & v = e^{-\frac{x^2}{2}}\end{array}\right| = \frac{2}{\sqrt{2\pi}}\lim_{A\to\infty}\left(\left. x^2e^{-\frac{x^2}{2}}\right|_0^A + \right.$$

$$\left. + \int_0^A e^{-\frac{x^2}{2}}dx\right) = \frac{2}{\sqrt{2\pi}}\lim_{A\to\infty} A^2e^{-\frac{A^2}{2}} + \frac{2}{\sqrt{2\pi}}\int_0^{\infty} e^{-\frac{x^2}{2}}dx = 1,$$

оскільки $\lim\limits_{A\to\infty} A^2e^{-\frac{A^2}{2}} = 0$. Нарешті, $\sigma(\xi) = 1$.

Числові характеристики загального нормального розподілу

$$E\xi = \int_{-\infty}^{\infty} x\rho_{a,\sigma}(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{-\frac{(x-a)^2}{2\sigma^2}}dx = \left|\begin{array}{l}\text{заміна}\quad \frac{x-a}{\sigma} = y \\ \qquad\qquad dx = \sigma\,dy\end{array}\right| =$$

$$=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma y+a)e^{-\frac{y^2}{2}}(\sigma dy)=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ye^{-\frac{y^2}{2}}dy+\frac{a}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy=$$

$$=0+1\cdot a=a.$$

$$D\xi=\int_{-\infty}^{\infty}x^2\rho_{a,\sigma}(x)dx-a^2.$$

Обчисливши останній інтеграл за допомогою аналогічної заміни змінних, отримаємо $D\xi=\sigma^2$.

Отже, параметр a — математичне сподівання нормально розподіленої випадкової величини, а параметр σ — середньоквадратичне відхилення нормально розподіленої випадкової величини.

Приклад 51. Знайти числові характеристики випадкової величини ξ, яка задана функцією розподілу

$$F(x)=\begin{cases}0, & x\le 0;\\ x^2, & 0<x\le 1;\\ 1, & 1<x.\end{cases}$$

Розв'язання. Знайдемо щільність розподілу випадкової величини ξ. Маємо

$$\rho(x)=\frac{dF}{dx}(x)=\begin{cases}2x, & 0\le x\le 1;\\ 0, & x\notin[0,1].\end{cases}$$

Обчислимо $E\xi$ і $D\xi$. Отримаємо

$$E\xi=\int_{-\infty}^{+\infty}x\rho_\xi(x)\,dx=\int_0^1 x2x\ dx=\frac{2}{3};$$

$$D\xi=E\xi^2-(E\xi)^2=\int_0^1 x^2 2x\,dx-\left(\frac{2}{3}\right)^2=\frac{1}{18};$$

$$\sigma(\xi)=\sqrt{D\xi}=\sqrt{\frac{1}{18}}\approx 0,24.$$

Приклад 52. Маса виготовленої на фармацевтичному підприємстві таблетки — нормально розподілена випадкова величина з параметрами $a=50$ мг і $\sigma^2=\frac{1}{625}$ мг^2. Визначити ймовірність виготовлення бракованої таблетки, якщо допустимі значення маси таблетки становлять $(50\pm0,1)$ мг.

Розв'язання. Ймовірності попадання нормально розподіленої випадкової величини в інтервал $(a-\delta,\ a+\delta)$ обчислюють за формулою (28) із використанням таблиці для стандартного інтеграла ймовірностей: $P=2\Phi\left(\frac{\delta}{\sigma}\right)=2\Phi(2,5)\approx0,9876$. Отже, шукана ймовірність браку становить $1-P\{a-\delta\le\xi\le a+\delta\}=0,0124$.

Приклад 53. Відомо, що зріст досліджуваних тварин розподілений за нормальним законом, причому $E\xi=a=75$ мм і $\sigma(\xi)=\sigma=6$ мм. Знайти ймовірність того, що хоча б одна з трьох навмання взятих тварин матиме зріст від 70 до 80 мм.

Розв'язання. Знайдемо ймовірність того, що зріст ξ однієї окремо взятої навмання піддослідної тварини знаходиться між 70 і 80 мм. Згідно з властивостями нормального розподілу маємо $p=P\{70\le\xi\le80\}=\Phi\left(\frac{80-75}{6}\right)-\Phi\left(\frac{70-75}{6}\right)=2\Phi\left(\frac{5}{6}\right)\approx0,59$ (див. табл. 1). Ймовірність P_3 шуканої події можна знайти за схемою Бернуллі: $P_3(1\le k\le3)=1-(1-p)^3=1-(0,41)^3=0,93$.

Визначення. Якщо ознака групи об'єктів описується нормально розподіленою випадковою величиною ξ, то ***нормованим відхиленням від середнього значення*** конкретного представника x цієї групи називають величину

$$t=\frac{x-a}{\sigma},$$

де $E\xi=a$ і $\sigma(\xi)=\sigma$.

Приклад 54. При обстеженні групи підлітків з'ясувалося, що їх зріст описується нормальним розподілом із параметрами $a=164,8$ см і $\sigma=5,8$ см. У групі виявився підліток, зріст якого $x=172,4$ см. Визначити нормоване відхилення його зросту від середнього зросту обстежених підлітків.

Розв'язання. Шукане відхилення визначимо за формулою

$$t=\frac{x-a}{\sigma}=\frac{172,4-164,8}{5,8}=+1,31\ \text{см}.$$

Зауваження. Знак відхилення вказує на те, в який бік від середнього значення відхиляється параметр конкретного представника групи: знак "+" — більше за середнє значення; знак "–" — менше за середнє значення.

Числові характеристики χ^2-розподілу Пірсона

Числові характеристики χ^2-розподілу Пірсона з n ступенями вільності такі:

$$E\chi^2 = n;\ D\chi^2 = 2n;\ \sigma(\chi^2) = \sqrt{2n}.$$

Числові характеристики t-розподілу Стьюдента

Числові характеристики t-розподілу Стьюдента з n ступенями вільності такі:

$$Et = 0;\ Dt = \frac{n}{n-2};\ \sigma(t) = \sqrt{\frac{n}{n-2}}.$$

Числові характеристики F-розподілу Фішера

Числові характеристики F-розподілу Фішера такі:

$$EF = \frac{n}{n-2},\ \ n > 2;\ \ DF = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)},\ \ n > 4;$$

$$\sigma(F) = \sqrt{\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}}.$$

§ 8. ГРАНИЧНІ ТЕОРЕМИ ТЕОРІЇ ЙМОВІРНОСТЕЙ

У цьому параграфі наведено деякі граничні теореми теорії ймовірностей, які дають уявлення про поведінку статистичних характеристик моделі у випадку значного збільшення кількості результатів вимірювання.

Важливість граничних теорем теорії ймовірностей обумовлена двома обставинами. По-перше, статистичні параметри, отримані в результаті вибіркового дослідження, наближаються до теоретичних ймовірнісних показників. Умови такої збіжності формулюють у вигляді граничних теорем. По-друге, відповідність теоретичної гіпотези про випадкову модель піддослідного явища отриманим результатам обстеження цього явища як правило перевіряють за допомогою конструкцій, які будують у вигляді граничних теорем.

Класичний підхід до означення ймовірності не завжди безпосередньо дає належний результат. Досить часто він здійснюється асимптотично, тобто зі збільшенням числа випробувань. У таких випадках події мають частоту, відхилення якої від їхньої ймовірності прямує до нуля. Чітким прикладом такої події є співвідношення статі в потомстві багатьох тварин і людини. Відомо, що стать потомства визначається в момент запліднення, коли в зиготу потрапляють або XX або XY хромосоми. Отже, ймовірність появи у потомстві особин чоловічої й жіночої статі одна і та ж: $p = 1/2$. Однак статистика свідчить, що, наприклад, на 1000 новонароджених за певний проміжок часу 52 % припадає на чоловічу стать і 48 % — на жіночу. Зрозуміло, що зі збільшенням кількості новонароджених (числа випробувань) відхилення появи осіб чоловічої статі від жіночої зменшуватиметься, тобто ймовірність їх появи прямуватиме до $p = 1/2$. У цьому факті виявляється дія так званого ***закону великих чисел***, який наведено наприкінці цього параграфа.

Як уже зазначалось, велике практичне значення має *схема випробувань Бернуллі*. Нагадаємо, що ймовірність k успіхів у n незалежних випробуваннях за схемою Бернуллі з фіксованою ймовірністю успіху p в одному випробуванні обчислюють за формулою

$$P_n(k) = P\{\xi_n = k\} = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)} p^k q^{n-k},\ \ q = 1 - p.$$

Для великих n і k ця формула надзвичайно громіздка. Якщо, наприклад, $n = 100$, $k = 50$, то для обчислення $P_{100}(50)$ необхідно виконати більш як 200 арифметичних операцій. Для спрощення обчислень застосовують теорему 1.

Теорема 1 (локальна теорема Муавра—Лапласа). Нехай ймовірність успіху в кожному випробуванні за схемою Бернуллі дорівнює p, $0 < p < 1$. Тоді ймовірність $P_n(k)$ того, що в n незалежних випробуваннях настане k успіхів, задовольняє співвідношення

$$\sqrt{npq}\ P_n(k) = \varphi(x)(1 + \alpha_n(x)),$$

де $x = \dfrac{k - np}{\sqrt{npq}}$, $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $\lim\limits_{n\to\infty} \alpha_n(x) = 0$.

На практиці локальну теорему Муавра—Лапласа застосовують у вигляді наближеної формули

$$P_n(k) \approx \frac{1}{\sqrt{npq}}\varphi\left(\frac{k-np}{\sqrt{npq}}\right), \quad (n \ge 50), \tag{31}$$

яка є тим точнішою, чим ближче p, а отже і q, до 0,5.

Приклад 55. Знайти ймовірність того, що з 600 виробів, виготовлених на підприємстві, 250 будуть вищого сорту, якщо ймовірність того, що виріб має вищий сорт дорівнює 0,4.

Розв'язання. За умовою, $n = 600$, $k = 250$, $p = 0,4$, $q = 0,6$. Тому $x = \frac{k-np}{\sqrt{npq}} = \frac{250 - 600\cdot 0,4}{\sqrt{600\cdot 0,4\cdot 0,6}} = \frac{10}{12} \approx 0,833$. За даними табл. 1 знаходимо $\varphi(0,833) = 0,282$.

Відповідно до локальної граничної теореми, ймовірність того, що 250 виробів із 600 будуть вищого сорту, дорівнює

$$P_{600}(250) = \frac{1}{\sqrt{600\cdot 0,4\cdot 0,6}}\cdot 0,282 = \frac{0,282}{12} = 0,0235.$$

Якщо p близьке до 0 чи 1, то, застосувавши вираз (31) навіть для великих n, можна отримати значну похибку. У цих випадках потрібно використовувати граничну теорему (формулу) Пуассона.

Нехай проведено n серій випробувань за схемою Бернуллі, по n випробувань у кожній серії, причому ймовірність успіху в i-й серії дорівнює $p_i = \lambda/i$, де λ — фіксоване число, $\lambda \le i \le n$.

Теорема 2 (теорема Пуассона). Якщо $P_n(k)$ — ймовірність k успіхів у n-й серії випробувань за схемою Бернуллі з ймовірністю успіху $p_i = \lambda/i$ у i-й серії, то

$$\lim_{n\to\infty} P_n(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Доведення. Відповідно до формули біноміального розподілу маємо

$$P_n(k) = C_n^k (p_n)^k (1-p_n)^{n-k} = \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} =$$

$$= \frac{1}{k!}n(n-1)(n-2)\ldots(n-k+1)\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} =$$

$$= \frac{1}{k!}\frac{n}{n}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\lambda^k\left(1-\frac{\lambda}{n}\right)^{n-k}.$$

Якщо $n \to \infty$, а число k фіксоване, то кожен співмножник $\left(1-\frac{1}{n}\right), \left(1-\frac{2}{n}\right), \ldots, \left(1-\frac{k-1}{n}\right)$ прямує до 1. Крім того,

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-k} = \lim_{n\to\infty}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} = \frac{e^{-\lambda}}{1} = e^{-\lambda}.$$ Тому $\lim\limits_{n\to\infty} P_n(k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

Це означає, що граничний розподіл числа успіхів описується законом Пуассона.

Дослідження цієї формули показало, що її можна застосовувати для таких випробувань за схемою Бернуллі, в яких p є малим $(p \le 0,1)$, n — досить великим $(n \ge 50)$ і $np \le 10$. Тоді

$$P_n(k) \approx \frac{e^{-np}(np)^k}{k!}. \tag{32}$$

Приклад 56. Ймовірність виготовлення бракованої вакцини дорівнює 0,004. Знайти ймовірність того, що з 1000 доз виготовлених вакцин 5 будуть бракованими.

Розв'язання. За умовою, $n = 1000$, $k = 5$, $p = 0,004$, $np = 4$. Ці числа задовольняють вимоги $p \le 0,1$, $n \ge 50$, $np \le 10$. Щоб знайти ймовірності $P_{1000}(5)$, застосуємо формулу (32):

$$P_{1000}(5) = \frac{e^{-4}4^5}{5!} \approx 0,1563.$$

Якщо ймовірність $P_{1000}(5)$ обчислювати за формулою (31), то вона приблизно дорівнюватиме 0,1763, а її справжнє значення, отримане згідно з біноміальним розподілом, близьке до 0,1552. Отже, відносна похибка обчислення за формулою (31) становить 13,6 %, за формулою (32) — 0,7 %, що значно менше.

У багатьох практичних задачах при проведенні випробувань за схемою Бернуллі важливо знати ймовірність попадання числа успіхів у деякий інтервал. Так, партію з 1000 виробів контролер приймає, якщо число бракованих виробів не перевищує 50; серію зі 100 незалежних експериментів вважають успішною, якщо більш як 80 із них підтверджують деяке припущення; гравця вважають переможцем у шаховому турнірі з 24 партій, якщо він виграв більш як 12 партій і т. д. Шукати ймовірності попадання числа успіхів в інтервал за допомогою локальної теореми Муавра—Лапласа чи за теоремою Пуассона незручно, тому що доведеться обчислювати і додавати велике число ймовірностей, й отже, значно зросте похибка обчислень. У цьому разі застосовують інтегральну теорему.

Теорема (інтегральна теорема Муавра—Лапласа). Нехай імовірність успіху в кожному випробуванні за схемою Бернуллі дорівнює p, $0 < p < 1$; ξ_n — число успіхів у n випробуваннях за схемою Бернуллі. Тоді для будь-яких $a < b$

$$\lim_{n\to\infty} P\left\{a \le \frac{\xi_n - np}{\sqrt{npq}} \le b\right\} = \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{y^2}{2}}dy. \quad (33)$$

На практиці інтегральну теорему Муавра—Лапласа застосовують у вигляді наближеної рівності. Нехай $k_2 > k_1$ — цілі числа, тоді

$$P\{\ k_1 \le \xi_n < k_2\ \} \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right), \quad (34)$$

де $\Phi(x)$ — стандартний інтеграл ймовірностей (див. табл. 1).

Приклад 57. Підприємство випускає в середньому 4 % бракованих вакцин. Яка ймовірність того, що число бракованих вакцин у партії з 4000 вакцин не перевищить 170?

Розв'язання. Маємо $n = 4000$, $k_1 = 0$, $k_2 = 170$, $p = 0,04$, $q = 0,96$. Якщо ξ_{4000} — число бракованих вакцин, то за формулою (34) отримаємо

$$P\{\ 0 \le \xi_{4000} < 170\ \} = \Phi\left(\frac{170 - 4000\cdot 0,04}{\sqrt{4000\cdot 0,04\cdot 0,96}}\right) -$$

$$-\Phi\left(\frac{0 - 4000\cdot 0,04}{\sqrt{4000\cdot 0,04\cdot 0,96}}\right) \approx \Phi(0,806) + \Phi(13) \approx 0,79.$$

Нехай проведено n випробувань за схемою Бернуллі, ξ_n — число успіхів. Відносною частотою успіху називають величину ξ_n/n. Перетворимо формулу (33) до вигляду, зручного для вивчення поведінки ξ_n/n:

$$\lim_{n\to\infty} P\left\{a\sqrt{\frac{pq}{n}} \le \frac{\xi_n}{n} - p \le b\sqrt{\frac{pq}{n}}\right\} = \Phi(b) - \Phi(a).$$

Поклавши $a = -\varepsilon\sqrt{n/pq}$, $b = \varepsilon\sqrt{n/pq}$, отримаємо наближене співвідношення:

$$P\left\{\left|\frac{\xi_n}{n} - p\right| \le \varepsilon\right\} \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right). \quad (35)$$

Формулу (35) використовують для розв'язування трьох типів завдань.

1. Задано число випробувань n, ймовірності p та q і відхилення ε. Знайти ймовірність того, що відносна частота ξ_n/n відхилиться від ймовірності успіху p не більш як на ε.

2. Задано число випробувань n, ймовірності p та q і число α, $0 < \alpha < 1$. Знайти найбільше значення відхилення ε, для якого $P\left\{\left|\frac{\xi_n}{n} - p\right| < \varepsilon\right\} = \alpha$.

Шукаємо в табл. 1 таке t, щоб $2\Phi(t) = \alpha$; за формулою (35) обчислюємо: $\varepsilon = t\sqrt{\frac{pq}{n}}$.

3. Задано числа ε і α, $0 < \alpha < 1$; ймовірності p і q. Знайти найменше n, для якого $P\left\{\left|\frac{\xi_n}{n} - p\right| < \varepsilon\right\} = \alpha$.

Шукаємо за табл. 1 таке t, щоб $2\Phi(t) = \alpha$. За формулою (35) знаходимо $\varepsilon = t\sqrt{\frac{pq}{n}}$, звідки $n = \frac{pqt^2}{\varepsilon^2}$.

8.1. Центральна гранична теорема

Згідно з формулою (33), граничний розподіл випадкової величини $\frac{\xi_n - np}{\sqrt{npq}}$, де ξ_n — число успіхів у n випробуваннях за схемою Бернуллі, є стандартним нормальним. Зазначимо, що ξ_n можна подати у вигляді суми незалежних випадкових величин $\xi_n = \eta_1 + \eta_2 + \ldots + \eta_n$, де $\eta_i = \begin{cases} 1, & \text{якщо в } i\text{-му експерименті настає успіх;} \\ 0, & \text{якщо в } i\text{-му експерименті настає невдача.} \end{cases}$

При цьому $E\xi_n = E\eta_1 + E\eta_2 + ... + E\eta_n = np$, $D\xi_n = D\eta_1 + D\eta_2 + ... + D\eta_n = npq$.

Тому ***центральну граничну теорему*** можна сформулювати так: якщо $\eta_1, \eta_2, ..., \eta_n$ — незалежні випадкові величини, що набувають значення 1 з імовірністю p і значення 0 з імовірністю q, то

$$\lim_{n\to\infty} P\left\{ a < \frac{\sum_{i=1}^{n}\eta_i - \sum_{i=1}^{n} E\eta_i}{\sqrt{\sum_{i=1}^{n} D\eta_i}} < b \right\} = \Phi(b) - \Phi(a).$$

Аналогічне твердження, за наявності додаткових умов, є справедливим для довільних незалежних випадкових величин. Ці теореми вперше довели А.А. Марков і А.М. Ляпунов у XIX ст.

Теорема 1 (теорема Ляпунова). Якщо для послідовності взаємно незалежних випадкових величин $\eta_1, \eta_2, ...\eta_n, ...$ існує таке $\delta > 0$, що

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} E(|\eta_i - E\eta_i|^{2+\delta})}{\left(\sqrt{\sum_{i=1}^{n} D\eta_i}\right)^{2+\delta}} = 0,$$

тоді рівномірно по x

$$\lim_{n\to\infty} P\left\{ \frac{\sum_{i=1}^{n}\eta_i - \sum_{i=1}^{n} E\eta_i}{\sqrt{\sum_{i=1}^{n} D\eta_i}} < x \right\} = \Phi(x).$$

Умова теореми Ляпунова виконується, зокрема, якщо η_1, $\eta_2, ..., \eta_n, ...$ — незалежні й однаково розподілені, тобто функції їх розподілу однакові. Це означає, що їх математичні сподівання і дисперсії однакові.

Теорема 2 (теорема Ліндерберга—Леві). Нехай випадкові величини $\eta_1, \eta_2, ..., \eta_n, ...$ однаково розподілені, $E\eta_i = a$, $D\eta_i = \sigma^2$, $i = 1$, 2,... . Тоді рівномірно по x

$$\lim_{n\to\infty} P\left\{ \frac{\sum_{i=1}^{n}\eta_i - na}{\sigma\sqrt{n}} < x \right\} = \Phi(x).$$

Зміст центральної граничної теореми полягає в тому, що сума великої кількості незалежних випадкових доданків за певного нормування має граничний стандартний нормальний розподіл, тобто не залежить від розподілів окремих доданків.

8.2. Закон великих чисел

Як зазначалось раніше, відносні частоти деяких подій мають статистичну стійкість при збільшенні числа випробувань. Цей емпіричний факт необхідно обґрунтувати. Наведемо підтвердження у вигляді закону великих чисел.

Лема 1. Нехай ξ — невід'ємна випадкова величина: $\xi = \xi(\omega) \geq 0$ за будь-якого $\omega \in \Omega$. Якщо існує математичне сподівання $E\xi$, то $P\{\xi \geq 1\} \leq E\xi$.

Доведення. Нехай ξ — неперервна випадкова величина (для дискретних величин доведення аналогічне). Тоді

$$P\{\xi \geq 1\} = \int_1^\infty \rho_\xi(x)dx \leq \int_1^\infty x\rho_\xi(x)\,dx \int_0^\infty x\rho_\xi(x)\,dx = E\xi.$$

Лема 2 (нерівність Чебишова). Нехай ξ — випадкова величина, що має математичне сподівання $E\xi$ і дисперсію $D\xi$. Тоді для будь-якого $\varepsilon > 0$ справедлива нерівність

$$P\{|\xi - E\xi| \geq \varepsilon\} \leq \frac{D\zeta}{\varepsilon^2}.$$

Доведення. Розглянемо випадкову величину $\zeta = \frac{|\xi - E\xi|^2}{\varepsilon^2}$. Очевидно, що $\zeta \geq 0$ і $E\zeta = \frac{E\left([\xi - E\xi]^2\right)}{\varepsilon^2} = \frac{D\xi}{\varepsilon^2}$ існує. Тому ζ задовольняє умови леми 1, й отже, $P\{\zeta \geq 1\} \leq E\zeta$, але $P\{\zeta \geq 1\} = = P\left\{\frac{|\xi - E\xi|^2}{\varepsilon^2} \geq 1\right\} = P\{|\xi - E\xi| \geq \varepsilon\} \leq E\zeta = \frac{D\xi}{\varepsilon^2}$. Отже, $P\{|\xi - E\xi| \geq \geq \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}$, що і треба було довести.

У курсі теорії ймовірностей вивчають різні поняття збіжності. Розглянемо деякі з них.

Визначення. Послідовність випадкових величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ *збіжна за ймовірністю* до випадкової величини ξ ($\xi_n \xrightarrow{p} \xi$ за умови $n \to \infty$), якщо для кожного $\varepsilon > 0$ $P\{|\xi_n - \xi| > \varepsilon\} \to 0$ за $n \to \infty$.

Очевидно, $\xi_n \xrightarrow{p} \xi$ тоді і тільки тоді, коли $\xi_n - \xi \xrightarrow{p} 0$ за $n \to \infty$. Випадковою величиною ξ може бути стала a. Збіжність за ймовірністю означає, що відхилення випадкових величин ξ_n від граничної величини ξ на будь-яке значення ε зі зростанням n стає дедалі менш ймовірним.

Теорема 1 (закон великих чисел). Якщо випадкові величини $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — незалежні й

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} D\xi_i = 0,$$

то

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i - \frac{1}{n}\sum_{i=1}^{n} E\xi_i \xrightarrow{p} 0, \quad n \to \infty.$$

Доведення. Позначимо $\zeta_n = \frac{1}{n}\sum_{i=1}^{n} \xi_i$. Тоді

$$E\zeta_n = \frac{1}{n}\sum_{i=1}^{n} E\xi_i, \quad D\zeta_n = \frac{1}{n^2}\sum_{i=1}^{n} D\xi_i.$$

Відповідно до нерівності Чебишова

$$P\{|\zeta_n - E\zeta_n| \geq \varepsilon\} \leq \frac{D\zeta_n}{\varepsilon^2}.$$

За умовою теореми $D\zeta_n \to 0$. Тому

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} \xi_i - \frac{1}{n}\sum_{i=1}^{n} E\xi_i\right| \geq \varepsilon\right\} \to 0$$

при $n \to \infty$, що і доводить теорему.

Значення закону великих чисел полягає в тому, що за виконання умови теореми 1 середньоарифметичне незалежних випадкових величин, яке є випадковою величиною, за ймовірністю близьке до середньоарифметичного їх математичних сподівань.

Справедливі такі наслідки теореми 1.

Теорема 2 (теорема Чебишова). Якщо випадкові величини $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — незалежні та їх дисперсії обмежені $D\xi_i \leq C$, де C — стала, то

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i - \frac{1}{n}\sum_{i=1}^{n} E\xi_i \xrightarrow{p} 0, \quad n \to \infty.$$

Доведення. Випливає з теореми 1, оскільки при $D\xi_i \leq C$, $i = 1, 2, \ldots$ маємо

$$\frac{1}{n^2}\sum_{i=1}^{n} D\xi_i \leq \frac{nC}{n^2} = \frac{C}{n} \to 0, \quad n \to \infty.$$

Теорема 3. Якщо випадкові величини $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — незалежні, однаково розподілені і мають скінченне математичне сподівання $E\xi_i = a$ і дисперсію $D\xi_i = \sigma^2$, то

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i \xrightarrow{p} a,$$

якщо $n \to \infty$.

Доведення. Випливає з теореми 2, оскільки у цьому випадку всі дисперсії рівні, й отже, є обмеженими, тому

$$\frac{1}{n}\sum_{i=1}^{n} E\xi_i = \frac{1}{n} na = a.$$

Теорема 4 (теорема Бернуллі). Нехай ξ_n — число успіхів у n випробуваннях за схемою Бернуллі, p — ймовірність успіху в одному випробуванні. Тоді відносна частота успіху збігається за ймовірністю до ймовірності успіху

$$\frac{\xi_n}{n} \xrightarrow{p} p,$$

якщо $n \to \infty$.

Доведення. Подамо ξ_n за формулою біноміального розподілу: $\xi_n = \eta_1 + \eta_2 + ... + \eta_n$, де випадкові величини $\xi_1, \xi_2, ..., \xi_n, ...$ — незалежні й однаково розподілені, $E\eta_i = p$ для всіх $i = 1, 2, ...$, $D\eta_i = pq < \infty$. Отже, умови теореми 3 виконуються.

Однак має місце сильніше твердження, яке називають посиленим законом великих чисел.

Теорема 5 (теорема Бореля). Нехай p — ймовірність успіху в одному випробуванні, ξ_n — число успіхів у n випробуваннях за схемою Бернуллі. Тоді $P\left\{\lim\limits_{n\to\infty} \frac{\xi_n}{n} = p\right\} = 1$.

РОЗДІЛ II

ОСНОВИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

У попередньому розділі розглянуто ймовірнісні характеристики випадкових явищ, до яких належать ймовірність випадкових подій, функції розподілу випадкових величин, математичне сподівання, дисперсія та ін. На практиці значення числових характеристик зазвичай невідомі і підлягають визначенню на підставі результатів спостережень і відповідних вимірювань.

Математична статистика — розділ математики, який вивчає методи відбору та обробки результатів спостережень і випадкових явищ для виявлення їхніх закономірностей та визначення ймовірнісних характеристик.

За даними спостережень можна оцінити невідомі ймовірнісні характеристики й висунути гіпотезу, що спостережуване випадкове явище може бути описане конкретною математичною моделлю. Ця гіпотеза підлягає перевірці за допомогою математичних методів. Саме та математична модель, яка витримує перевірку гіпотез, тобто не суперечить фактичним даним, здатна описати математичні закономірності випадкового явища. Отже, для з'ясування математичних закономірностей, яким підпорядковані випадкові явища, насамперед необхідно вивчити фактичні статистичні дані, тобто числові відомості про спостережувані явища.

Основними завданнями математичної статистики є статистична перевірка гіпотез, оцінювання розподілу статистичних ймовірностей та його параметрів, вивчення статистичної залежності, визначення основних числових характеристик випадкових вибірок.

Математична статистика виникла у XVII ст. і почала розвиватися разом із теорією ймовірностей. Своїм розвитком вона зобов'язана таким відомим математикам, як К.Ф. Гаусс, Ф. Гальтон, К. Пірсон, П.Л. Чебишов, А.А. Марков, О.М. Ляпунов, А.М. Колмогоров та ін.

§ 9. ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ

9.1. Вибірковий метод. Варіаційні ряди

Нехай необхідно вивчити певні ознаки деякої сукупності об'єктів, наприклад, термін придатності продукції, яку випускає підприємство. Якщо воно виробляє ампули з вакциною, то не можна проконтролювати кожну ампулу, оскільки у цьому разі всі вони будуть зіпсовані. Потрібно обрати партію ампул із загального їх числа, щоб, з одного боку, вибірка досить повно представляла властивості всієї продукції, а з іншого — не була занадто великою.

Визначення. *Генеральною сукупністю* називають усю сукупність досліджуваних об'єктів.

Визначення. *Вибіркою*, або *контрольно-вибірковою сукупністю*, називають ту частину генеральної сукупності, яку піддають безпосередньому спостереженню (вимірюванню).

Визначення. *Об'ємом*, або *обсягом*, *вибірки* називають кількість об'єктів у вибірці.

Визначення. *Репрезентативністю вибірки* називають властивість вибіркової сукупності достовірно і повно відображати параметри генеральної сукупності, частиною якої вона є.

Зауваження. Якщо вибірка нерепрезентативна, то вона не відображає ознак генеральної сукупності, з якої отримана.

Приклад 58. На підприємстві працює 300 робітників. Для вивчення продуктивності праці обрали 32 робітники й підрахували кількість виробів, які ті виготовили за зміну. Отримано таку таблицю:

26	32	32	34	30	32	26	36
26	32	26	40	36	32	34	30
34	32	32	30	40	32	36	34
32	30	32	36	34	28	26	40

Згрупуємо ці дані й розмістимо їх у порядку зростання. Таку дію називають *ранжуванням статистичних даних.* У результаті отримаємо *ранжований числовий ряд:*

Кількість виробів x_i	26	28	30	32	34	36	40
Число робітників n_i, які виготовили x_i виробів	5	1	4	10	5	4	3

Визначення. Різні значення x_i, $1 \le i \le k$, ознаки, які спостерігаються, називають *варіантами*, а числа n_i, $1 \le i \le k$, які показують, скільки разів з'явилась варіанта у вибірці — *частотою варіанти*. Послідовні суми $n_1 + n_2 + ... + n_i$, $1 \le i \le k$ частот називають *накопиченими частотами*. Очевидно, $\sum_{i=1}^{k} n_i = n$. Сукупність варіант разом з їх частотами утворюють *варіаційний ряд*, який зручно зображати у вигляді таблиці:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

Визначення. *Варіаційний ряд є дискретним*, якщо значення ознаки відрізняються одне від одного не менш як на деяке фіксоване число, і *неперервним*, якщо значення ознаки можуть відрізнятись на як завгодно малу величину.

Приклад 59 (*неперервний варіаційний ряд*). У результаті вимірювання діаметра виробів, які виробляються на підприємстві, отримано ряд:

Діаметр виробу, мм	Кількість виробів n
95—100 100—102 105—110	180 340 90

Варіаційні ряди можна інтерпретувати графічно.

1. Нехай задано *дискретний варіаційний ряд* (див. приклад 58). У декартовій прямокутній системі координат нанесемо точки з координатами (x_i, n_i) і з'єднаємо їх послідовно відрізками прямої, а крайні точки з'єднаємо з віссю абсцис. Отримаємо фігуру, яку називають *полігоном розподілу* (рис. 18).

2. Нехай задано *неперервний варіаційний ряд* (див. приклад 59). У декартовій прямокутній системі координат по осі абсцис відкладемо інтервали фіксованої величини і на цих відрізках як на основах побудуємо прямокутники з висотами, що дорівнюють часто-

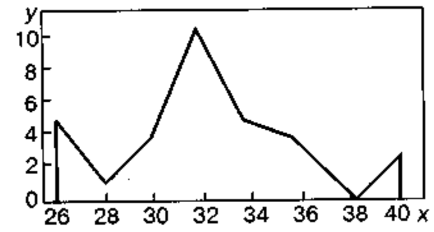


Рис. 18. Полігон розподілу

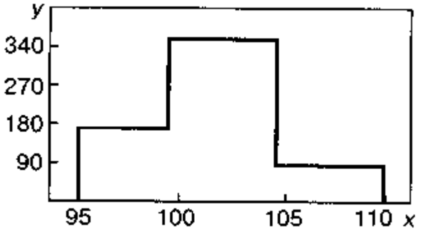


Рис. 19. Гістограма розподілу

там інтервалу. Отриману східчасту фігуру називають *гістограмою* (рис. 19).

Можна також побудувати *кумулятивну криву* за накопиченими частотами.

9.2. Числові характеристики статистичних даних

Варіаційний ряд — це початкова форма вивчення статистичних даних. Розглянемо докладніше їх числові характеристики.

Зазначимо, що з погляду теорії ймовірностей досліджувана ознака є випадковою величиною ξ з невідомою функцією розподілу $F(x)$; $x_1,...,x_n$ — її значення. Припускають, що експерименти з визначення ознаки проведені за однакових умов і незалежно один від одного. Тому результати експериментів можна вважати незалежними випадковими величинами з однією й тією ж функцією розподілу $F(x)$.

Нехай x — будь-яке дійсне число. Тоді значення $x_1, x_2,..., x_n$ можна розбити на дві множини: на ті значення x_i, які менші за x, і ті, які більші за x.

Позначимо через $\nu_i = \dfrac{n_i}{n}$, $1 \le i \le k$, *відносні частоти* варіанти x_i. Тоді варіаційний ряд набуває вигляду

x_i	x_1	x_2	...	x_k
ν_i	ν_1	ν_2	...	ν_k

Властивості відносних частот

1. $0 < \nu_i < 1$.
2. $\sum\limits_{i=1}^{n} \nu_i = 1$.

Скориставшись відносними частотами, вибірку з прикладу 58 можна зобразити так:

x_i	26	28	30	32	34	36	40
n_i	5	1	4	10	5	4	3
ν_i	5/32	1/32	4/32	10/32	5/32	4/32	3/32

Визначення. *Емпіричною функцією розподілу* називають функцію $F_n(x)$, що дорівнює частці тих членів вибірки об'єму n, в яких досліджувана ознака виявилась меншою за x, тобто

$$F_n(x) = \frac{n(x)}{n},$$

де $n(x)$ — число членів вибірки, менших за x.

Очевидно, що значення $n(x)$ збігаються з накопиченими частотами.

Емпіричну функцію розподілу можна виписати у явному вигляді:

$$F_n(x) = \begin{cases} 0, & \text{якщо} \quad x \le x_1; \\ \nu_1, & \text{якщо} \quad x_1 < x \le x_2; \\ \nu_1 + \nu_2, & \text{якщо} \quad x_2 < x \le x_3; \\ \dots\dots & \dots\dots \quad \dots\dots\dots\dots \\ \nu_1 + \nu_2 + \ldots + \nu_{k-1}, & \text{якщо} \quad x_{k-1} < x \le x_k; \\ 1, & \text{якщо} \quad x_k < x. \end{cases}$$

Графік емпіричної функції розподілу має вигляд східців, які ведуть від рівня 0 до рівня 1 так, що i-й східець $x_i < x \le x_{i+1}$ має висоту ν_i (рис. 20). Емпірична функція розподілу неперервна зліва.

Далі припускають, що проведено n незалежних експериментів, у результаті яких випадкова величина набуває відповідних значень $x_1, x_2,...,x_n$.

Визначення. *Вибірковим середнім* $\bar{x}$ називають середньоарифметичне всіх значень вибірки:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{k} n_i x_i . \tag{36}$$

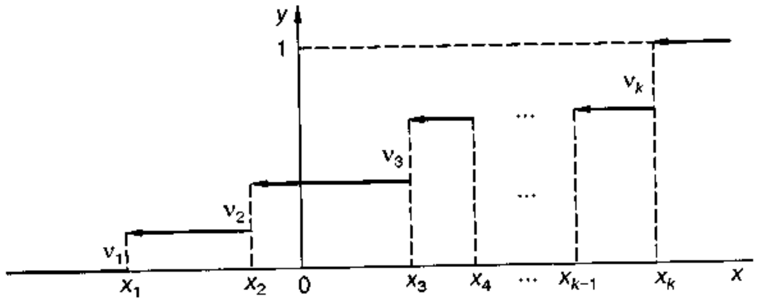


Рис. 20. Графік емпіричної функції розподілу

Визначення. *Вибірковою дисперсією* S_n^2 є величина

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \overline{x^2} - \left(\overline{x}\right)^2, \qquad (37)$$

де $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2$.

Визначення. *Виправленою вибірковою дисперсією* $\hat{S}_n^2$ називають величину

$$\hat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2, \text{ тобто } \hat{S}_n^2 = \frac{n}{n-1}S_n^2. \qquad (38)$$

Визначення. *Модою Mo* вибірки є варіанта x_i, $1 \le i \le k$, яка має найбільшу частоту появи n_i, $1 \le i \le k$.

Визначення. *Медіаною Me* вибірки є варіанта x_i, $1 \le i \le k$, яка поділяє варіаційний ряд на дві частини, рівні за кількістю варіант.

Визначення. *Відхиленням варіанти* x_i, $1 \le i \le k$ називають різницю $x_i - \overline{x}$, $1 \le i \le k$.

Визначення. *Розмахом R вибірки* називають різницю між найбільшою та найменшою варіантами варіаційного ряду:

$$R = \max_{1 \le i \le k} x_i - \min_{1 \le i \le k} x_i.$$

Визначення. *Вибірковим середньоквадратичним відхиленням* S_n називають величину

$$S_n = \sqrt{S_n^2}. \qquad (39)$$

Визначення. *Виправленим вибірковим середньоквадратичним відхиленням* $\hat{S}_n$ є величина

$$\hat{S}_n = \sqrt{\hat{S}_n^2}. \qquad (40)$$

Припустимо, що генеральна сукупність наділена ознаками X та Y, залежність між якими потрібно дослідити. Позначимо через (x_i, y_i), $i = \overline{1, n}$ вибіркові значення пари ознак X та Y.

Визначення. *Вибірковим коефіцієнтом кореляції* між ознаками X та Y називають величину

$$r_n = r_n(X, Y) = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}}, \qquad (41)$$

де $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Вибірковий коефіцієнт кореляції — це кількісна характеристика залежності між ознаками X та Y.

Зауваження. Позначимо через $x_1, x_2, \ldots, x_k$ варіанти, яких набуває ознака X за результатами вибіркового дослідження, через $y_1, y_2, \ldots, y_m$ — варіанти, яких набуває ознака Y. У цьому випадку вибірку можна зобразити у вигляді таблиці

$Y \backslash X$	x_1	x_2	...	x_k
y_1	n_{11}	n_{12}	...	n_{1k}
y_2	n_{21}	n_{22}	...	n_{2k}
...	...	...	...	...
y_m	n_{m1}	n_{m2}	...	n_{mk}

де n_{ij}, $i = \overline{1, k}$; $j = \overline{1, m}$ — частота пари (x_i, y_j) у вибірці. Тоді формула вибіркового коефіцієнта кореляції набуде вигляду

$$r_n = r_n(X, Y) = \frac{\overline{xy} - \overline{x}\,\overline{y}}{S_n(x) S_n(y)},$$

де $\overline{xy} = \frac{1}{n}\sum_{i=1, j=1}^{k, m} n_{ij} x_i y_j$, $\hat{n}_i = \sum_{j=1}^{m} n_{ji}$, $\overline{x} = \frac{1}{n}\sum_{i=1}^{k}\hat{n}_i x_i$, $\tilde{n}_j = \sum_{i=1}^{k} n_{ji}$, $\overline{y} = \frac{1}{n}\sum_{j=1}^{m}\tilde{n}_j y_j$, $S_n^2(x) = \overline{x^2} - \left(\overline{x}\right)^2$, $S_n^2(y) = \overline{y^2} - \left(\overline{y}\right)^2$, $S_n(x) = \sqrt{S_n^2(x)}$, $S_n(y) = \sqrt{S_n^2(y)}$.

Основні властивості вибіркового коефіцієнта кореляції

1. $|r_n| \le 1$.
2. Якщо X та Y — незалежні випадкові величини, то $r_n \to 0$, якщо $n \to \infty$, тобто вони є некорельованими.

3. Якщо між X та Y існує лінійна залежність $Y = aX + b$, де a і b — деякі сталі, то $|r_n| \to 1$, якщо $n \to \infty$.

Якщо $0 < r_n \le 1$, за умови $n \to \infty$ маємо додатну кореляцію. Це означає, що при зростанні однієї з величин інша так само у середньому зростатиме. Якщо $-1 \le r_n < 0$, за умови $n \to \infty$ спостерігається від'ємна кореляція, коли при збільшенні однієї з величин інша має тенденцію в середньому до спадання. Зазвичай вважають, що, наприклад за додатної кореляції, значення $0{,}7 \le r_n \le 1$ свідчать про високий ступінь зв'язку між величинами X та Y; $0{,}3 \le r_n < 0{,}7$ — про середній; $0 < r_n < 0{,}3$ — про слабкий зв'язок між величинами X та Y.

Приклад 60. Проведено вибіркове визначення вмісту крохмалю у картоплі. Отримано таблицю (дані Хальда). Обчислити середнє $\bar{x}$ та виправлене середньоквадратичне відхилення $\tilde{S}_n$.

n	X, %	n	X, %
1	21,7	6	15,6
2	13,7	7	17,7
3	18,3	8	16,6
4	17,5	9	14
5	18,5	10	17,2

Розв'язання. За формулами (36), (38), (40), отримаємо:

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 17{,}1;\quad \tilde{S}_n = \sqrt{\frac{10}{9}\left[\overline{x^2} - \left(\bar{x}\right)^2\right]} = 2{,}3.$$

Приклад 61. Знайти вибірковий коефіцієнт кореляції залежності між вмістом сухої речовини X у свіжому шпинаті та вмістом аскорбінової кислоти Y за результатами, представленими у таблиці (дані Петерсена):

n	1	2	3	4	5	6
X, %	10	8,9	8,9	9,2	7,8	10,1
Y, %	70,9	74	58,6	80,6	69,4	76

Розв'язання. Визначимо вибіркові середні для X та Y: $\bar{x} = \frac{1}{6}\sum_{i=1}^{6} x_i = 9{,}2$; $\bar{y} = \frac{1}{6}\sum_{i=1}^{6} y_i = 71{,}6$. Тоді, згідно з формулою (41),

маємо

$$r_6 = r_6(X, Y) = \frac{\overline{xy} - \bar{x}\,\bar{y}}{S_6(x)\,S_6(y)} = 0{,}3.$$

Отже, можна стверджувати, що між ознаками X та Y існує слабкий зв'язок.

§ 10. ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛУ

10.1. Точкові оцінки

Емпірична функція розподілу та вибіркові характеристики можуть слугувати оцінками відповідно функції розподілу та числових характеристик досліджуваної величини ξ. Розглянемо деякі загальні поняття теорії оцінок.

Нехай a — деякий невідомий параметр розподілу (це може бути числова характеристика — математичне сподівання, дисперсія і т. д.).

Визначення. *Оцінкою параметра* a називають таку функцію від спостережень $\tilde{a}_n = \tilde{a}(x_1, x_2, \ldots, x_n)$, яка у певному сенсі близька до a.

Зрозуміло, що можна розглядати різні оцінки одного й того самого параметра.

Визначення. *Оцінку* $\tilde{a}_n$ параметра a називають *незміщеною*, якщо $E\tilde{a}_n = a$.

Незміщеність оцінки означає, що при заміні справжнього значення параметра a на його оцінку $\tilde{a}_n$ ми не отримаємо систематичних односторонніх похибок.

Визначення. *Оцінку* $\tilde{a}_n$ параметра a називають *спроможною*, якщо $\tilde{a}_n \xrightarrow{P} a$ при $n \to \infty$.

Теорема 1. Емпірична функція розподілу $F_n(x)$ є незміщеною і спроможною оцінкою теоретичної функції розподілу $F(x)$:

$$E(F_n(x)) = F(x);\quad F_n(x) \xrightarrow{\ P\ } F(x),\ n \to \infty.$$

Доведення. 1. Нехай $x_1, x_2, \ldots, x_n$ — незалежні однаково розподілені результати спостережень із теоретичною функцією розподілу $F(x)$. Нехай подія $\{n(x) = i\}$ полягає в тому, що i величин виявились меншими за x, $(n - i)$ величин — більшими за x. Тому

$$P\{\, n(x) = i \,\} = C_n^i (F(x))^i (1 - F(x))^{n-i},$$

тобто $n(x)$ має біноміальний розподіл з параметром $p = F(x)$. За формулою математичного сподівання для біноміального розподілу (див. розділ I) маємо

$$E(F_n(x)) = \frac{1}{n} E\,(n(x)) = \frac{1}{n} nF(x) = F(x).$$

2. За теоремою Бернуллі (див. розділ I), стосовно поведінки відносної частоти маємо

$$F_n(x) = \frac{n(x)}{n} \xrightarrow{p} F(x)$$

при $n \to \infty$.

Теорема 2. Вибіркове математичне сподівання $\bar{x}$ є незміщеною і спроможною оцінкою математичного сподівання $m = E\xi$:

$$E\bar{x} = m, \quad \bar{x} \xrightarrow{p} m,$$

якщо $n \to \infty$.

Доведення. 1. В силу того, що випадкові величини $x_1, x_2, \ldots, x_n$ однаково розподілені, отримаємо:

$$E\bar{x} = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) \frac{1}{n} \sum_{i=1}^{n} Ex_i = \frac{1}{n} nm = m.$$

2. Нехай дисперсія ξ дорівнює σ^2. Тоді, згідно із законом великих чисел (див. розділ I), маємо

$$\frac{1}{n}(x_1 + x_2 + \ldots + x_n) \xrightarrow{p} m$$

при $n \to \infty$, що і треба було довести.

Теорема 3. Вибіркова дисперсія S_n^2 є зміщеною і спроможною оцінкою дисперсії $D\xi = \sigma^2$. Величина

$$\hat{S}_n^2 = \frac{n}{n-1} S_n^2$$

буде незміщеною і спроможною оцінкою цієї дисперсії.

Доведення. 1. Розглянемо випадкові величини $y_i = x_i - m$. Тоді

$$\bar{y} = \bar{x} - m,\ Ey_i = 0,\ Dy_i = Ey_i^2 = \sigma^2,\ E(S_n^2) = \frac{1}{n} E\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right) =$$

$$= \frac{1}{n} E\left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} y_i^2\right) -$$

$$- \frac{2}{n} E\left(\sum_{i=1}^{n} y_i\right)\bar{y} + E(\bar{y})^2 = \frac{1}{n} n\sigma^2 - 2E(\bar{y})^2 + E(\bar{y})^2 = \sigma^2 - E(\bar{y})^2.$$

Однак

$$E(\bar{y})^2 = E\left(\frac{1}{n}(y_1 + y_2 + \ldots + y_n)^2\right) = \frac{1}{n^2}\left[E\left(\sum_{i=1}^{n} y_i^2\right) + \sum_{i \neq j} E(y_i y_j)\right] =$$

$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n},$$

тому $E(S_n^2) = \frac{n-1}{n}\sigma^2$. Це означає, що $E(\hat{S}_n^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$.

2. Спроможність оцінок S_n^2 і $\hat{S}_n^2$ доводиться, як і в теоремі 2.

10.2. Інтервальне оцінювання. Довірчі інтервали та їх побудова

Вище було розглянуто точкові оцінки невідомого параметра розподілу a, тобто оцінки одним числом. Часто, особливо за малих об'ємів вибірок, точкове оцінювання ненадійне. У таких випадках використовують інтервальне оцінювання, тобто за наявними даними будують інтервал, у який із заданою довірчою ймовірністю (рівнем довіри) γ потрапляє оцінюваний параметр a.

Визначення. *Довірчим інтервалом* рівня довіри γ, $0 < \gamma < 1$, називають такий випадковий інтервал $[\hat{\theta}_1, \hat{\theta}_2]$, де $\hat{\theta}_1 \equiv \hat{\theta}_1(x_1, x_2, \ldots, x_n)$ і $\hat{\theta}_2 \equiv \hat{\theta}_2(x_1, x_2, \ldots, x_n)$, що $P\{\hat{\theta}_1 \le \theta \le \hat{\theta}_2\} \ge \gamma$.

Зауваження. У біологічних дослідженнях найчастіше довірчу ймовірність обирають такою, що дорівнює $\gamma = 0{,}95$.

Врахувавши важливість Гауссівського розподілу, розглянемо довірчі інтервали для кожного з його параметрів a та σ^2 за наявності або відсутності інформації про інший параметр. Нехай $x_1, x_2, \ldots, x_n$ — вибірка з нормально розподіленої $N(a, \sigma^2)$ генеральної сукупності. Тоді x_i, $1 \le i \le n$, утворюють множину незалежних $N(a, \sigma^2)$ випадкових величин.

Довірчий інтервал для оцінки математичного сподівання а нормального розподілу за відомої дисперсії σ^2

Параметри вибіркового середнього $\overline{x}$ становлять $E\overline{x} = a$ і $D\overline{x} = \sigma^2/n$. Тоді випадкова величина $(\overline{x} - a)\sqrt{n}/\sigma$ має розподіл $N(0, 1)$, звідки $P\left\{-g \le (\overline{x} - a)\sqrt{n}/\sigma \le g\right\} = 2\Phi(g)$. Отже, $P\left\{\overline{x} - g\sigma/\sqrt{n} \le a \le \overline{x} - g\sigma/\sqrt{n}\right\} = 2\Phi(g)$.

Нехай рівень довіри γ заданий. Знайдемо таке g, щоб $2\Phi(g) = \gamma$. Тоді $\left[\overline{x} - \frac{1}{\sqrt{n}}\sigma g,\ \overline{x} + \frac{1}{\sqrt{n}}\sigma g\right]$ утворює довірчий інтервал для a рівня довіри γ. Число $\Delta = \frac{1}{\sqrt{n}}\sigma g$ називають *точністю оцінки* середнього a.

Зауваження. З формули довірчого інтервалу випливає, що зі збільшенням об'єму вибірки n ($n \to \infty$) величина $\frac{1}{\sqrt{n}}\sigma g$ зменшується, тобто точність оцінки зростає.

Довірчий інтервал для оцінки дисперсії σ^2 нормального розподілу за відомого математичного сподівання а

Позначимо $\breve{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2$. З теорії ймовірностей відомо, що випадкова величина $\eta = \frac{n\breve{S}_n^2}{\sigma^2}$ має χ^2*-розподіл Пірсона* з n ступенями вільності (див. п. 7.4.1). Для нього складено таблиці значень (табл. 2), за якими для заданого рівня довіри γ можна знайти такі значення χ_1^2 і χ_2^2, що $P\left\{\chi_1^2 \le \frac{n\breve{S}_n^2}{\sigma^2} \le \chi_2^2\right\} = \gamma$. Тоді $\left[\frac{n\breve{S}_n^2}{\chi_2^2},\ \frac{n\breve{S}_n^2}{\chi_1^2}\right]$ буде довірчим інтервалом для σ^2 рівня довіри γ. Оскільки розподіл χ^2 не симетричний (рис. 15), то для знаходження значень χ_1^2 і χ_2^2 використовують поняття p-квантилів (від англ. quantile) розподілу χ^2.

Таблиця 2. **p-Квантилі розподілу χ^2**

n	p						
	0,025	0,3	0,5	0,7	0,9	0,95	0,975
1	0,00	0,14	0,45	1,07	2,71	3,84	5,02
2	0,05	0,71	1,39	2,41	4,61	5,99	7,38
3	0,22	1,42	2,37	3,67	6,25	7,82	9,35
4	0,48	2,20	3,36	4,88	7,78	9,49	11,14
5	0,83	3,00	4,35	6,06	9,24	11,1	12,83
6	1,24	3,83	5,35	7,23	10,6	12,6	14,45
7	1,70	4,67	6,36	8,38	12,0	14,1	16,01
8	2,20	5,53	7,34	9,52	13,4	15,5	17,53
9	2,70	6,39	8,34	10,7	14,7	16,9	19,02
10	3,25	7,27	9,34	11,8	16,0	18,3	20,48
11	3,81	8,15	10,3	12,9	17,3	19,7	21,92
12	4,40	9,03	11,3	14,0	18,5	21,0	23,34
13	5,01	9,93	12,3	15,1	19,8	22,4	24,74
14	5,63	10,0	13,3	16,2	21,1	23,7	26,12
15	6,26	11,7	14,3	17,3	22,3	25,0	27,49
16	6,91	12,6	15,3	18,4	23,5	26,3	28,85
17	7,56	13,5	16,3	19,5	24,8	27,6	30,19
18	8,23	14,4	17,3	206	26,0	28,9	31,53
19	8,91	15,4	18,3	21,7	27,2	30,1	32,85
20	9,59	16,3	19,3	22,8	28,4	31,4	34,17
21	10,28	17,2	20,3	23,9	29,6	32,7	35,48
22	10,98	18,1	21,3	24,9	30,8	33,9	36,78
23	11,69	19,0	22,3	26,0	32,0	35,2	38,08
24	12,40	19,9	23,3	27,1	33,2	36,4	39,36
25	13,12	20,9	24,3	28,2	34,3	37,7	40,64
26	13,84	21,8	25,3	29,2	35,6	38,9	41,92
27	14,57	22,7	26,3	30,3	36,7	40,1	43,19
28	15,31	23,6	27,3	31,4	37,9	41,3	44,46
29	16,05	24,6	28,3	32,5	39,1	42,6	45,72
30	16,79	25,5	29,3	33,5	40,3	43,8	46,98

Нехай випадкова величина ξ має неперервну функцію розподілу $F(x)$, а число $p \in (0, 1)$. Позначимо через $\Gamma(\xi)$ розподіл випадкової величини ξ.

Визначення. *p-Квантилем розподілу* $\Gamma(\xi)$ називають розв'язок ζ_p рівняння $p = F(\zeta)$, $0 < p < 1$.

Позначимо через $F_n(x)$ функцію розподілу $\chi^2(n)$. Вона неперервна. Графік щільності розподілу $\chi^2(n)$ зображено на рис. 21, де незаштрихована площа під кривою дорівнює γ. Тоді для χ_1^2 і

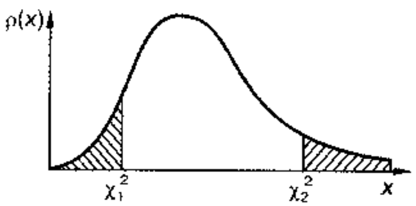


Рис. 21. Критична область критерію

χ_2^2 отримаємо співвідношення $F_n(\chi_2^2) - F_n(\chi_1^2) = \gamma$, яке зазвичай розв'язують так, щоб $F_n(\chi_2^2) = \frac{1+\gamma}{2}$ і $F_n(\chi_1^2) = \frac{1-\gamma}{2}$.

Позначимо через $\chi^2_{n;p}$ p-квантиль розподілу $\chi^2(n)$. Тоді $\chi_1^2 = \chi^2_{n;\frac{1-\gamma}{2}}$ і $\chi_2^2 = \chi^2_{n;\frac{1+\gamma}{2}}$ — $\frac{1-\gamma}{2}$ і $\frac{1+\gamma}{2}$ — квантилі розподілу χ^2 з n ступенями вільності. Остаточно інтервал довіри для σ^2 рівня довіри γ має вигляд $\left[\frac{n\breve{S}_n^2}{\chi^2_{n;\frac{1+\gamma}{2}}},\ \frac{n\breve{S}_n^2}{\chi^2_{n;\frac{1-\gamma}{2}}}\right]$.

Приклад 62. При перевірці апарата, що наливає по 1 мл розчину в пробірки, з'ясувалось, що внаслідок випадкових коливань струму цей пристрій допускає Гауссівське відхилення відносно середнього. В результаті тестування продуктивності апарата було отримано вибірку:

мл	0,98	0,99	1,00	1,01	1,02
n_i	10	3	2	2	10

Оцінити довірчим інтервалом за надійності $\gamma = 0,95$ дисперсію випадкового відхилення відносно відомого середнього 1 мл.

Розв'язання. Математичне сподівання нормального розподілу відоме $a = 1$. Об'єм вибірки $n = 27$. $\breve{S}_{27}^2 = \frac{1}{27}\sum_{i=1}^{n}(x_i - 1)^2 = \frac{1}{27} \times \left(10(-0,02)^2 + 3(-0,01)^2 + 2(0,01)^2 + 10(0,02)^2\right) = \frac{0,0086}{27}$. Далі $\frac{1+\gamma}{2} = \frac{1+0,95}{2} = 0,975$ і $\frac{1-\gamma}{2} = \frac{1-0,95}{2} = 0,025$. Згідно з даними табл. 2, маємо: $\chi^2_{27;0,975} = 43,19$; $\chi^2_{27;0,025} = 14,57$. Оскільки $\frac{27\breve{S}_{27}^2}{\chi^2_{27;0,975}} = \frac{0,0086}{43,19} = 0,00019912$ і $\frac{27\breve{S}_{27}^2}{\chi^2_{27;0,025}} = \frac{0,0086}{14,57} = 0,000590254$, то $0,00019 < \sigma^2 < 0,0006$ з надійністю $\gamma = 0,95$.

Довірчий інтервал для оцінки математичного сподівання a нормального розподілу за невідомої дисперсії σ^2

З теорії ймовірностей відомо, що випадкова величина

$$\eta = \frac{1}{\tilde{S}_n}\left(\bar{x} - a\right)\sqrt{n}$$

має *t-розподіл Стьюдента* з $n-1$ ступенем вільності (див. п. 7.4.2), для якого складено таблиці значень (табл. 3). Оскільки розподіл t симетричний відносно своєї середньої точки, то для заданого рівня довіри γ можна знайти єдине $t_{n-1;\gamma}$, таке що $P\{|\eta| \le t_{n-1;\gamma}\} = \gamma$. Тоді $\left[\bar{x} - \frac{1}{\sqrt{n}}t_{n-1;\gamma}\tilde{S}_n,\ \bar{x} + \frac{1}{\sqrt{n}}t_{n-1;\gamma}\tilde{S}_n\right]$ утворює довірчий інтервал для a рівня довіри γ. Число

$$\Delta = \frac{1}{\sqrt{n}}t_{n-1;\gamma}\tilde{S}_n \qquad (42)$$

називають *точністю оцінки* середнього a.

Зауваження. Рекомендовано користуватись розподілом Стьюдента при побудові довірчих інтервалів для малих вибірок.

Таблиця 3. **Коефіцієнти Стьюдента $t_{n\gamma}$**

n	γ				
	0,90	0,95	0,98	0,99	0,999
3	2,92	4,30	6,97	9,93	31,60
4	2,35	3,18	4,54	5,84	12,94
5	2,13	2,78	3,75	4,60	8,61
6	2,02	2,57	3,37	4,03	6,86
7	1,94	2,45	3,14	3,71	5,96
8	1,90	2,37	3,00	3,50	5,41
9	1,86	2,31	2,90	3,36	5,04
10	1,83	2,26	2,82	3,25	4,78
11	1,81	2,23	2,76	3,17	4,59
12	1,80	2,20	2,72	3,11	4,44
13	1,78	2,18	2,68	3,06	4,32
14	1,77	2,16	2,65	3,01	4,22
15	1,76	2,15	2,62	2,98	4,14
20	1,73	2,09	2,54	2,86	3,88
∞	1,64	1,96	2,33	2,58	3,29

Приклад 63. Випадкова величина розподілена за нормальним законом із параметром $\sigma = 0,5$. Знайти мінімальний об'єм n вибірки, щоб з довірчою ймовірністю $\gamma = 0,95$ і точністю $\Delta = 0,1$ виконувалась рівність $a \approx \bar{x}$.

Розв'язання. За уточненою табл. 1 для $\gamma = 0,95$ отримаємо $g = 1,96$. У результаті матимемо: $n = (g\sigma/\Delta)^2 \approx 96$.

Приклад 64. Відомо, що генеральна сукупність, з якої отримана вибірка, що представлена таблицею у прикладі 60, має нормальний розподіл, параметри якого невідомі. Визначити довірчий інтервал для середнього a рівня довіри $\gamma = 0,95$.

Розв'язання. Врахувавши, що $\bar{x} = 17,1$, $\tilde{S}_{10} = 2,3$ (див. приклад 60), а також $t_{9;0,95} = 2,31$ (див. табл. 3), отримаємо

$$17,1 - \frac{2,31 \cdot 2,3}{\sqrt{10}} = 15,5 \text{ і } 17,1 + \frac{2,31 \cdot 2,3}{\sqrt{10}} = 18,7.$$

Отже, шукана величина знаходиться в довірчому інтервалі $15,5 \le a \le 18,7$ з довірчою ймовірністю $\gamma = 0,95$.

Довірчий інтервал для оцінки дисперсії σ^2 нормального розподілу за невідомого математичного сподівання a

З теорії ймовірностей відомо, що випадкова величина $\eta = \frac{nS_n^2}{\sigma^2}$ має χ^2-*розподіл Пірсона* з $n - 1$ ступенем вільності.

За допомогою міркувань, аналогічним тим, що наведені в п. 7.4.2, дійдемо висновку, що довірчий інтервал за рівня довіри γ для невідомої дисперсії σ^2 за невідомого математичного сподівання a має вигляд $\left[\frac{nS_n^2}{\chi^2_{n-1;\frac{1+\gamma}{2}}}, \frac{nS_n^2}{\chi^2_{n-1;\frac{1-\gamma}{2}}}\right]$, де $\chi^2_{n-1;\frac{1+\gamma}{2}}$ і $\chi^2_{n-1;\frac{1-\gamma}{2}}$ — $\frac{1+\gamma}{2}$ і $\frac{1-\gamma}{2}$ — квантилі розподілу χ^2 з $n - 1$ ступенем вільності.

Розглянемо тепер випадок, коли розподіл генеральної сукупності невідомий. Нехай $x_1, x_2, \ldots, x_n$ — вибірка з генеральної сукупності з довільною функцією розподілу, невідомим середнім m і відомою дисперсією σ^2.

Довірчий інтервал для оцінки математичного сподівання m довільного розподілу за відомої дисперсії σ^2

Згідно з центральною граничною теоремою (див. розд. I),

$$P\left\{-g \le \frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{n} x_i - nm\right) \le g\right\} \to 2\Phi(g), \; n \to \infty,$$

або

$$P\left\{\left|\bar{x} - m\right| \le \frac{1}{\sqrt{n}}\sigma g\right\} \approx 2\Phi(g).$$

Нехай рівень довіри γ заданий. Знайдемо таке g, що $2\Phi(g) = \gamma$. Тоді $\left[\bar{x} - \frac{1}{\sqrt{n}}\sigma g, \; \bar{x} + \frac{1}{\sqrt{n}}\sigma g\right]$ є довірчим інтервалом для m рівня довіри γ. Число $\Delta = \frac{1}{\sqrt{n}}\sigma g$ називають *точністю оцінки* середнього m.

Наведемо два приклади побудови довірчого інтервалу, в яких застосовують методики, характерні для біологічних досліджень.

Приклад 65. Відомо, що кількість кальцію у сироватці крові колонії мавп має нормальний розподіл. Оцінити основні характеристики вибірки обстежених мавп із використанням *біометричної обробки даних за Н.А. Плохінським (1961)*, якщо число мавп — 100 особин, а кількість кальцію в їх крові змінюється від 9 до 14,7 мг.

Розв'язання. Наближені значення середнього й середньоквадратичного відхилень визначають за такими формулами:

$$\bar{x} = \frac{x_{\min} + x_{\max}}{2}, \quad \tilde{S}_n = \frac{x_{\max} + x_{\min}}{K}, \qquad (43)$$

де $x_{\min} = 9$, $x_{\max} = 14,7$. Значення коефіцієнта K залежно від об'єму вибірки наведено в табл. 4.

Таблиця 4. **Значення коефіцієнта K за Н.А. Плохінським (1961)**

n	2—5	6—15	16—49	50—200	201—1000	>1000
K	2	3	4	5	6	7

У нашому випадку $K = 5$. Отже, маємо: $\bar{x} = 11,85$ мг і $S_{100} = 1,14$ мг.

Приклад 66. На групі з 10 лабораторних мишей випробовували дію отруйної речовини. За допомогою *методу Спірмена—Кербера* розрахувати середню дозу ефекту $\bar{m}$ і побудувати довірчий інтервал для генеральної середньої. Результати дослідів наведено у таблиці:

Доза, мг/кг	110	120	130	140	150	160	170	180	Сумарна частка
Число тварин, які загинули	0	1	3	4	6	7	9	10	40
Частка тварин, які загинули	0	0,1	0,3	0,4	0,6	0,7	0,9	1	4
Накопичена частка тварин, які загинули	0	0,1	0,4	0,8	1,4	2,1	3	4	11,8

Розв'язання. При випробуванні радіоактивних та інших біологічно активних речовин з'ясовано, що особини однорідної групи реагують на одну і ту саму дозу по-різному і що різні дози можуть викликати однаковий ефект у цілої групи індивідів. Звідси випливає, що про силу дії на організм біологічно активних речовин можна судити лише за середнім результатом.

Середню дозу ефекту визначимо за *формулою Спірмена—Кербера*

$$\bar{m} = m - d(P_1 - 0,5),$$

де m — мінімальна доза, яка викликає ефект у 100 % піддослідних індивідів; d — різниця між дозами; P_1 — загальна (сумарна) частка індивідів, які реагують на дози.

Середньоквадратичне відхилення розраховують за формулою

$$S_n = d\sqrt{2P_2 - P_1^2 - P_1 - 1/12},$$

де P_2 — сума ряду накопичених часток індивідів, які реагують на дози.

За умовою задачі $n = 10$, $d = 10$, $m = 180$ мг/кг, $P_1 = 4$, $P_2 = 1,18$. Тоді $\bar{m} = 145$ мг/кг і $S_n = 18,75$. Розраховані величини дають можливість побудувати довірчий інтервал для генеральної

середньої дози ефекту, тобто: $\bar{m} \pm \Delta$, де $\Delta = \dfrac{t_{9;0,95} S_n}{\sqrt{n}} = 13,4$ мг/кг, $t_{9;0,95} = 2,31$ (див. табл. 3) — величина граничної помилки середньої $\bar{m}$.

§ 11. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

Визначення. *Статистичною гіпотезою* називають довільне припущення про генеральну сукупність, яке перевіряється по вибірці.

Гіпотези бувають непараметричними (наприклад, гіпотеза про рівність двох функцій розподілу) і *параметричними* (наприклад, гіпотеза про те, що середнє випадкової величини дорівнює нулю).

Нехай $\rho_\theta(x)$ — щільність розподілу випадкової величини, яка залежить від параметра θ. За даними спостереження висувають гіпотезу, що значення параметра θ дорівнює θ_0: $\theta = \theta_0$. Позначимо цю гіпотезу через H_0. Альтернативою гіпотезі H_0 можна висунути гіпотезу H_1 про те, що $\theta = \theta_1$, або гіпотезу $H_2: \theta \neq \theta_0$, або гіпотезу $H_3: \theta > \theta_0$ (одностороння гіпотеза). Нехай, наприклад, альтернативою є гіпотеза H_1. Необхідно перевірити гіпотезу H_0 на основі деякої вибірки $\{x_1, x_2,...,x_n\}$. Ті значення $\{x_1, x_2,...,x_n\}$, за яких гіпотеза H_0 приймається, називають *областю прийняття гіпотези* H_0; ті ж значення $\{x_1, x_2,...,x_n\}$, за яких гіпотеза H_0 відкидається, називають *критичною областю.*

При перевірці гіпотези H_0 можна припуститись двох типів помилок: помилка 1-го типу — гіпотеза H_0 відкинута, коли вона є правильною; помилка 2-го типу — гіпотеза H_0 прийнята, коли вона є неправильною.

Бажано обрати критичну область (у цьому випадку область прийняття гіпотези визначається автоматично) так, щоб зменшити ймовірності α і β помилок обох типів. На практиці можна (за фіксованого об'єму вибірки) зробити як завгодно малим або α, або β. Число α (β) називають *рівнем значущості гіпотези.* Зазначимо, що узгодження даних спостережень з гіпотезою H_0 не означає, що ця гіпотеза є правильною: цілком можливо, що дані узгоджуються також і з іншою гіпотезою. У цьому разі можна тільки стверджувати, що дані не суперечать гіпотезі H_0.

Розглянемо два важливих випадки перевірки статистичних гіпотез.

11.1. Перевірка гіпотези про закон розподілу. Критерій згоди χ^2

Нехай ξ — дискретна випадкова величина. Висувається гіпотеза H_0 про те, що закон розподілу ξ має вигляд

ξ	x_1	x_2	...	x_m
p	p_1	p_2	...	p_m

Нехай значенням варіанти $x_1, x_2,...,x_m$ після спостереження відповідають частоти $n_1, n_2, ..., n_m$, $\sum_{i=1}^{m} n_i = n$. Побудуємо функцію χ_n^2, яка характеризує відхилення спостережуваних частот від теоретичних ймовірностей:

$$\chi_n^2 = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i}.$$

З теорії ймовірностей відомо, що при $n \to \infty$ функція χ_n^2 розподілена за законом χ^2(хі-квадрат) з $(m - 1)$ ступенями вільності (цей розподіл табульовано). Позначимо через k число невідомих параметрів розподілу ξ. Якщо рівень значущості α заданий, тоді шукаємо за табл. 2 $(1 - \alpha)$-квантиль розподілу χ^2 з $m - 1 - k$ ступенями вільності $\chi^2_{m-1-k;1-\alpha}$, тобто таке значення $\chi^2_{m-1-k;1-\alpha}$, що $P\{\,\chi^2(m-1-k) \le \chi^2_{m-1-k;1-\alpha}\} = 1 - \alpha$.

Критерій перевірки гіпотези H_0 формулюють так. Ті значення χ_n^2, для яких $\chi_n^2 \le \chi^2_{m-1-k;1-\alpha}$, становлять *область допустимих значень* прийняття гіпотези H_0, тобто для таких χ_n^2 гіпотеза H_0 приймається; натомість, ті значення χ_n^2, для яких $\chi_n^2 > \chi^2_{m-1-k;1-\alpha}$, утворюють *критичну область* прийняття гіпотези H_0, тобто для таких χ_n^2 гіпотезу H_0 відхиляють.

Критерій згоди χ^2 використовують також і для перевірки гіпотези про те, що випадкова величина ξ неперервна, тобто має неперервну функцію розподілу $F(x)$. У цьому разі числову вісь розбивають на r інтервалів $\Delta_1, \Delta_2,...,\Delta_r$, які не перетинаються; n_i — кількість елементів вибірки, що потрапили в інтервал Δ_i $i = \overline{1, r}$;

$p_i = P\{\,\xi \in \Delta_i / H_0\,\}$ — теоретична частота попадання ξ в інтервал Δ_i за умови, що гіпотеза справедлива.

Приклад 67. Результати вибіркового обстеження наведено нижче:

Інтервал	[0; 4)	[4; 6)	[6; 8)	[8; 10)
n_i	20	40	30	10

Для рівня значущості $\alpha = 0{,}05$ за допомогою критерію χ^2 Пірсона перевірити гіпотезу про те, що ця вибірка отримана з нормально розподіленої генеральної сукупності.

Розв'язання. Об'єм вибірки $n = 100$. Обчислимо оцінки параметрів нормального розподілу відповідно до вибірки, а саме вибіркове середнє $\overline{x}$ та виправлену вибіркову дисперсію $\tilde{S}_n^2$. Для підрахунку $\overline{x}$ та $\tilde{S}_n^2$ за значення варіанти братимемо середини інтервалів:

$$\overline{x} = \frac{1}{100}(2 \cdot 20 + 5 \cdot 40 + 7 \cdot 30 + 9 \cdot 10) = 5{,}4;$$

$$\overline{x^2} = \frac{1}{100}\left(2^2 \cdot 20 + 5^2 \cdot 40 + 7^2 \cdot 30 + 9^2 \cdot 10\right) = 33{,}6;$$

$$\tilde{S}_n^2 = \frac{100}{99}\left[\overline{x^2} - \left(\overline{x}\right)^2\right] = 1{,}01[33{,}6 - 29{,}16] = 4{,}4844;\ \tilde{S}_n = 2{,}12.$$

Знайдемо p_i, $i = \overline{1, 4}$. Нехай $\Phi(x)$ — функція Лапласа (див. формулу (27), табл. 1). Маємо

$$p_1 = P\{\,0 \le \xi < 4\} = P\left\{\frac{0 - 5{,}4}{2{,}12} \le \frac{\xi - m}{\sigma} < \frac{4 - 5{,}4}{2{,}12}\right\} =$$

$$= P\left\{-2{,}55 \le \frac{\xi - m}{\sigma} < -0{,}66\right\} = \Phi(2{,}55) - \Phi(0{,}66) =$$

$$= 0{,}4945 - 0{,}2454 = 0{,}2491.$$

$$p_2 = P\{\,4 \le \xi < 6\} = P\left\{\frac{4 - 5{,}4}{2{,}12} \le \frac{\xi - m}{\sigma} < \frac{6 - 5{,}4}{2{,}12}\right\} =$$

$$= P\left\{-0{,}66 \le \frac{\xi - m}{\sigma} < 0{,}283\right\} =$$

$$=\Phi(0,283)+\Phi(0,66)=0,1120+0,2454=0,3611.$$

$$p_3=P\{6\le\xi<8\}=P\left\{\frac{6-5,4}{2,12}\le\frac{\xi-m}{\sigma}<\frac{8-5,4}{2,12}\right\}=$$

$$=P\left\{0,283\le\frac{\xi-m}{\sigma}<1,226\right\}=\Phi(1,226)-\Phi(0,283)=$$

$$=0,3895-0,1120=0,2775.$$

$$p_4=P\{8\le\xi<10\}=P\left\{\frac{8-5,4}{2,12}\le\frac{\xi-m}{\sigma}<\frac{10-5,4}{2,12}\right\}=$$

$$=P\left\{1,226\le\frac{\xi-m}{\sigma}<2,17\right\}=\Phi(2,17)-\Phi(1,226)=$$

$$=0,4850-0,3895=0,0955.$$

Результати підрахунку χ^2_{100} зведемо в таблицю:

Інтервал	[0; 4)	[4; 6)	[6; 8)	[8; 10)
n_i	20	40	30	10
p_i	0,2491	0,3611	0,2775	0,0955
np_i	24,91	36,11	27,75	9,55
$(n_i-np_i)^2$	24,11	15,13	5,06	0,20
$\frac{(n_i-np_i)^2}{np_i}$	0,97	0,42	0,18	0,02

Маємо $\chi^2_{100}=0,97+0,42+0,18+0,02=1,59$. Оскільки параметри гіпотетичного нормального розподілу $E\xi$ та $D\xi$ невідомі, а їх число $k=2$, тоді $m-1-k=4-2-1=1$. За табл. 2 знаходимо: $\chi^2_{1;0,95}=\chi^2_{1;0,95}=3,84$. Оскільки $\chi^2_{100}<\chi^2_{1;0,95}$, гіпотеза про те, що вибірка отримана з нормально розподіленої генеральної сукупності приймається.

11.2. Перевірка гіпотези про рівність середніх значень двох нормальних розподілів

Нехай ξ і η — дві випадкові величини, які мають нормальний розподіл з параметрами (a_1, σ_1) і (a_2, σ_2), де a_1, a_2 — невідомі середні, σ_1, σ_2 — відомі середньоквадратичні відхилення. Висувається гіпотеза H_0 про те, що $a_1=a_2$ всупереч альтернативі $H_1: a_1\ne a_2$. Відомо, що оцінкою середніх a_1, a_2 є вибіркові середні $\overline{x}$, $\overline{y}$. З теорії ймовірностей відомо, що $\overline{x}$ та $\overline{y}$ також мають нормальний розподіл з параметрами $\left(a_1,\ \sigma_1/\sqrt{n_1}\right)$ і $\left(a_2,\ \sigma_2/\sqrt{n_2}\right)$, де n_1, n_2 — об'єми вибірок. При цьому $\overline{x}$ та $\overline{y}$ — незалежні і $\overline{x}-\overline{y}$ також має нормальний розподіл з параметрами a_1-a_2, σ;

$$\sigma^2=\frac{1}{n_1}\sigma_1^2+\frac{1}{n_2}\sigma_2^2.$$

Отже, величина $z=\frac{1}{\sigma}\left(\overline{x}-\overline{y}\right)$ за умови справедливості гіпотези H_0 має стандартний нормальний розподіл.

Нехай задано рівень значущості α (на практиці зазвичай покладають $\alpha=0,05$ (у біологічних дослідженнях); 0,01 або 0,001). Виберемо за таблицями значень функції щільності стандартного нормального розподілу таке значення z_α, що $P\{|z|\le z_\alpha\}=1-\alpha$. Тоді ті значення z, для яких $|z|\le z_\alpha$, утворюють область допустимих значень, а ті значення z, для яких $|z|>z_\alpha$, утворюють критичну область (рис. 22).

Оскільки при перевірці гіпотези H_0 контролювалась лише помилка 1-го типу (дорівнює α), то нічого не можна сказати про ймовірність прийняття неправильної гіпотези.

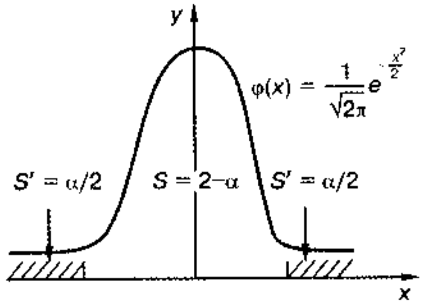


Рис. 22. Критична область критерію

§ 12. БАГАТОВИМІРНІ ДАНІ. АНАЛІЗ СТАТИСТИЧНИХ ЗВ'ЯЗКІВ

Кожне природне явище описується, як правило, не одним показником, а їх комплексом. Від того, наскільки повно і точно показники відображають сутність явища, залежить результат аналізу. В такого роду задачах результати спостереження матимуть багатовимірний характер. Для аналізу таких даних важливо вміти виявляти факторні й результативні показники, основні та другорядні співвідношення, визначити форми зв'язку. Для аналізу статистичних зв'язків застосовують регресійний, дискримінантний, дисперсійний та інші види статистичного аналізу.

12.1. Поняття про регресійний аналіз

У багатьох практичних експериментах вимірюють одночасно значення двох ознак. Такі *експерименти* називають *парними*. Мета парних експериментів полягає у вивченні залежності між отриманими значеннями двох випадкових величин X та Y. Якщо залежність існує, то можна обчислити її міру і знайти відповідне рівняння. У випадку, коли всі пари значень точно задовольняють деяке рівняння, кажуть, що *випадкові величини повністю корельовані*. Якщо зобразити пари значень (x_i, y_i) як точки з координатами (x_i, y_i) на площині, то у випадку повної кореляції всі вони будуть розміщені на одній кривій.

Якщо точки (x_i, y_i) розміщені на площині хаотично, то *випадкові величини вважають некорельованими*. Можливий проміжний випадок *часткової кореляції*: точки (x_i, y_i) розміщені поблизу деякої кривої.

Найпростішим випадком є *лінійна кореляція*. У розділі 1 було введено *коефіцієнт кореляції* $\rho = \rho(X, Y)$, $(0 \le |\rho(X, Y)| \le 1)$ між випадковими величинами X та Y, причому $\rho = 0$, якщо X та Y незалежні і $\rho = \pm 1$ тоді і тільки тоді, коли $Y = aX + b$ (лінійно залежні). Якщо ж $0 < |\rho(X, Y)| < 1$, то можна спробувати знайти таку пряму, від якої сумісні значення (X, Y) відхиляються найменше.

Визначення. Пряму $y = \theta_1 x + \theta_0$ називають *прямою регресії Y на X*, якщо серед усіх можливих дійсних значень α і β квадратичне відхилення $E(Y - \alpha X - \beta)^2$ досягає свого мінімального значення за умови $\alpha = \theta_1$ і $\beta = \theta_0$.

Коефіцієнт θ_1 називають *коефіцієнтом регресії Y на X*.

Зауваження. Рівняння для знаходження θ_1 і θ_0 можна записати у вигляді $\min\limits_{\alpha,\beta} E(Y - \alpha X - \beta)^2 = E(Y - \theta_1 X - \theta_0)^2$. Значення θ_1 і θ_0 можна знайти *методом найменших квадратів*. Для цього перепишемо $E(Y - \alpha X - \beta)^2$ у вигляді $E(Y - \alpha X - \beta)^2 = \left(\rho\sqrt{DY} - \alpha\sqrt{DX}\right)^2 + \left(1 - \rho^2\right)\sqrt{DY} + (EY - \alpha EX - \beta)^2$, де $\rho = \rho(X, Y)$. Оскільки другий доданок останнього виразу не залежить від α і β, потрібний мінімум досягається тоді, коли перший і останній доданки дорівнюють нулю, тобто при $\rho\sqrt{DY} - \alpha\sqrt{DX} = 0$ і $EY - \alpha EX - \beta = 0$, звідки $\theta_1 = \rho\dfrac{\sigma(Y)}{\sigma(X)}$, $\theta_0 = EY - \rho\dfrac{\sigma(Y)}{\sigma(X)}EX$, де позначено $\sigma(X) = \sqrt{DX}$ і $\sigma(Y) = \sqrt{DY}$. Отже, рівняння прямої регресії має вигляд

$$y - EY = \rho\frac{\sigma(Y)}{\sigma(X)}(x - EX).$$

Тепер розглянемо задачу побудови прямої регресії за результатами вибіркового обстеження. Подамо вибірку у вигляді таблиці

Y \ X	x_1	x_2	...	x_k
y_1	n_{11}	n_{12}	...	n_{1k}
y_2	n_{21}	n_{22}	...	n_{2k}
...	...	...	...	...
y_m	n_{m1}	n_{m2}	...	n_{mk}

де n_{ij}, $i = \overline{1,k}; j = \overline{1,m}$ — частота пари (x_i, y_j) у вибірці.

Опишемо залежність між y_j та x_i співвідношенням

$$y_j = \theta_1 x_i + \theta_0 + e_{ij}, \quad i = \overline{1,k}; j = \overline{1,m},$$

де e_{ij} — значення однаково розподілених незалежних випадкових величин, які описують відхилення точок (x_i, y_j) від лінійної моделі так, що $Ee_{ij} = 0$ і $De_{ij} = \sigma^2$.

Реалізуємо метод найменших квадратів у цій ситуації. Позначимо через $R(\theta_0, \theta_1)$ сумарне квадратичне відхилення вибіркових

значень y_j від прямої регресії $y = \theta_1 x + \theta_0$, тобто $R(\theta_0, \theta_1) = $
$= \sum_{j=1}^{m}\sum_{i=1}^{k} e_{ij}^2 = \sum_{j=1}^{m}\sum_{i=1}^{k}\left(y_j - \theta_1 x_i - \theta_0 \right)^2 n_{ij}$.

Значення θ_0 і θ_1, як і в попередньому випадку, знайдемо з умови мінімуму $R(\theta_0, \theta_1)$: $\min\limits_{\alpha,\beta} R(\alpha, \beta) = R(\theta_0, \theta_1)$. Необхідна умова екстремуму функції $R(\theta_0, \theta_1)$ визначає для θ_0 і θ_1 таку систему рівнянь:

$$\begin{cases} \partial R(\theta_0, \theta_1)/\partial \theta_0 = 0 \\ \partial R(\theta_0, \theta_1)/\partial \theta_1 = 0. \end{cases}$$

Можна довести, що ця система завжди має єдиний розв'язок, а саме: $\theta_0 = \overline{y} - r_n \dfrac{S_n(y)}{S_n(x)} \overline{x}$, $\theta_1 = r_n \dfrac{S_n(y)}{S_n(x)}$, $r_n = r_n(X, Y)$ — вибірковий коефіцієнт кореляції.

Отже, рівняння прямої регресії Y на X набуде вигляду

$$y = \overline{y} + r_n \frac{S_n(y)}{S_n(x)}\left(x - \overline{x} \right). \tag{44}$$

Розглянемо детальніше випадок, коли відхилення від лінійної моделі $e_{ij}, i = \overline{1,k}; j = \overline{1,m}$ мають нормальний розподіл із параметрами $Ee_i = 0$ і $De_i = \sigma^2$, тобто $e_{ij} \sim N(0, \sigma^2)$. У цьому разі параметри θ_0, θ_1 і σ^2 можна оцінити не тільки точково, як у формулі (44), а й за допомогою інтервалів довіри. Позначимо

$$\tilde{\theta}_1 = \frac{\sum_{j=1}^{m}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)\left(y_j - \overline{y}\right)n_{ij}}{\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 \tilde{n}_i}, \quad \tilde{n}_i = \sum_{j=1}^{m} n_{ji}, \quad \tilde{\theta}_0 = \overline{y} - \tilde{\theta}_1 \overline{x},$$

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{k}\left(y_j - \tilde{\theta}_0 - \tilde{\theta}_1 x_i \right)^2 n_{ij}, \quad S_n^2(x) = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x} \right)^2 \tilde{n}_i. \tag{45}$$

Теорема. Величини $\tilde{\theta}_0$, $\tilde{\theta}_1$, $\tilde{\sigma}^2$ є незміщеними і незалежними між собою оцінками параметрів θ_0, θ_1, σ^2. Довірчі інтервали для θ_0, θ_1 і σ^2 рівня довіри γ мають такий вигляд:

$$\tilde{\theta}_0 - t_{\gamma;n-2}\frac{\tilde{\sigma}}{\sqrt{n-2}} < \theta_0 < \tilde{\theta}_0 + t_{\gamma;n-2}\frac{\tilde{\sigma}}{\sqrt{n-2}};$$

$$\tilde{\theta}_1 - t_{\gamma;n-2}\frac{\tilde{\sigma}}{S_n(x)\sqrt{n-2}} < \theta_1 < \tilde{\theta}_1 + t_{\gamma;n-2}\frac{\tilde{\sigma}}{S_n(x)\sqrt{n-2}};$$

$$\frac{n\tilde{\sigma}^2}{\chi^2_{n-2;\frac{1+\gamma}{2}}} < \sigma^2 < \frac{n\tilde{\sigma}^2}{\chi^2_{n-2;\frac{1-\gamma}{2}}},$$

де $t_{\gamma;n-2}$ — симетричний γ-довірчий рівень для розподілу Стьюдента з $n-2$ ступенями вільності $P\left\{ |t| < t_{\gamma;n-2} \right\} = \gamma$; $\chi^2_{n-2;\frac{1-\gamma}{2}}$ та $\chi^2_{n-2;\frac{1+\gamma}{2}}$ — $\dfrac{1-\gamma}{2}$ і $\dfrac{1+\gamma}{2}$ — квантилі розподілу χ^2 з $n-2$ ступенями свободи, відповідно.

Наслідок. Нерівність для θ_1 теореми можна використати для побудови *статистичного t-критерію* перевірки статистичної гіпотези $H_0: \theta_1 = a$. Позначимо $t = \dfrac{\tilde{\theta}_1 - a}{\sqrt{S^2 / S_n^2(x)}}$, де $S^2 = \tilde{\sigma}^2/(n-2)$.

Якщо $|t| < t_{\gamma;n-2}$, то гіпотеза H_0 приймається з рівнем значущості $\alpha = 1 - \gamma$. В іншому разі гіпотеза H_0 відкидається.

Приклад 68. Побудувати лінію регресії для даних, наведених у таблиці до прикладу 61. Перевірити гіпотезу $H_0: \theta_1 = 0$ з рівнем значущості 0,05.

Розв'язання. Згідно з формулами (45), маємо

$$\tilde{\theta}_1 = \frac{\sum_{i=1}^{6}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^{6}\left(x_i - \overline{x}\right)^2} = 2{,}7 \text{ і } \tilde{\theta}_0 = \overline{y} - \tilde{\theta}_1 \overline{x} = 46{,}8.$$

Отже, шукана лінія регресії описується рівнянням $y = 46{,}8 + 2{,}7x$.

Гіпотезу H_0 перевіримо, скориставшись *статистичним t-критерієм*. Обчислимо t згідно з наслідком теореми. Отримаємо $t = 0{,}7$. За табл. 3 $t_{4;0{,}95} = 3{,}18$. Оскільки розраховане значення t набуває значень з області прийняття гіпотези H_0: $0{,}7 \in (-3{,}18;\ 3{,}18)$, то гіпотеза H_0 узгоджується з експериментальними даними.

12.2. Елементи дисперсійного аналізу

Дисперсійний аналіз — це статистичний метод аналізу результатів спостережень, які залежать від різних одночасно діючих чинників, вибору найважливіших із них та оцінки їхнього впливу.

12.2.1. Однофакторний дисперсійний аналіз

Припустимо, що на кількісну нормально розподілену ознаку Y впливає фактор X, який має p різних сталих рівнів. На кожному i-му рівні здійснюють n_i вимірювань. Припустимо, що результати спостереження досліджуваної величини, мають таку структуру:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad j = \overline{1, n_i}, \ i = \overline{1, p}, \quad n = n_1 + \ldots + n_p,$$

де y_{ij} — спостереження досліджуваної величини; n_i — число спостережень величини за i-го значення фактора; μ — загальне середнє нормально розподіленої ознаки Y; α_i — ефект фактора на i-му рівні; ε_{ij} — похибка спостережень.

Припустимо, що ε_{ij} — незалежні й $\varepsilon_{ij} \sim N(0, \sigma^2)$, тобто нормально розподілені випадкові величини з параметрами $E\varepsilon_{ij} = 0$ і $D\varepsilon_{ij} = \sigma^2$.

Задача однофакторного дисперсійного аналізу полягає в перевірці гіпотези:

$H_0 : \alpha_i = 0, \ i = \overline{1, p}$ — фактор не впливає на результати спостережень.

Введемо такі позначення:

$\overline{y} = \frac{1}{n} \sum_{i=1, j=1}^{p, n_i} y_{ij}$ — загальне середнє, яке слугує оцінкою μ;

$\overline{y_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ — групове середнє.

Нехай $Q = \frac{1}{n} \sum_{i=1, j=1}^{p, n_i} \left(y_{ij} - \overline{y} \right)^2$ — сума квадратів відхилень спостережень від загального середнього. Тоді $Q = Q_1 + Q_2$, де $Q_1 = \sum_{i=1}^{p} n_i \left(\overline{y_i} - \overline{y} \right)^2$ — факторна сума квадратів відхилень групових середніх від загального середнього, що характеризує розсіяння між групами; $Q_2 = \sum_{i=1, j=1}^{p, n_i} \left(y_{ij} - \overline{y_i} \right)^2$ — залишкова сума квадратів

Таблиця 5. **Таблиця Фішера—Снедекора** $F_{(\alpha, k_1, k_2)}$, $\alpha = 0,05$

k_2 \ k_1	1	2	3	4	5	9	∞
1	161	200	216	225	230	241	254
2	18,5	190	19,2	19,3	19,3	19,4	19,5
3	10,1	9,55	9,28	9,12	9,01	8,81	8,53
4	7,71	6,94	6,59	6,39	6,26	6,00	5,63
5	6,61	5,79	5,41	5,19	5,05	4,78	4,36
6	5,99	5,14	4,76	4,53	4,39	4,10	3,67
7	5,59	4,74	4,35	4,12	3,97	3,68	3,23
8	5,32	4,46	4,07	3,84	3,69	3,39	2,93
9	5,12	4,26	3,86	3,63	3,48	3,18	2,71
∞	3,84	2,99	2,60	2,37	2,21	1,88	1

відхилень спостережуваних значень групи від свого групового середнього.

Для перевірки гіпотези H_0 про те, що фактор не впливає на результати спостережень, використовують критерій Фішера—Снедекора.

Позначимо $S_1^2 = \frac{Q_1}{p-1}$ і $S_2^2 = \frac{Q_2}{n-p}$. Тоді $S^2 = \frac{Q}{n-1}$ — незміщена оцінка σ^2. Нехай $F = \frac{S_1^2}{S_2^2}$, $k_1 = p - 1$, $k_2 = n - p$ і $\alpha = 1 - \gamma$.

Критерій Фішера—Снедекора для перевірки гіпотези H_0 формулюється так: якщо $F < F_{(\alpha, k_1, k_2)}$, то гіпотеза приймається; якщо $F \geq F_{(\alpha, k_1, k_2)}$, то гіпотеза відхиляється. Величину $F_{(\alpha, k_1, k_2)}$ знаходять за таблицею Фішера—Снедекора з (k_1, k_2) ступенями вільності (табл. 5).

РОЗДІЛ III

ТЕОРІЯ ПОХИБОК ВИМІРЮВАННЯ

§ 13. ТИПИ ПОХИБОК

За характером отримання результатів експерименту вимірювання фізичних величин можна поділити на дві категорії: *прямі* та *непрямі вимірювання*.

Визначення. *Прямими* називають вимірювання, за яких шукане значення фізичної величини отримують безпосередньо з експерименту шляхом порівняння її з еталонною мірою або за допомогою того чи іншого засобу вимірювання, градуйованого в одиницях відповідної міри.

Наприклад, вимірювання довжини — лінійкою, часу — секундоміром, сили струму в електричному колі — за допомогою амперметра тощо.

Визначення. *Непрямими* називають вимірювання, за яких шукане значення фізичної величини знаходять на основі відомої залежності між цією величиною та величинами, отриманими за допомогою прямих вимірювань.

Наприклад, вимірювання площі фігури (S) за результатами прямих вимірювань її довжини (a) і ширини (b). Шукану площу обчислюють за формулою $S = ab$. Вимірювання опору резистора (R) за результатами прямих вимірювань напруги (U) і сили струму (I) в електричному колі (закон Ома). Шуканий опір резистора обчислюють за формулою $R = U/I$.

Визначення. *Істинним значенням* фізичної величини називають таке її значення, яке в якісному і кількісному відношеннях ідеально відображає відповідну властивість об'єкта.

Досвід показав, що довільні вимірювання фізичної величини завжди супроводжуються відхиленнями від істинного значення вимірюваної величини.

Визначення. *Похибка вимірювання* фізичної величини — це відхилення результату вимірювання від істинного значення вимірюваної фізичної величини.

Кількісно похибки вимірювання поділяють на абсолютні та відносні.

Визначення. *Абсолютною похибкою* вимірювання називають похибку, яка виражена в одиницях вимірюваної величини і дорівнює модулю різниці між результатом вимірювання (x_i) та істинним значенням вимірюваної величини (X):

$$\Delta X = |x_i - X|. \tag{46}$$

Визначення. *Відносною похибкою* вимірювання називають похибку, яка дорівнює відношенню абсолютної похибки вимірювання до істинного значення вимірюваної величини. Її виражають у відносних одиницях або у відсотках:

$$E = \frac{\Delta X}{X}(100\ \%). \tag{47}$$

Оскільки істинне значення фізичної величини X невідоме, тому у формулах (46) і (47) замість величини X підставляють її наближене значення. Як це зробити, буде сказано далі.

§ 14. ПОХИБКИ ПРИ ПРЯМИХ ВИМІРЮВАННЯХ

Абсолютна похибка при прямих вимірюваннях складається з двох компонент: *систематичної* та *випадкової* похибок.

Визначення. *Систематична похибка вимірювань* — це така похибка, яка залишається сталою або закономірно змінюється при повторних вимірюваннях однієї й тієї ж величини, що виконуються за незмінних умов.

Систематична складова абсолютної похибки зумовлена головним чином похибками засобів вимірювання (*інструментальними похибками*). Наприклад, розміри тіла вимірюють за допомогою міліметрової лінійки. Похибка вимірювання визначається точністю градуювання, тобто систематична абсолютна похибка дорівнює ціні поділки вимірювального приладу (в нашому випадку — ціні поділки лінійки): $\Delta X_{сист} = 1$ мм. Систематичну абсолютну похибку мінімізують удосконаленням методу вимірювання, а також ретельного налаштування вимірювального приладу.

Визначення. *Випадкова похибка вимірювань* — це така похибка, яка змінюється випадковим чином при проведенні повторних вимірювань однієї і тієї ж величини.

Випадкова складова обумовлена такими факторами, які змінюються при кожному вимірюванні і які неможливо контролювати. Наприклад, температура середовища, атмосферний тиск, фонові зміни напруги в електромережі, нестабільна робота елементів приладу вимірювання тощо.

Розглянемо математичні аспекти мінімізації випадкових похибок вимірювання, а саме *метод статистичної обробки результатів за допомогою коефіцієнтів Стьюдента,* який передбачає багаторазове вимірювання однієї й тієї ж величини. Цей метод застосовують у випадку числа вимірювань $n \le 30$.

Нехай n — число вимірювань, x_i — вимірюване значення шуканої фізичної величини, $x_c = \frac{1}{n}\sum_{i=1}^{n} x_i$ — середнє значення серії проведених вимірювань, $\hat{S}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - x_c)^2}$ — середньоквадратична похибка методу (стандартна похибка), $S_{X_c} = \hat{S}_n/\sqrt{n}$ — середньоквадратична похибка результату (похибка репрезентативності), $\Delta X_{\text{вип}} = t_{\gamma,n-1} S_{X_c}$ — випадкова похибка середнього x_c (довірча границя вимірювання), яка відповідає довірчій ймовірності γ; $t_{\gamma,n}$ — коефіцієнт Стьюдента (див. табл. 3).

Остаточний результат для істинного значення фізичної величини такий

$$X = x_c \pm \Delta X_{\text{вип}} \tag{48}$$

з довірчою ймовірністю γ.

14.1. Кількісне співвідношення систематичної та випадкової похибок

При проведенні прямих вимірювань можуть виникати такі кількісні співвідношення між систематичною та випадковою похибками:

1) систематична похибка набагато більша за випадкову похибку: $\Delta X_{\text{сист}} >> \Delta X_{\text{вип}}$; за такого співвідношення випадковою похибкою нехтують, систематичну похибку вимірювання беруть за абсолютну, тобто $X = x_c \pm \Delta X_{\text{сист}}$;

2) випадкова похибка набагато більша за систематичну похибку: $\Delta X_{\text{вип}} >> \Delta X_{\text{сист}}$; за такого співвідношення систематичною похибкою нехтують, випадкову похибку вимірювання беруть за абсолютну, тобто $X = x_c \pm \Delta X_{\text{вип}}$;

3) систематична й випадкова похибки мають один порядок: $\Delta X_{\text{сист}} \sim \Delta X_{\text{вип}}$; за такого співвідношення абсолютну похибку вимірювання визначають за формулою

$$\Delta X = \sqrt{\Delta X^2_{\text{сист}} + \Delta X^2_{\text{вип}}}\,. \tag{49}$$

Шукане значення фізичної величини та її відносна похибка будуть такими:

$$X = x_c \pm \Delta X \;\text{ і }\; E = \frac{\text{Д}X}{x_c}(100\,\%)\,. \tag{50}$$

Зауваження. Випадкова похибка зменшується зі збільшенням кількості вимірювань n, оскільки $\Delta X_{\text{вип}} \sim 1/\sqrt{n}$. Зменшення випадкової похибки внаслідок збільшення числа випробувань n обмежене величиною систематичної похибки в силу залежності (49). Отже, кількість випробувань n варто збільшувати до такого числа, коли $\Delta X_{\text{вип}}$ зменшиться до одного порядку з $\Delta X_{\text{сист}}$, яке визначають методом вимірювання.

§ 15. ПОХИБКИ ПРИ НЕПРЯМИХ ВИМІРЮВАННЯХ

При непрямих вимірюваннях шукану фізичну величину Y задають як функцію інших фізичних величин $X_1, X_2, \ldots, X_m$, значення яких отримують прямим вимірюванням: $Y = f(X_1, X_2, \ldots, X_m)$. Оскільки при вимірюванні кожної з величин X_k виникає похибка ΔX_k, то для похибки непрямих вимірювань ΔY за формулою обчислення приросту функції має місце наближення: $\Delta Y \approx \frac{\partial f}{\partial x_1}\Delta x_1 + \frac{\partial f}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial f}{\partial x_m}\Delta x_m$. При цьому, згідно з теорією похибок, обчислювати похибку непрямих вимірювань ΔY потрібно за формулою

$$\Delta Y = \sqrt{\sum_{k=1}^{m}\left(\partial f/\partial x_k\right)^2 \Delta x_k^2}\,. \tag{51}$$

Розглянемо процедуру обчислення похибок при непрямих вимірюваннях на простому прикладі.

Приклад 69. Під час дослідження електричних характеристик мембрани клітини за допомогою мікроелектродної техніки виникає потреба у визначенні опору мікроелектрода R. Скласти план обчислення R.

Розв'язання. За законом Ома

$$R = U/I, \tag{52}$$

де I та U — відповідно сила струму і різниця потенціалів, які реєструють в експерименті.

Прологарифмуємо цей вираз:

$$\ln R = \ln U - \ln I. \tag{53}$$

Продиференціюємо вираз (53), вважаючи U та I незалежними змінними. Врахувавши, що $d\ln x = \dfrac{dx}{x}$, отримаємо:

$$\frac{dR}{R} = \frac{dU}{U} - \frac{dI}{I}. \tag{54}$$

Щоб отримати формулу непрямих обчислень похибки ΔR замінимо диференціали усіх величин на абсолютні похибки зі знаками "+", а самі величини — на їх середні значення:

$$\frac{\Delta R}{R_c} = \frac{\Delta U}{U_c} + \frac{\Delta I}{I_c}. \tag{55}$$

Як бачимо, відносна похибка шуканої величини R є сумою відносних похибок величин U та I, які визначають із прямих вимірювань.

Зауваження 1. При диференціюванні вихідного виразу, що містить сталі і табличні величини, потрібно враховувати таке:

1) похідні усіх сталих величин дорівнюють нулю;

2) похибку табличної величини визначають як половину одиниці останнього значущого розряду; наприклад, якщо взяти табличне значення величини $\pi = 3{,}14$, то $\Delta\pi = 0{,}005$.

З виразу (55) знаходимо абсолютну похибку вимірювання ΔR шуканої величини R

$$\Delta R = R_c\left(\frac{\Delta U}{U_c} + \frac{\Delta I}{I_c}\right), \tag{56}$$

де ΔU і ΔI — абсолютні похибки величин U та I, які визначають із прямих вимірювань.

Середнє значення опору R_c знаходимо шляхом підстановки у формулу (52) середніх значень результатів прямих вимірювань

величин U та I:

$$R_c = \frac{U_c}{I_c}. \tag{57}$$

Остаточно для шуканої величини R запишемо відповідь:

$$R = (R_c \pm \Delta R)\ \text{Ом і}\ E = \frac{\Delta R}{R_c}(100\ \%). \tag{58}$$

Зауваження 2. При розрахунках усіх величин за відповідними формулами потрібно завжди зберігати однакову точність обчислень, а саме, залишати однакову кількість значущих знаків після коми. Крім того, при розрахунках значення шуканої фізичної величини за відповідною формулою усі розмірні величини, які фігурують у ній, потрібно брати за міжнародною системою одиниць СІ.

Зауваження 3. Врахувавши, що $Y = f(X_1, X_2) \Rightarrow R = f(U, I)$, за формулою (51) маємо

$$\Delta R = R_c\sqrt{\left(\frac{\Delta U}{U_c}\right)^2 + \left(\frac{\Delta I}{I_c}\right)^2} =$$

$$= R_c\sqrt{\left(\frac{\Delta U}{U_c} + \frac{\Delta I}{I_c}\right)^2 - \frac{2\Delta U \Delta I}{U_c I_c}} \approx R_c\left(\frac{\Delta U}{U_c} + \frac{\Delta I}{I_c}\right),$$

що узгоджується з отриманою вище формулою (56) для функції двох змінних.

Приклад 70. Визначити константу швидкості гідролітичного розщеплення сахарози.

Складемо загальний план проведення експерименту:

1) за допомогою напівтіньового поляриметра визначити кут повороту площини поляризації α_t; провести 20 вимірювань через кожні 3 хв;

2) визначити кут обертання $\alpha_з$, який відповідає закінченню реакції гідролізу;

3) за методом найменших квадратів побудувати графік залежності $y(t) = \lg(\alpha_t - \alpha_з)$; екстраполюванням цієї лінії до точки $t = 0$ отримаємо відрізок на осі ординат, який дорівнює $\lg(\alpha_0 - \alpha_з)$, де α_0 — кут обертання площини поляризації на початку реакції;

4) розрахувати константу швидкості реакції k за кімнатної температури для кожного моменту часу t за формулою

$$k = \frac{2,303}{t} \lg \frac{\alpha_t - \alpha_\infty}{\alpha_0 - \alpha_\infty};$$

5) результати проведених експериментів і розрахунків записати в таблицю; знайти похибки вимірювань фізичних величин.

Розв'язання. **1.** Результати експерименту 1) та розрахунків 3) запишемо в таблицю, врахувавши результат експерименту 2) ($\alpha_з = -15°$):

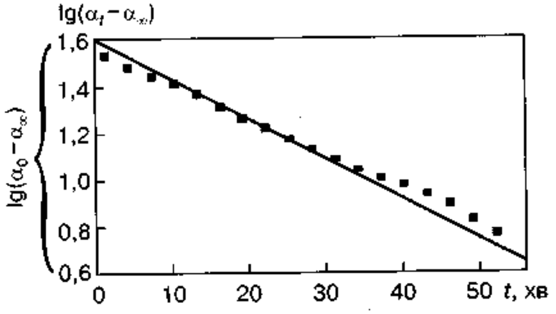
n	t, хв	α_t, град	$\alpha_t - \alpha_з$, град	$y(t) = \lg(\alpha_t - \alpha_з)$
1	1	19,3	34,3	1,54
2	4	16,2	31,2	1,49
3	7	13,2	28,2	1,45
4	10	11,1	26,1	1,42
5	13	8,2	23,2	1,37
6	16	5,6	20,6	1,31
7	19	3,4	18,4	1,26
8	22	1,6	16,6	1,22
9	25	0	15	1,18
10	28	–1,5	13,5	1,13
11	31	–2,8	12,2	1,09
12	34	–3,9	11,1	1,05
13	37	–4,7	10,3	1,01
14	40	–5,5	9,5	0,98
15	43	–6,2	8,8	0,94
16	46	–7	8	0,9
17	49	–8,3	6,7	0,83
18	52	–9,1	5,9	0,77
19	55	–11,4	3,6	0,56
0	58	–12,6	2,4	0,38

2. Шукатимемо функціональну залежність між $\lg(\alpha_t - \alpha_з)$ і t у вигляді апроксимувальної лінійної функції $\lg(\alpha_t - \alpha_з) \approx at + b$ методом найменших квадратів (див. розд. II, п. 12.1).

За методом найменших квадратів з урахуванням розрахунків $\sum_{i=1}^{20} y_i t_i = 543,97$, $\sum_{i=1}^{20} t_i^2 = 23390$, $\sum_{i=1}^{20} t_i = 590$, $\sum_{i=1}^{20} y_i = 21,88$ знаходимо систему лінійних рівнянь для визначення невідомих a і b:

$$\begin{cases} 23390a + 590b = 543,97 \\ 590a + 20b = 21,88. \end{cases}$$

Розв'язавши цю систему, отримаємо: $a = -0,017$, $b = 1,596$, й отже, $y(t) = -0,017t + 1,596$ — рівняння шуканої апроксимувальної прямої, зображеної на рисунку



Нарешті $\lg(\alpha_0 - \alpha_з) = \lim_{t \to 0} y(t) = b = 1,596$, звідки $\alpha_0 - \alpha_з = 39,4°$, $\alpha_0 = 24,4°$.

Зауваження 1. Залежність $y(t) = \lg(\alpha_t - \alpha_з)$ побудована за допомогою пакета програми "Origin".

Зауваження 2. Всі однотипні розрахунки, у тому числі й числові характеристики вибіркових даних, можна ефективно виконати за допомогою пакета програми "Excel".

3. Результати розрахунку константи швидкості реакції k за кімнатної температури для кожного моменту часу t занесемо у таблицю:

n	t, хв	k, хв$^{-1}$	n	t, хв	k, хв$^{-1}$
1	1	0,139	11	31	0,038
2	4	0,058	12	34	0,037
3	7	0,021	13	37	0,036
4	10	0,041	14	40	0,036
5	13	0,041	15	43	0,035
6	16	0,041	16	46	0,035
7	19	0,040	17	49	0,036
8	22	0,039	18	52	0,037
9	25	0,039	19	55	0,044
10	28	0,038	20	58	0,048

Визначимо $k_c = \bar{k} = \frac{1}{20}\sum_{i=1}^{20} k_i = 0,041$ хв$^{-1}$ — середнє значення шуканої величини, потім $S_{20} = \sqrt{\frac{1}{19}\sum_{i=1}^{20}(k_c - k_i)^2} = 0{,}023$ хв$^{-1}$ — середньоквадратичну похибку методу, $S_{k_c} = \frac{S_{20}}{\sqrt{20}} = 0{,}005$ хв$^{-1}$ — середньоквадратичну похибку результату і $\Delta k_{\text{вип}} = t_{0,95,19} S_{k_c} = 0{,}011$ хв$^{-1}$ — довірчий інтервал вимірювання (абсолютна похибка; $t_{0,95,19} = 2{,}09$ за табл. 3). Отже, шукана величина становить: $k = k_c \pm \Delta k_{\text{вип}} = (0{,}041 \pm 0{,}011)$ хв$^{-1}$.

Завдання для самостійної роботи. Визначити пульс P як кількість ударів N у кровоносній судині за проміжок часу Δt (наприклад, за 60 с): $P = N/\Delta t$. Слід виміряти пульс у групах людей одного віку, кожна з яких складається з п'яти осіб, за таких умов: 1) у стані спокою; 2) після 10 присідань; 3) після 20 присідань. Обчислити середню частоту пульсу, абсолютну і відносну похибки вимірювань та побудувати графік залежності частоти пульсу людини від кількості присідань.

РОЗДІЛ IV

КОМП'ЮТЕРНЕ РОЗВ'ЯЗУВАННЯ ЗАДАЧ З ТЕОРІЇ ЙМОВІРНОСТЕЙ І БІОЛОГІЧНОЇ СТАТИСТИКИ

§ 16. РОЗВ'ЯЗУВАННЯ ЗАДАЧ З ВИКОРИСТАННЯМ ПРОГРАМИ "EXCEL"

У програмному забезпеченні "Excel" передбачена можливість розв'язування багатьох важливих задач з теорії ймовірностей і математичної статистики. При цьому досягається висока точність розрахунків, можливість роботи з великими обсягами статистичних даних. Умовно виділено два рівні використання "Excel" у задачах біометрії:

1) вмонтованих в "Excel" спеціальних функцій зі статистики;
2) вмонтованих в "Excel" пакета "Статистичний аналіз".

В "Excel" є 78 функцій для проведення статистичних розрахунків. Щоб ознайомитися з набором цих функцій, у вікні "Excel" потрібно обрати в меню опцію "Вставка", а в меню, що з'явиться — опцію "Функція". Далі у вікні меню, яке з'явилося, обрати опцію "Статистичні" (рис. 22). За допомогою клавіші прокручування можна обрати будь-яку зі згаданих функцій. Наприклад, на рис. 22 обрано функцію СРЗНАЧ, яка розраховує вибіркове середнє. Опис функції отримують, натиснувши клавішу зі знаком "?" (Допомога).

До складу "Excel" входить пакет аналізу даних, призначений для розв'язування складних статистичних задач. Щоб ознайомитися з цим пакетом, у вікні "Excel" слід обрати в меню опцію "Сервіс", а в меню, що з'явиться — опцію "Аналіз даних". За допомогою клавіші прокручування можна обрати будь-яку з наведених функцій аналізу. Наприклад, якщо обрано опцію "Вибірка" (рис. 23), на екрані з'явиться меню "Вибірка". За допомогою цієї функції аналізу здійснюється побудова простої випадкової вибірки заданого об'єму за генеральною сукупністю.

Приклад 71. Розглянемо методику знаходження числових характеристик вибіркових даних для константи швидкості гідролітичного розщеплення сахарози k, наведених у прикладі 70, за допомогою статистичних функцій.

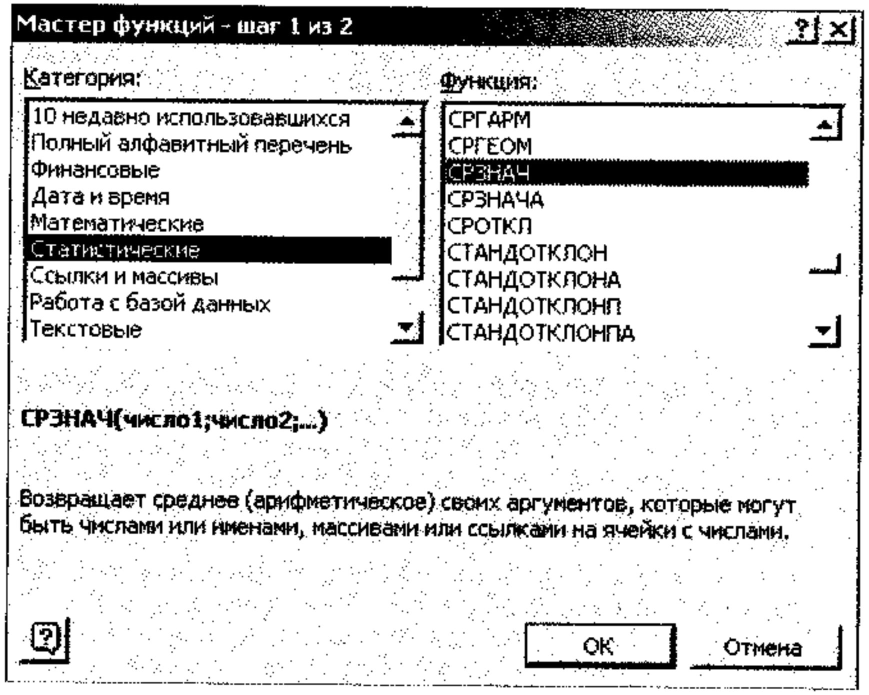


Рис. 22

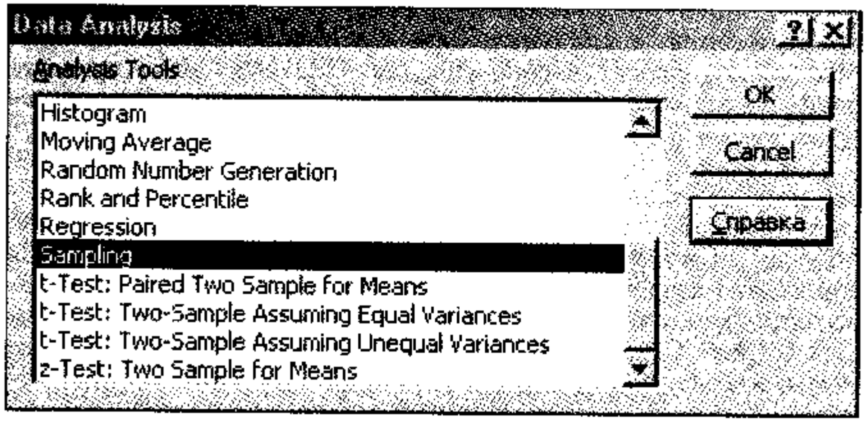
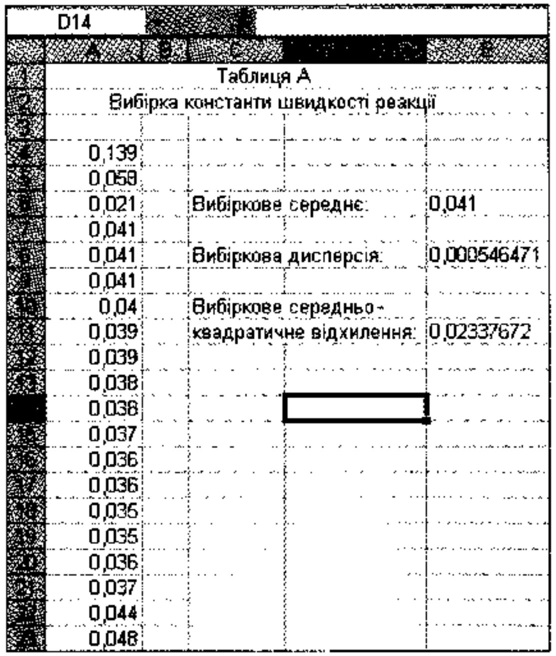


Рис. 23

Розв'язання. Подамо дані для константи швидкості реакції у вигляді таблиці (рис. 24). У відповідних комірках вводимо назви:

"Вибіркове середнє" (C6-D6); "Вибіркова дисперсія" (C8-D8); "Вибіркове середньоквадратичне відхилення" (C10-D10; C11-D11).

Вибіркове середнє $\bar{k} = \frac{1}{n}\sum_{i=1}^{n} k_i$ обчислюємо за допомогою функції СРЗНАЧ (число 1; число 2;...), де "число 1", "число 2", ... — це від 1 до 30 числових аргументів, які відповідають вибірці з генеральної сукупності. Далі активуємо комірку E6 і введемо функцію (рис. 25): = (A4 : A23). Після натискання клавіші "Enter" результат розрахунку автоматично заноситься в комірку E6 (див. рис. 24). Аналогічно розрахуємо виправлену вибіркову дисперсію $\bar{S}^2 = \frac{1}{n-1}\sum_{i=1}^{20}(k_i - \bar{k})^2$ і виправлене вибіркове середньоквадратичне відхилення $\bar{S}^2 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(k_i - \bar{k})^2}$, скориставшись відповідно функціями ДИСП (число 1; число 2;...) та СТАНДОТКЛ (число 1; число 2;...). Результати розрахунків автоматично занесуться у попередньо активовані комірки E8 та E11 (рис. 24).



	A	B	C	D	E
1	Таблиця А				
2	Вибірка константи швидкості реакції				
3					
4	0,139				
5	0,058				
6	0,021		Вибіркове середнє:		0,041
7	0,041				
8	0,041		Вибіркова дисперсія:		0,000546471
9	0,041				
10	0,04		Вибіркове середньо-		
11	0,039		квадратичне відхилення:		0,02337672
12	0,039				
13	0,038				
14	0,038				
15	0,037				
16	0,036				
17	0,036				
18	0,036				
19	0,035				
20	0,036				
21	0,037				
22	0,044				
23	0,048				

Рис. 24

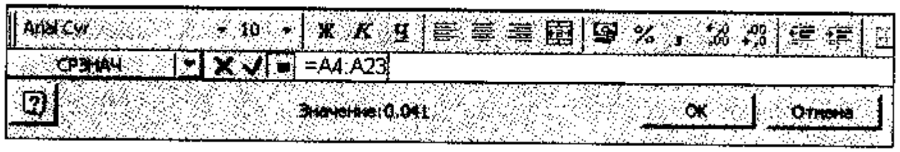


Рис. 25

	A	B	C	D
1		Задача а)	Задача б)	Задача в)
2	Вага тварини	менше 4.8 кг	більше 5.1 кг	в межах від 4.8 до 5.1 кг
3	Ймовірність			

Рис. 26

Приклад 72. Маса тіла деяких тварин, які мешкають на різних територіях, розподілена за нормальним законом із середнім значенням 5 кг і стандартним відхиленням 0,1 кг. Знайти ймовірність того, що навмання взята тварина: а) важить менш як 4,8 кг; б) важить більш як 5,1 кг; в) маса тіла тварини знаходиться в межах 4,8—5,1 кг.

Розв'язання. У випадку нормального розподілу випадкової величини скористаємось функцією НОРМРАСП (x, середнє, станд_відх, інтегральна), де x — це значення, для якого будують розподіл; "середнє" — математичне сподівання m (середньоарифметичне розподілу); "станд_відх" — середньоквадратичне відхилення σ (стандартне відхилення розподілу); "інтегральна" — логічне значення, яке визначає форму функції: $f(x) =$ $= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ (щільність розподілу), якщо "інтегральна" = ЛОЖЬ і $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}} dy$ (функція розподілу), якщо "інтегральна" = ИСТИНА. За умовою задачі "середнє" = $m = 5$ і "станд_відх" = $\sigma = 0,1$. Побудуємо таблицю (рис. 26). Далі розглянемо такі випадки:

а) для підрахунку ймовірності $P\{X < 4,8\}$ події, що навмання взята тварина важить менш як 4,8 кг, скористаємось функцією НОРМРАСП за умови, що "інтегральна" = ИСТИНА. Для цього активуємо комірку B3; наберемо команди: "Вставка",

"Функція", "Статистичні" і НОРМРАСП, у вікні меню функції НОРМРАСП, що з'явилося, наберемо параметри задачі (рис. 27, 28); результат $P\{X < 4,8\} = 0,02275$ можна прочитати в меню функції; після натискання клавіші "ОК" він також автоматично занесеться в комірку B3 (див. рис. 27);

б) для підрахунку ймовірності $P\{X \ge 5,1\}$ події, що навмання взята тварина важить більш як 5,1 кг, скористаємось співвідношенням $P\{X \ge 5,1\} = 1 - P\{X < 5,1\}$; для цього активуємо комірку C3, далі наберемо функцію (рис. 29): = 1 — НОРМРАСП (5,1; 5; 0,1; ИСТИНА); після натискання клавіші "Enter", результат $P\{X \ge 5,1\} = 0,158655$ автоматично занесеться в комірку C3 (рис. 29);

B3 =НОРМРАСП(4,8;5;0,1;ИСТИНА)

	A	B	C	D
1		Задача а)	Задача б)	Задача в)
2	Вага тварини	менше 4.8 кг	більше 5.1 кг	в межах від 4.8 до 5.1 кг
3	Ймовірність	0,02275		

Рис. 27

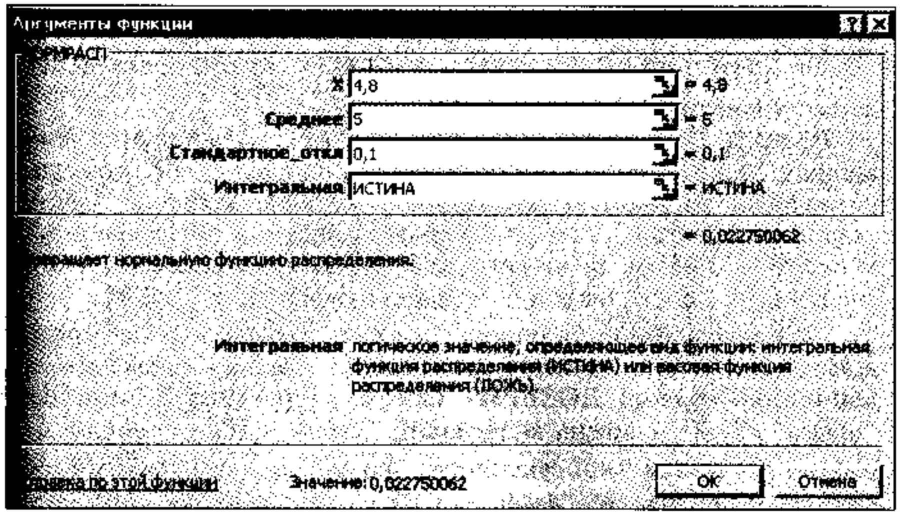


Рис. 28

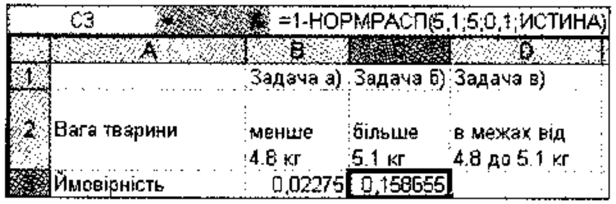


Рис. 29

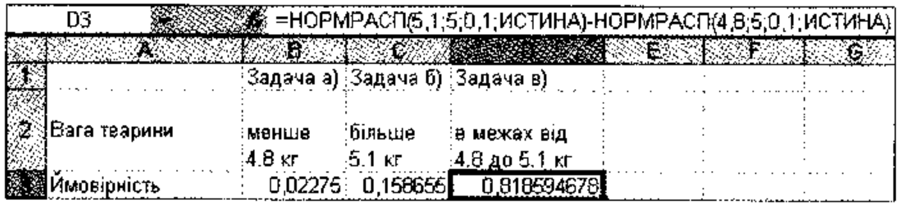


Рис. 30

в) для підрахунку ймовірності $P\{4,8 < X < 5,1\}$ події, що вага навмання взятої тварини знаходиться в межах 4,8—5,1 кг, скористаємось співвідношенням $P\{4,8 < X < 5,1\} = P\{X < 5,1\} - -P\{X < 4,8\}$; для цього активуємо комірку D3, потім наберемо функцію (рис. 30): =НОРМРАСП (5,1;5;0,1;ИСТИНА) — НОРМРАСП (4,8;5;0,1;ИСТИНА); після натискання клавіші "Enter" результат $P\{4,8 < X < 5,1\} = 0,8186$ автоматично занесеться у комірку D3 (див. рис. 30).

Приклад 73. У результаті обстеження групи тварин отримано вибірку обсягом $n = 100$. Випадкова величина — зріст особини, розподілена за нормальним законом. Середнє значення зросту тварин становить 45,5 см, стандартне відхилення генеральної сукупності — 3,24 см. Визначити довірчий інтервал для середнього значення m зросту тварин із заданою ймовірністю $\gamma = 0,95$.

Розв'язання. Точність оцінки Δ, яка пов'язана з визначенням довірчого інтервалу шуканої величини ($x = m \pm \Delta$), знаходять за допомогою функції ДОВЕРИТ (альфа; станд_відх; розмір), де "альфа" $= 1 - \gamma$ — рівень значущості; "станд_відх" — стандартне відхилення генеральної сукупності; "розмір" — обсяг вибірки. За умовою задачі маємо: "альфа" $= 0,05$; "станд_відх" = 3,24; "розмір" = 100. Побудуємо таблицю "Точність оцінки" та активуємо

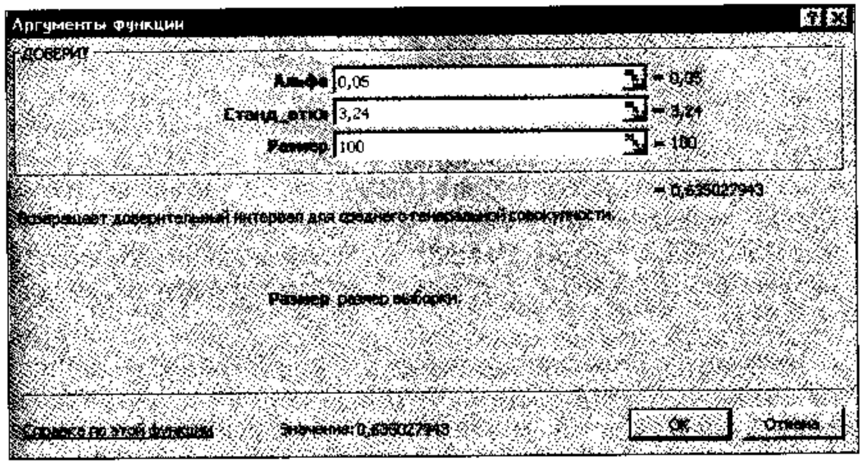


Рис. 31

комірку A2. Наберемо команди: "Вставка", "Функція", "Статистичні" і ДОВЕРИТ. У вікні меню функції ДОВЕРИТ, що з'явилося, наберемо числові значення її аргументів (рис. 31). Результат $\Delta = 0,635$ можна прочитати у вікні меню цієї функції (див. рис. 31). Після натискання клавіші "ОК" він також автоматично занесеться у комірку A2 (рис. 32).

Завдання 1. Обчислити вибіркове середнє, вибіркову дисперсію та вибіркове середньоквадратичне відхилення для такої вибірки (відсотки виконання навчальної програми до зимової сесії у вибірці з $n = 200$ студентів):

Виконання програми, %	Кількість студентів, n
<10	2
10—20	3
20—30	10
30—40	15
40—50	25
50—60	55
70—80	90
80—90	150
90—100	50

Завдання 2. За допомогою функцій ФАКТР(число), ЧИСЛКОМБ(число; число_вибраних) та ПЕРЕСТ(число; число_вибра-

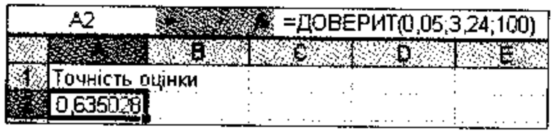


Рис. 32

них), де "число" — ціле число, що задає кількість об'єктів n; "число_вибраних" — ціле число, що задає кількість об'єктів у кожній комбінації або перестановці m, розрахувати відповідно $10!$, C_{10}^{3} та A_{10}^{3}.

Завдання 3. Використовуючи функцію БИНОМРАСП (число_успіхів; число_випробувань; ймовірність_успіху; інтегральна), де "число_успіхів" — кількість успішних випробувань m; "число_випробувань" — число незалежних випробувань n; "ймовірність_успіху" — ймовірність успіху кожного випробування p; "інтегральна" — логічне значення, що визначає форму функції: $P_n(m) = C_n^m p^m (1-p)^{n-m}$ (ймовірність події, яка полягає в тому, що настане рівно m успіхів у n незалежних випробуваннях), якщо "інтегральна" = ЛОЖЬ і $P_n(k \le m) = \sum_{r=0}^{m} P_n(r)$ (ймовірність події, яка полягає в тому, що настане не більш як m успіхів у n випробуваннях), якщо "інтегральна" = ИСТИНА, розглянути наведене нижче завдання. Після багаторічних спостережень з'ясувалося, що ймовірність піддослідної тварини інфікуватися вірусом гепатиту становить $0,4$. За припущення, що в певний день у випадковий спосіб відібрано 25 тварин, визначити ймовірність таких подій: а) вірусом гепатиту буде інфіковано рівно 10 тварин; б) не більш як 10 тварин; в) не менш як 10 тварин; г) від 10 до 15 тварин.

§ 17. РОЗВ'ЯЗУВАННЯ ЗАДАЧ З ВИКОРИСТАННЯМ ПРОГРАМИ "OriginPro"

Обов'язковою умовою дослідницької діяльності є вміння спланувати і провести експерименти, а потім коректно проаналізувати отримані результати. Задачі попереднього планування досліджень і кінцевого аналізу їх результатів розв'язує статистичний аналіз. Після проведення деякої кількості експериментів дослідник отримує дані, тобто показники тих параметрів,

які його цікавлять. Статистичний аналіз можна застосовувати до числових і нечислових даних.

Пакет програм "OriginPro 8.0" — потужний, повнофункціональний науковий пакет для аналізу даних. Він є універсальним засобом для обробки математичних і статистичних функцій, призначений для роботи з числовими і графічними науковими даними, дає змогу працювати з великими обсягами інформації, використовувати різні типи даних, автоматизувати рутинні розрахункові задачі. До того ж "Origin" працює з файлами різних пакетів і програм математичного спрямування та електронних таблиць, зокрема з "Excel", що значно спрощує можливості їх сумісного використання.

Пакет "Origin" містить потужні засоби для аналізу даних, які просто переглянути в меню **Statistics** (рис. 33) та **Analysis** (за активного вікна таблиць, згрупованих у робочу книгу — Book) (рис. 34) або лише **Analysis** (за активного вікна графіків — Graph) (рис. 35). Він дає змогу систематизувати дані та результати їх аналізу в багатоаркушні книги (за замовчуванням вони мають назву Book із наданням відповідного порядкового номера).

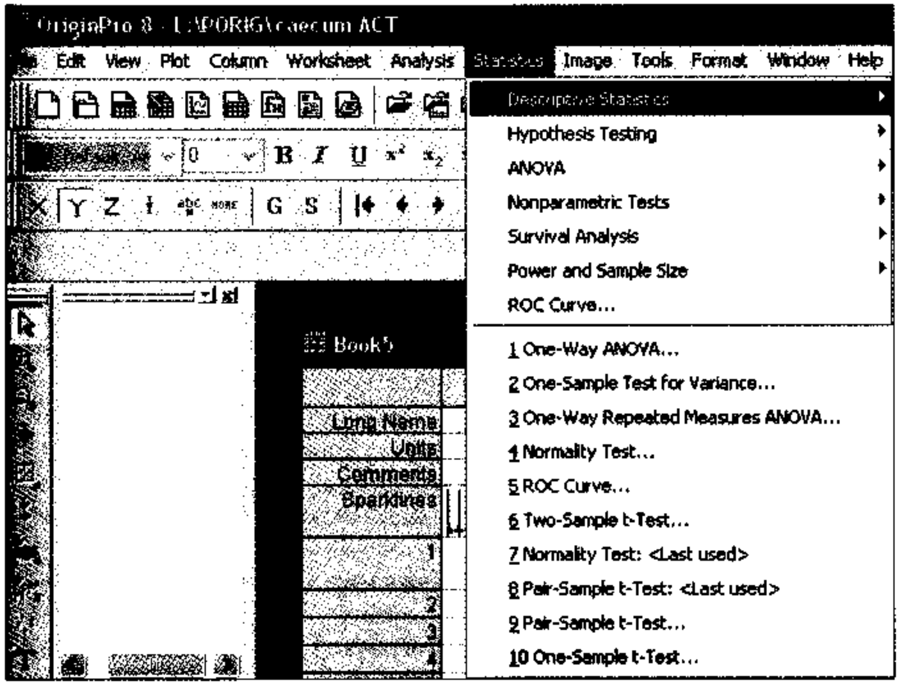


Рис. 33

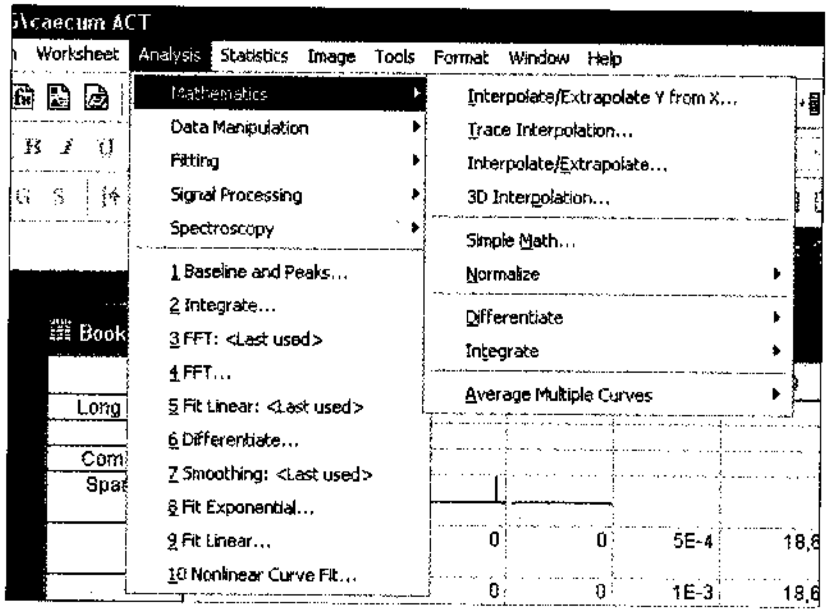


Рис. 34

Після проведення експериментів та отримання кількісних даних, дослідник перш за все має обрати спосіб їх опису. Для цього первинною процедурою статистичного аналізу є перевірка, чи відповідає вид розподілу отриманих експериментальних значень дослідженого показника, наприклад, закону нормального розподілу. Суть аналізу закону розподілу зводиться до визначення, чи отримана вибірка взята з генеральної сукупності, в якій показник розподілений за нормальним законом. Для цього перевіряються **статистичні гіпотези** щодо виду розподілу:

• **нульова гіпотеза** — розподіл досліджуваної ознаки x_i у генеральній сукупності відповідає закону нормального розподілу;

• **альтернативна гіпотеза** — розподіл досліджуваної ознаки x_i у генеральній сукупності не відповідає закону нормального розподілу.

Для перевірки цих гіпотез можна скористатися такими **статистичними критеріями**:

• **Колмогорова—Смирнова** (застосовують, коли середнє значення і середньоквадратичне відхилення ознаки відомі апріорі, а не визначаються з вибірки);

• **Ліллієфорса** (застосовують, коли середнє значення і середньоквадратичне відхилення ознаки апріорі не відомі й визначаються з вибірки);

• **Шапіро—Уілка** (застосовують, коли середнє значення і середньоквадратичне відхилення ознаки апріорі не відомі; найпотужніший критерій, особливо корисний при роботі з малими вибірками $n < 25$).

Приклад 74. Розрахувати середні значення показників активності природних радіоактивних речовин, наприклад ^{40}K і деревний попіл з Чорнобильської зони відчуження під час виконання лабораторної роботи з радіаційної біології.

За умовою методу, для кожної речовини необхідно провести не менш як 10 вимірювань, кожне з яких передбачає підрахунок кількості імпульсів упродовж 100 с.

Розв'язання. Нехай за допомогою бета-радіометра РУБ-ОІП було досліджено природний фон, а також природні радіоактивні сполуки — сипкі речовини (сіль KCl з вмістом ^{40}K і деревний попіл, зібраний на радіоактивно забрудненій території після Чорнобильської катастрофи). Проведено по 15 вимірювань, дані внесено до таблиці в стовпчики B, C і D (рис. 36).

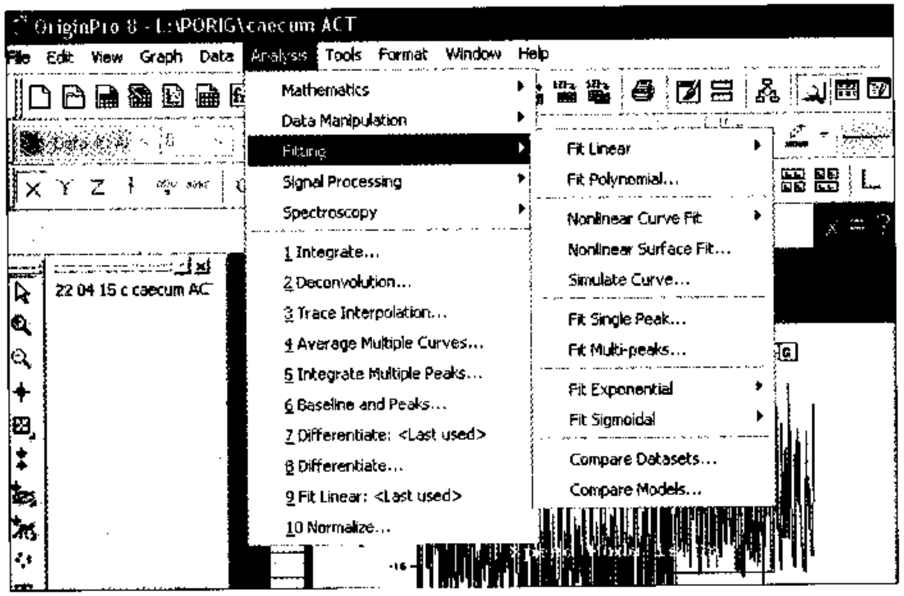


Рис. 35

З метою коректного використання саме середньоарифметичного варто перевірити всі дані на їх належність до нормально розподілених сукупностей. Це здійснюють виконанням команд меню **Statistics—Descriptive Statistics—Normality Test—Open Dialog** (Статистики—Описові Статистики—Тест на нормальність—Відкрити діалог). У діалоговому вікні **Normality Test** потрібно вказати діапазон вхідних даних (**Data Range**) — виділити стовпчики з даними, обрати тест на нормальність (ліпше тест Шапіро—Уілка — **Shapiro—Wilk**) або обрати побудову графіків (**Histograms** і **Box Chats**). Оскільки показники ймовірностей (стовпчик Prob < W) перевищують рівень $\alpha = 0,05$, результати проведення тесту на нормальність дають змогу прийняти статистичну гіпотезу про належність даних до нормально розподілених сукупностей (рис. 37).

Важливо зауважити, що методи параметричного статистичного аналізу застосовують лише тоді, коли для всіх груп даних доведена їх належність до нормально розподілених генеральних сукупностей; інакше — їх непараметричні аналоги.

При розрахунку вибіркових характеристик, що ініціюються командами меню **Statistics—Descriptive Statistics—Statistics on Column**, отримуємо зокрема середні значення (стовпчик **Mean**) та їх стандартні похибки (стовпчик **SE of mean**) (рис. 38).

Критерій Фішера для перевірки гіпотези про рівність генеральних дисперсій. Найпоширенішим статистичним завданням, з яким постійно стикаються дослідники, є зіставлення даних двох або більше вибірок з метою підтвердження або відхилення передбачення про рівність їх дисперсій і середніх значень.

У випадку, коли групи даних незалежні й розподілені за нормальним законом, статистичну гіпотезу про рівність генеральних дисперсій перевіряють за допомогою F-критерію Фішера або критерію Левена. **Статистичні гіпотези** формулюють так:

• **нульова гіпотеза** — дисперсії двох нормально розподілених генеральних сукупностей однакові;

• **альтернативна гіпотеза** — дисперсії двох нормально розподілених генеральних сукупностей неоднакові.

Book1

	A(X)	B(Y)	C(Y)	D(Y)
Long Name		фон	К	поліп
Units				
Comments				
1	1	14	107	97
2	2	23	102	96
3	3	20	112	89
4	4	18	123	88
5	5	15	111	91
6	6	19	109	95
7	7	17	122	89
8	8	19	116	91
9	9	20	117	95
10	10	15	112	98
11	11	16	109	88
12	12	18	110	92
13	13	19	115	94
14	14	20	117	93
15	15	16	116	91

Рис. 36

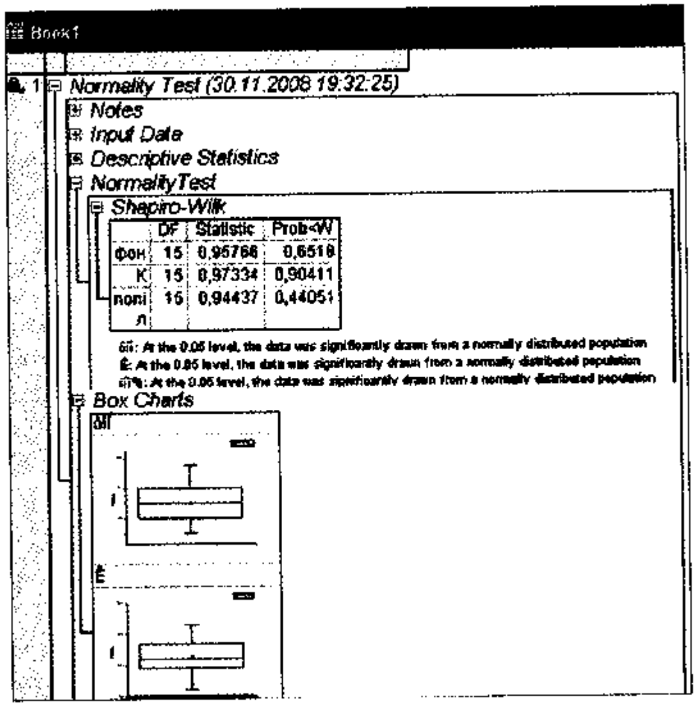


Рис. 37

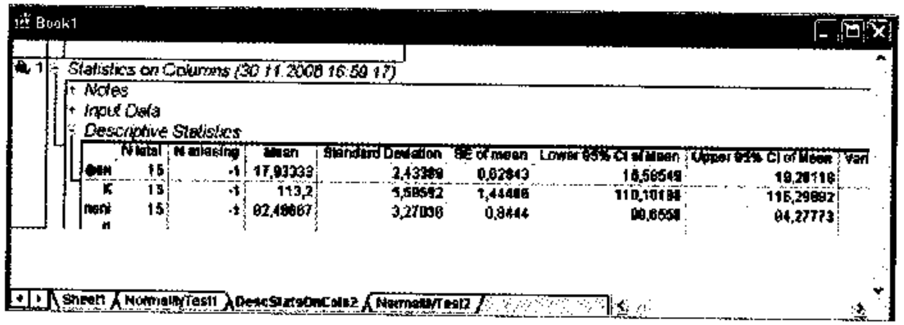


Рис. 38

Приклад 75. Розроблено нову, швидшу порівняно з відомою, тест-систему на наявність забруднення нітратами харчових продуктів, яку перевіряють за допомогою стандартного розчину, що містить ізосорбід мононітрат у концентрації 10 мг/л. Обома (відомою і новою) тест-системами проведено 8 вимірювань тестового розчину й отримано такі дані:

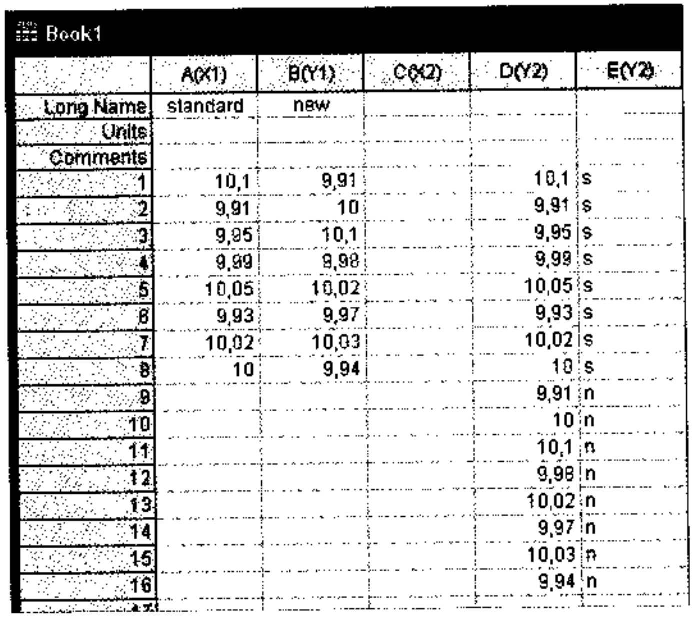
Відома тест-система	10,1	9,91	9,95	9,99	10,05	9,93	10,02	10,0
Нова тест-система	9,91	10,0	10,1	9,98	10,02	9,97	10,03	9,94

Потрібно встановити, чи в однаковій мірі варіюють покази цих тест-систем.

Розв'язання. Перш за все згрупуємо дані (рис. 39), вказавши в одній колонці усі числові значення, в іншій — групувальні коди, за якими програма ідентифікуватиме належність значень до тієї чи іншої вибірки.

Переконаємось, що обидві вибірки отримано з нормально розподілених генеральних сукупностей. За тестом Шапіро—Уілка в обох випадках (для перевірки використали незгруповані дані) підтверджено нульову гіпотезу (рис. 40).

Тепер для застосування тесту Фішера скористаємось командами: **Statistics—Hypothesis Testing—Two-sample Test for Variance** (рис. 41). У вікні **Statistic\Hypothesis Testing: TwoSample TestVar** вкажемо діапазон даних і групувальних кодів та альтернативну гіпотезу (рис. 42). За результатами тесту (рис. 43) приймаємо нульову гіпотезу. Отже, результати показів обох тест-систем варіюють в однаковій мірі.



Book1

	A(X1)	B(Y1)	C(X2)	D(Y2)	E(Y2)
Long Name	standard	new			
Units					
Comments					
1	10,1	9,91		10,1	s
2	9,91	10		9,91	s
3	9,95	10,1		9,95	s
4	9,99	9,98		9,99	s
5	10,05	10,02		10,05	s
6	9,93	9,97		9,93	s
7	10,02	10,03		10,02	s
8	10	9,94		10	s
9				9,91	n
10				10	n
11				10,1	n
12				9,98	n
13				10,02	n
14				9,97	n
15				10,03	n
16				9,94	n

Рис. 39

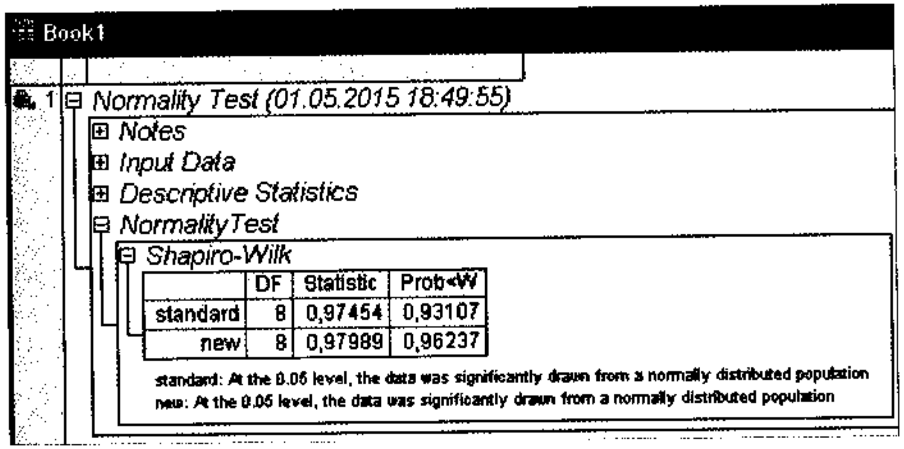


Рис. 40

Критерій Стьюдента (t-критерій) для перевірки гіпотези про рівність генеральних середніх двох вибірок (або вибірки з популяційним середнім). Практично всі біологічні дослідження потребують

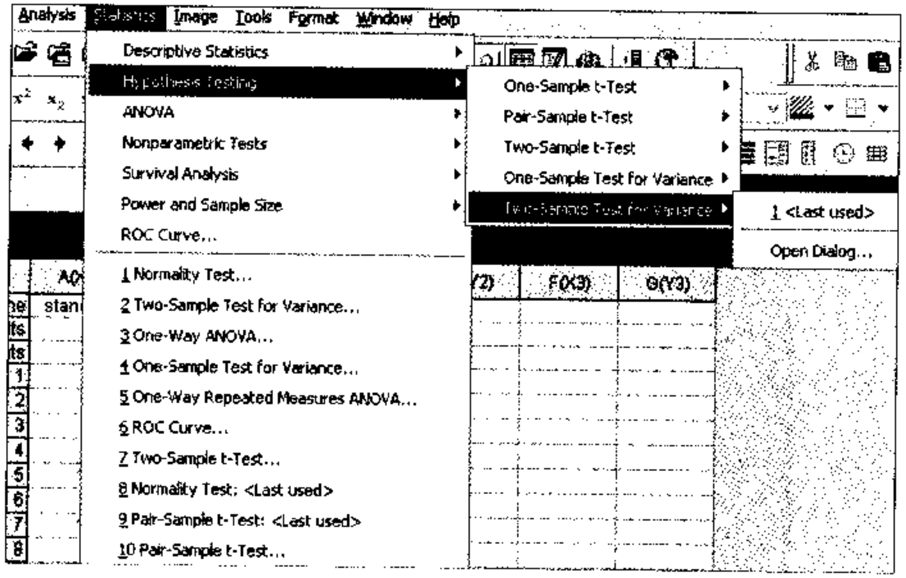


Рис. 41

порівняння результатів експериментів, коли необхідно перевірити наявність чи відсутність відмінностей середніх значень показників, зареєстрованих у контролі та за змін умов проведення дослідження (так звана проба). Таким чином, параметричний статистичний аналіз зводиться до перевірки гіпотези про рівність генеральних середніх за критерієм Стьюдента (t-критерієм):

- **нульова гіпотеза** — генеральні середні двох нормально розподілених генеральних сукупностей однакові;
- **альтернативна гіпотеза** — генеральні середні двох нормально розподілених генеральних сукупностей неоднакові (**ненаправлена альтернативна гіпотеза**) або генеральна середня першої (чи другої) генеральної сукупності більша (**направлена альтернативна гіпотеза**).

Застосування критерію Стьюдента можливе у чотирьох різних випадках: для залежних вибірок (парний критерій), для незалежних вибірок з однаковими дисперсіями, для незалежних вибірок з нерівними дисперсіями та для порівняння вибірки з популяційним середнім (див. рис. 41). В усіх версіях t-критерію використовуються різні формули для розрахунку фактичного значення критерію.

Приклад 76. У рамках попереднього завдання вирішимо ще одне завдання статистичного аналізу — перевіримо, чи відрізняються між собою покази стандартної та нової тест-систем на нітрати. Спробуємо також дати відповідь на таке питання: чи можна вважати обидві (або якусь одну) тест-систему такою, що адекватно вимірює концентрацію нітратів (порівняно з тестовим розчином ізосорбід мононітрату).

Розв'язання. Для вирішення першої частини завдання треба застосувати ***критерій Стьюдента для двох незалежних вибірок з однаковими дисперсіями*** (останнє твердження — за результатами тесту Фішера). Виберемо в меню послідовно: **Statistics—Hypothesis Testing—Two-Sample t-Test—Open Dialog**. У діалоговому вікні **Statistics\Hypothesis Testing: TwoSampletTest** вкажемо групувальний діапазон (**Group Range**), діапазон даних (**Data Range**) та альтернативну гіпотезу (у цьому разі оберемо ненаправлену: **Mean1—Mean2 < > 0**) (рис. 44).

У вікні виведення результатів (рис. 45) бачимо, що ймовірність (значення p-рівня) практично дорівнює 1, тобто не існує

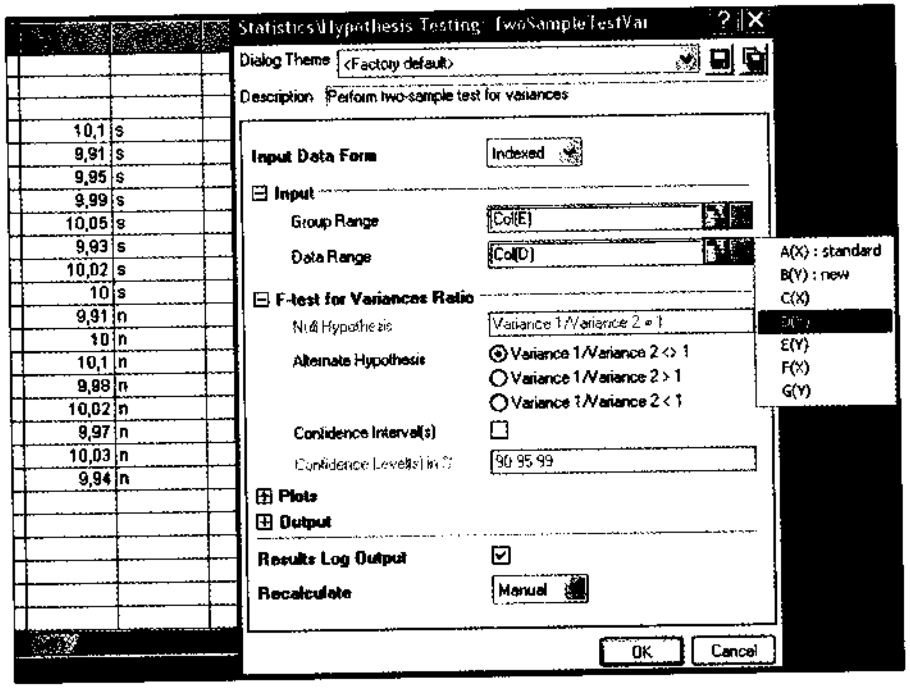


Рис. 42

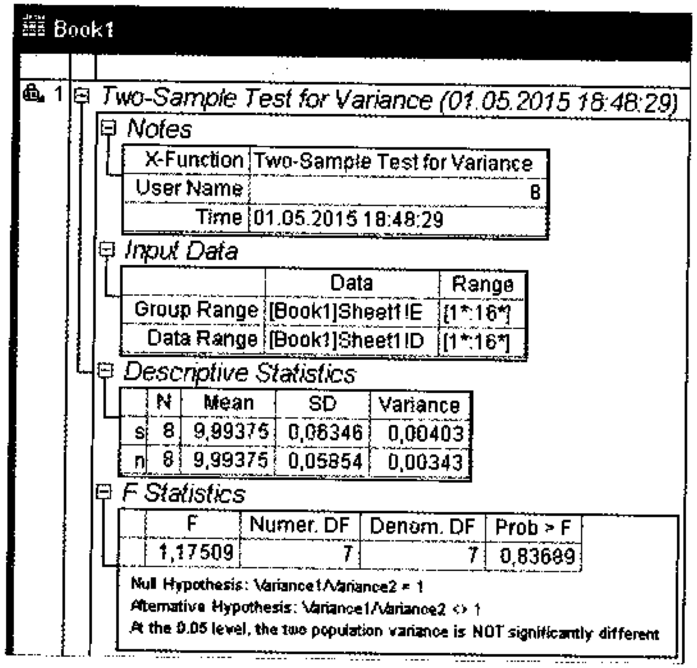


Рис. 43

жодних підстав відхилити нульову гіпотезу. *Висновок*: точність вимірювання відомою й новою тест-системами не відрізняється.

Тепер, після підтвердження гомогенності роботи обох тест-систем, скористаємось ***одновибірковим критерієм Стьюдента***, щоб з'ясувати відповідь на другу частину завдання, тобто перевіримо, чи можна вважати тест-системи такими, що адекватно вимірюють концентрацію нітратів (порівняно з тестовим розчином ізосорбід мононітрату). Для цього виділимо обидві колонки незгрупованих даних і виберемо в меню послідовно: **Statistics—Hypothesis Testing—One-Sample t-Test—Open Dialog**. У діалоговому вікні **Statistics\Hypothesis Testing: OneSampletTest** вкажемо вхідний діапазон даних (**Input**), популяційну середню (**Test Mean**) та альтернативну гіпотези (у цьому разі оберемо ненаправлену: **Mean1—Mean2 < > 0**) (рис. 46).

Згідно з даними таблиці результатів тесту ($p_{\text{standard}} \approx 0{,}79$ та $p_{\text{new}} \approx 0{,}77$) (рис. 47), приймаємо нульову гіпотезу. *Висновок*: обидві тест-системи є такими, що адекватно визначають концентрацію нітратів у розчинах.

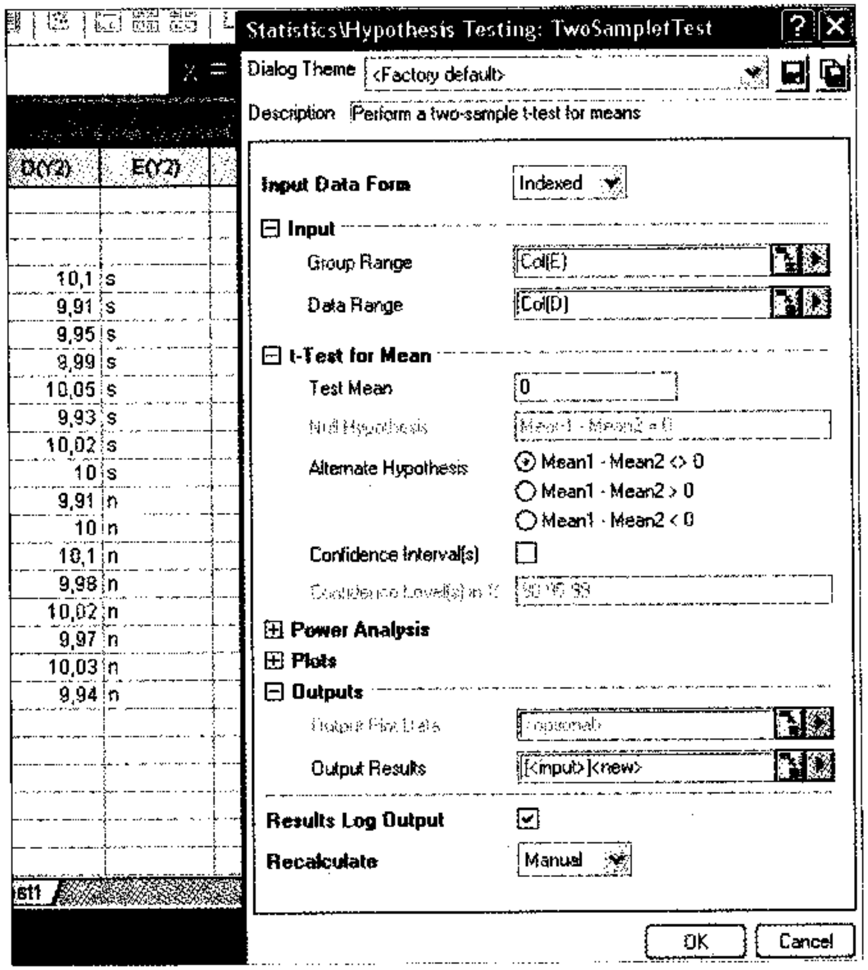


Рис. 44

Приклад 77. У відділенні лікарні перебувають 7 пацієнтів, хворих на запалення легенів, у яких на момент прийняття у стаціонар та через тиждень після початку лікування було визначено активність аспартатамінотрансферази (АСТ, яка є одним із маркерів запалення легенів). Дані дослідження наведено в таблиці:

Початок лікування	55,4	52,5	54,9	57,1	59,2	58,5	57,3
Через тиждень лікування	50,3	45,9	48,1	44,8	47,7	48,1	49,2

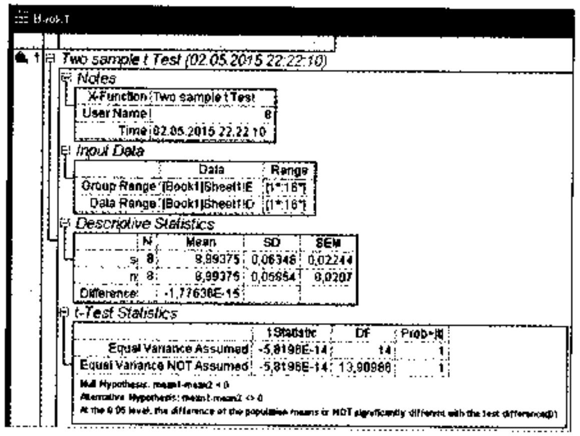


Рис. 45

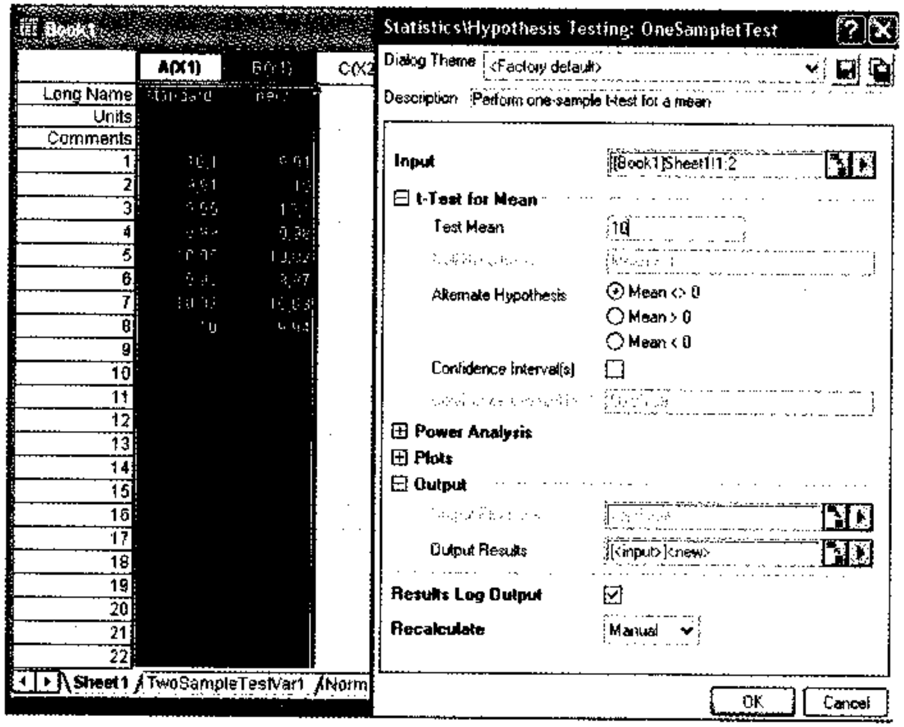


Рис. 46

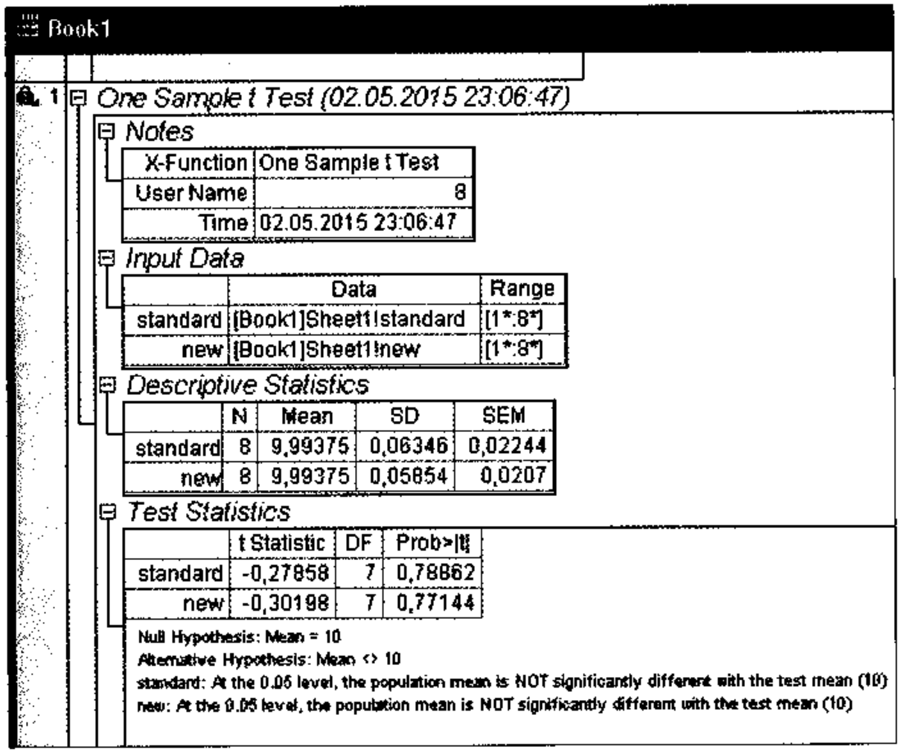


Рис. 47

Треба перевірити, чи знизився рівень АСТ (у нормі верхня його межа становить 47,0 Од/л) у результаті лікування пневмонії.

Розв'язання. Перевірка даних на нормальність за тестом Шапіро—Уілка підтвердила, що обидві групи відповідають нормальному розподілу. Оскільки вибірки не є незалежними (дослідження динаміки активності ензиму в крові одних і тих самих пацієнтів у процесі лікування), застосуємо критерій Стьюдента для залежних вибірок. Для цього виберемо в меню послідовно: **Statistics—Hypothesis Testing—Pair-Sample t-Test—Open Dialog**. У діалоговому вікні **Statistics\Hypothesis Testing: PairSampletTest** вкажемо вхідний діапазон даних (**Input: 1st Data Range, 2st Data Range**), нульову гіпотезу залишимо без змін та оберемо альтернативну гіпотезу (у цьому разі оберемо ненаправлену: **Mean1—Mean2 < > 0**) (рис. 48).

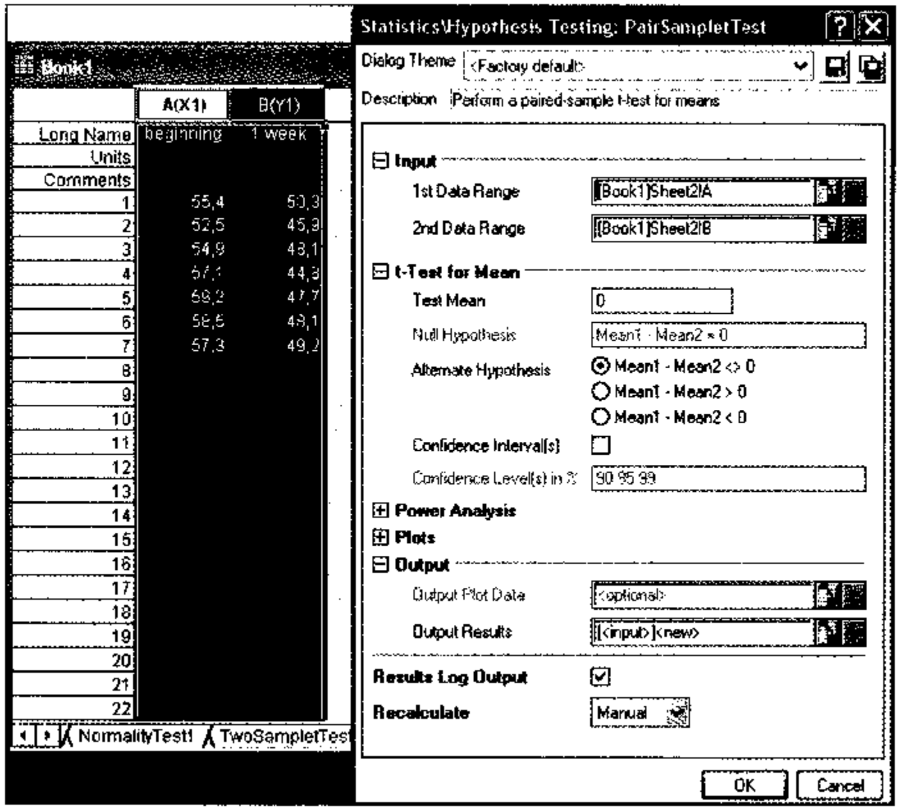


Рис. 48

У вікні результатів побачимо (рис. 49), що показник ймовірності (стовпчик Prob>|t|) становить $1{,}56 \cdot 10^{-4}$, тобто є значно нижчим від рівня $\alpha = 0{,}05$. Отже, результати тесту дають підставу відхилити статистичну гіпотезу про рівність генеральних середніх. *Висновок*: у процесі лікування хворих на запалення легенів рівень ензиматичної активності АСТ знизився.

Однофакторний дисперсійний аналіз. У випадку, коли потрібно порівняти більш як дві (три і більше) групи, що відрізняються між собою за кількісною ознакою, застосовують однофакторний дисперсійний аналіз **(One-Way analysis of variance—One-Way ANOVA)** (рис. 50). Зокрема цей тип статистичного аналізу використовують, коли необхідно встановити відмінності активності ензиму NO-синтази залежно від стадії розвитку запалення тканин, або коли по-

трібно дослідити, чи відрізняються за ефективністю дії однотипні фармакологічні препарати за антигіпертензивної терапії, тощо.

Нагадаємо, що **фактор** — це чинник, який має впливати на результат експерименту, **рівні фактора** — це значення, яких набуває фактор (наприклад, концентрації речовини, стадії захворювання). Залежно від кількості факторів розрізняють однофакторний **(One-Way ANOVA)**, двофакторний **(Two-Way ANOVA)** та багатофакторний дисперсійний аналізи. Наприклад, дисперсійний аналіз застосовують, щоб перевірити, чи залежить активність ферменту протеїнкінази С у клітинах печінки від стадії захворювання на гепатит (тут враховуємо один фактор — стадія захворювання), від стадії захворювання і віку пацієнтів (два фактори — стадія захворювання і вік), від стадії захворювання, віку пацієнтів і методів терапії (три фактори — стадія захворювання, вік та спосіб лікування).

Обов'язковими умовами проведення параметричного дисперсійного аналізу є: по-перше, порівняння стосується кількісного розподіленого показника, причому доведено його нормальний роз-

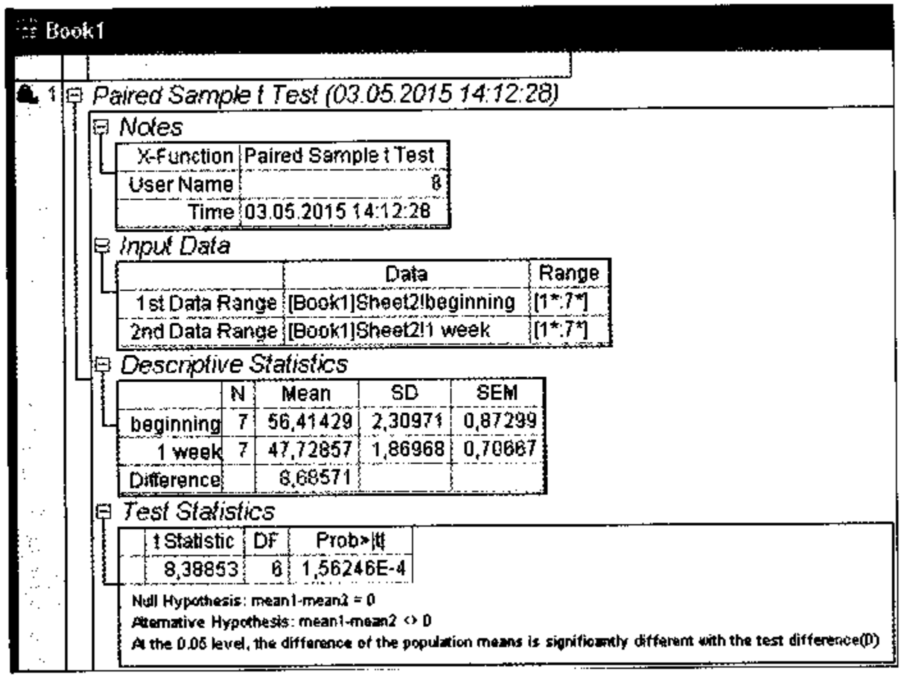


Рис. 49

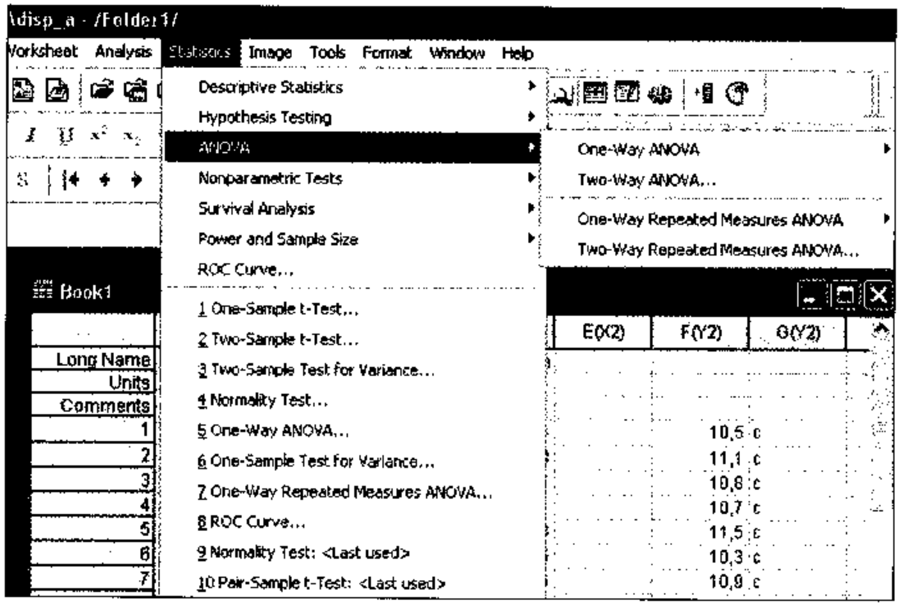


Рис. 50

поділ для всіх груп даних; по-друге, доведена рівність дисперсій усіх груп даних.

Параметричний однофакторний дисперсійний аналіз проводять за такими етапами:

1) перевіряють гіпотези про нормальний розподіл кожної групи даних;

2) перевіряють гіпотези про однаковість дисперсій;

3) проводять, власне, аналіз варіацій;

4) виконують апостеріорний аналіз (попарне порівняння груп даних спеціальними тестами).

Приклад 78. У лабораторії досліджували закономірності участі оксиду азоту (NO) в регуляції тонусу (вираженому в мН) аорти щурів. Відомо, що у тканині стінки аорти є фермент синтаза оксиду азоту, яка синтезує NO з амінокислоти *L*-аргініну. Було використано препарат-донор оксиду азоту нітрогліцерин і блокатор NO-синтази сполуку L-NAME (обидві речовини застосовували у концентрації 100 мкМ). Постановка експерименту передбачала формування 3 груп щурів (по 9 тварин у кожній): контрольна група тварин, в яких вимірювали тонус препаратів аорти, що омивався фізіологічним розчином; група щурів, препарати аорти яких

упродовж 30 хв витримували у розчині, що містив L-NAME і потім реєстрували тонус м'язових препаратів; третя група щурів, препарати аорти яких упродовж 30 хв витримували у фізіологічному розчині з додаванням нітрогліцерину, після чого реєстрували тонус м'язів. У результаті отримано такі дані:

Контроль	10,5	11,1	10,8	10,7	11,5	10,3	10,9	11,2	11,0
L-NAME	11,8	12,5	12,9	11,7	12,4	12,7	11,9	12,6	12,8
Нітрогліцерин	9,4	8,8	8,7	9,1	9,3	8,5	8,6	9,0	9,2

Проаналізуємо ці результати і спробуємо дати відповіді на такі запитання: чи впливає NO на тонус аорти щурів? Чи можна стверджувати, що в нормі NO синтезується у тканині аорти? Чи можна вважати нітрогліцерин ефективним препаратом щодо тонусу аорти?

Розв'язання. Насамперед визначимося з блоком статистичних критеріїв, які можна застосовувати до даних цього експерименту. По-перше, незалежно один від одного перевіримо дані трьох груп (контроль, L-NAME та нітрогліцерин) щодо їх нормального розподілу. Оскільки у цьому випадку маємо невеликі вибірки, скористаємось критерієм Шапіро—Уілка.

За результатами тесту Шапіро—Уілка приймаємо нульову гіпотезу: в усіх випадках значення колонки (Prob < W) перевищують критичний рівень α (також під таблицею результатів тесту Шапіро—Уілка містяться висновки про те, що дані кожної з вибірок вірогідно належать до нормально розподілених генеральних сукупностей).

Далі із застосуванням критерію Левена потрібно перевірити гіпотези про рівність дисперсій. Варто зазначити, що в "Origin" усі наступні етапи дисперсійного аналізу — від тесту на рівність дисперсій до апостеріорного порівняння даних — задаються одночасно, але дослідник має враховувати результати аналізу в потрібній, коректній послідовності. Отже, виберемо в меню послідовно: **Statistics—Hypothesis Testing—ANOVA—One-Way ANOVA—Open Dialog**. У діалоговому вікні **ANOVAOne-Way** вкажемо вхідний діапазон даних (**Input Data: Factor** — посилання на згруповану колонку з кодами і **Data** — посилання на колонку зі зведеними числовими даними), критичний рівень — **Significance Level** (0,05) залишимо без змін, оберемо тест для апостеріорного аналізу **Means Comparison** (наприклад, тест Тьюкі — **Tukey**) та вкажемо необхідний тест на рівність дисперсій **Test for Equal Variance (Levene)** (рис. 51).

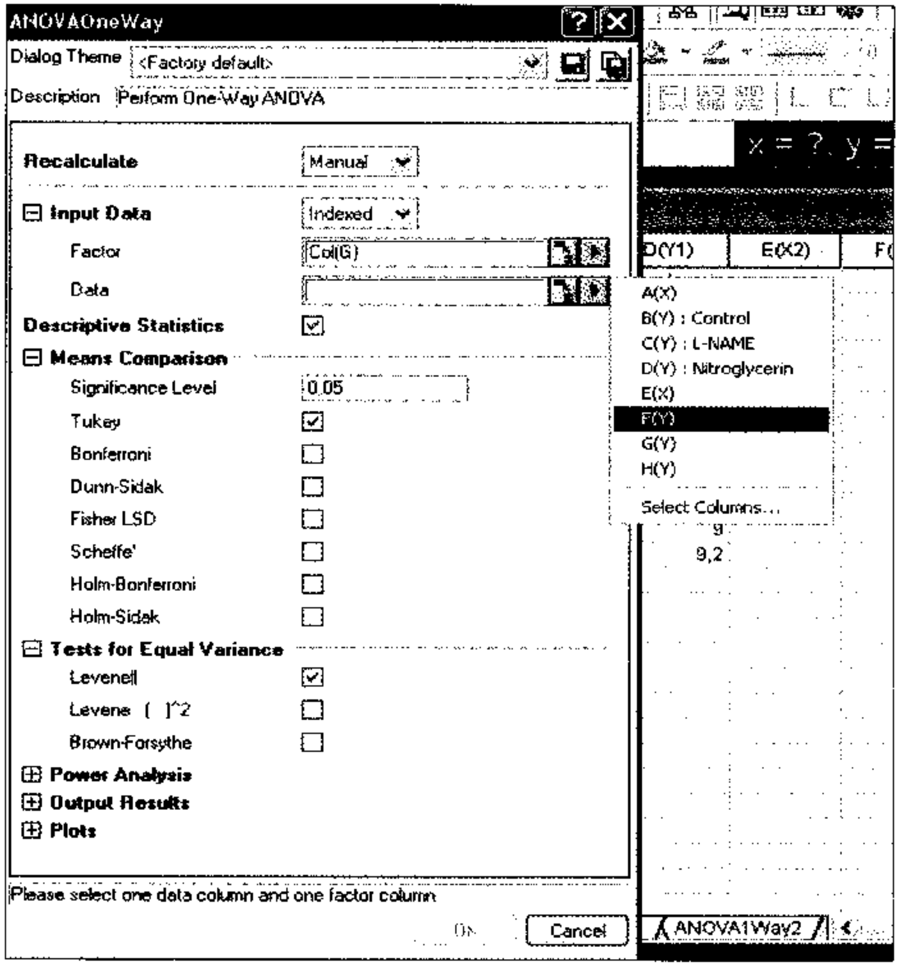


Рис. 51

В аркуші результатів дисперсійного аналізу, який автоматично додав пакет "Origin", бачимо результати обраних тестів (рис. 52). Тест Левена підтвердив однаковість дисперсій (показник **(Prob > F)** становить близько 0,44, тобто понад 0,05), цим виконана остання необхідна умова для використання параметричного дисперсійного аналізу.

Власне аналіз варіацій **(Overall ANOVA)** встановив вплив фактора, тобто між груповими середніми показниками у масиві да-

них існує відмінність (значення колонки **(Prob > F)** становить близько $3{,}8 \cdot 10^{-15}$, тобто менш як 0,05) і тому маємо підстави розглянути результати апостеріорного аналізу, щоб з'ясувати, які саме групи відрізняються. Як видно з результатів тесту Тьюкі **(Tukey test)** (колонка **Prob)**, підтверджено відмінності між усіма груповими середніми.

Пакет "Origin" містить засоби для здійснення ***кореляційного і регресійного аналізів***. Параметричний кореляційний аналіз (Пірсона) застосовують для дослідження зв'язку нормально розподілених кількісних ознак (рис. 53). Непараметричні методи кореляційного аналізу (критерії Спірмена, Кендалла, гамма) використовують або незалежно від характеру розподілу даних, або для пошуку зв'язку між кількісним і якісним показниками, або у випадку двох порядкових показників.

	Sample Size	Mean	Standard Deviation	SE of Mean
c	9	10,88889	0,36553	0,12184
l	9	12,36667	0,45277	0,15092
n	9	8,95556	0,32059	0,10686

One Way ANOVA

Overall ANOVA

	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	52,67185	26,33593	178,99811	3,77476E-15
Error	24	3,53111	0,14713		
Total	26	56,20296			

Null Hypothesis: The means of all levels are equal
Alternative Hypothesis: The means of one or more levels are different
At the 0.05 level, the population means are significantly different.

Fit Statistics

R-Square	Coeff Var	Root MSE	Data Mean
0,93717	0,03572	0,38357	10,73704

Means Comparisons

Tukey Test

	MeanDiff	SEM	q Value	Prob	Alpha	Sig	LCL	UCL
l c	1,47778	0,18082	11,55794	0	0,05	1	1,02622	1,92933
n c	-1,93333	0,18082	15,12091	0	0,05	1	-2,38489	-1,48178
n l	-3,41111	0,18082	26,67865	0	0,05	1	-3,86267	-2,95955

Sig equals 1 indicates that the means difference is significant at the 0,05 level.
Sig equals 0 indicates that the means difference is not significant at the 0,05 level.

Homogeneity of Variance Test

Levene's Test(Absolute Deviations)

	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	0,06325	0,03162	0,86039	0,43564
Error	24	0,88211	0,03675		

At the 0.05 level, the population variations are not significantly different.

Рис. 52

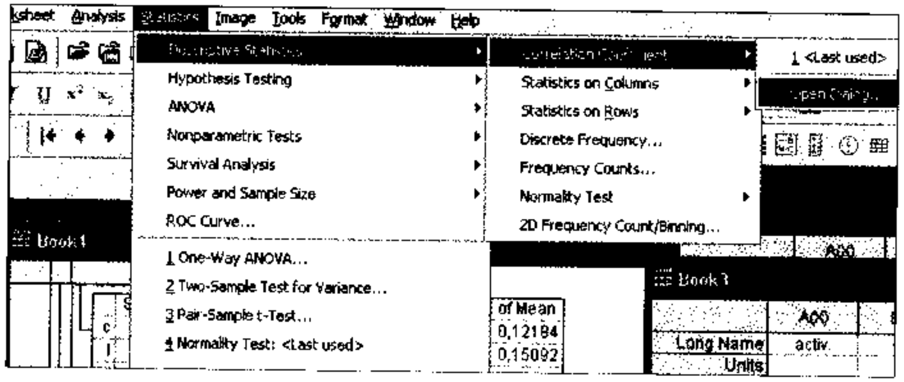


Рис. 53

Приклад 79. У лабораторії проводили дослідження властивостей транспорту іонів Ca^{2+} в ендоплазматичному ретикулумі гепатоцитів щура. Досліджено 9 проб печінки тварин, причому одночасно реєстрували два параметри: активність Ca^{2+}-АТФази мембран ендоплазматичного ретикулума (ензим, який здійснює енергозалежне транспортування іонів Ca^{2+} із цитоплазми у ретикулум і сприяє його накопиченню там) і концентрацію іонів Ca^{2+} в цих органелах.

Активність Ca^{2+}-АТФази	1,08	1,1	1,12	1,14	1,15	1,25	1,36	1,38	1,4
Концентрація Ca^{2+}, мкМ	257	264	279	287	311	324	340	353	364

Проаналізуємо отримані результати і спробуємо відповісти на такі запитання: чи існує лінійний зв'язок між показниками активності Ca^{2+}-АТФази ендоплазматичного ретикулума і внутрішньоретикулярної концентрації іонів Ca^{2+}? Яким рівнянням, за наявності лінійного зв'язку, можна описати залежність між активністю ензиму і концентрацією Ca^{2+}?

Розв'язання. На початку аналізу визначимося з блоком статистичних критеріїв, які можна застосовувати до даних цього експерименту. Оскільки необхідно перевірити результати експерименту на наявність лінійного зв'язку, а отримані в експерименті дані є кількісними інтервальними величинами, застосуємо параметричний критерій Пірсона (див. рис. 53). Виділимо обидві колонки з даними концентрації іонів кальцію та ензиматичної

активності, а потім скористаємось меню: **Statistics—Descriptive Statistics—Correlation Coefficient**. У вікні **Statistics\Descriptive Statistics: corrcoef** (рис. 54) оберемо тип кореляції Пірсона.

Як видно з результатів дослідження (рис. 55), між досліджуваними показниками існує лінійний зв'язок ($r = 0,96352$, $p \approx 2,9 \cdot 10^{-5}$).

Тепер, оскільки наявність лінійного зв'язку доведено, дамо відповідь на друге запитання, встановивши рівняння, яке пов'язує активність Ca^{2+}-АТФази ендоплазматичного ретикулума і внутрішньоретикулярну концентрацію іонів Ca^{2+}. Для цього побудуємо графік у координатах відповідних величин (рис. 56) та виконаємо регресійний аналіз, скориставшись меню: **Analysis—Fitting—Fit Linear** (рис. 57).

Внаслідок виконаної процедури у вікні графіка з'явились лінія регресії і таблиця зі зведеними результатами регресійного аналізу (рис. 58). Як бачимо, параметр a (Intercept) становить –47,1, тоді як b (Slope) дорівнює 291,7.

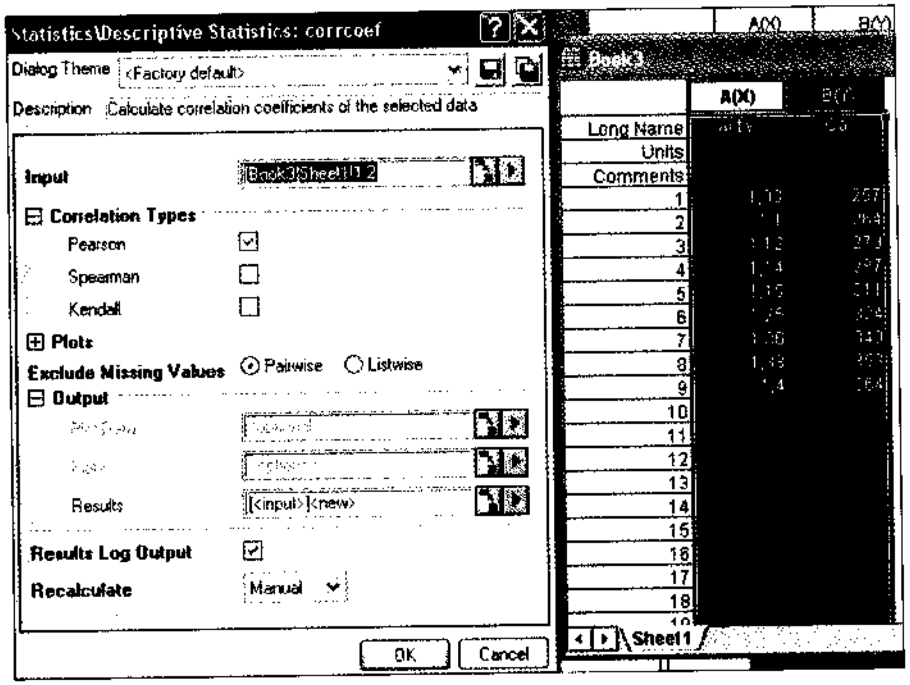
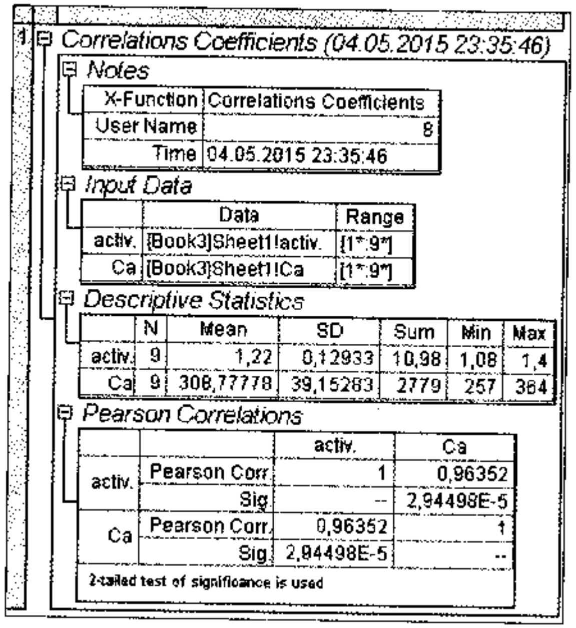


Рис. 54



Correlations Coefficients (04.05.2015 23:35:46)

Notes

X-Function	Correlations Coefficients
User Name	8
Time	04.05.2015 23:35:46

Input Data

	Data	Range
activ.	[Book3]Sheet1!activ.	[1*:9*]
Ca	[Book3]Sheet1!Ca	[1*:9*]

Descriptive Statistics

	N	Mean	SD	Sum	Min	Max
activ.	9	1,22	0,12933	10,98	1,08	1,4
Ca	9	308,77778	39,15283	2779	257	364

Pearson Correlations

		activ.	Ca
activ.	Pearson Corr.	1	0,96352
	Sig	--	2,94498E-5
Ca	Pearson Corr.	0,96352	1
	Sig	2,94498E-5	--

2-tailed test of significance is used

Рис. 55

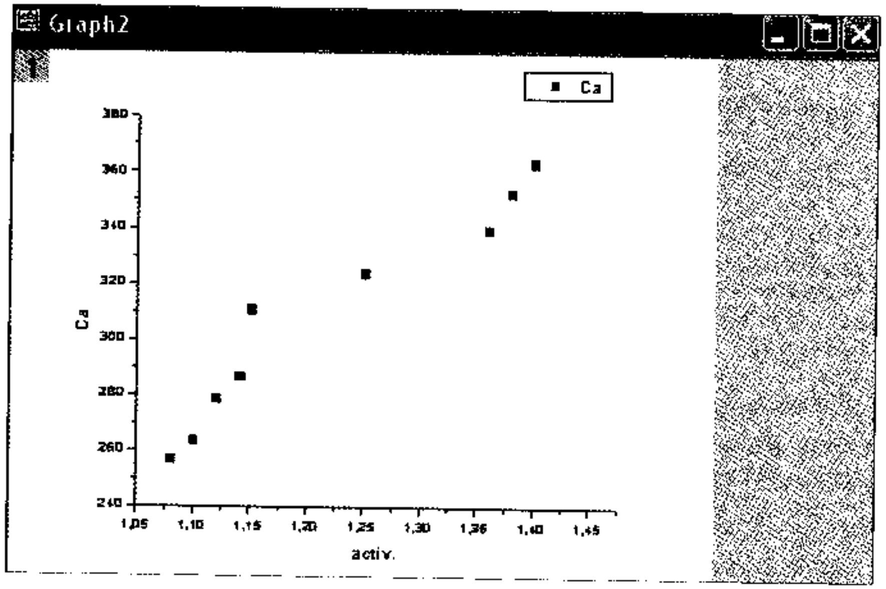


Рис. 56

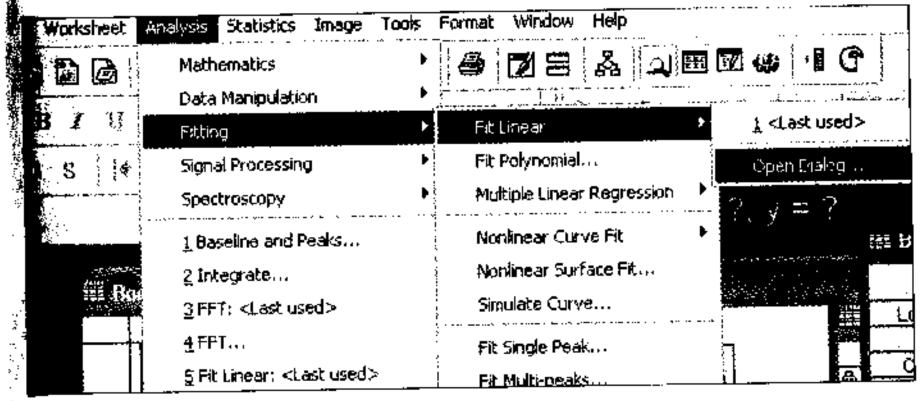


Рис. 57

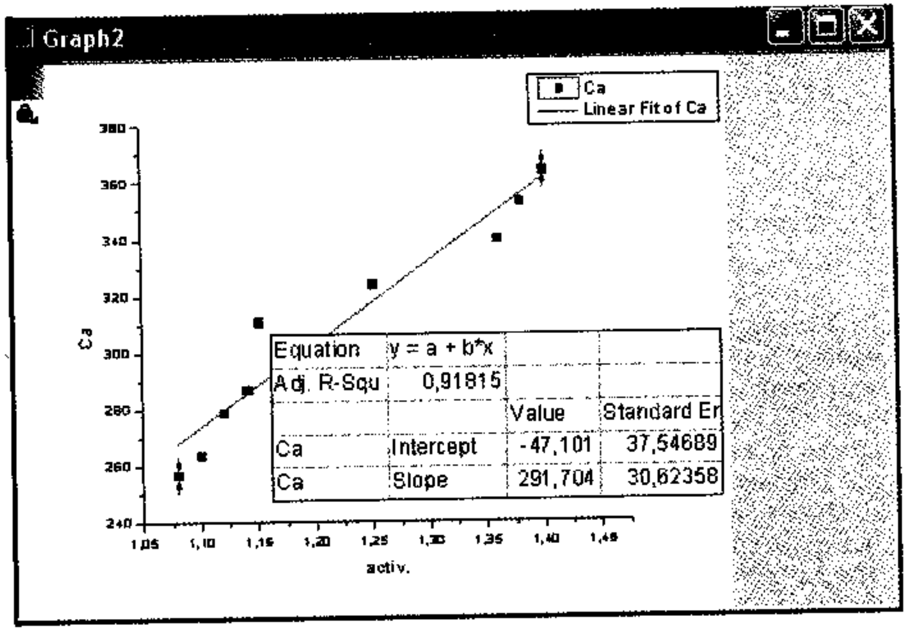
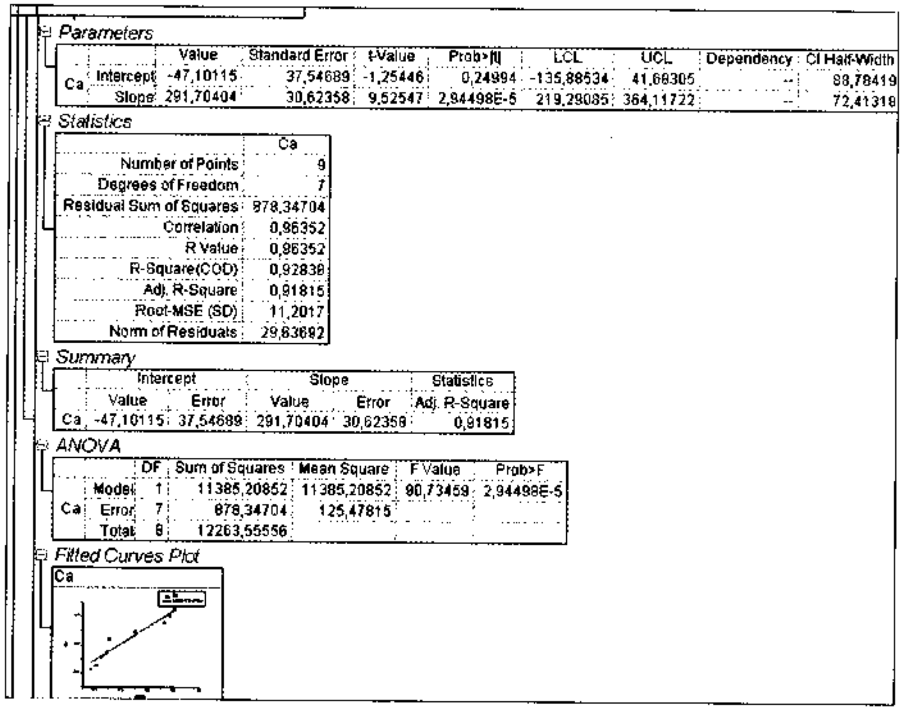


Рис. 58

Також унаслідок виконаної процедури до робочої книги (Book) було автоматично додано аркуш **FitLinear1**, який містить зведені дані проведеного регресійного аналізу, зокрема дані щодо статистичної значущості відповідних показників (рис. 59).



Parameters

		Value	Standard Error	t-Value	Prob>|t|	LCL	UCL	Dependency	CI Half-Width
Ca	Intercept	-47,10115	37,54689	-1,25446	0,24994	-135,88534	41,68305	--	88,78419
	Slope	291,70404	30,62358	9,52547	2,94498E-5	219,29085	364,11722	--	72,41318

Statistics

	Ca
Number of Points	9
Degrees of Freedom	7
Residual Sum of Squares	878,34704
Correlation	0,96352
R Value	0,96352
R-Square(COD)	0,92838
Adj. R-Square	0,91815
Root-MSE (SD)	11,2017
Norm of Residuals	29,63692

Summary

	Intercept		Slope		Statistics
	Value	Error	Value	Error	Adj. R-Square
Ca	-47,10115	37,54689	291,70404	30,62358	0,91815

ANOVA

		DF	Sum of Squares	Mean Square	F Value	Prob>F
Ca	Model	1	11385,20852	11385,20852	90,73459	2,94498E-5
	Error	7	878,34704	125,47815		
	Total	8	12263,55556			

Fitted Curves Plot

Ca

Рис. 59

Якість регресійної моделі можна поліпшити (наприклад, усунувши проблему зі значущістю параметра Intercept у рівнянні) вилученням із розрахунків точок, які значно відхиляються (рис. 60). Для цього слугують засоби маскування обраних точок. Отже, виключивши ці точки з аналізу, ми значно поліпшимо побудовану модель (рис. 61, 62).

Непараметричний аналіз. У випадку, коли результати експериментів представлені якісними даними, або при перевірці кількісних даних щодо їх нормального розподілу нульову гіпотезу було відхилено хоча б для однієї групи даних, статистичний аналіз необхідно проводити з використанням непараметричних критеріїв (рис. 63). Зокрема, непараметричними аналогами критерію Стьюдента для залежних вибірок є критерій Уілкоксона та критерій знаків; для незалежних вибірок найчастіше застосовують критерій Манна—Уітні або критерій Колмогорова—Смірнова. Коли потрібно провести дисперсійний аналіз, але дані унеможливлюють використання параметричних засобів, застосовують їх не-

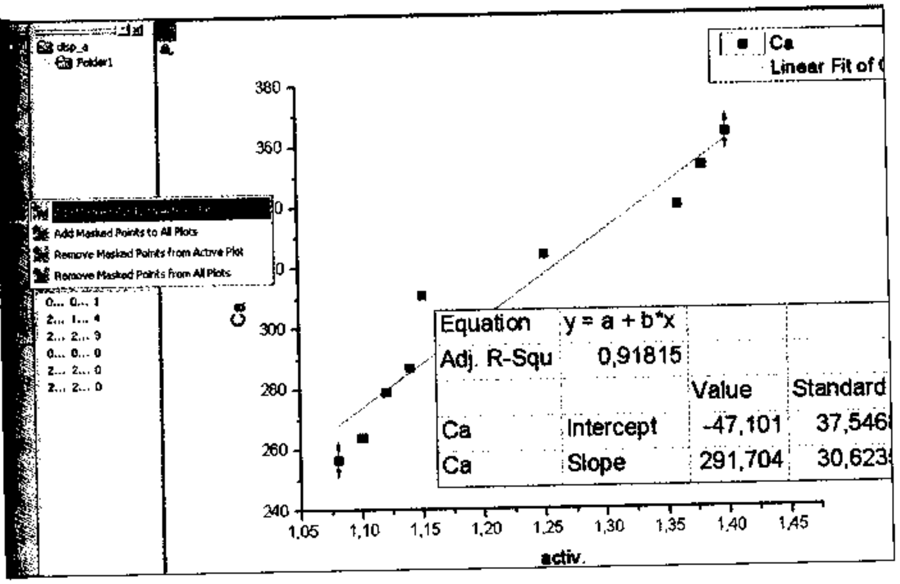


Рис. 60

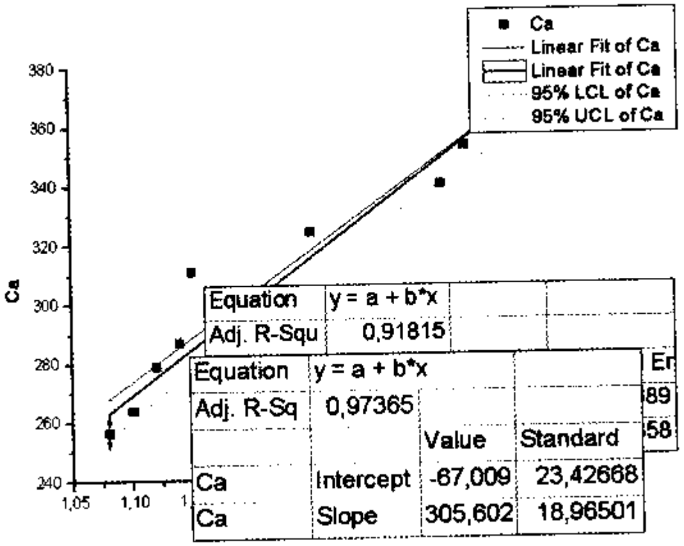


Рис. 61

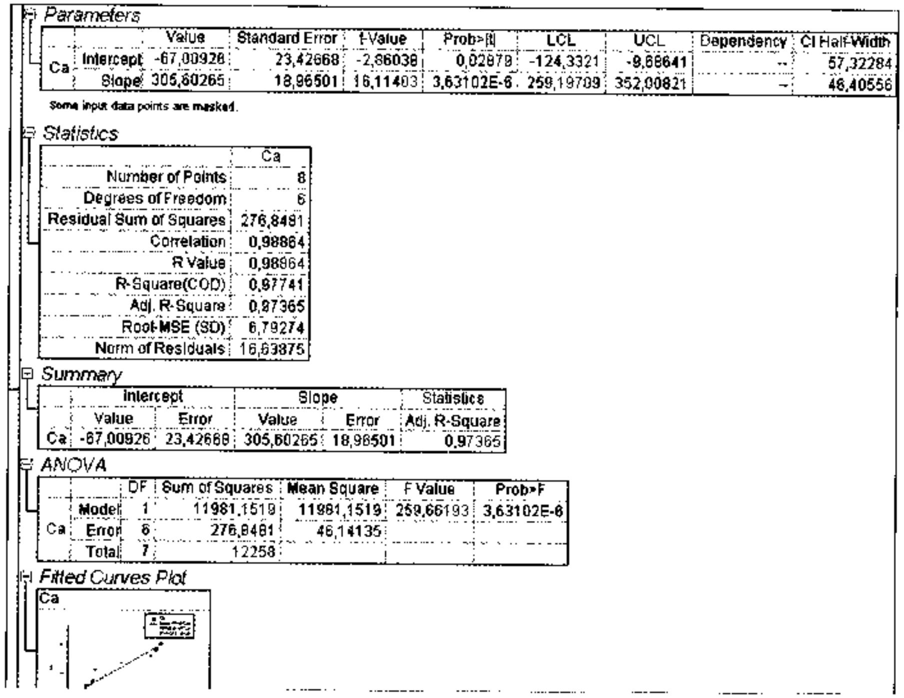


Рис. 62

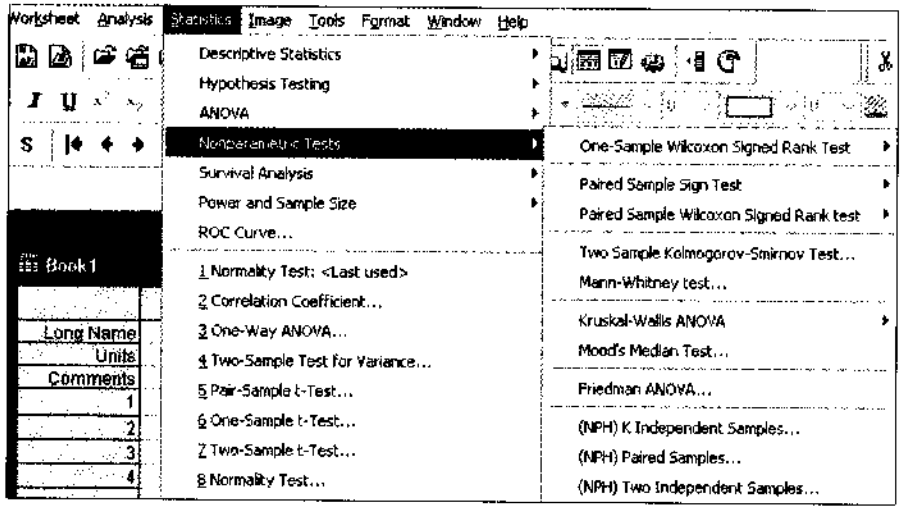


Рис. 63

параметричні аналоги. Також широко застосовують непараметричні показники кореляції (Спірмена, Кендалла, гамма).

Непараметричні тести для порівняння двох залежних вибірок. У випадку, коли хоча б одна з двох вибірок не задовольняє умову нормального розподілу, або коли дані є якісними, використовують тест Уілкоксона і тест знаків (останній застосовують для якісних даних, коли проведення експерименту дає змогу встановити лише напрямок зміни досліджуваного параметра).

Нульовою гіпотезою тестів є: вибірки належать до однієї генеральної сукупності або двох генеральних сукупностей з однаковими параметрами.

Альтернативна гіпотеза: вибірки взяті з генеральних сукупностей, параметри яких відрізняються.

Приклад 80. Треба перевірити ефективність спеціального тренінгу на поліпшення пам'яті у студентів. Для цього обрано 10 студентів-добровольців, яких протестували на запам'ятовування числових даних до тренінгу, а потім тестом аналогічної складності — після тренінгу. Отримано такі дані (бали):

До тренінгу	85	90	92	93	93	94	95	95	140	150
Після тренінгу	86	95	96	93	95	96	98	101	133	155

Необхідно проаналізувати отримані результати й відповісти на запитання: як змінився рівень пам'яті у студентів після проведення цього тренінгу?

Розв'язання. На початку аналізу визначимося з блоком статистичних критеріїв, які можна застосовувати до наведених даних.

Для обох вибірок при аналізі тестом Шапіро—Уілка відхилено гіпотезу щодо нормального розподілу генеральних сукупностей, з яких їх було відібрано. Отже, хоча дві вибірки отримано при проведенні експерименту, який сплановано як парний (пари значень "до тренінгу"—"після тренінгу"), ми не можемо застосовувати параметричний критерій Стьюдента для залежних даних. Тому варто скористатись непараметричним ранговим критерієм Уілкоксона. Для цього виділимо обидві колонки з даними та скористаємось меню: **Statistics—Nonparametric Tests—Paired Sample Wilcoxon Signed Rank test** (див. рис. 63). У вікні **Statistics\Nonparametric Tests: signrank2** (рис. 64) переконаємось, що вхідні дані **Input (1st Data Range, 2nd Data Range)** вказані правильно (колонки з даними вибірок). У цьому разі можна обрати на-

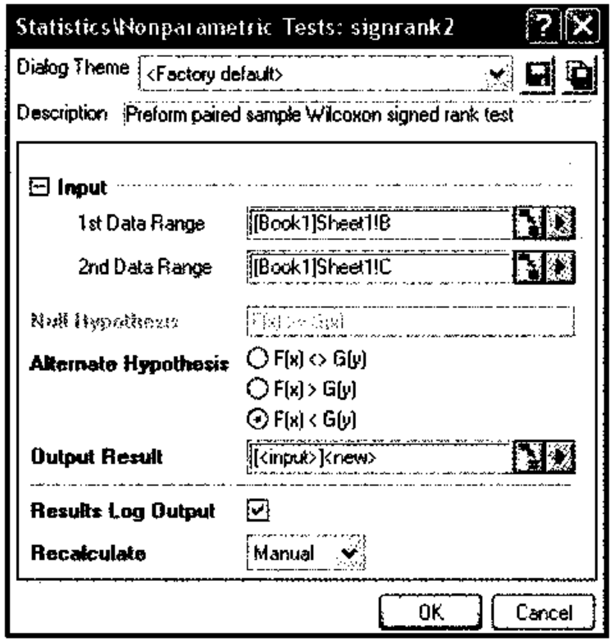


Рис. 64

правлену альтернативну гіпотезу (**Alternative Hypothesis: F(x)<G(y)**), адже за умовою постановки експерименту тренінг мав на меті поліпшити показники пам'яті студентів.

Як видно з результатів застосування критерію Уілкоксона (рис. 65), проведення тренінгу не привело до поліпшення пам'яті і цей тренінг не можна вважати ефективним.

Приклад 81. Досліджували вплив полівітамінних харчових добавок на стресостійкість у групі студентів-добровольців. Одразу після написання модульної контрольної роботи проведено опитування, в якому студенти могли оцінити ступінь стресу в межах від 0 (повністю спокійний) до 3 (сильне хвилювання). Упродовж наступного місяця студенти цієї групи вживали полівітамінні препарати, після чого при написанні наступного модуля знову оцінювали ступінь стресу за тією ж шкалою. Результати опитувань наведені в таблиці:

Без полівітамінів	0	2	3	1	1	2	2	2	3	0
З полівітамінами	1	2	2	0	0	1	1	1	1	0

Потрібно проаналізувати отримані результати та відповісти на запитання: як змінився рівень стресостійкості студентів після вживання полівітамінних харчових добавок (інакше, чи можна рекомендувати полівітамінні засоби як такі, що знижують рівень стресу)?

Розв'язання. На початку аналізу визначимося зі статистичним критерієм, який можна застосувати до наведених даних. Оскільки ці дані є залежними і якісними, ми можемо лише встановити напрямок зміни показника рівня стресу і для їх порівняння оберемо критерій знаків. Для цього виділимо обидві колонки з даними та скористаємось меню: **Statistics—Nonparametric Tests—Paired Sample Sign Test** (рис. 66).

У вікні **Statistics\Nonparametric Tests: sign2** (рис. 67) переконаємось, що вхідні дані **Input (1st Data Range, 2nd Data Range)** вка-

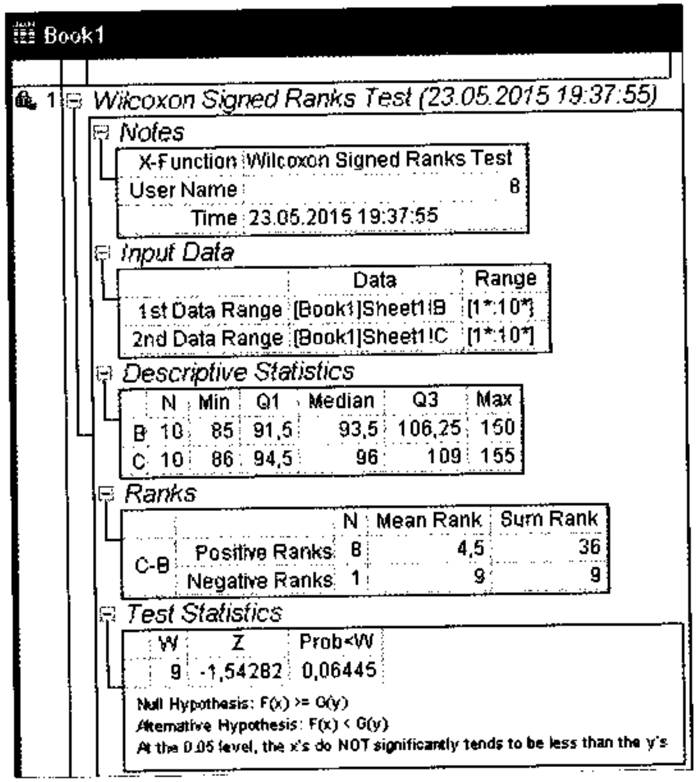


Рис. 65

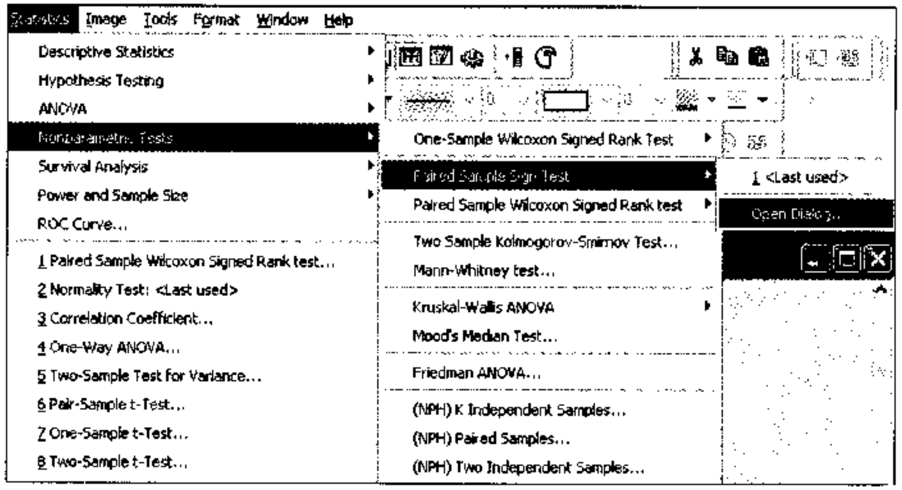


Рис. 66

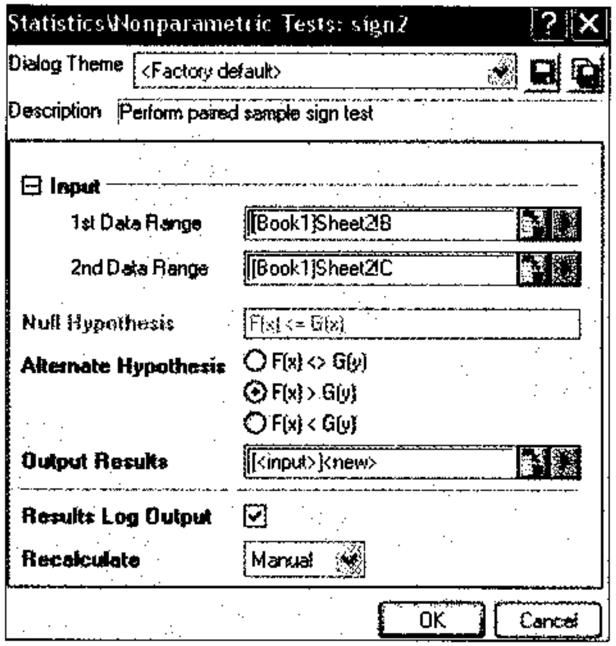


Рис. 67

зані правильно (колонки з даними вибірок). У цьому випадку, як і в попередньому, можемо обрати направлену альтернативну гіпотезу (**Alternative Hypothesis**: але цього разу — **F(x)>G(y)**), адже за умовою постановки експерименту вживання вітамінів мало за мету знизити рівень стресу студентів.

Як видно з результатів застосування критерію знаків (рис. 68), вживання мультивітамінних харчових добавок привело до зниження рівня суб'єктивного відчуття стресу у студентів, тому їх використання можна вважати ефективним.

Порівняння двох незалежних вибірок непараметричними методами. В OriginPro8 непараметричні методи для порівняння двох незалежних вибірок представлені тестами Манна—Уітні та Колмогорова—Смірнова. В обох випадках критерії перевіряють гіпотези про рівність параметрів генеральних сукупностей, з яких отримано вибірки.

Нульова гіпотеза: вибірки належать до однієї генеральної сукупності або двох генеральних сукупностей з однаковими параметрами.

Альтернативна гіпотеза: вибірки взяті з генеральних сукупностей, параметри яких відрізняються.

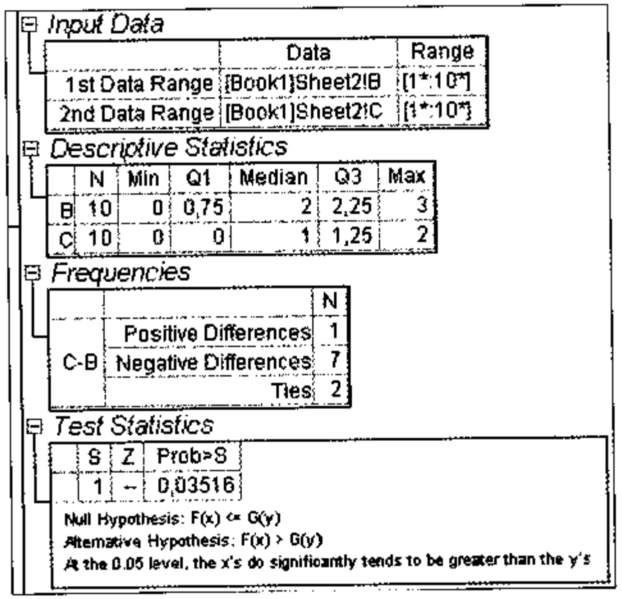


Рис. 68

Приклад 82. Досліджували швидкість ферментативного гідролізу АТФ міозиновою АТФазою із серцевого та скелетного м'язів. Отримані дані наведено в таблиці:

Серцевий м'яз	1,56	1,48	1,61	1,42	1,14	1,38	1,35	1,44	1,5
Скелетний м'яз	1,87	1,11	1,82	1,59	1,15	1,75	1,19	1,67	1,12

Потрібно встановити, чи відрізняється ензиматична активність міозину в м'язах різного типу.

Розв'язання. Використання тесту Шапіро—Уілка дало підставу відхилити гіпотезу щодо нормального розподілу групи, з якої було обрано вибірку ензиматичної активності скелетних м'язів. Отже, статистичний аналіз потрібно проводити непараметричними методами. Хоча вибірки однакового розміру, вони є незалежними. Застосуємо критерій Манна—Уітні. Попередньо згрупуємо дані та скористаємось меню: **Statistics—Nonparametric Tests—Mann-Whitney Test** (рис. 69).

У вікні **Statistics\Nonparametric Tests: mwtest** (рис. 70) переконаємось, що вхідні дані **Input** (**Group Range** — колонка з групувальними кодами, **Data Range** — колонка зі згрупованими даними) вказані правильно. У цьому разі можемо скористатись ненаправленою альтернативною гіпотезою (**Alternative Hypothesis: F(x)<>G(y)**).

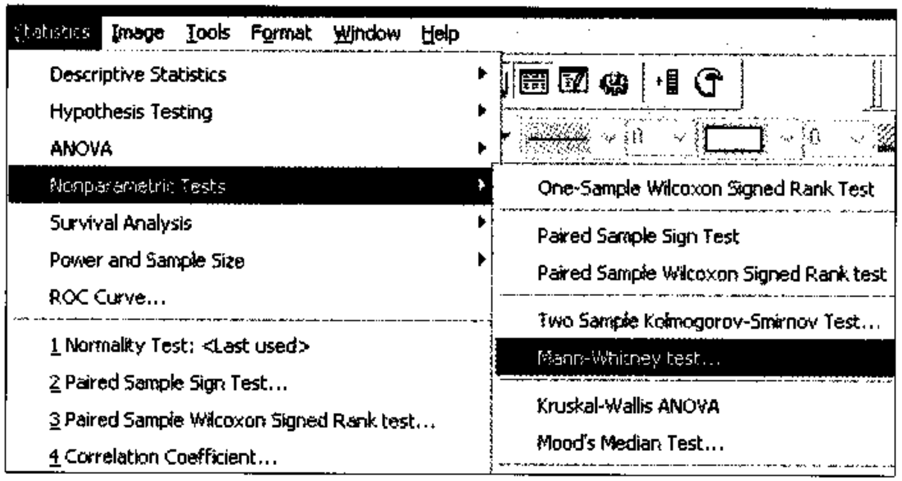


Рис. 69

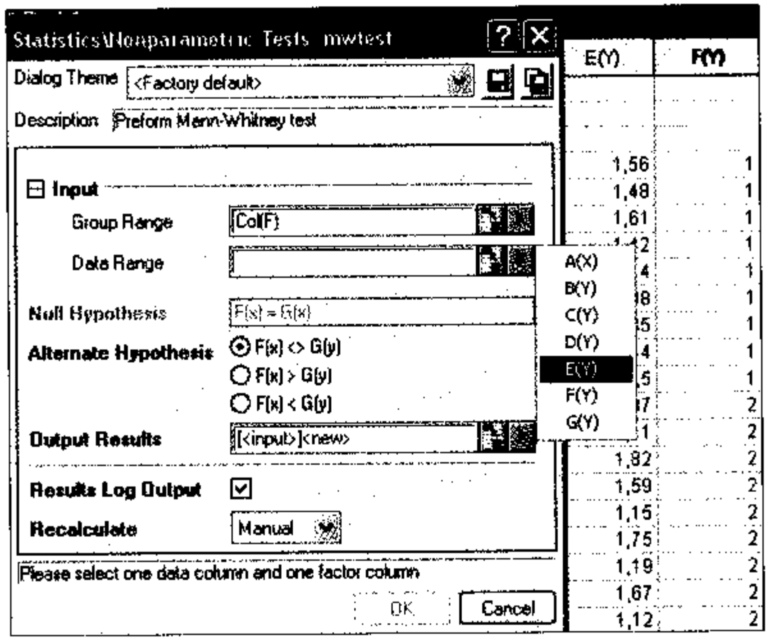


Рис. 70

Як видно з результатів застосування критерію Манна—Уітні (рис. 71), приймаємо нульову гіпотезу щодо гомогенності параметрів розподілів генеральних сукупностей, з яких було обрано вибірки. Отже, відмінність між ензиматичною активністю міозину серцевого і скелетного м'язів відсутня.

Непараметричний однофакторний дисперсійний аналіз. Коли потрібно провести однофакторний дисперсійний аналіз, але дані є якісними, або за наявності кількісних даних не дотримуються умови застосування параметричних критеріїв (хоча б для однієї з груп, які потрібно порівняти, відхилено гіпотезу щодо нормальності розподілу або встановлено нерівність внутрішньогрупових дисперсій), використовують непараметричні критерії. У разі, коли вибірки незалежні, застосовують критерій Краскела—Уолліса, коли залежні — критерій Фрідмена. Перевіряють статистичні гіпотези:

- нульова гіпотеза — кожна група має однаковий розподіл величин у генеральній сукупності;
- альтернативна гіпотеза — групи відрізняються за розподілом величин у генеральній сукупності.

Mann-Whitney Test (06.06.2015 19:34:24)

Notes

X-Function	Mann-Whitney Test
User Name	8
Time	06.06.2015 19:34:24

Input Data

	Data	Range
Group Range	[Book1]Sheet3!F	[1*:18*]
Data Range	[Book1]Sheet3!E	[1*:18*]

Descriptive Statistics

	N	Min	Q1	Median	Q3	Max
1	9	1,14	1,365	1,44	1,53	1,61
2	9	1,11	1,135	1,59	1,785	1,87

Ranks

	N	Mean Rank	Sum Rank
1	9	8,88889	80
2	9	10,11111	91

Test Statistics

U	Z	Prob>|U|
35	-0,44151	0,65884

Null Hypothesis: F(x) = G(y)
Alternative Hypothesis: F(x) <> G(y)
At the 0.05 level, the two distributions are NOT significantly different.

Рис. 71

Приклад 83. Досліджували спадкову схильність до набирання маси тіла лабораторними щурами. Із трьох ліній щурів — L, S і R — відбирали по 8 самців 10-місячного віку із подібною масою тіла й упродовж одного місяця утримували на харчовому раціоні з підвищеним вмістом жирів. Після закінчення експерименту тварин зважували. Встановлено такий відносний приріст маси тіла тварин (г):

Лінія L	108	96	99	104	119	89	75	83
Лінія S	77	58	49	91	51	78	56	25
Лінія R	64	67	65	54	56	53	61	58

Потрібно перевірити, чи існує генетична схильність щурів до набирання маси тіла.

Розв'язання. Перевіримо дані щодо нормального закону розподілу і гомогенності дисперсій сукупностей, з яких їх відібрано. Тест Шапіро—Уілка підтвердив нормальний розподіл усіх груп, але тест Левена встановив неоднорідність дисперсій набирання маси тіла лініями щурів (рис. 72). Отже, на наступному етапі статистичного аналізу слід врахувати непараметричні методи. Оскільки групи незалежні, застосуємо тест Краскела—Уолліса. Для цього скористаємось меню **Statistics—Nonparametric Tests—Kruskal-Wallis ANOVA** (див. рис. 63).

У вікні **Statistics\Nonparametric Tests: kwanova** (рис. 73) вкажемо вхідні дані **Input** (**Group Range** — колонка з групувальними кодами, **Data Range** — колонка зі згрупованими даними).

Згідно з результатами дисперсійного аналізу (рис. 74), нульову гіпотезу щодо гомогенності параметрів розподілів генеральних

ANOVAOneWay (07.06.2015 16:34:10)

Notes

Input Data

Descriptive Statistics

	Sample Size	Mean	Standard Deviation	SE of Mean
l	8	96,625	14,1516	5,00335
s	8	60,625	20,76355	7,34102
r	8	59,75	5,28475	1,86844

One Way ANOVA

Overall ANOVA

	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	7084,08333	3542,04167	16,11676	5,73269E-5
Error	21	4615,25	219,77381		
Total	23	11699,33333			

Null Hypothesis: The means of all levels are equal
Alternative Hypothesis: The means of one or more levels are different
At the 0.05 level, the population means are significantly different.

Fit Statistics

R-Square	Coeff Var	Root MSE	Data Mean
0,60551	0,20495	14,82477	72,33333

Homogeneity of Variance Test

Levene's Test(Absolute Deviations)

	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	533,85938	266,92969	3,8629	0,03728
Error	21	1451,11719	69,10082		

At the 0.05 level, the population variations are significantly different.

Рис. 72

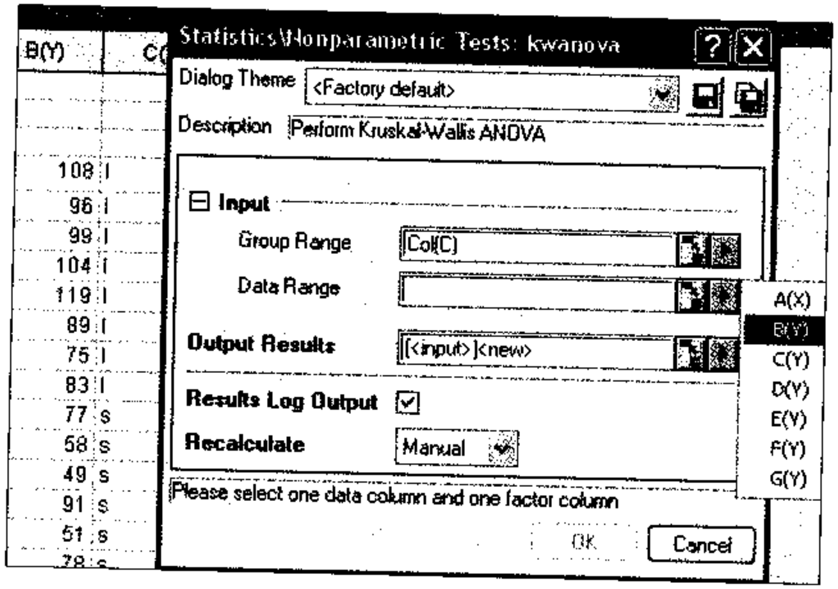


Рис. 73

сукупностей, з яких було обрано вибірки, відхиляємо. Отже, в експерименті встановлено існування відмінності між мірою збільшення маси тіла у різних генетичних лініях щурів.

Встановивши відмінність між параметрами розподілів груп, надалі маємо виявити, які саме групи між собою відрізняються. Для цього треба попарно апостеріорно порівняти дані в усіх групах. Виконаємо це порівняння (непараметричний дисперсійний аналіз незалежних груп) із використанням критерію Манна—Уітні. Проте варто пам'ятати, що використання критеріїв, спрямованих на порівняння двох груп даних, для аналізу трьох і більше груп можливе лише за зміни критичного рівня значущості α. У цьому разі потрібно використати новий рівень $\alpha = 0{,}05/3 = 0{,}0167$.

Для попарного порівняння показників набирання маси тіла щурами різних генетичних ліній спочатку треба розділити дані колонок так, щоб у цифрові значення **Input** (**Group Range** та **Data Range**) увійшли первинні дані стосовно двох генетичних ліній (рис. 75).

Повторивши тричі поспіль процедуру тесту Манна—Уітні та порівнявши вміст колонки **Prob>|U|** із критичним рівнем 0,0167, переконаємось, що відмінності показників набирання маси тіла існують між генетичними лініями L і R ($p = 9{,}4 \cdot 10^{-4}$), а також L і S ($p = 5{,}4 \cdot 10^{-3}$) та не існують між R і S ($p = 0{,}875$).

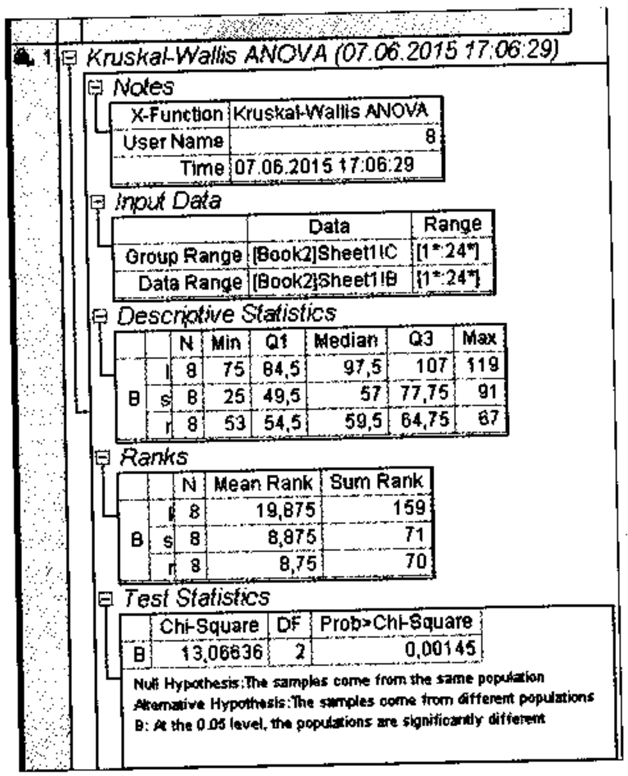


Рис. 74

Приклад 84. Упродовж лікування хворих на гепатит А пацієнтів-чоловіків інфекційного відділення визначали активність ензиму аланінамінотрансферази (АлАТ) у їх крові (відомо, що нормальна активність цього ензиму в крові чоловіків знаходиться в межах 10—40 Од/л). Вимірювання проводили перед лікуванням, через місяць і через три місяці після початку лікування. Потрібно перевірити, чи змінюється активність АлАТ у крові пацієнтів, які лікуються від гепатиту А. Якщо так, на якому етапі відбуваються ці зміни? Дані ензиматичної активності АлАТ (Од/л) наведено в таблиці:

Перед лікуванням	114	98	87	104	119	89	75	84
Через 1 місяць	65	63	49	67	62	50	51	49
Через 3 місяці	42	40	37	35	43	30	33	38

D(Y)	E(Y)	F(Y)	G(Y)	H(Y)	I(Y)
108	l	77	s	108	l
96	l	58	s	96	l
99	l	49	s	99	l
104	l	91	s	104	l
119	l	51	s	119	l
89	l	78	s	89	l
75	l	56	s	75	l
83	l	25	s	83	l
64	r	64	r	77	s
67	r	67	r	58	s
65	r	65	r	49	s
54	r	54	r	91	s
56	r	56	r	51	s
53	r	53	r	78	s
61	r	61	r	56	s
58	r	58	r	25	s

Рис. 75

Розв'язання. Перевіркою даних щодо нормального закону розподілу встановлено, що група значень, отриманих через 1 місяць після початку лікування хворих, не відповідає цьому закону розподілу. Отже, наступний аналіз потрібно виконувати з використанням непараметричних критеріїв. Оскільки групи залежні (вимірювання проводили у динаміці лікування одних і тих самих пацієнтів), варто скористатись критерієм Фрідмена.

Попередньо згрупуємо дані, причому крім колонки даних (**Data Range**) і колонки кодів (**Factor Range**) маємо ввести колонку з ідентифікатором — номером пацієнта (**Subject Range**). Скористаємось меню **Statistics—Nonparametric Tests—Subject ANOVA**. У вікні **Statistics\Nonparametric Tests: friedman** (рис. 76) вкажемо вхідні дані **Input** (**Data Range**, **Factor Range** і **Subject Range**).

Як випливає з результатів проведеного аналізу (рис. 77), нульову гіпотезу слід відхилити, тобто у процесі лікування справді змінюється (зменшується, що відбивається у зниженні показників медіани **Median** і квартилів **Q1**, **Q3** у таблиці **Descriptive Statistics** аркуша результатів аналізу) ензиматична активність АлАТ. Застосувавши далі апостеріорний аналіз, ми маємо виявити групи, які відрізняються між собою. Як і в разі критерію Краскела—

Уолліса, для критерію Фрідмена апостеріорний аналіз проводять попарним порівнянням усіх груп даних між собою із застосуванням критерію Уілкоксона (але з обов'язковою зміною критичного рівня α аналогічно попередньому прикладу). Скориставшись поправкою Бонферроні, отримаємо новий рівень $\alpha = 0{,}05/3 = 0{,}0167$.

Після попарного порівняння даних із застосуванням критерію Уілкоксона дійдемо висновку, що активність АлАТ відрізняється на всіх етапах лікування гепатиту А (в усіх випадках навіть за ненаправленої альтернативної гіпотези маємо значення колонки **Prob>|W|** 0,00781).

Непараметричні методи кореляційного аналізу. Непараметричні методи кореляційного аналізу (Спірмена, Кендалла, гамма) застосовують для дослідження зв'язку між кількісними показниками незалежно від виду їх розподілу, між кількісним та якісним порядковим показником або між двома порядковими показниками.

Приклад 85. У групі курсантів військового факультету за допомогою психофізіологічних тестів визначали показники агре-

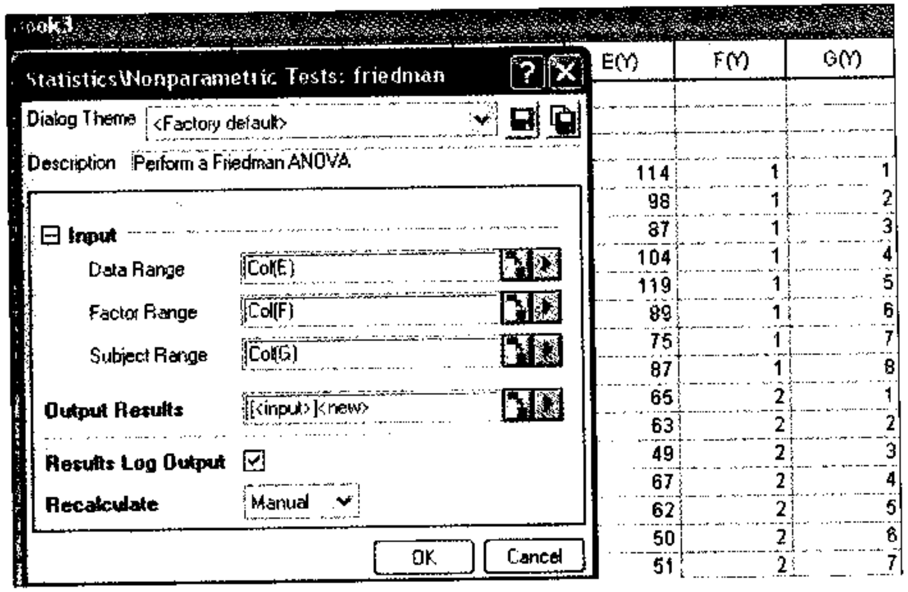
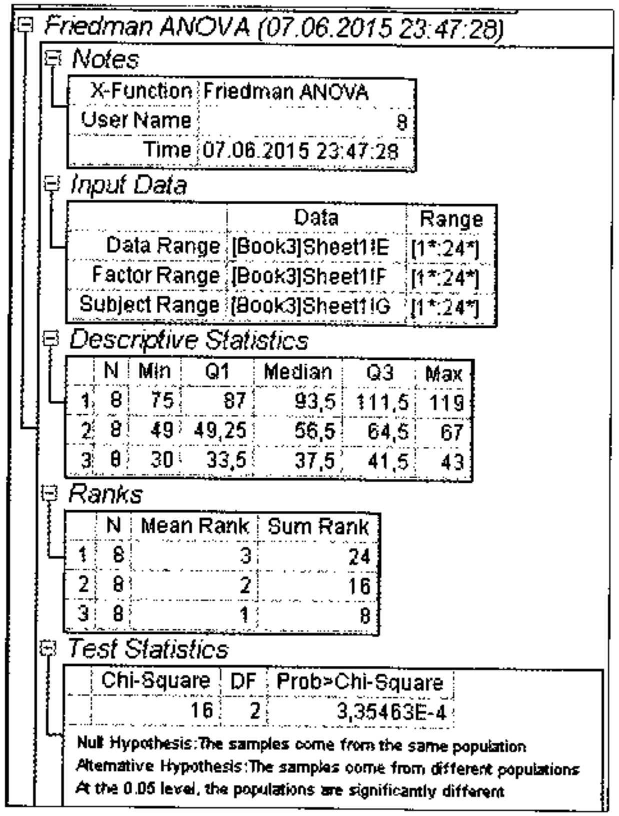


Рис. 76



Friedman ANOVA (07.06.2015 23:47:28)

Notes

X-Function	Friedman ANOVA
User Name	8
Time	07.06.2015 23:47:28

Input Data

	Data	Range
Data Range	[Book3]Sheet1!E	[1*:24*]
Factor Range	[Book3]Sheet1!F	[1*:24*]
Subject Range	[Book3]Sheet1!G	[1*:24*]

Descriptive Statistics

	N	Min	Q1	Median	Q3	Max
1	8	75	87	93,5	111,5	119
2	8	49	49,25	56,5	64,5	67
3	8	30	33,5	37,5	41,5	43

Ranks

	N	Mean Rank	Sum Rank
1	8	3	24
2	8	2	16
3	8	1	8

Test Statistics

Chi-Square	DF	Prob>Chi-Square
16	2	3,35463E-4

Null Hypothesis:The samples come from the same population
Alternative Hypothesis:The samples come from different populations
At the 0.05 level, the populations are significantly different

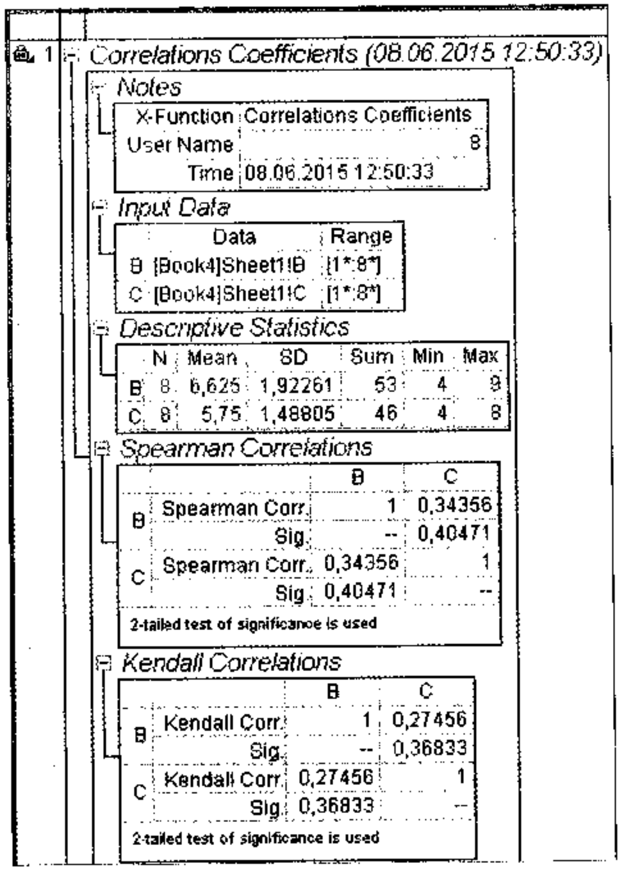
Рис. 77

сивності та схильності до лідерства, які виражали у балах від 1 до 10. Результати експерименту наведено в таблиці:

Агресивність	8	7	9	6	5	9	5	4
Схильність до лідерства	5	7	6	4	4	8	5	4

Потрібно перевірити, чи існує зв'язок між цими показниками.

Розв'язання. Відповідно до особливостей досліджених показників, застосуємо непараметричний кореляційний аналіз. Для цього скористаємось меню **Statistics—Descriptive Statistics—Correlation Coefficient** (див. рис. 53). У вікні **Statistics\Descriptive Statistics: corrcoef** (див. рис. 54) оберемо непараметричні тести: мож-



1 Correlations Coefficients (08.06.2015 12:50:33)

Notes

X-Function	Correlations Coefficients
User Name	8
Time	08.06.2015 12:50:33

Input Data

	Data	Range
B	[Book4]Sheet1!B	[1*:8*]
C	[Book4]Sheet1!C	[1*:8*]

Descriptive Statistics

	N	Mean	SD	Sum	Min	Max
B	8	6,625	1,92261	53	4	9
C	8	5,75	1,48805	46	4	8

Spearman Correlations

		B	C
B	Spearman Corr.	1	0,34356
	Sig.	--	0,40471
C	Spearman Corr.	0,34356	1
	Sig.	0,40471	--

2-tailed test of significance is used

Kendall Correlations

		B	C
B	Kendall Corr.	1	0,27456
	Sig.	--	0,36833
C	Kendall Corr.	0,27456	1
	Sig.	0,36833	--

2-tailed test of significance is used

Рис. 78

на скористатись обома представленими тестами, а саме, Спірмена (**Spearman**) і Кендалла (**Kendall**).

Як випливає з результатів аналізу (зверніть увагу на вміст колонок **sig**, які в обох випадках кореляційних матриць перевищують критичний рівень α), зв'язку між показниками агресивності та схильності до лідерства у курсантів військового факультету не існує (рис. 78).

§ 18. РОЗВ'ЯЗУВАННЯ ЗАДАЧ ІЗ ВИКОРИСТАННЯМ ПРОГРАМИ "STATISTICA"

Найуживанішим статистичним програмним пакетом є Statistica (фірма StatSoft Inc., США), головними перевагами якого є простий інтерфейс, аналіз великих масивів даних і широкі графічні можливості. У ньому також передбачена можливість аналізу даних, які попередньо були систематизовані в електронних таблицях Excel.

У цьому параграфі ми зупинимось переважно на прикладах застосування програми "Statistica" для аналізу біологічних даних.

Після запуску програми на екрані з'являється вікно-запрошення **Welcome to Statistica**, яке дає змогу швидко перейти до робочих файлів користувача (рис. 79). Таблиці даних (**Spreadsheet**) формуються шляхом заповнення стовпчиків (**Vars**, змінні) та рядків (**Cases**, спостереження).

Засоби для аналізу даних у програмному пакеті Statistica 7.0 можна переглянути у меню **Statistics** (рис. 80).

Приклади параметричного аналізу даних

Приклад 86. Досліджували закономірності дії екстракту наперстянки пурпурової на внутрішньоклітинну концентрацію іонів Ca^{2+} у клітинах серцевого м'яза щурів. Відомо, що рослина наперстянка пурпурова є джерелом глікозидів, які вибірково пригнічують роботу Na^{+}, K^{+}-АТФази плазматичної мембрани клітин. За цих умов зростає внутрішньоклітинна концентрація іонів Na^{+} і знижується — іонів K^{+}. Відомо також, що у кардіоміоцитах транспорт іонів Na^{+} і Ca^{2+} спряжений, тому ймовірна зміна внутрішньоклітинних концентрацій цих катіонів. Використано екстракт наперстянки пурпурової у концентрації 10 мг/мл. Згідно з постановкою експерименту, отриману суспензію кардіоміоцитів ділили на дві частини: у першій реєстрували контрольні значення концентрації іонів Ca^{2+}, до другої додавали аліквоту екстракту наперстянки пурпурової і вже потім реєстрували концентрацію іонів Ca^{2+}. Отримано такі дані (у нмоль/л):

Контроль	105,2	121,1	109,8	104,7	110,5	110,3	120,9	117,2	119,0
Додавання наперстянки пурпурової	141,8	152,5	132,9	141,7	128,4	129,7	131,9	142,6	142,8

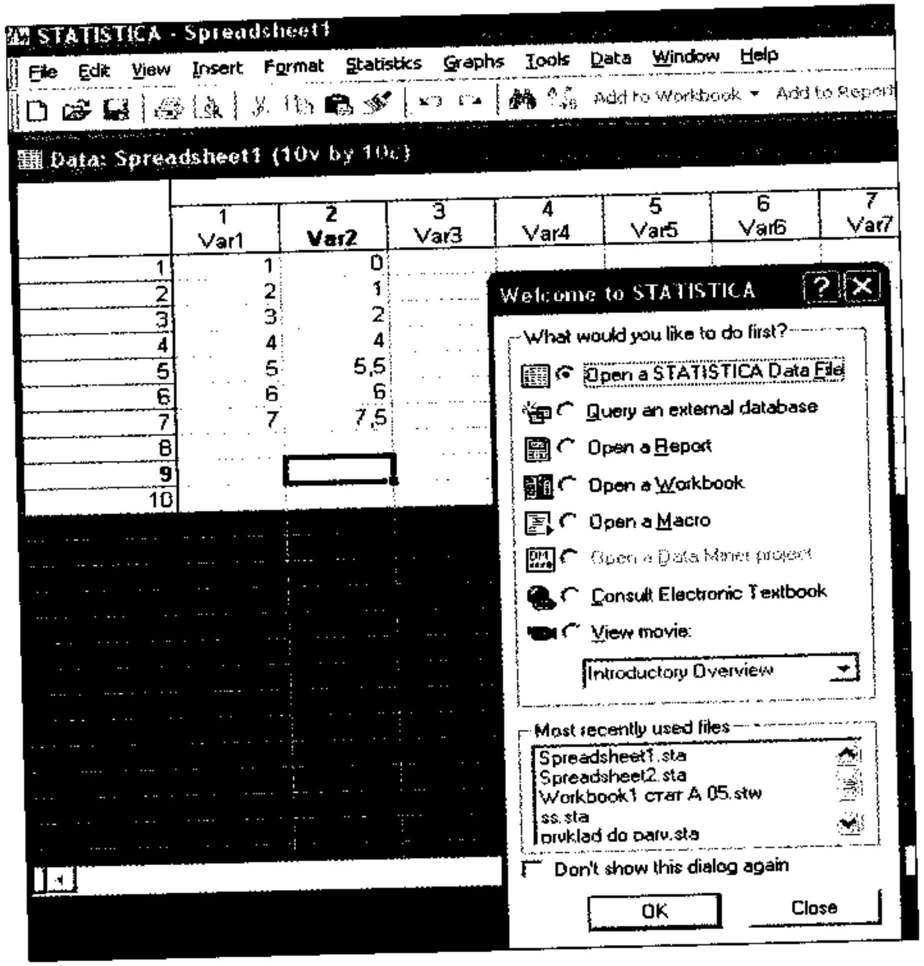


Рис. 79

Потрібно проаналізувати результати та відповісти на запитання: чи впливає екстракт наперстянки пурпурової на базальний рівень внутрішньоклітинної концентрації іонів Ca^{2+}?

Розв'язання. Насамперед визначимо статистичний критерій, який можна застосовувати до даних експерименту. Для цього перевіримо дані двох груп (контрольні вимірювання та вимірювання за наявності екстракту наперстянки пурпурової) на належність до нормально розподілених генеральних сукупностей. Оскільки маємо невеликі вибірки ($n_{\text{контр}} = n_{\text{нап. пурп}} = 9$), найопти-

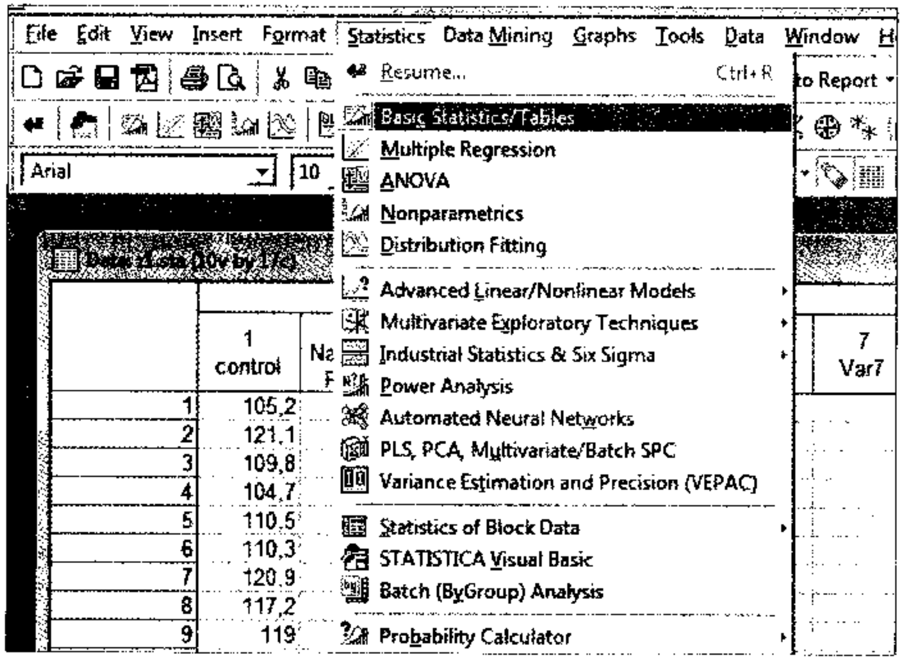


Рис. 80

мальнішим критерієм (тестом) для перевірки гіпотез про нормальність є тест Шапіро—Уілка. Виділимо обидві колонки з даними, а потім скористаємось меню **Statistics—Basic Statistics/Tables** (див. рис. 80).

У вікні **Basic Statistics and Tables** (рис. 81) оберемо **Descriptive Statistics** та натиснемо кнопку OK. У вікні **Descriptive Statistics** перейдемо на вкладку **Normality** і переконаємось, що поряд із кнопкою **Variables** вказані потрібні стовпчики (коли ні — натиснемо на кнопку **Variables** та у вікні **Selected the variables for the analysis** оберемо необхідні змінні). Далі оберемо критерій Шапіро—Уілка (поряд із Shapiro—Wilk's W test поставимо прапорець) (рис. 82). Після цього необхідно натиснути кнопку **Histograms** або **Frequency tables** і, таким чином, програма створить книгу аналізу (**Workbook**), в яку автоматично заноситимуться результати виконаних статистичних процедур. У нашому випадку ми можемо переглянути два аркуші з гістограмами, які містять результати тесту Шапіро—Уілка (рис. 83). В обох випадках за результатами експерименту маємо $p > 0{,}05$ й, отже, приймаємо нульову гіпотезу щодо нормального закону розподілу.

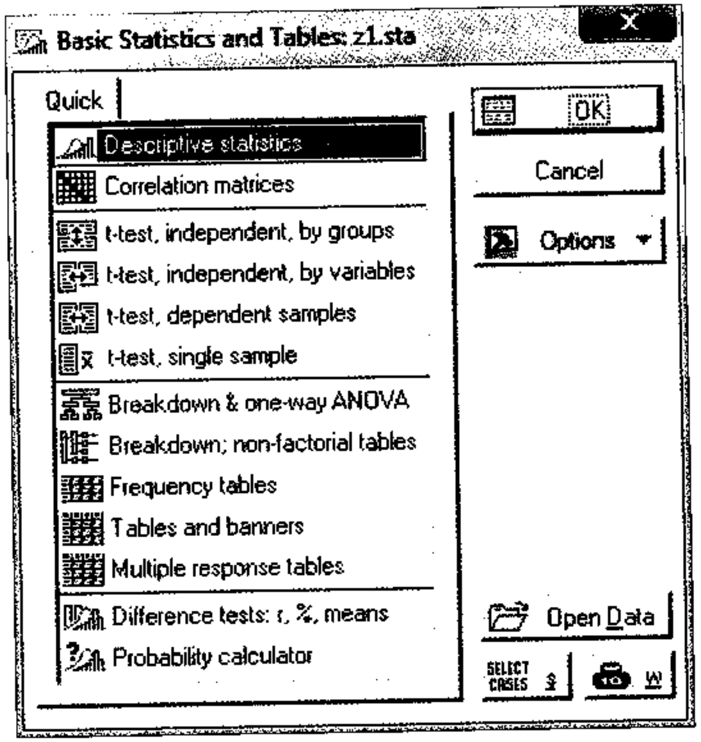


Рис. 81

Тепер ми маємо усі підстави для використання параметричного статистичного аналізу (у цьому випадку — критерій Стьюдента для незалежних вибірок) (рис. 84). Варто звернути увагу, що цей тест (**t-test, independent**) можна застосовувати як для групованих даних (процедура групування аналогічна описаній у попередньому параграфі; **by groups** — але в цьому разі змінні матимуть назви **Dependent** — масив чисел, **Grouping** — масив кодів), так і негруповані дані (**by variables** — у цьому разі змінні матимуть назви **First variable** — перша змінна, **Second variable** — друга змінна).

У вікні **T-test for Independent Samples by Variables** у вкладці Опції (**Options**) встановимо прапорці **Levene's test** (тест на гомогенність дисперсій) і **t-test with separate variance estimates** (на випадок, коли тестами на гомогенність дисперсій нульова гіпотеза буде відхилена) й натиснемо кнопку **Summary** (рис. 85).

У згаданій таблиці **Workbook** з'явилась нова папка **T-test for...** з аркушем-таблицею відповідного аналізу (рис. 86). У вікні про-

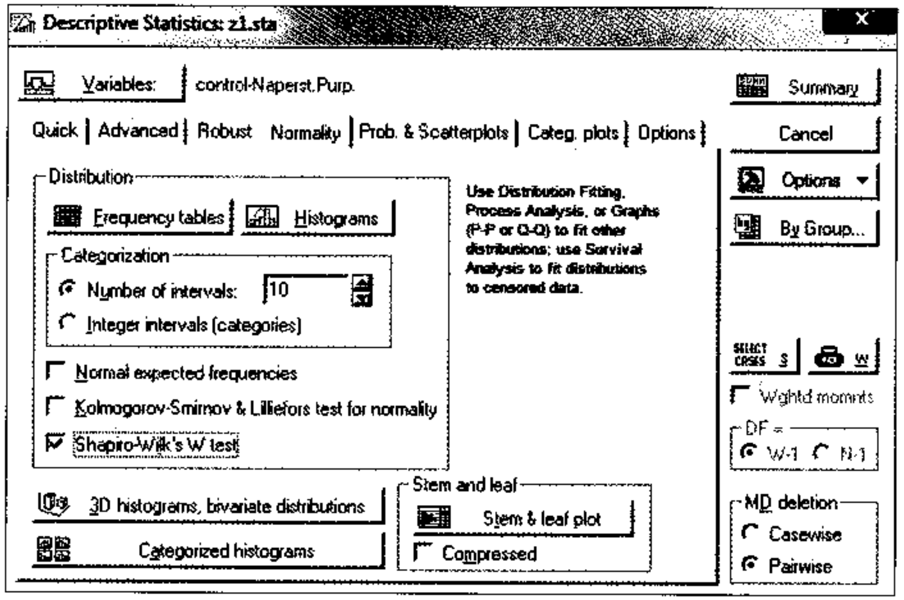


Рис. 82

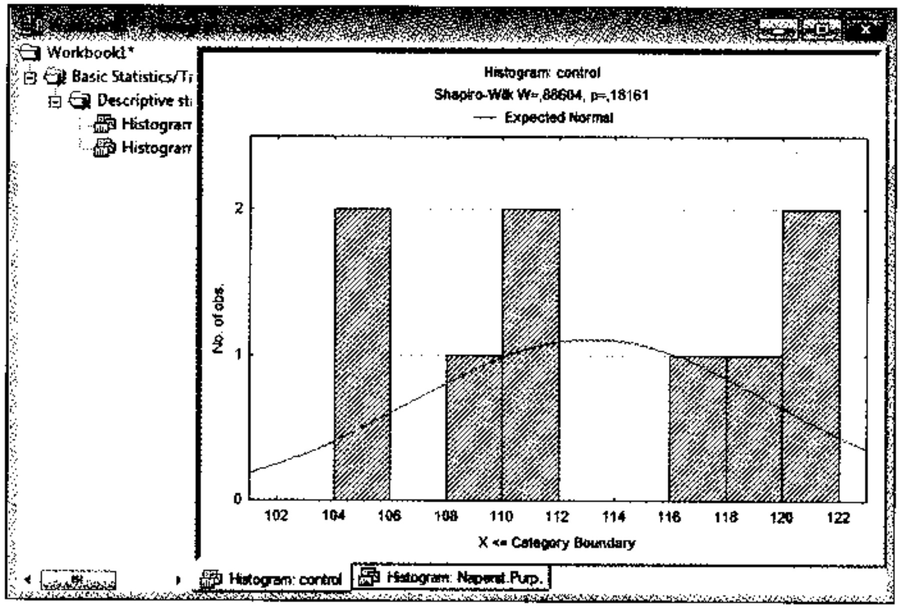


Рис. 83

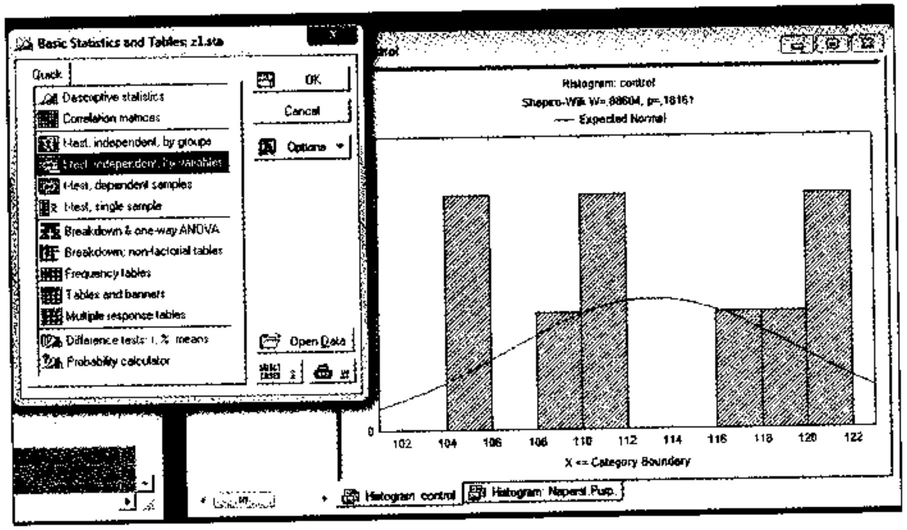


Рис. 84

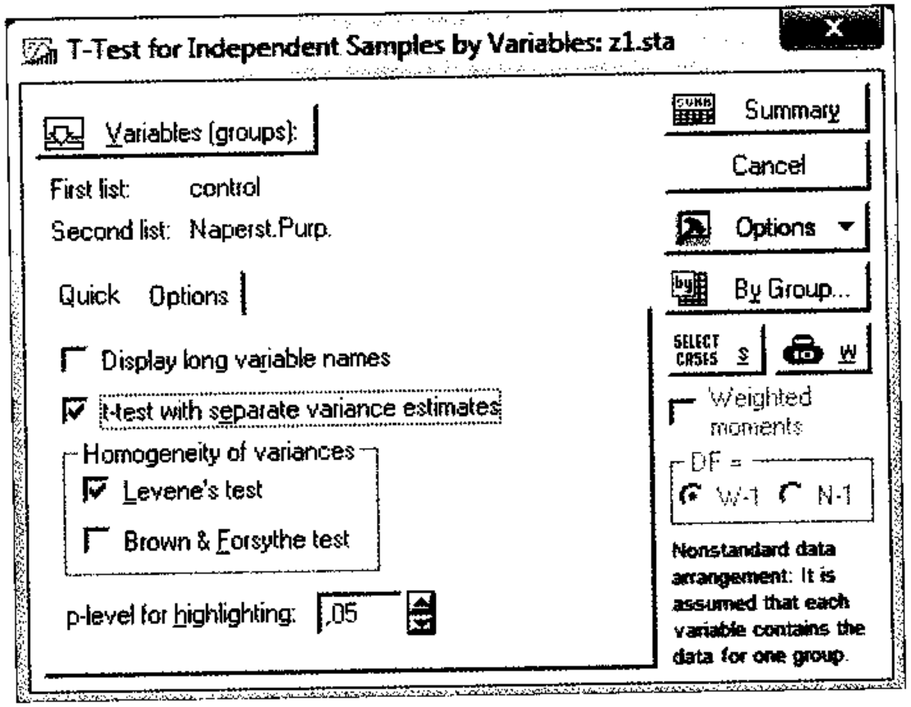
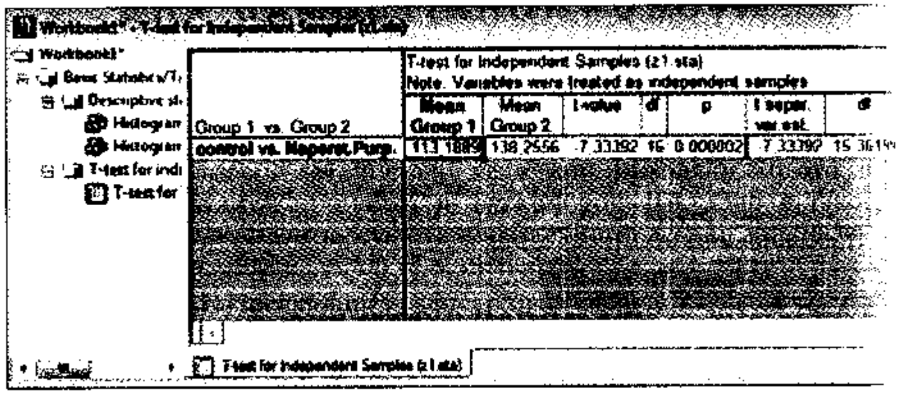


Рис. 85



грами привертає увагу червоний колір цифр, що вказує на необхідність відхилити нульову гіпотезу критерію Стьюдента. Оскільки в цій таблиці тести на рівність дисперсій Левена (колонка **p Levene**) та Фішера (**p Variances**) дали змогу прийняти нульову гіпотезу (маємо випадок критерію Стьюдента для незалежних вибірок із рівними дисперсіями), скористаємось вмістом колонки **p**, аби переконатися, що за результатами тесту Стьюдента нульову гіпотезу відхилено. Отже, можна дійти висновку, що екстракт наперстянки пурпурової спричинює зростання внутрішньоклітинної концентрації іонів Ca^{2+} у клітинах серцевого м'яза.

Приклад 87. Досліджували інтенсивність росту паростків пшениці за температури 10 і 15 °С. Експеримент сплановано так: спочатку реєстрували приріст паростків за 10 °С за 1 добу, далі температуру підвищували до 15 °С, утримували її упродовж 1 доби і знову вимірювали відносний приріст паростків. Отримані результати (мм) наведено в таблиці:

10 °С	18,5	17,9	19,8	18,9	18,3	18,5	17,8	19,5	19,1
15 °С	22,5	24,6	23,8	23,4	24,1	23,9	22,9	24,2	23,7

Потрібно перевірити, чи вплинула зміна температури на інтенсивність росту паростків пшениці.

Розв'язання. Виконавши дії, зазначені у прикладі 86, перевіримо дані щодо належності до нормально розподілених сукупностей із використанням критерію Шапіро—Уілка. В обох випадках приймаємо нульову гіпотезу.

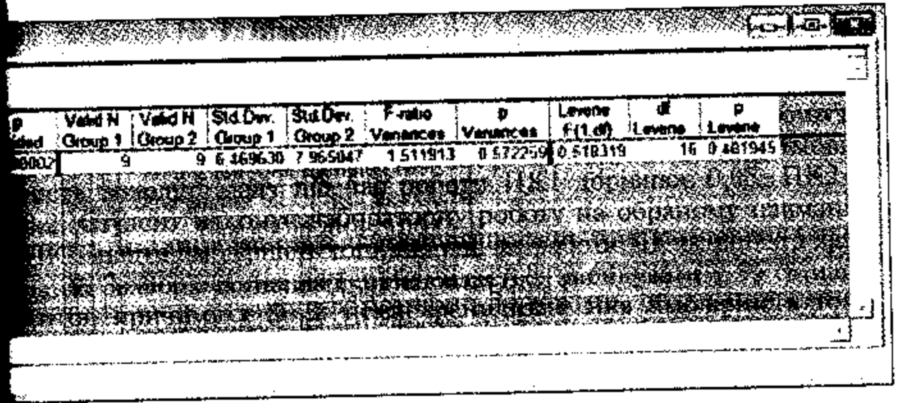


Рис. 86

Оскільки експеримент проведений на одних і тих самих об'єктах, застосуємо критерій Стьюдента для залежних вибірок (**t-test, dependent samples**) (див. рис. 81). Зауважимо, що в цьому разі порівняння проводиться лише за змінними, тому групування є зайвим. У вікні **T-test for Dependent Samples** переконаємось, що змінні для аналізу обрано правильно (проведемо порівняння за змінними: **First variable list** — стовпчик, який відповідає даним за 10 °С, **Second variable list** — стовпчик, який відповідає даним за 15 °С) і натиснемо кнопку **Summary** (рис. 87).

Згідно з результатами проведеного аналізу (рис. 88), підвищення температури впливає на інтенсивність росту паростків пшениці.

Приклад 88. Перед забором крові виміряно артеріальний тиск у 15 донорів. Отримано такі результати:

120	145	135	110	109	114	118	119	125	122	140	137	129	124	126

Треба встановити, чи можна вважати обстежену групу донорів нормотоніками, якщо середній показник систолічного артеріального тиску в нормі становить 120 мм рт. ст.?

Розв'язання. Перевірка вибірки на її належність до нормально розподіленої генеральної сукупності тестом Шапіро—Уілка дала підставу прийняти нульову гіпотезу. Отже, наступний аналіз зводиться до застосування ***одновибіркового критерію Стьюдента***. Для цього виділимо стовпчик із даними і скористаємось меню

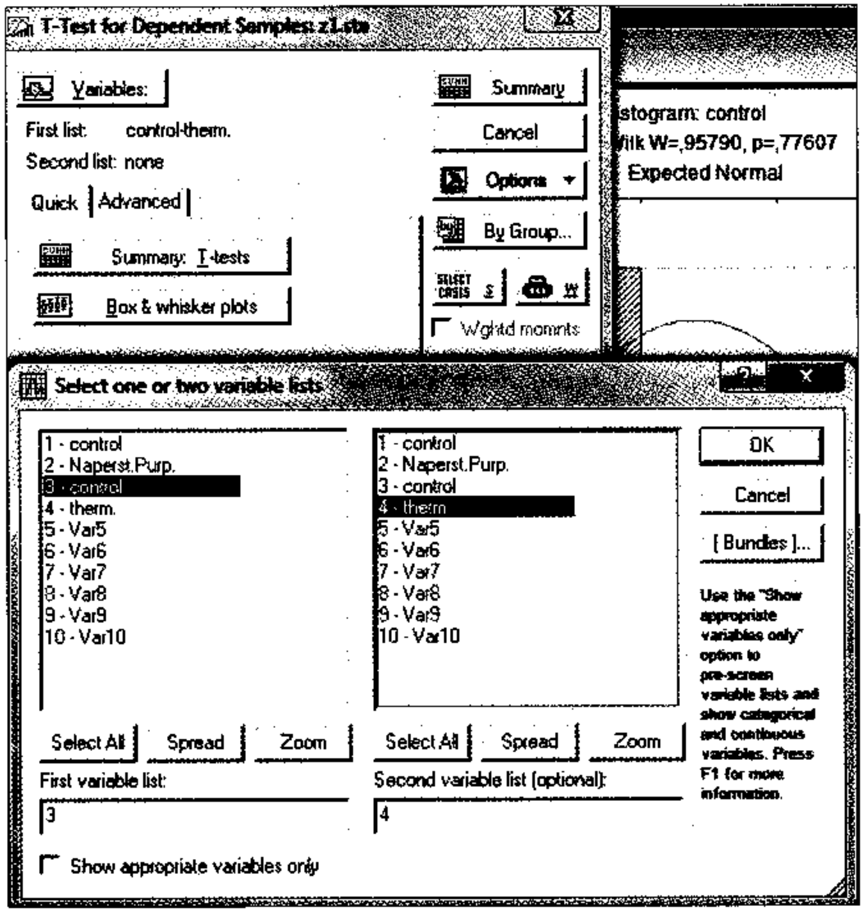


Рис. 87

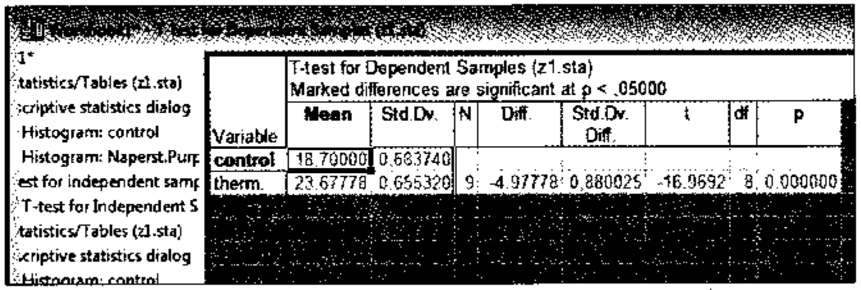


Рис. 88

Statistics—Basic Statistics/Tables; у вікні **Basic Statistics and Tables** оберемо **t-test, single sample** та натиснемо кнопку **OK** (див. рис. 81). У вікні **T-Test for Single Means** перейдемо до вкладки **Advanced** і вкажемо референтне значення систолічного артеріального тиску в полі **Reference values** (прапорець біля **Test all means against**) 120; натиснемо **Summary** (рис. 89).

Згідно з результатами одновибіркового критерію Стьюдента (рис. 90), приймаємо нульову гіпотезу (значення колонки p на рівні 0,1). Отже, є всі підстави віднести групу донорів до нормотоніків.

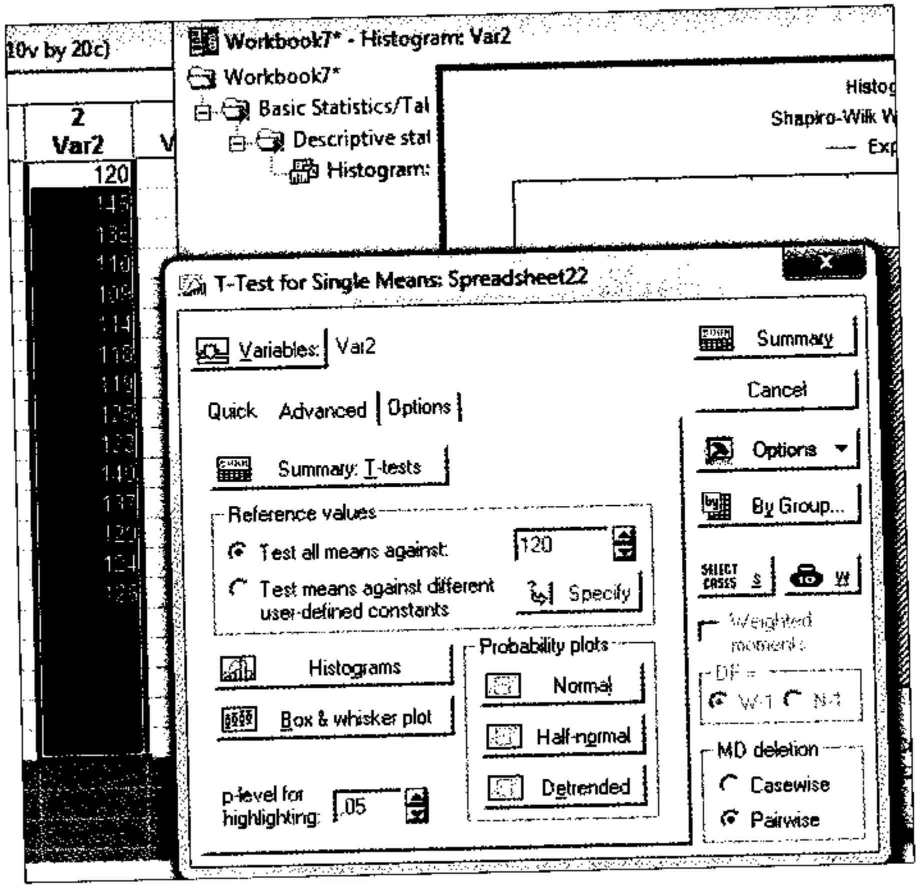


Рис. 89

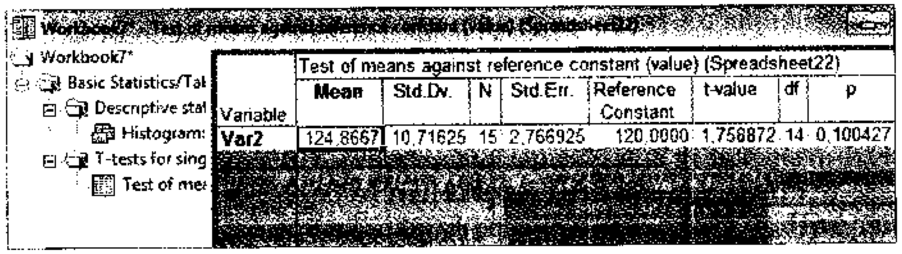


Рис. 90

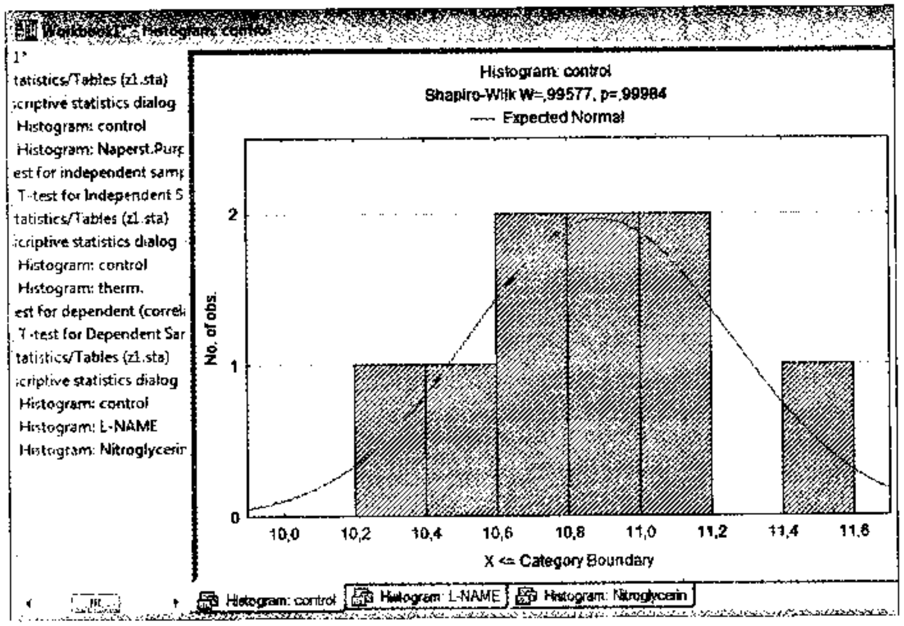


Рис. 91

Приклад 89. Розглянемо поетапно проведення однофакторного дисперсійного аналізу засобами програми "Statistica" з використанням завдання з прикладу 78.

Розв'язання. Перевіримо дані груп (контроль, L-NAME, нітрогліцерин) щодо їх нормального розподілу (тест Шапіро—Уілка). Для всіх груп приймемо нульову гіпотезу (рис. 91).

На наступному етапі аналізу із застосуванням критерію Левена перевіримо гіпотезу щодо рівності дисперсій. Для цього попередньо згрупуємо дані і скористаємось меню **Statistics—Basic**

Statistics/Tables; у вікні **Basic Statistics and Tables** оберемо **Breakdown & One-Way ANOVA** й натиснемо кнопку **OK** (рис. 92).

У вікні **Statistics by Groups (Breakdown)** (рис. 93) перейдемо до вкладки **Individual tables** і вкажемо змінні (**Dependent** — колонка з числовими даними, **Grouping** — колонка з кодами) (рис. 94); натиснемо кнопку **Codes of grouping variables** та оберемо усі коди кнопкою **All** (рис. 95); натиснемо кнопку **OK**; у вікні **Statistics by Groups (Breakdown)** знову натиснемо кнопку **OK**.

У вікні **Statistics by Groups Results** (рис. 96) перейдемо до вкладки **ANOVA & tests** і натиснемо кнопку **Levene tests**.

Як видно з результатів тесту Левена (рис. 97), дисперсії груп гомогенні. Отже, проведемо параметричний дисперсійний аналіз. Для цього повернемось у вікно **Statistics by Groups Results** (вкладка **ANOVA & tests**) і натиснемо на кнопку **Analysis of Variance** (див. рис. 96).

Згідно з результатами дисперсійного аналізу (рис. 98), фактор (концентрація оксиду азоту) впливає на тонус аорти щурів.

Проведемо апостеріорний аналіз, щоб з'ясувати, які саме групи відрізняються. Для цього повернемось у вікно **Statistics by**

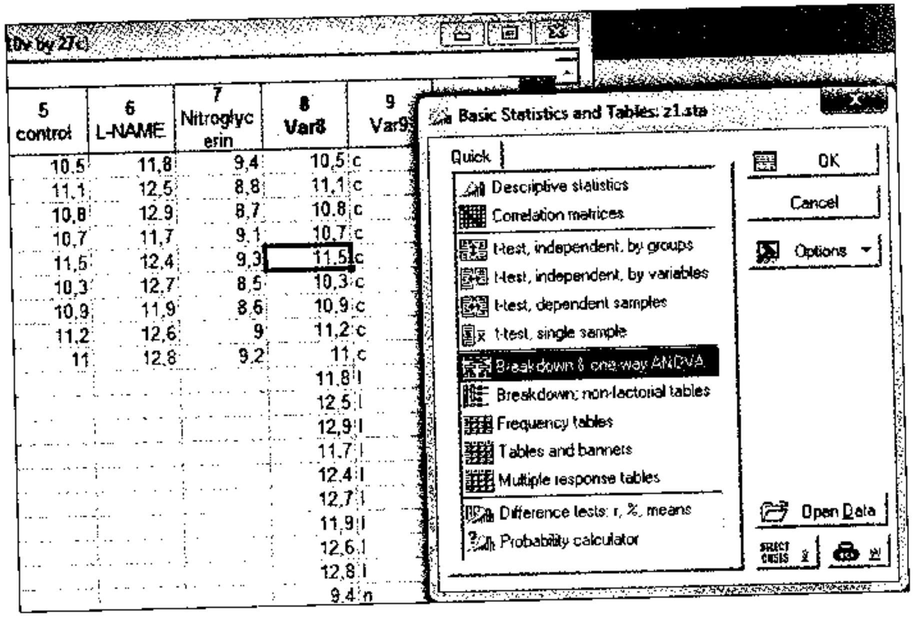


Рис. 92

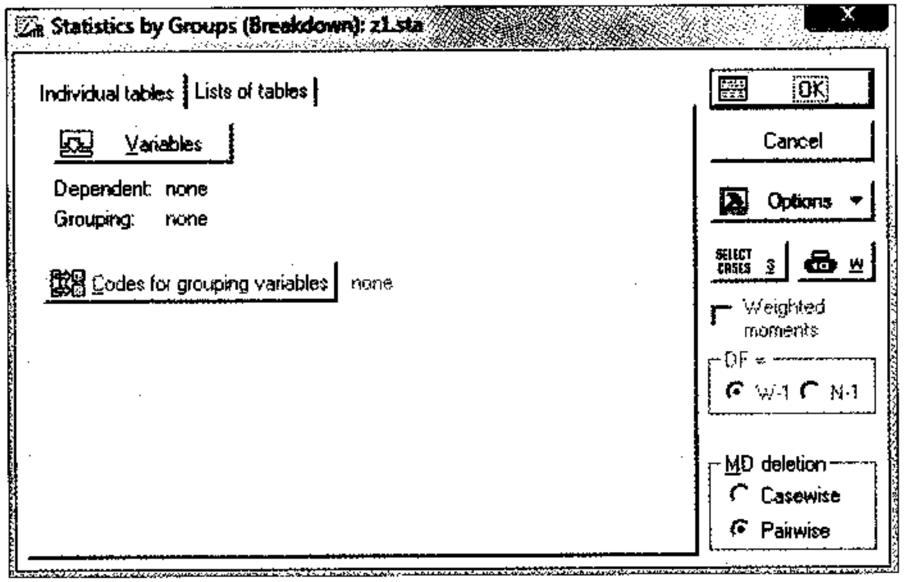


Рис. 93

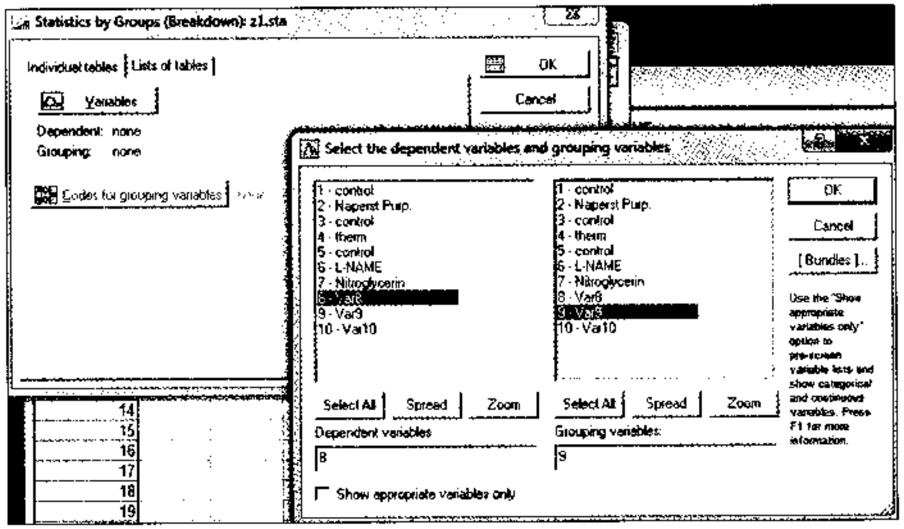


Рис. 94

Groups Results (перейдемо до вкладки **Post-hoc**) і натиснемо кнопку обраного тесту (рис. 99).

Результати апостеріорного аналізу (рис. 100) підтвердили відмінності між усіма груповими середніми.

Приклад 90. Проведемо параметричний кореляційний аналіз даних з урахуванням умови прикладу 79.

Розв'язання. Для перевірки наявності лінійного зв'язку між внутрішньоретикулярною концентрацією кальцію та активністю Ca^{2+}-АТФази і розрахунку коефіцієнта кореляції скористаємось меню **Statistics—Basic Statistics/Tables**; у вікні **Basic Statistics and Tables** оберемо **Correlation matrices** і натиснемо кнопку **OK**. У вікні **Product-Moment and Partial Correlation** перейдемо до вкладки **Options**, оберемо колонки зі змінними і формат відображення

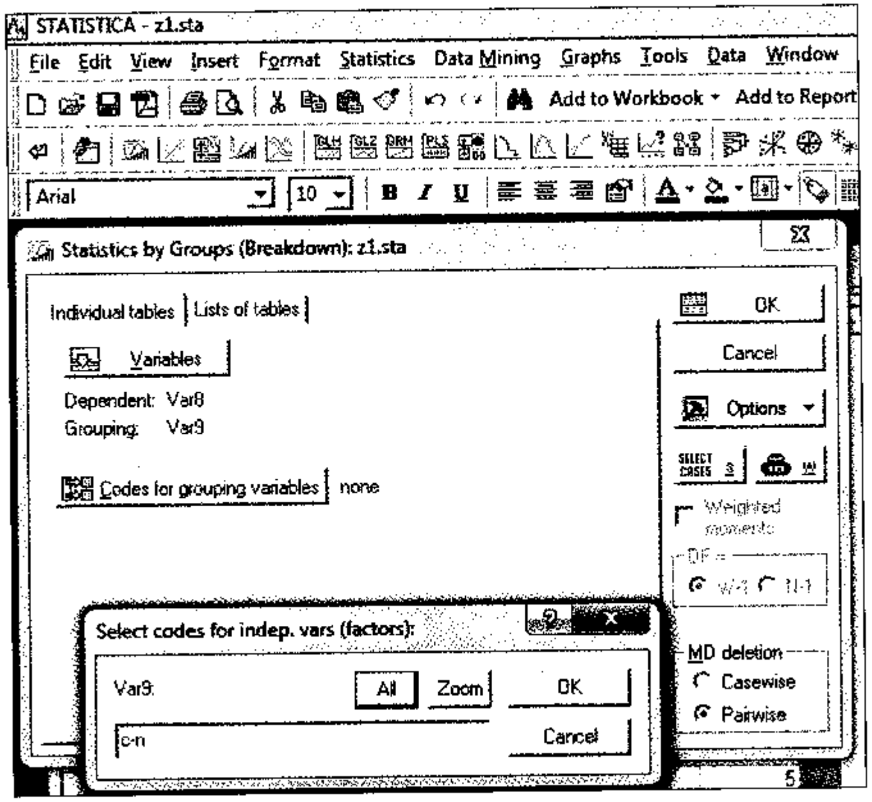


Рис. 95

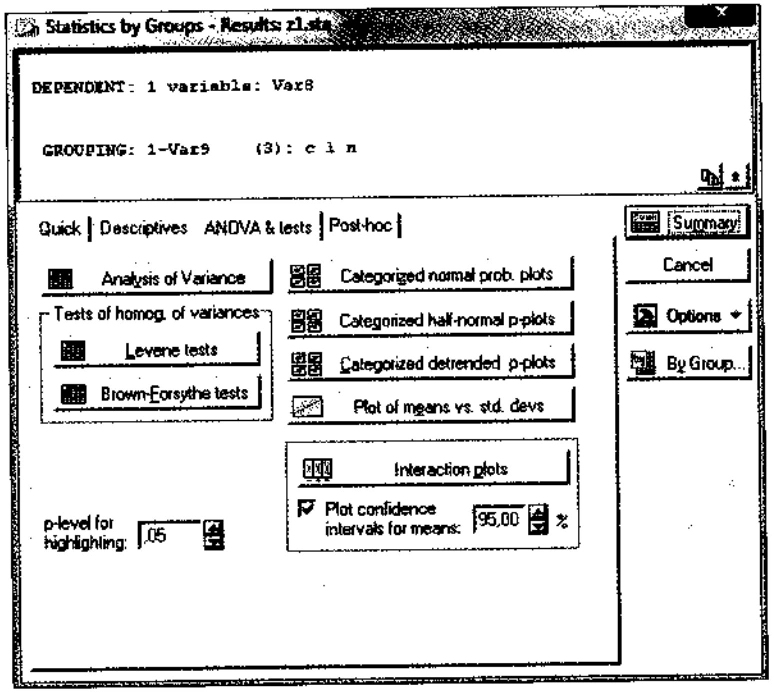
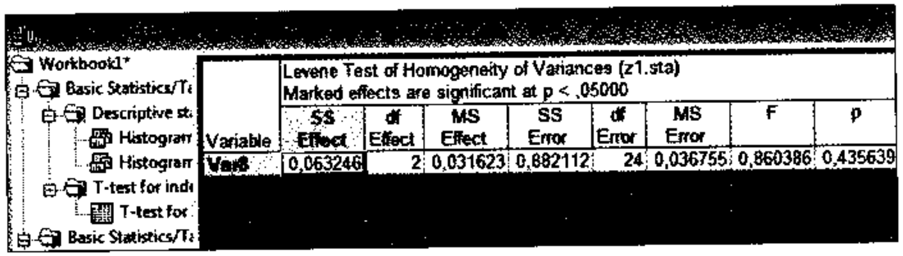


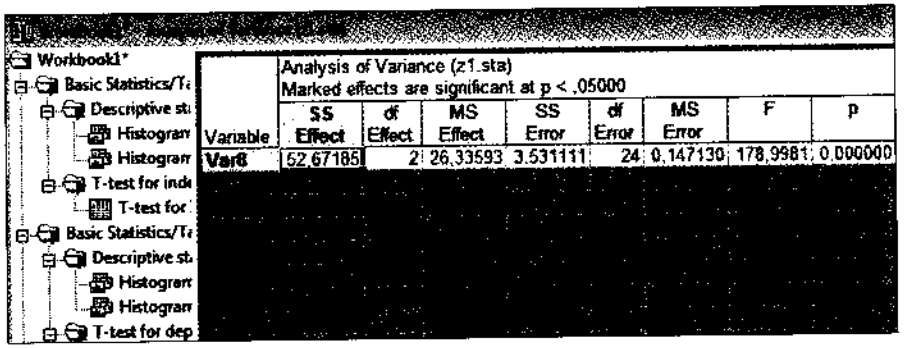
Рис. 96



Levene Test of Homogeneity of Variances (z1.sta)
Marked effects are significant at p < ,05000

Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Var8	0,063246	2	0,031623	0,882112	24	0,036755	0,860386	0,435639

Рис. 97



Analysis of Variance (z1.sta)
Marked effects are significant at p < ,05000

Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Var8	52,67185	2	26,33593	3,531111	24	0,147130	178,9981	0,000000

Рис. 98

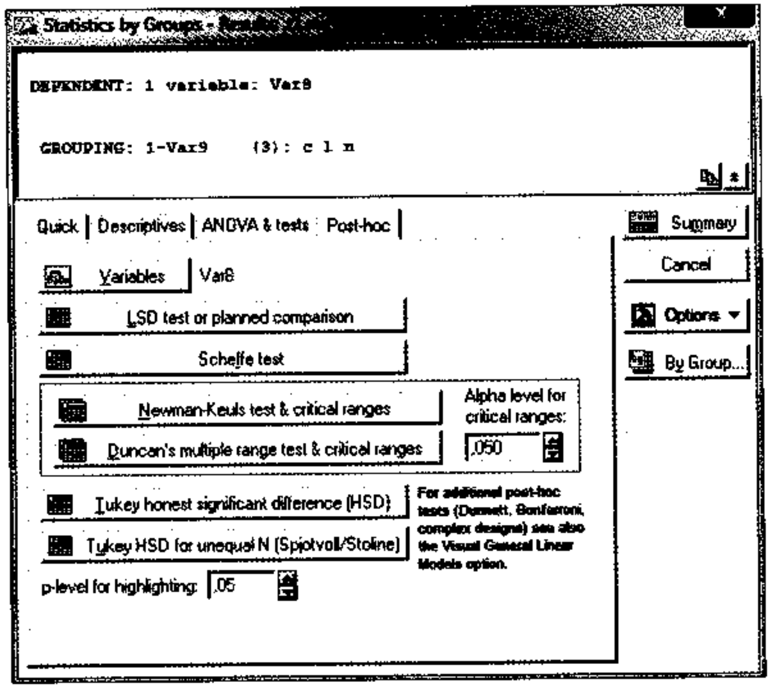


Рис. 99

кореляційної матриці — поставимо прапорець **Display r, p-levels, and N's** (рис. 101), потім натиснемо кнопку **Summary**.

Результат кореляційного аналізу (рис. 102) свідчить (червоний колір цифр у матриці та p = 0,000...), що між досліджуваними параметрами існує лінійний зв'язок (r = 0,9635).

На наступному етапі встановимо рівняння, яке пов'язує активність Ca^{2+}-АТФази ендоплазматичного ретикулума і внутрішньоретикулярну концентрацію іонів Ca^{2+}. Для цього проведемо регресійний аналіз: у вікні **Product-Moment and Partial Correlation** перейдемо до вкладки **Advanced/Plot** (рис. 103) і натиснемо кнопку **2D scatterplots**. У книжці результатів аналізу отримаємо графік і рівняння лінійної регресії (над графіком) (рис. 104).

Приклади непараметричного аналізу даних

Приклад 91. В інфекційному відділенні лікарні перебуває 9 пацієнтів, хворих на запалення легенів. За сукупністю ознак стан хворих визначили як: 4 — тяжкий стан; 3 — стан середньої тяж-

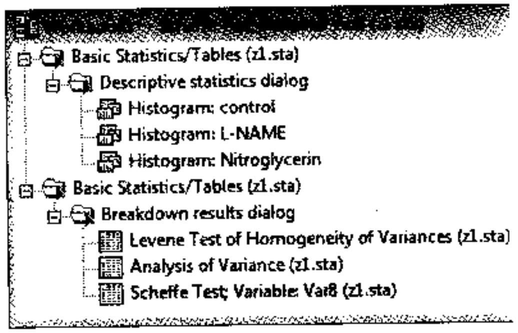


Рис. 100

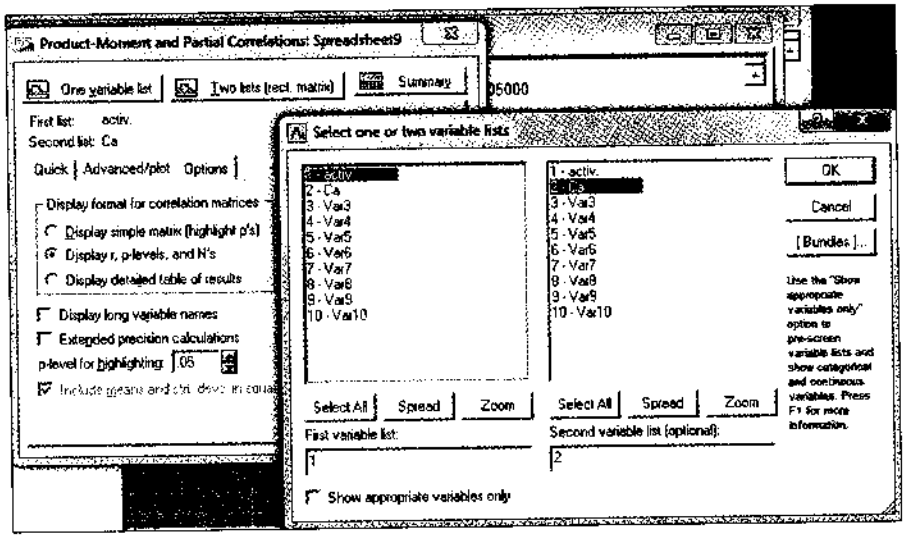


Рис. 101

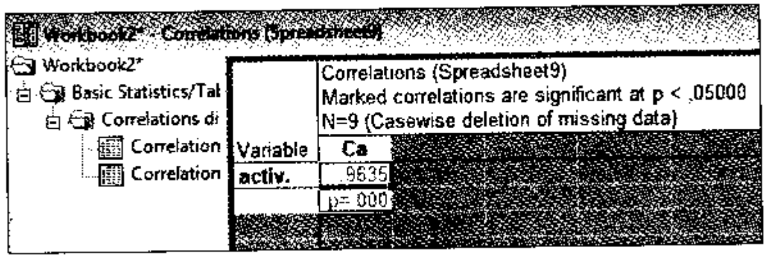
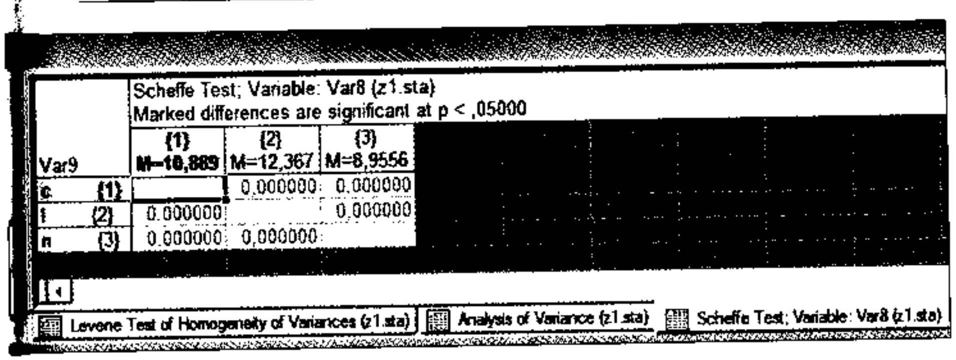


Рис. 102



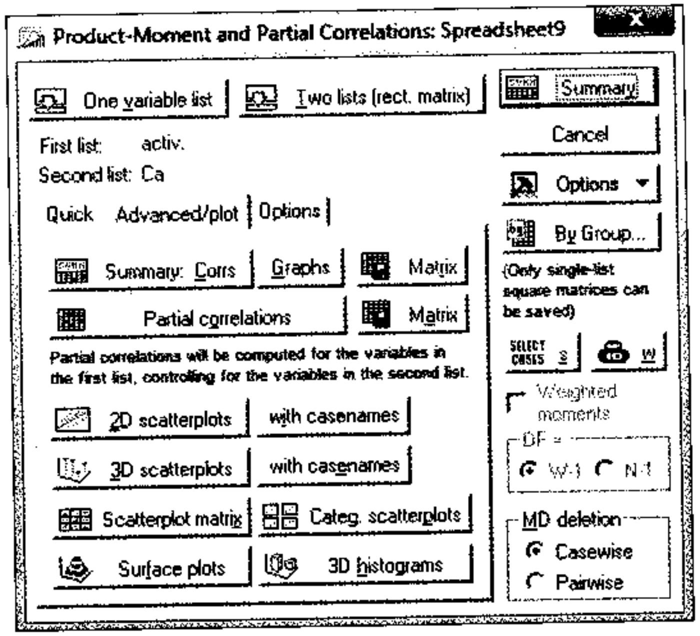


Рис. 103

кості; 2 — легке захворювання; 1 — немає клінічних проявів захворювання. Через тиждень після початку лікування провели повторне обстеження пацієнтів, й отримали такі дані:

До лікування	4	3	4	3	4	3	3	4	2
Після тижня лікування	2	3	3	2	4	2	2	2	1

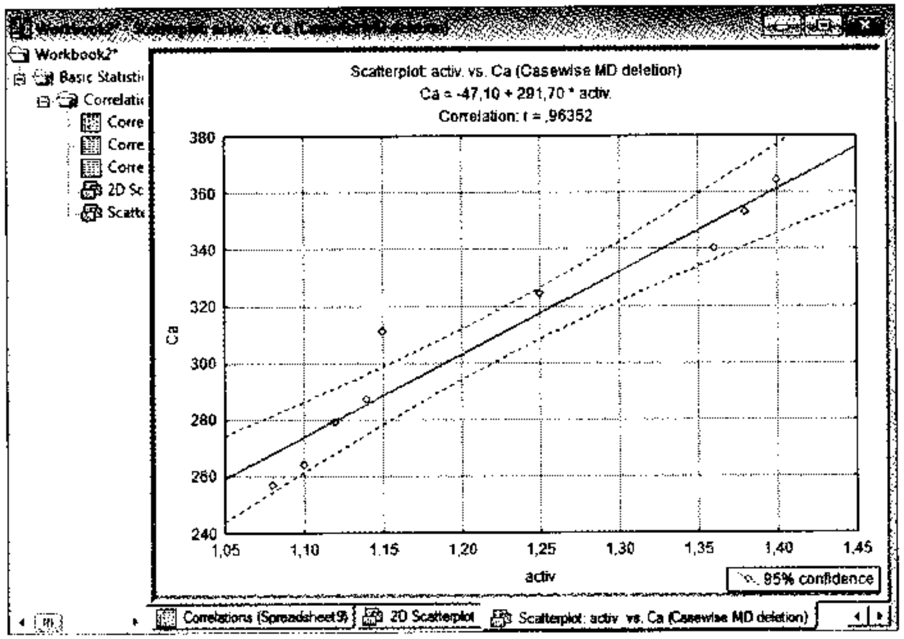


Рис. 104

Треба перевірити, чи є тижневе лікування ефективним.

Розв'язання. Для порівняння залежних вибірок скористаємось непараметричним критерієм знаків: меню **Statistics—Nonparametrics** (рис. 105); у вікні **Nonparametric Statistics** оберемо **Comparing two dependent samples (variables)** і натиснемо кнопку **OK** (рис. 106). У вікні **Comparing two variables** оберемо колонки зі змінними і натиснемо кнопку **Sign test** (рис. 107).

Згідно з результатами тесту знаків (рис. 108), тижневе лікування пацієнтів, хворих на запалення легенів, загалом було ефективним.

Приклад 92. Серед школярів перших класів випробовували дві альтернативні методики навчання англійської мови. Ефективність навчання визначали за швидкістю читання текстів (кількість слів за 3 хв). Отримані результати наведено в таблиці:

Методика 1	10	90	40	18	15	78	83	79	92
Методика 2	63	92	78	65	71	85	88	82	91

Треба встановити, чи відрізняються методики за ефективністю і чи можна вважати якусь із них ліпшою.

Розв'язання. Перевірка вибірок на нормальність тестом Шапіро—Уілка показала, що для першої групи характерне відхилення нульової гіпотези ($p = 0{,}035$), отже, подальший аналіз треба проводити із застосуванням непараметричних критеріїв. Оскільки вибірки незалежні, спочатку згрупуємо дані, а потім скористаємось критерієм Манна—Уітні: меню **Statistics—Nonparametrics** (див. рис. 105); у вікні **Nonparametric Statistics** оберемо **Comparing two independent samples (groups)** і натиснемо кнопку **OK** (див. рис. 106). У вікні **Comparing two groups** вкажемо колонки з даними **Dependent variable list** і кодами груп (**Indep. (grouping) variable**), натиснемо кнопку **M-W U test** (рис. 109).

Згідно з результатами тесту знаків (рис. 110), не можна віддати перевагу якійсь одній методиці викладання англійської мови, оскільки між показниками читання дітей двох груп відмінності немає.

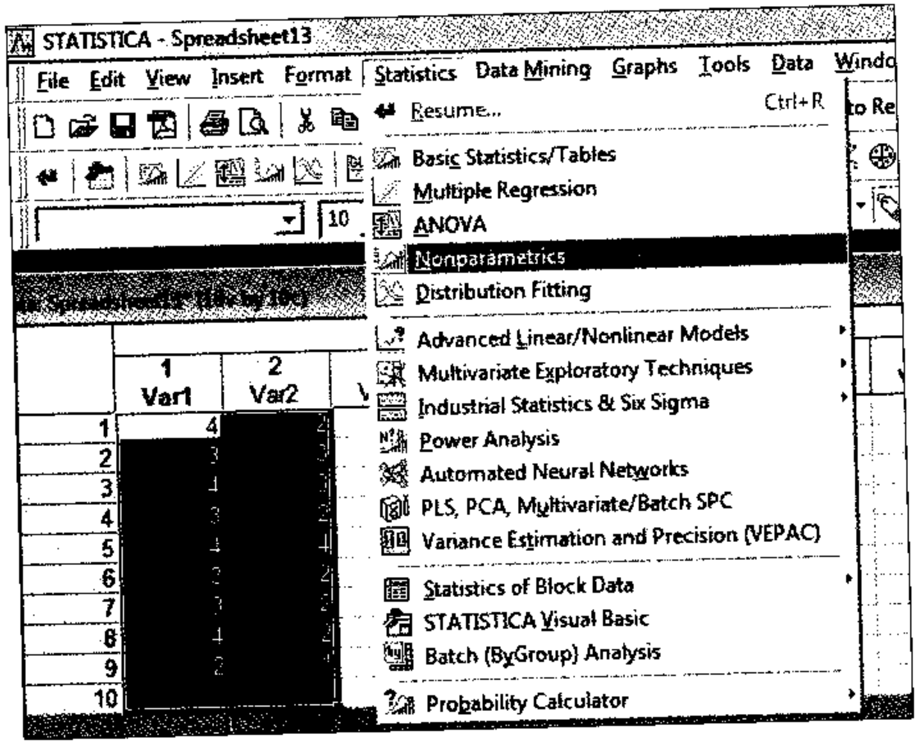


Рис. 105

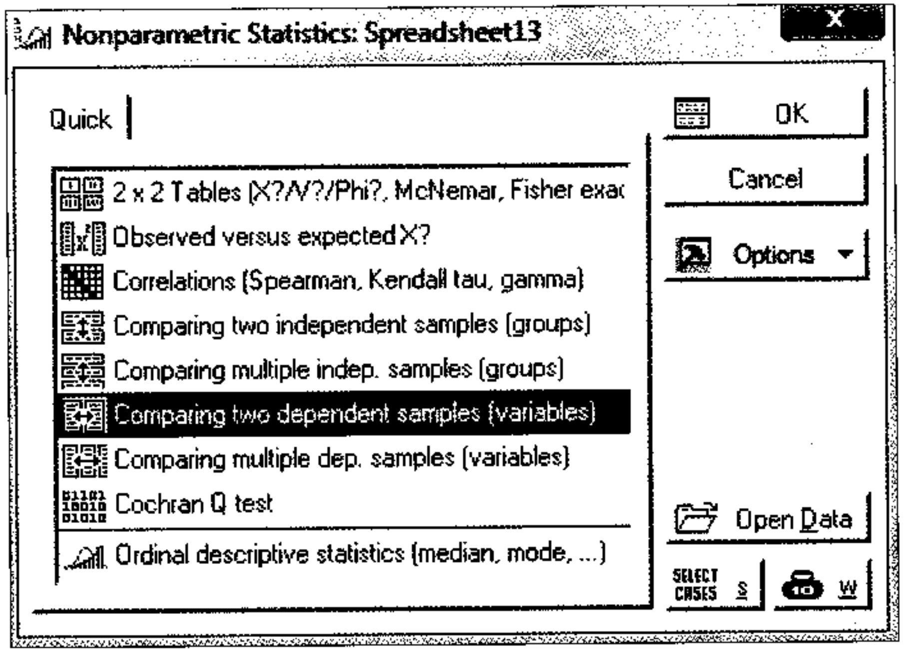


Рис. 106

Приклад 93. В інфекційній клініці перебувають хворі на гепатит A, B і C. Перед початком лікування в них було визначено активність маркерного ензиму — аланінамінотрансферази, Од/л:

A	78	75	79	81	88	91	90	65	70
B	80	79	71	85	91	84	95	71	68
C	65	80	72	74	90	82	87	22	78

Треба перевірити, чи відрізняється активність АлАТ у крові хворих із різними вірусами гепатиту.

Розв'язання. Перевіркою даних щодо нормального закону розподілу встановлено, що група показників, отриманих від хворих на гепатит C, не відповідає цьому закону розподілу, отже, подальші етапи аналізу слід проводити з непараметричними критеріями. Оскільки групи незалежні, скористаємось критерієм Краскела—Уолліса.

Спочатку згрупуємо дані й скористаємось меню **Statistics—Nonparametrics** (див. рис. 105); у вікні **Nonparametric Statistics** оберемо **Comparing multiple indep. samples (groups)** і натиснемо кнопку **OK** (див. рис. 106). У вікні **Kruskal-Wallis ANOVA and Median Test** вкажемо колонки з даними (**Dependent variable list**) і кодами груп (**Indep. (grouping) variable**), потім оберемо всі коди груп (рис. 111). Проведення тесту ініціюємо опцією **Summary: Kruskal-Wallis ANOVA & Median Test**.

Згідно з результатами виконаного аналізу (рис. 112) ($p = 0,697$), приймемо нульову гіпотезу, тобто вважатимемо, що активність АлАТ у крові хворих на гепатит A, B і C не відрізняється, а є лише маркером вірусного ураження печінки.

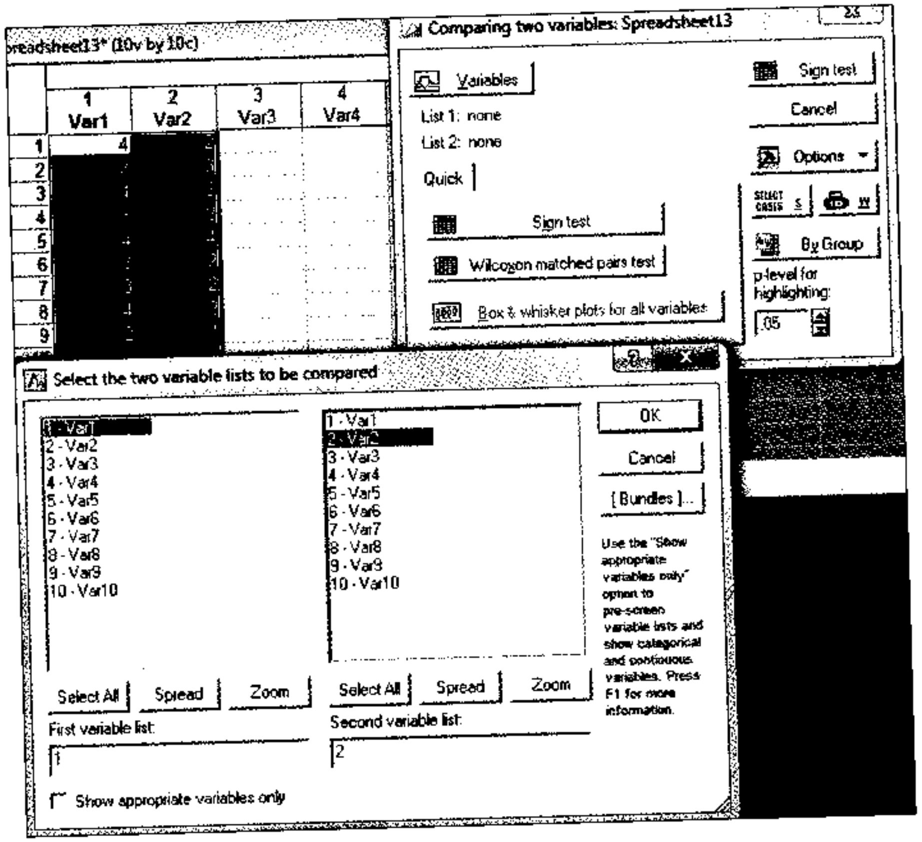


Рис. 107

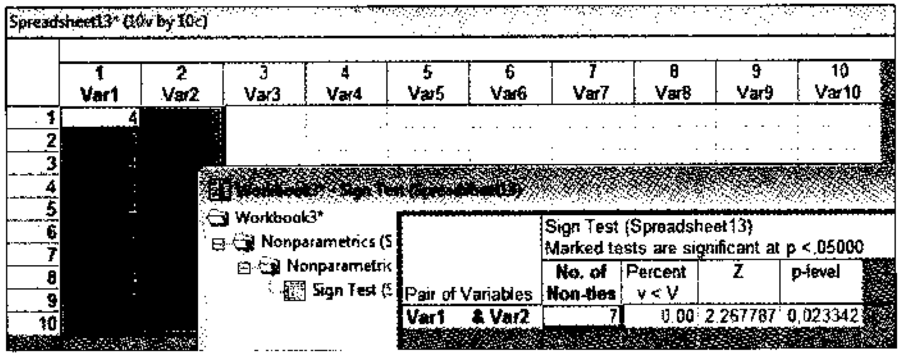


Рис. 108

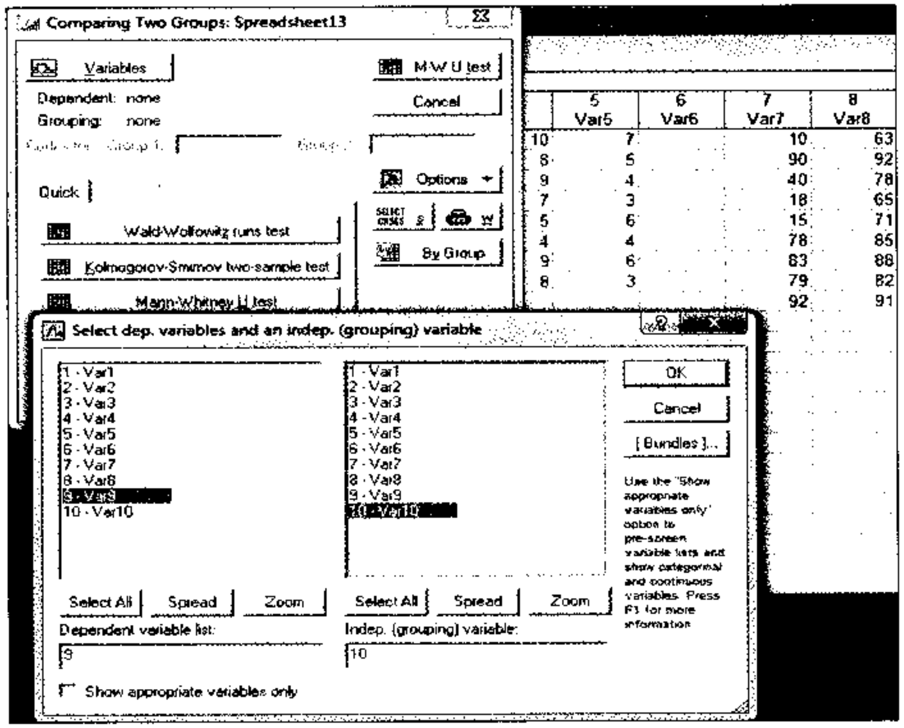


Рис. 109

Варто зазначити, що апостеріорний аналіз (опція **Multiple comparisons of mean ranks for all groups** у вікні **Kruskal-Wallis ANOVA and Median Test**, див. рис. 111) у цьому випадку проводити не потрібно.

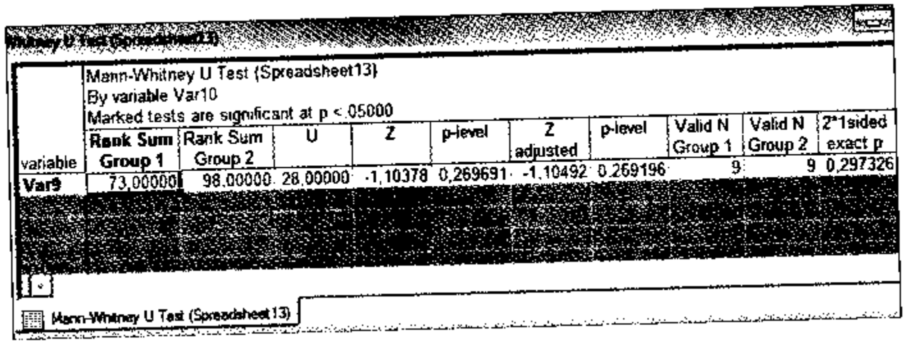


Рис. 110

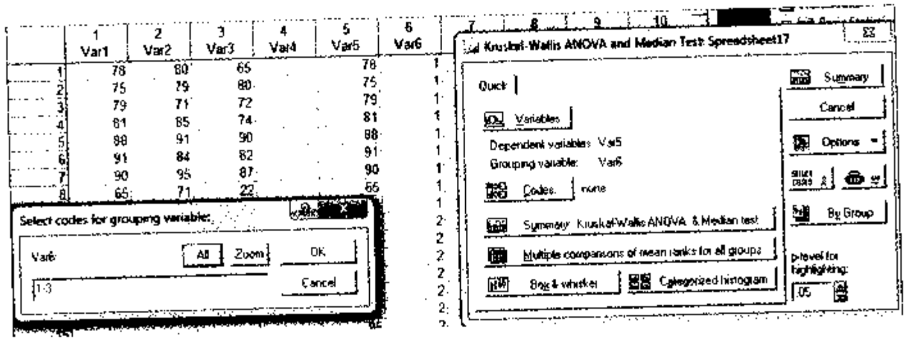


Рис. 111

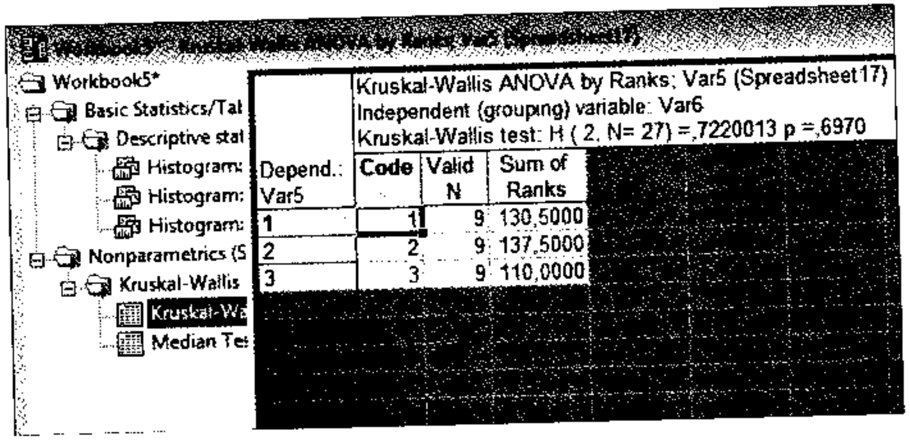


Рис. 112

Приклад 94. У групі дітей дошкільного віку, яких навчали англійській мові, спостерігали динаміку (через один, два і три місяці після початку навчання) швидкості читання (кількість слів за 3 хв). Результати наведено в таблиці:

1 місяць	10	11	11	9	8	12	15	25	10	13
2 місяці	15	25	18	18	23	35	34	39	28	29
3 місяці	41	29	35	39	37	55	42	57	38	52

Треба перевірити, чи можна вважати навчання ефективним і, якщо так, на якому етапі спостерігаються зміни.

Розв'язання. Перевіркою даних щодо нормального закону розподілу виявлено, що у випадку першої вибірки (1 місяць) нульова гіпотеза відхиляється. Отже, далі аналіз треба проводити з використанням непараметричних критеріїв. Оскільки вибірки залежні, варто застосувати критерій Фрідмена. Зазначимо, що в цьому випадку групувати дані не потрібно.

Виділимо колонки з даними і скористаємось меню **Statistics—Nonparametrics**; у вікні **Nonparametric Statistics** оберемо **Comparing multiple dep. samples (variables)** і натиснемо кнопку **OK**. У вікні **Friedman ANOVA by Ranks** ініціюємо проведення тесту опцією **Summary: Friedman ANOVA & Kendall's concordance** (рис. 113).

Згідно з результатами аналізу (рис. 114) ($p = 0{,}0005$), нульову гіпотезу відхиляємо, тобто навчання було ефективним.

Щоб з'ясувати, на якому етапі вивчення англійської мови настають зміни у швидкості читання, проведемо апостеріорний аналіз. Для цього як критерієм скористаємось парним тестом Уілкоксона (обов'язково слід враховувати поправку Бонферроні для встановлення нового рівня α; у нашому випадку $\alpha = 0{,}05/3 =$ $= 0{,}0167$) (див. рис. 114).

Як видно з отриманих результатів (рис. 115), позитивна динаміка спостерігається на всіх етапах навчання дітей.

Приклад 95. Досліджували ефективність обробки зорової та звукової інформації групою студентів (обидва показники виражали в балах від 1 до 10). Отримано такі результати:

Зорова інформація	10	9	8	8	10	9	7	5	6	8
Слухова інформація	7	7	5	6	7	10	7	4	5	7

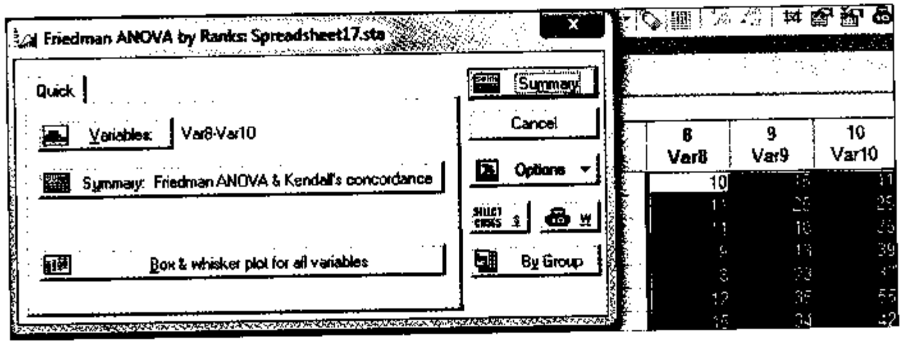


Рис. 113

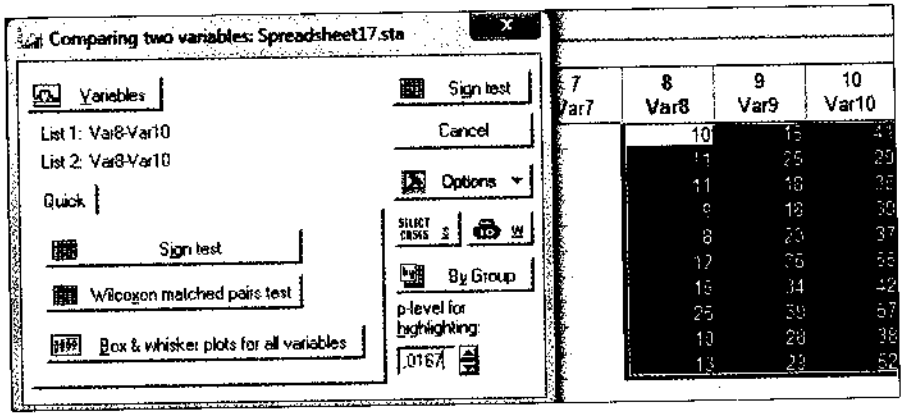


Рис. 114

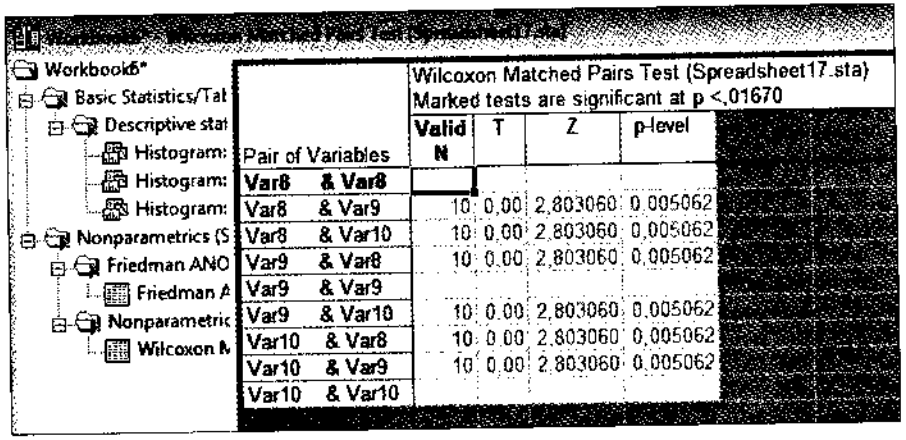


Рис. 115

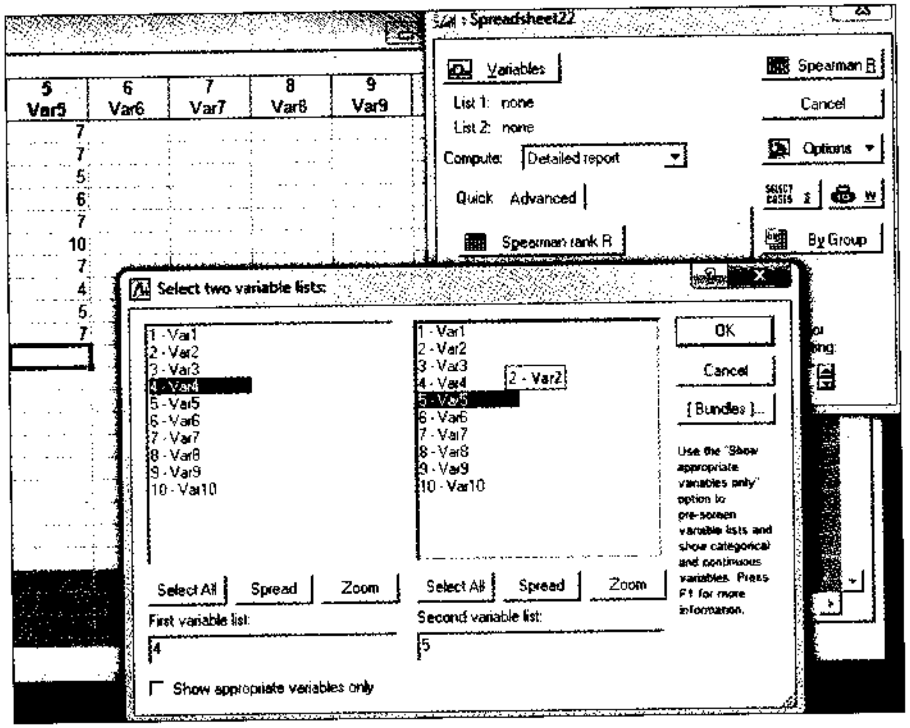


Рис. 116

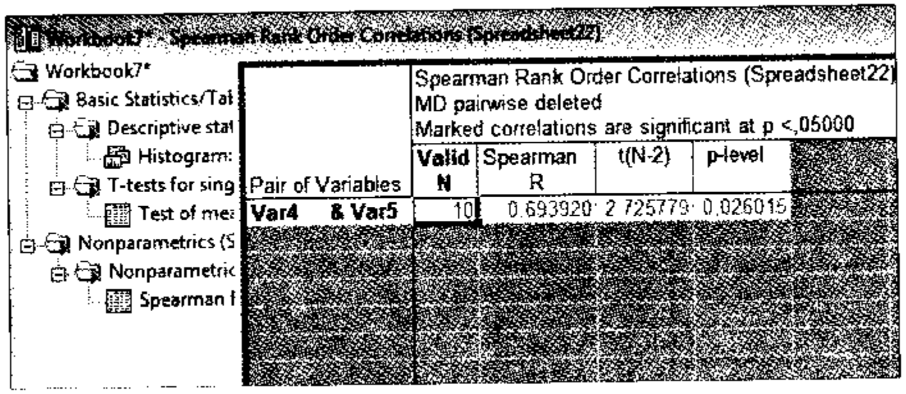


Рис. 117

Треба перевірити, чи корелюють показники ефективності обробки студентами зорової та звукової інформації.

Розв'язання. Щоб відповісти на поставлене запитання, скористаємось непараметричним критерієм Спірмена: меню **Statistics—Nonparametrics**; у вікні **Nonparametric Statistics** оберемо **Correlations (Spearman, Kendall tau, gamma)** і натиснемо кнопку **OK**. У вікні **Spreadsheet** (вкладка **Advanced**) вкажемо змінні й оберемо з випадаючого списку **Compute**: **Detailed report**; ініціюємо проведення тесту опцією **Spearman R** (рис. 116).

Як видно з отриманих результатів, показники ефективності обробки студентами зорової і звукової інформації пов'язані прямим помірним зв'язком (рис. 117).

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

А. АУДИТОРНІ ЗАВДАННЯ

1. Нехай події A і B полягають у тому, що молекули α і β здатні зв'язуватися з активним сайтом білкового комплексу. Виразити через A і B такі події: 1) C — обидві молекули зв'язалися із сайтом; 2) D — хоча б одна молекула зв'язалася із сайтом; 3) E — лише одна молекула зв'язалася із сайтом.

2. У книжковій шафі знаходяться підручники з математики, біології і хімії. Студент навмання бере 2 підручники. Скласти повну групу подій, які у цьому випадку можуть відбутися.

3. Знайти ймовірність $P(A)$, якщо відомо: $P(A \cap B) = 0{,}72$ і $P(A \cap \bar{B}) = 0{,}18$.

4. Подія A може статися за умови появи однієї з несумісних подій B_1, B_2, B_3 які утворюють повну групу подій. Відомі умовні ймовірності: $P(B_1 / A) = 0{,}6$ і $P(B_2 / A) = 0{,}3$. Знайти умовну ймовірність $P(B_3 / A)$.

5. У лабораторію доставили 24 однакові за зовнішнім виглядом ампули з вакциною, 10 з яких виготовлені на фармзаводі № 1, а інші — на фармзаводі № 2. Вакцину потрібно ввести чотирьом тваринам. Яка ймовірність того, що спочатку першим трьом тваринам буде введена вакцина, виготовлена на фармзаводі № 2, а четвертій тварині — на фармзаводі № 1?

6. Вірусологічна лабораторія перевіряє тварин на штами вірусу за двома параметрами. Встановлено, що у 8 з 25 тварин не підтверджено лише перший параметр, у 6 тварин — лише другий, у 3 тварин — обидва параметри. Навмання для досліджень беруть одну тварину. Яка ймовірність того, що вона не інфікована вірусом?

7. У лабораторії випробовують 40 ампул вакцин, виготовлених на фармзаводі № 1, 35 — на фармзаводі № 2 і 25 — на

фармзаводі № 3. Відомо, що 2, 3 і 5 % вакцин цих фармзаводів неякісні. Яка ймовірність того, що навмання взята вакцина є неякісною?

8. В обчислювальному центрі є шість персональних комп'ютерів першого класу (ПК1) і чотири — другого (ПК2). Ймовірність виходу з ладу під час роботи ПК1 дорівнює 0,05, ПК2 — 0,2. Студент виконує лабораторну роботу на обраному навмання ПК. Знайти ймовірність того, що під час роботи ПК не вийде з ладу.

9. Захворювання на гепатит вдається вилікувати у 95 % пацієнтів, причому у 80 % немає рецидивів. Яка ймовірність того, що у навмання обраного для обстеження хворого не буде рецидивів?

10. У лікарню поступило 50 % хворих на грип, 30 % — на ангіну і 20 % — на запалення легенів. Ймовірність повного одужання від грипу дорівнює — 0,7, від ангіни — 0,8, від запалення легенів — 0,9. Виписано хворого, який повністю одужав. Знайти ймовірність того, що він був хворий на грип.

11. На іспит запропоновано 30 екзаменаційних квитків, серед яких 5 "щасливих". Кому вигідніше тягнути квиток — першому чи другому студенту?

12. У рибалки є три улюблені місця, куди він приходить рибалити з однаковою ймовірністю. Ймовірність клювання на першому місці дорівнює 1/3, на другому — 1/2, на третьому — 1/4. Рибалка закинув вудку у навмання обраному місці і риба клюнула. Знайти ймовірність того, що рибалка закинув вудку на першому місці.

13. Група піддослідних тварин складається з 20 кролів, 6 щурів і 4 мишей. Ймовірність інфікуватися вірусом для кроля дорівнює 0,9, для щура — 0,8, для миші — 0,75. Знайти ймовірність того, що навмання взята тварина буде інфікованою.

14. У першій клітці знаходиться 10 щурів, з яких 7 білих; у другій — 12 щурів, з яких 10 білих; у третій — 10 щурів, з яких 5 білих. З кожної клітки втекло по одному щуру. Зловити вдалось лише одного. Яка ймовірність того, що впійманий щур виявиться білим?

15. Три групи студентів виконали однакову контрольну роботу. У першій групі, де 30 студентів, 8 робіт, виконаних на "відмінно", у другій групі, де 28 студентів — 6 таких робіт, у третій групі, де 27 студентів — 9 робіт. Знайти ймовірність того, що навмання взята робота не виконана на "відмінно" (контрольні роботи розкладені по групах).

16. Два дослідники проводять біохімічний аналіз на виявлення гепатиту А. Більш досвідчений з них обробляє в середньому 60 % аналізів за день, менш досвідчений — 40 %. Ймовірність того, що досвідчений дослідник допустить помилку при обробці аналізу дорівнює 0,03, а для менш досвідченого дослідника така ймовірність становить 0,05. Взятий на контроль аналіз виявився помилковим. Яка ймовірність того, що помилився більш досвідчений дослідник?

17. Серед донорів, що здають кров, 8 осіб мають першу групу, 6 — другу, 4 — третю і 3 — четверту. Яка ймовірність того, що із двох донорів: 1) обидва мали четверту групу; 2) хоча б один мав третю групу крові?

18. У ПК є дві ключові мікросхеми. Ймовірність поламки першої мікросхеми дорівнює 0,3, другої — 0,1 (за деякий проміжок часу). Відомо, що з ладу вийшла одна з них. Яка ймовірність того, що це друга мікросхема?

19. Скількома способами на 15 сторінках газети можна так розмістити 5 фотографій, щоб на кожній сторінці містилась лише одна фотографія?

20. У групі 15 студентів, серед яких лише три хлопці. Скількома способами можна скласти списки в деканат по: 1) 7 студентів; 2) 5 студенток; 3) 10 студентів, серед яких два хлопці?

21. У групі навчається 30 студентів. Яка ймовірність того, що хоча б у двох студентів збігаються дні народження?

22. Скільки існує способів опитування 10 студентів на одному занятті, якщо жоден з них не може бути опитаний двічі, причому порядок опитування студентів не важливий?

23. З групи піддослідних тварин, яка складається з 6 самців і 4 самок, потрібно відібрати 2 тварини. Яка ймовірність того, що навмання взяті тварини є самками?

24. У бібліотеці зберігається 15 журналів "Cell" за 2015 рік, пронумеровані від 1 до 12 (номерів 8, 9 і 10 — по два). Навмання взяли 6 журналів. Яка ймовірність того, що серед них виявляться журнали з номерами 7 і 8?

25. У клітці знаходиться 15 білих, 9 сірих і 6 чорних мишей. Для досліджень навмання взяли 6 мишей. Знайти ймовірність того, що 1 з них буде чорною, 2 — сірими і 3 — білими?

26. У кожному з 25 екзаменаційних квитків є по 2 питання, які не повторюються. Студент знає відповіді лише на 30 питань. Яка ймовірність того, що у взятому студентом екзаменаційному квитку містяться питання, на які він може відповісти?

27. У клітці є 20 кролів, з яких 5 — альбіноси. Навмання взяли 3 кролі. Яка ймовірність того, що серед них: а) усі не альбіноси; б) усі альбіноси; в) один альбінос і два не альбіноси?

28. Серед 20 пацієнтів у трьох негативний резус-фактор. Яка ймовірність того, що у двох навмання обраних для обстеження пацієнтів резус виявиться негативним?

29. У ПК три основні мікросхеми. Ймовірність безвідмовної роботи кожної з них упродовж деякого часу становить 5/6. Мікросхеми виходять з ладу незалежно одна від одної. За відмови хоча б однієї мікросхеми перестає працювати ПК. Знайти ймовірність поламки ПК.

30. Для сіяння пшениці заготовлено 95 % насіння першого ґатунку, 3 % — другого і 2 % — третього. Ймовірність того, що з насіння виросте колосся, яке містить 50 насінин, для першого ґатунку становить 0,5, для другого — 0,2, для третього — 0,1. Знайти ймовірність того, що навмання взяте колосся матиме не менш як 50 насінин.

31. У клітці знаходяться тварини двох підвидів: першого і другого, причому тварин другого підвиду в 1,5 раза більше. Знайти ймовірність того, що серед трьох навмання взятих тварин хоча б одна буде першого підвиду.

32. У папці міститься 10 файлів, серед яких 6 графічних. Навмання відкривають 5 файлів. Яка ймовірність того, що не менш як 4 з них будуть графічними?

33. Проростання пшеничного насіння становить 85 %. Знайти ймовірність того, що з 2000 посіяних насінин проросте від 1880 до 1920.

34. Ймовірність присутності одного студента на лекції дорівнює 0,8. Знайти ймовірність того, що зі 100 студентів на лекції будуть присутні не менш як 75 і не більш як 90.

35. Ймовірність носія вірусу герпесу в популяції людей становить 0,92. Яка ймовірність того, що серед 10 обстежених осіб буде виявлено більш як 6 носіїв цього вірусу?

36. Залоза складається з 4500 клітин. Ймовірність мутації клітини становить 0,005. Знайти ймовірність того, що мутують менш як 5 клітин.

37. Ймовірність виходу з ладу мікросхеми ПК упродовж 1 год роботи дорівнює 0,004. Яка ймовірність того, що за 1000 год роботи ПК доведеться 5 разів міняти мікросхему?

38. Дискретна випадкова величина X може набувати лише двох значень — x_1 і x_2 ($x_1 < x_2$). Відомі ймовірність $p_1 = 0{,}2$

появи значення x_1, математичне сподівання $EX = 3{,}8$ і дисперсія $DX = 0{,}16$. Знайти закон розподілу цієї випадкової величини.

39. Знайти числові характеристики випадкової величини *X*, яка задана законом:

X	-5	0	4	5
p	0,125	0,5	0,25	0,125

40. Функція розподілу випадкової величини *X* задана виразом

$$F(x) = \begin{cases} 0, & \text{якщо } x < -\pi/4; \\ \alpha\sin(x - \pi/4) + 1/2 & \text{за } -\pi/4 \le x \le 3\pi/4; \\ 1, & \text{якщо } x \ge 3\pi/4. \end{cases}$$

Знайти: коефіцієнт α; ймовірність потрапляння значення випадкової величини *X* внаслідок випробовування в інтервал $(\pi/4; 3\pi/4)$; щільність розподілу *X* та побудувати її графік.

41. Випадкова величина *X* задана щільністю розподілу

$$\rho(x) = \begin{cases} 0, & \text{якщо } x < 1; \\ x - 1, & \text{якщо } 1 \le x < 2; \\ 3 - x, & \text{якщо } 2 \le x < 3; \\ 0, & \text{якщо } x \ge 3. \end{cases}$$

Знайти числові характеристики цієї випадкової величини.

42. Вибіркова сукупність задана такими значеннями і частотами:

x_i	1	2	3	4
n_i	15	10	20	5

Знайти вибіркові характеристики.

43. Знайти вибіркові характеристики ряду розподілу кальцію у сироватці крові мавп і побудувати 95%-й довірчий інтервал для генеральної середньої цього розподілу:

x_i, мг	9,0	9,8	10,6	11,4	12,2	13,0	13,8	14,6
n_i	2	6	15	23	25	17	7	5

44. Випадкова величина розподілена за нормальним законом із параметром $\sigma = 2$. Зроблено вибірку об'єму $n = 25$. З довірчою ймовірністю $\gamma = 0{,}95$ знайти довірчий інтервал невідомого параметра цього розподілу.

45. Розмір тіла комах розподілений нормально з параметрами $a = 5$ см і $\sigma = 4$ см. Знайти ймовірність того, що виміряний зріст відхиляється від справжнього значення більше як на 2 см.

46. Для нормально розподіленої випадкової величини *X* знайти $P\{5 < X < 15\}$, якщо $EX = 20$ і $DX = 25$.

47. Випадкова величина розподілена за нормальним законом з математичним сподіванням $a = 8$ і середньоквадратичним відхиленням $\sigma = 1$. Знайти ймовірність того, що навмання взяті два значення величини з трьох потраплять в інтервал (8; 13).

48. За дослідними даними методом найменших квадратів знайти параметри лінійної залежності $y = ax + b$:

x_i	1	2	3	4
y_i	1,1	0,48	$-0{,}1$	$-0{,}8$

49. Побудувати лінію регресії для такої вибірки:

x_i	1	3	5	7	9
y_i	0,48	1,26	2,85	3,96	5,15

За допомогою *t*-критерію перевірити гіпотезу $H_0 : \theta_1 = 0$ з довірчою ймовірністю $\gamma = 0{,}95$.

50. Вивчено залежність між масами тіл мавп-матерів та їх новонароджених дитинчат. Проводили лабораторний нагляд за 20 мавпами. Отримано такі результати:

Номер	1	2	3	4	5	6	7	8	9	10
Маса тіла мавпи-матері x_i, кг	10	10,8	11,3	10	10,1	11,1	11,3	10,2	13,5	12,3
Маса тіла новонародженого дитинчати y_i, кг	0,7	0,73	0,75	0,7	0,65	0,65	0,7	0,61	0,7	0,63

Номер	11	12	13	14	15	16	17	18	19	20
Маса тіла мавпи-матері x_i, кг	14,5	11	12	11,8	13,4	11,4	12	15,6	13	12,1
Маса тіла новонародженого дитинчати y_i, кг	0,7	0,65	0,72	0,69	0,78	0,7	0,6	0,85	0,8	0,75

Розрахувати коефіцієнт кореляції між масою тіла мавп-матерів і масою тіла їх новонароджених дитинчат. Побудувати лінію регресії.

51. Вивчали показники АТФазної активності (швидкості гідролітичного розщеплення АТФ ферментом АТФазою, мкмоль P_i/хв) серцевого м'яза великої рогатої худоби за різних температур. Отримано такі результати:

Активність	1,08	1,1	1,12	1,14	1,15	1,25	1,36	1,38	1,4	1,42
Температура	10	12	14	16	18	20	22	24	26	28

Чи залежить активність АТФази від температури? Перевірте це і вкажіть значення використаних статистичних критеріїв.

52. Полярографічна активність фільтрату сироватки крові (F, ум. од.) на різних етапах лікування 10 хворих на гепатит А становила:

F_1	98	152	168	143	128	162	96	157	162	184
F_2	90	79	101	86	171	142	82	104	108	64
F_3	69	71	116	102	110	64	120	141	92	76
F_4	51	51	71	94	64	66	72	135	75	124

Проаналізуйте, чи залежить активність фільтрату від тривалості захворювання. Вкажіть значення використаних статистичних критеріїв.

53. Досліджували вплив кави на серцево-судинну систему у групі добровольців. Їх тестували три дні підряд так, щоб на один день припадало одне дослідження. Тиск у нормі (Н) після вживання кофеїновмісної кави (КК) і безкофеїнової кави (БК) становив:

Н	121	135	118	125	129	110	118	120	117	115
КК	126	145	137	126	137	127	126	139	143	145
БК	122	143	115	113	135	115	117	130	125	123

Перевірте, чи залежить тиск від вживання певного напою і вкажіть значення використаних статистичних критеріїв.

Б. Домашні завдання

1. Довести, що достовірною є така подія:

$$[(A\cup\overline{B})\cap(\overline{A}\cup B)]\cup[(A\cup B)\cap(\overline{A}\cup\overline{B})].$$

2. Довести, що подія $(A\cup B)\cap(\overline{A}\cup B)\cap(A\cup\overline{B})\cap(\overline{A}\cup\overline{B})$ є неможливою.

3. Нехай A, B і C — три довільні події. Опишіть події, які полягають у тому, що: а) відбулися всі три події; б) відбулася рівно одна подія; в) відбулося рівно дві події; г) відбулася хоча б одна подія; д) відбулися не більш як дві події; е) не відбулося жодної події.

4. У групі 30 студентів, із них 21 вивчає німецьку мову, 14 — англійську і 10 — обидві мови. Яка ймовірність того, що обраний навмання студент вивчає хоча б одну мову?

5. Всередину кола радіусом R навмання кинута точка. Знайти ймовірність того, що точка виявиться всередині вписаного в коло: а) квадрата; б) правильного трикутника. Припустіть, що ймовірність влучення точки в частину кола пропорційна площі цієї частини і не залежить від її розміщення всередині кола.

6. Навмання взято два додатних числа x і y, кожне з яких не більше за одиницю. Знайти ймовірність того, що сума $x+y$ не перевищує одиницю, а добуток xy не менший за $0{,}009$.

7. У змаганні беруть участь 18 команд, із них 5 команд з вищої ліги. Навмання формують дві підгрупи по 9 команд у кожній. Знайти ймовірності таких подій: а) усі команди з вищої ліги потраплять в одну й ту саму підгрупу; б) 2 команди з вищої ліги потраплять в одну з підгруп, а 3 — в іншу.

8. У ліфт дев'ятиповерхового будинку на першому поверсі зайшло 6 людей. Кожна з них з однаковою ймовірністю може вийти на будь-якому поверсі, починаючи з другого. Знайти ймовірності таких подій: а) усі люди вийдуть на четвертому поверсі; б) усі люди вийдуть на одному й тому ж поверсі; в) усі люди вийдуть на різних поверхах.

9. У групі навчаються 12 студентів, із них 7 — відмінники. За списком навмання обрано 8 студентів. Знайти ймовірність того, що серед них 5 відмінників.

10. Прилад містить три незалежно працюючі елементи. Ймовірності відмови елементів дорівнюють відповідно 0,05; 0,08; 0,10. Знайти ймовірність відмови приладу, якщо він виходить з ладу в разі відмови: а) хоча б одного елемента; б) двох елементів; в) усіх елементів.

11. Ймовірність хоча б одного влучення у ціль стрільцем, що зробив три постріли, дорівнює 0,875. Знайти ймовірність влучення в ціль при одному пострілі.

12. З першого заводу надходить 40 %, з другого — 35, з третього — 25 % компонентів фармпрепарату. Серед компонентів, виготовлених на першому заводі, 0,6 % бракованих, на другому — 0,3, на третьому — 0,5 %. Знайти ймовірність того, що виготовлений фармпрепарат бракований.

13. В акваріумі живуть 12 риб-неонів, 20 золотих рибок і 18 скалярій. Ймовірність того, що до годівниці припливе неон становить 0,9, золота рибка — 0,6, скалярія — 0,7. До годівниці підпливла риба. Яка ймовірність, що це неон?

14. Ймовірності відмови кожного з трьох незалежно працюючих приладів дорівнюють відповідно 0,3; 0,1; 0,25. Відомо, що два прилади вийшли з ладу. Знайти ймовірність того, що відмовили другий і третій прилади.

15. У діагностичний центр в однакових кількостях потрапляють пацієнти з трьох консультативних центрів. Ймовірність того, що діагноз буде підтверджений для пацієнтів з направленням першого центру становить 0,7, з другого — 0,85, з третього — 0,55. Яка ймовірність того, що пацієнт, у якого діагноз підтвердився, був направлений другим центром?

16. Для лікування певного захворювання можуть бути використані два лікарські препарати. Ймовірність позитивного результату при застосуванні першого препарату становить 0,9, другого — 0,8. Яка ймовірність того, що навмання обраний препарат приведе до позитивного результату?

17. У клітці 7 мишей лінії А і 9 — лінії В. З клітки навмання взяли 3 миші. Скласти закон розподілу числа мишей лінії В серед відібраних.

18. Ймовірність влучення в ціль при одному пострілі дорівнює 0,8. Знайти ймовірність того, що при 10 пострілах в ціль буде влучено 7 разів.

19. За даними технічного контролю, в середньому 7 % виготовлених на заводі фармпрепаратів браковані. Знайти ймовірність того, що з 8 виготовлених на заводі фармпрепаратів 2 браковані.

20. У ботанічному музеї зібрано гербарії 34 видів злакових, причому 10 із них занесені до Червоної книги. Працівник музею довільно відбирає 5 зразків з гербарію. Яка ймовірність того, що хоча б 4 з обраних зразків гербарію занесені до Червоної книги?

21. У підручнику з біофізики допущено 50 помилок на 500 сторінках. Яка ймовірність того, що у розділі з 30 сторінок допущено: а) дві помилки; б) менш як дві помилки; в) дві або більше помилок; г) не допущено жодної помилки?

22. Ймовірність того, що стрілець влучить у ціль з одного пострілу дорівнює 0,8. Стрільцю видають набої доти, доки він не схибить: а) скласти закон розподілу випадкової величини ξ — числа набоїв, виданих стрільцю; б) знайти найімовірніше число виданих стрільцю набоїв.

23. Тривалість міжміської телефонної розмови вимірюється хвилинами і становить випадкову величину з геометричним розподілом. Яка ймовірність того, що розмова триватиме ще 3 хв, якщо вона вже тривала 10 хв. Параметр геометричного розподілу вважати таким, що дорівнює $p > 0$.

24. Середнє число викликів, що надходить на АТС за одну хвилину, дорівнює двом. Знайти ймовірність того, що за чотири хвилини надійде: а) три виклики; б) менш як три виклики; в) не менше як три виклики.

25. Скільки родзинок у середньому має бути в тісті, щоб з імовірністю $p = 0,95$ бути переконаним у тому, що в кожній випічці міститься хоча б одна родзинка?

26. Сумісний розподіл випадкових величин ξ і η заданий такими даними:

η \ ξ	-1	2	3
1	0,1	0,2	0
4	0,3	0,15	0,25

Знайти закон розподілу кожної з величин ξ і η, їх математичні сподівання, дисперсії та коефіцієнт кореляції між ними.

27. Випадкова величина ξ має рівномірний розподіл із математичним сподіванням $E\xi = 1$ і дисперсією $D\xi = 3$. Знайти щільність розподілу випадкової величини ξ.

28. Дискретна випадкова величина X може набувати лише трьох цілих значень x_1 з імовірністю $p_1 = 0,1$, x_2 з імовірністю p_2 і $x_3 = 1$ з імовірністю $p_3 = 0,3$. Відомі її математичне сподівання $E\xi = 2$ і дисперсія $D\xi = 5,2$. Знайти закон розподілу цієї випадкової величини.

29. Записати функцію розподілу часу безвідмовної роботи ПК упродовж проміжку часу $x = 2$.

30. Параметри нормально розподіленої випадкової величини ξ дорівнюють $a = 20$ і $\sigma = 5$. Знайти ймовірність того, що в результаті експерименту ξ набуде значення з інтервалу $(15;\ 25)$.

31. Автомат виготовляє деталі. Деталь вважають придатною, якщо відхилення її діаметра ξ від проектного розміру за абсолютною величиною менше за $0,7$ мм. Вважаючи, що випадкова величина ξ розподілена нормально з параметром $\sigma = 0,4$ мм, знайти кількість придатних деталей серед 100 виготовлених.

32. Знайти ймовірність того, що подія A настане рівно 70 разів у 243 випробуваннях, якщо ймовірність появи цієї події у кожному випробуванні дорівнює 0,25.

33. Ймовірність влучити у мішень при одному пострілі дорівнює 0,8. Знайти ймовірність того, що при 100 пострілах в мішень буде влучено рівно 75 разів.

34. Ймовірність того, що виріб не пройде контроль, дорівнює 0,0004. Знайти ймовірність того, що з 500 виробів не пройдуть контролю не менш як 3 вироби.

35. У результаті експерименту подія A настає з імовірністю $0,001$. Експеримент повторюють 2000 разів. Яка ймовірність того, що подія A настане 3 рази?

36. Ймовірність настання події у кожному з 2100 незалежних випробувань дорівнює $0,7$. Знайти ймовірність того, що подія настане: а) не менш як 1470 і не більш як 1500 разів; б) не менш як 1470 разів; в) не більш як 1469 разів.

37. У ПК, що містить 300 деталей, застосовані деталі з ймовірністю придатності 0,8. Знайти ймовірність того, що 400 таких деталей достатньо для комплектування цього ПК.

38. Ймовірність настання події у кожному з 900 незалежних випробувань дорівнює 0,5. Знайти ймовірність того, що відносна частота настання події відхилиться від її ймовірності за абсолютною величиною не більш як на 0,02.

39. Відділ контролю перевіряє 475 виробів на брак. Ймовірність того, що виріб бракований, дорівнює 0,05. Знайти з ймовірністю 0,9426 межі, в яких знаходитиметься число бракованих виробів.

40. В умовах заданого технологічного процесу 70 % виробів випускаються першого ґатунку і 30 % — другого. Скільки треба взяти готових виробів, щоб з імовірністю 0,9973 можна було стверджувати, що частка виробів першого ґатунку серед них відрізнятиметься від ймовірності виготовлення виробу першого ґатунку за абсолютною величиною не більш як на 0,05?

41. Прилад складається з 10 незалежно працюючих елементів. Ймовірність відмови кожного елемента за час T дорівнює $0,05$. За допомогою нерівності Чебишова оцінити ймовірність того, що абсолютна величина різниці між числом приладів, що відмовили, і середнім числом відмов за час T виявиться: 1) більшою за 2; 2) меншою за 2.

42. Ймовірність настання події A в кожному зі 150 випробувань дорівнює $0,2$. Скориставшись нерівністю Чебишова, оцінити ймовірність того, що відхилення числа настання події A від математичного сподівання буде більшим за 40.

43. Побудувати полігон розподілу за такою вибіркою:

x_i	5	10	15	20	25
n_i	3	8	12	4	2

44. Побудувати гістограму за заданою вибіркою:

Розмір виробу, мм	Кількість виробів n
250—260	100
260—275	240
275—290	260
290—300	120

45. Знайти емпіричну функцію розподілу за заданою вибіркою:

x_i	1	3	4	5	7
n_i	12	14	20	24	10

46. Знайти вибіркове середнє та вибіркову дисперсію за такою вибіркою:

x_i	0	1	2	3	4
n_i	5	8	7	4	6

47. Знайти довірчий інтервал рівня $\gamma = 0,95$ для невідомого математичного сподівання нормально розподіленої випадкової величини ξ, якщо її середньоквадратичне відхилення $\sigma = 2$, вибіркове середнє $\overline{x} = 5$, об'єм вибірки $n = 36$.

48. За двома незалежними вибірками, об'єми яких $n_1 = 40$ і $n_2 = 50$, здобутих з нормальних генеральних сукупностей, знайдено вибіркові середні $\overline{x} = 130$, $\overline{y} = 140$. Дисперсії відомі: $D\xi = 80$, $D\eta = 100$. Для рівня значущості $\alpha = 0,01$ перевірте гіпотезу H_0: $E\xi = E\eta$.

ТЕСТОВІ ЗАПИТАННЯ

1. Події поділяють на

А неможливі, достовірні та випадкові
Б неможливі, протилежні та випадкові
В елементарні, достовірні та випадкові
Г прості, складні та випадкові
Д достовірні, випадкові, складні

2. Об'єднанням двох подій А і В називають таку подію, що складається з усіх елементарних подій, які

А входять як в А, так і В
Б хоча б в одну з подій А або В
В не входять ні в А, ні в В
Г входять в А, але не входять у В
Д входять у В, але не входять в А

3. Перетином двох подій А і В називають таку подію, що складається з усіх елементарних подій, які

А входять як в А, так і В
Б хоча б в одну з подій А або В
В не входять ні в А, ні у В
Г входять в А, але не входять у В
Д входять у В, але не входять в А

4. Дві події А і В називають несумісними, якщо

А їх перетин є неможливою подією
Б їх об'єднання є неможливою подією
В поява події А виключає появу події В
Г події А і В відбуваються одночасно
Д події А і В є протилежними

5. Властивістю ймовірності кожної події А є

А $P(A) = 1$
Б $0 \le P(A) \le 1$
В $P(A) = 0$
Г $0 \le P(A) \le 0,1$
Д $P(A) \ge 1/2$

6. Формулу повної ймовірності застосовують, якщо

А подія А може відбутися лише разом з однією з несумісних між собою подій H_i
Б подія А може відбутися лише за умови, що відбулася будь-яка з подій H_i
В подія А може відбутися лише разом з однією з несумісних між собою подій H_i, які в об'єднанні становлять достовірну подію
Г подія А може відбутися лише разом з однією із сумісних між собою подій H_i
Д подія А може відбутися лише за умови, що не відбулася жодна з подій H_i

7. Формула Байєса визначає

А ймовірність однієї з несумісних між собою подій H_i

Б умовну ймовірність однієї з несумісних між собою подій H_i, які в об'єднанні становлять достовірну подію

В умовну ймовірність однієї з сумісних між собою подій H_i

Г ймовірність однієї з сумісних між собою подій H_i

Д умовну ймовірність однієї з несумісних між собою подій H_i

8. Сполученням з n елементів по k елементів називають комбінації, які

А відрізняються як порядком, так і елементами

Б відрізняються лише порядком елементів

В відрізняються хоча б одним елементом

Г відрізняються порядком і не відрізняються елементами

Д відрізняються елементами і не відрізняються порядком

9. Випадкові величини поділяють на

А скінченні та неперервні

Б скінченні та нескінченні

В дискретні та неперервні

Г скінченні та дискретні

Д злічені та незлічені

10. Властивістю щільності неперервної випадкової величини є

А $\rho(x) \ge 0$

Б $\rho(x) \le 1$

В $0 \le \rho(x) \le 1$

Г $\rho(x) \to 0$

Д $\rho(x) \to -\infty$

11. Числовою характеристикою випадкової величини є

А щільність розподілу

Б функція розподілу

В середньоквадратичне відхилення

Г відносна похибка

Д коефіцієнт кореляції

12. Випадкову величину називають нормально розподіленою з параметрами m і σ^2, якщо її щільність розподілу має вигляд

А $\rho(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$

Б $\rho(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x-m}{2\sigma}\right)$

В $\rho(x) = a + bx$

Г $\rho(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma}\right)$

Д $\rho(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$

13. Випадкова величина ξ розподілена за нормальним законом з параметрами m і σ^2, якщо наближено виконується умова

А $|\xi - m| < \sigma$

Б $|\xi - m| < 2\sigma$

В $|\xi - m| < 3\sigma$

Г $\sigma < |\xi - m| < 2\sigma$

Д $\sigma < |\xi - m| < 3\sigma$

14. Числовими характеристиками для t-розподілу (розподіл Стьюдента) з n ступенями вільності є

А $m = 0,\ \sigma^2 = \frac{n}{n-2}$

Б $m \ne 0,\ \sigma^2 = \frac{n}{n-2}$

В $m = 0,\ \sigma = 0$

Г $m = 0,\ \sigma = n$

Д $m \ne 0,\ \sigma = n$

15. Числовою характеристикою вибіркової сукупності є

А вибіркове середнє

Б вибіркове значення

В вибірковий коефіцієнт

Г відносна похибка

Д функція розподілу

16. Основною властивістю коефіцієнта кореляції є

А $-1 \le r(x,y) \le 1$

Б $r(x,y) > 1$

В $r(x,y) < -1$

Г $r(x,y) \to 0$

Д $r(x,y) \to \infty$

17. За допомогою коефіцієнта кореляції можна виявити зв'язок

А експоненційний

Б будь-який

В лінійний

Г лог-лінійний

Д не виявити жодного

18. Зв'язок між змінними вважають сильним за значення вибіркового коефіцієнта кореляції

А понад 0,6

Б понад 0,9

В менш як -0,5

Г менш як 0,3

Д понад -0,5

19. Точність оцінки шуканої величини, розподіленої за нормальним законом, збільшується в разі

А зростання об'єму вибірки

Б зменшення об'єму вибірки

В незалежно від об'єму вибірки

Г зростання дисперсії

Д зменшення математичного сподівання

20. У загальному випадку довірчу межу шуканої фізичної величини визначають за залежністю

А $\Delta X = \Delta X_{\text{сист}}$

Б $\Delta X = \Delta X_{\text{вип}}$

В $\Delta X = \sqrt{\Delta X_{\text{сист}}^2 + \Delta X_{\text{вип}}^2}$

Г $\Delta X = \Delta X_{\text{сист}} + \Delta X_{\text{вип}}$

Д $\Delta X = \frac{1}{2}(\Delta X_{\text{сист}} + \Delta X_{\text{вип}})$

21. Якщо довірча ймовірність становить 0,99, то рівень значущості дорівнює

А 0,01

Б 0,05

В 1

Г 0,9

Д 0,09

22. Для рівня значущості 0,05 довірча ймовірність дорівнює

А 0,2

Б 0,05

В 0,8

Г 0,95

Д 1

23. Для перевірки гіпотези про значущість вибіркового коефіцієнта кореляції формулювання H_0 таке:

А між двома показниками існує лінійний зв'язок

Б між двома показниками немає лінійного зв'язку

В між двома показниками існує зв'язок

Г вибірки взяті з однієї генеральної сукупності або з генеральних сукупностей з однаковим типом зв'язку між показниками

Д вибірки взяті з генеральних сукупностей з різним типом зв'язку між показниками

24. Формулюванням нульової гіпотези для критерію Стьюдента є

А генеральні дисперсії однакові

Б генеральні дисперсії не однакові

В генеральні середні однакові

Г фактор не впливає на досліджений показник

Д генеральні середні не однакові

25. Щоб застосовувати *t*-критерій Стьюдента мають дотримуватись умови

А вибірки представлені кодами
Б обидва параметри вибірок отримані на одних і тих самих об'єктах досліджень
В значення в обох вибірках мають різну розмірність
Г статистично підтверджено нормальний розподіл генеральних сукупностей, з яких отримані вибірки
Д статистично підтверджено нормальний розподіл генеральних сукупностей, з яких отримано хоча б одну вибірку

26. Параметричний дисперсійний аналіз можна проводити за умов, коли

А доведено, що всі вибірки отримані з нормально розподілених генеральних сукупностей
Б відхилено гіпотезу щодо рівності дисперсій усіх сукупностей
В вибірки залежні
Г хоча б одна вибірка отримана з нормально розподіленої сукупності
Д гіпотези щодо рівності дисперсій не розглядаються

27. Непараметричний дисперсійний аналіз можна проводити за умов, коли

А усі вибірки отримані з нормально розподілених генеральних сукупностей
Б вибірки незалежні
В доведено рівність дисперсій усіх сукупностей
Г відхилено гіпотезу щодо рівності дисперсій
Д вибірки великого об'єму

28. Метод Плохинського застосовують для

А перевірки даних на їх належність до нормально розподілених сукупностей
Б оцінювання впливу фактора на досліджуваний показник (у дисперсійному аналізі)
В тестування на рівність генеральних середніх
Г перевірки зв'язку між показниками
Д розрахунку *P*-квартиля

Відповіді на тестові запитання

Запитання	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Відповідь	А	Б	А	А	Б	В	Б	В	В	А	В	А	В	А
Запитання	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Відповідь	А	А	В	Б	А	В	А	Г	Б	В	Г	А	Г	Б

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. *Боровиков В.П.* Statistica. Статистический анализ и обработка данных в среде Windows / В.П. Боровиков, И.П. Боровиков. — М.: Информ.-изд. дом "Филин", 1997. — 608 с.
2. *Гланц С.* Медико-биологическая статистика / С. Гланц. — М.: Практика, 1998. — 459 с.
3. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике / В.Е. Гмурман. — М.: Высш. шк., 2004. — 404 с.
4. *Гмурман В.Е.* Теория вероятностей и математическая статистика / В.Е. Гмурман. — М.: Высш. шк., 2003. — 479 с.
5. *Гончаров А.* Microsoft Excel 97 в примерах / А. Гончаров. — СПб: Питер, 1997. — 328 с.
6. *Зайдель А.Н.* Погрешности измерений физических величин / А.Н. Зайдель. — Л.: Наука, 1985. — 112 с.
7. *Ільченко О.В.* Конспект лекцій з курсу "Основи теорії ймовірностей та математичної статистики" / О.В. Ільченко, С.В. Тищенко. — К.: Видав.-поліграф. центр "Київський університет", 2005. — 99 с.
8. *Лакин Г.Ф.* Биометрия / Г.Ф. Лакин. — М.: Высш. шк., 1990. — 352 с.
9. *Прилуцький Ю.І.* Біометрія: Методичні вказівки для студентів біологічного факультету / Ю.І. Прилуцький, О.В. Оглобля, Ю.П. Скляров. — К.: Видав.-поліграф. центр "Київський університет", 2003. — 74 с.
10. *Реброва С.В.* Статистический анализ медицинских данных. Применение пакета прикладных программ Statistica / С.В. Реброва, О.Ю. Орлова. — М.: МедиаСфера, 2002. — 312 с.
11. *Урбах В.Ю.* Статистический анализ в биологических и медицинских исследованиях / В.Ю. Урбах. — М.: Высш. шк., 1975. — 297 с.
12. *Черняк О.І.* Теорія ймовірностей та математична статистика: збірник задач: навч. посібник / О.І. Черняк, О.М. Обушна, А.В. Ставицький. — К.: Знання, КОО, 2002. — 199 с.
13. *Юнкеров В.И.* Математико-статистическая обработка данных медицинских исследований / В.И. Юнкеров, С.Г. Григорьев. — СПб: ВМедА, 2002. — 266 с.

ЗМІСТ

Навчальне видання

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ БІОХІМІЇ ім. О.В. ПАЛЛАДІНА

ПРИЛУЦЬКИЙ Юрій Іванович
ІЛЬЧЕНКО Олександр Вадимович
ЦИМБАЛЮК Ольга Володимирівна
КОСТЕРІН Сергій Олексійович

СТАТИСТИЧНІ МЕТОДИ В БІОЛОГІЇ

Київ, Науково-виробниче підприємство
«Видавництво "Наукова думка" НАН України», 2017

Художнє оформлення *М.А. Панасюк*
Технічний редактор *Т.С. Березяк*
Коректор *О.Є. Челок*
Оператори *В.Г. Каменькович, О.О. Пономаренко*
Комп'ютерна верстка *Т.О. Ценцеус*

Підп. до друку 25.12.2017. Формат 60×90/16. Папір офс. № 1.
Гарн. Таймс. Друк офс. Ум. друк. арк. 13,5. Ум. фарбо-відб. 14,0.
Обл.-вид. арк. 12,0. Тираж 200 прим. Зам. № ДФ 712

Оригінал-макет виготовлено
у НВП «Видавництво "Наукова думка" НАН України»
Свідоцтво про внесення суб'єкта видавничої справи
до Державного реєстру видавців, виготівників
і розповсюджувачів видавничої продукції
ДК № 2440 від 15.03.2006 р.
01601 Київ 1, вул. Терещенківська, 3

ПП «Видавництво "Фенікс"»
03680 Київ 680, вул. Шутова, 13б
Свідоцтво про внесення до Державного реєстру
серія ДК № 271 від 07.12.2000 р.