МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

В. М. Горбачук,
О. І. Кушлик-Дивульська

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Підручник

Затверджено Вченою радою КПІ ім. Ігоря Сікорського
як підручник для здобувачів ступеня бакалавра
за технічними та економічними спеціальностями

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2023

Рецензенти: Станжицький О. М., *доктор фіз.-мат. наук, професор,*
механіко-математичний факультет,
Київський національний університет імені Тараса Шевченка,
Пашко А. О., *доктор фіз.-мат. наук, професор,*
факультет комп'ютерних наук та кібернетики,
Київський національний університет імені Тараса Шевченка,
Крамаренко Т. Г., *канд. педаг. наук, доцент,*
фізико-математичний факультет,
Криворізький державний педагогічний університет

Відповідальний редактор Дудкін М. Є., *доктор фіз.-мат. наук, професор*

Гриф надано Вченою радою КПІ ім. Ігоря Сікорського
(протокол № 8 від 12.12.2022 р.)

Навчальне видання містить основні поняття, означення, формули, теореми з доведеннями і поясненнями теорії ймовірності та математичної статистики. Матеріал подано за розділами: «Випадкові події», «Випадкові величини», «Елементи математичної статистики», «Аналіз даних за допомогою пакету STATISTICA». В розділі «Випадкові величини» розглянуто основні закони розподілу дискретних та неперервних випадкових величин, також вивчаються багатовимірні випадкові величини. Значну увагу приділено застосуванням можливостей пакету Microsoft Excel для вирішення прикладних практичних завдань. Запропоновано для розв'язування задач роботу з пакетом STATISTICA.

Підручник відповідає навчальній програмі дисципліни «Теорія ймовірності» та «Вища математика» підготовки здобувачів вищої освіти ступеня бакалавр за спеціальностями 143 «Атомна енергетика», 186 «Видавництво та поліграфія», 073 «Менеджмент». Також буде корисним для науковців, викладачів, магістрантів, аспірантів, студентів закладів вищої освіти технічних та економічних спеціальностей, школярів та абітурієнтів, які вивчають теорію ймовірностей та математичну статистику.

Реєстр. № П 22/23-020. Обсяг 18,0 авт. арк.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
https://kpi.ua
Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

Зміст

Передмова

Теорія ймовірностей та математична статистика – математичні науки, які вивчають закономірності в масових випадкових явищах. Теорія ймовірностей – спеціальний розділ вищої математики, що вивчає закономірності випадкових явищ: випадкові події, випадкові величини, їхні функції, властивості та операції над ними. Математична статистика на основі експериментальних даних вивчає ймовірнісні закономірності масових явищ, її предметом є розробка методів збору та обробки статистичних даних для одержання наукових та практичних висновків.

Мета цього навчального видання – ознайомити здобувачів вищої освіти (студентів економічних і технічних спеціальностей), а також усіх зацікавлених осіб, з основними поняттями основ теорії ймовірностей: методами, теоремами та формулами теорії ймовірностей, також елементами математичної статистики: теорії оцінювання невідомих параметрів, перевірки статистичних гіпотез, елементами кореляційно-регресійного аналізу.

Підручник містить теоретичний матеріал, основну частину теорем та формул доведено. Наведено значну кількість розв’язування прикладів та прикладних задач, зокрема, більшість показано із використанням програмного пакету Microsoft Excel. Для практичного вирішення статистичних задач, які є трудомісткими через значну кількість математичних обчислень, показано використання програмного пакета STATISTICA. Викладання цього матеріалу розраховане на широке коло користувачів, які мають доступ до програмного забезпечення і початкові необхідні знання з теорії ймовірностей, зокрема елементів комбінаторики. Власне за допомогою таких знань [10], їх вміння реалізовувати, можливо розробити алгоритми для мінімізації похибок вимірювань, що є напрямком роботи фахівців, які працюють в сучасних лабораторіях [2], [5], впроваджують в життя сучасні інформаційні технології.

Достатньо і доступно викладено основний теоретичний матеріал, передбачений програмою для технічних спеціальностей вищих навчальних

закладів, який апробовано для підготовки студентів спеціальностей 143 «Атомна енергетика», «186 «Видавництво та поліграфія», також для економічної спеціальності 073 «Менеджмент».

Підручник складається з двох частин, чотирьох розділів, кожен з яких подано за темами. Розподіл матеріалу за темами, питаннями та підпитаннями дає змогу виділити головне, закцентувати на ньому увагу. Для зручності нумерацію формул, прикладів, рисунків, таблиць виконано окремо для кожного розділу. Після кожної теми є запитання для самоконтролю (теоретичні питання та задачі для самостійного розв'язування), також типові завдання для індивідуальної роботи. Додаток містить усі таблиці, необхідні для розв'язання задач. Зміст і структура відображають новітні тенденції у питаннях навчання та підготовки кваліфікованих спеціалістів.

Підручник розроблено на основі матеріалу навчального посібника «Теорія ймовірності та математична статистика» [6] (2012), рекомендованого Міністерством освіти і науки як навчальний посібник для студентів технічних та економічних спеціальностей вищих навчальних закладів, буде доцільним використання практикумів для розв'язування задач [7], [8].

Автори вважають, що підручник буде корисний не тільки для студентів технічних та економічних спеціальностей різних форм навчання, але й для фінансистів, бізнесменів, співробітників податкової інспекції та страхових компаній, соціологів та політологів.

Частина І. Теорія ймовірностей

Розділ 1. Випадкові події

1.1. Основні поняття теорії ймовірностей та комбінаторики

Вивчення курсу «Теорія ймовірностей» передбачає знання основних розділів курсу «Вища математика». Початковий розділ, в якому розглянуто означення основних понять теорії ймовірностей, вимагає знання алгебри висловлювань, теорії множин, комбінаторики.

1.1.1. Коротка історична довідка

Теорія ймовірностей – математична наука, що вивчає закономірності в масових випадкових явищах.

Основні поняття, методи, теореми та формули теорії ймовірностей ефективно використовують у науці, техніці, економіці, зокрема теоріях надійності та масового обслуговування, в плануванні та організації виробництва, страховій та податковій справах, соціології та політології, демографії та охороні здоров’я.

Теорія ймовірностей як наука зародилась в середині XVII ст., хоча перші роботи, в яких з’явились основні поняття теорії ймовірностей (XV – XVI ст.), виникли як спроби побудови теорії азартних ігор і належать таким видатним вченим, як Б. Спіноза, Дж. Кардано, Г. Галілей. Наступний етап розвитку теорії ймовірностей пов’язаний з роботами Б. Паскаля, П. Ферма, К. Гаусса, С. Пуассона, А. Муавра, П. Лапласа, Т. Байєса. Подальші успіхи розвитку теорії ймовірностей пов'язані з іменем Я. Бернуллі (1654-1918), який зробив перші теоретичні обґрунтування зібраних раніше матеріалів (закон великих чисел). Аксіоматичне означення ймовірності (1933) належить А. М. Колмогорову.

Вимоги різних технічних потреб суспільства дали поштовх до розвитку низки прикладних розділів теорії ймовірностей. Це статистична теорія зв’язку, теорія інформації, теорія масового обслуговування, теорія надійності, економетрія, математична статистика тощо.

Відомі слова засновника української школи теорії ймовірностей академіка Б. В. Гнєденка: «Теорія ймовірностей з дедалі більшою швидкістю почала посідати основні позиції в прикладній математиці й одночасно зміцнювати своє становище як одна з центральних математичних дисциплін».

1.1.2. Елементи комбінаторики

Комбінаторикою називають розділ математики, що вивчає правила групування елементів множини. Ров'язання задач ґрунтується на основному правилі та принципах комбінаторики [3], конкретних способах підрахунку кількості підмножин, які можна утворити з певної множини елементів, що розглянемо далі.

1.1.2.1. Два основні принципи комбінаторики

Основне правило комбінаторики. Між скінченними множинами A та B можна встановити взаємно однозначну відповідність тоді і тільки тоді, коли вони мають однакову кількість елементів.

Основне правило комбінаторики дозволяє виконувати обчислення кількості елементів множини фактично зведенням до множини з відомою кількістю елементів або до множини, для якої кількість можна підрахувати дещо простіше. У більшості випадків його використання не акцентується, а приймається як дещо зрозуміле.

Значна кількість формул і теорем комбінаторики ґрунтуються на двох основних елементарних принципах, які називаються принципами суми та добутку.

Принцип суми. Нехай $n(A)$, $n(B)$ – кількості елементів скінченних неперетинних множин A та B відповідно. Тоді для множини $C = A \bigcup B$ кількість елементів

$$n(C) = n(A) + n(B). \qquad (1.1)$$

Принцип добутку. Нехай потрібно послідовно виконати k дій, причому першу дію можна виконати n_1 способом, після чого другу дію – n_2 способами і

так далі до *k*-ї дії, яку можна виконати n_k способами. Тоді всі *k* дій можна виконати $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ способами.

1.1.2.2. Розміщення, перестановки, комбінації

Нехай *M* – множина, що містить *n* елементів.

Означення. **Розміщенням** із *n* елементів по *k* називають довільну впорядковану підмножину з *k* елементів із множини *M*.

Два розміщення вважають різними не тільки тоді, коли вони складаються з різних елементів, але й тоді, коли вони складаються з однакових елементів, але відрізняються порядком їх розташування.

Кількість розміщень із *n* елементів по *k* $(k \le n)$ позначають A_n^k і обчислюють за формулою

$$A_n^k = \frac{n!}{(n-k)!} = n(n-1)(n-2)\cdot\ldots(n-k+1). \qquad (1.2)$$

Означення. **Розміщенням з повтореннями** з *n* елементів по *k* називають довільну впорядковану підмножину з *k* елементів із множини *M* (елементи не обов'язково різні).

Кількість розміщень із повтореннями з *n* елементів по *k* обчислюють за формулою

$$\overline{A_n^k} = n^k. \qquad (1.3)$$

Зауваження. $\overline{A_n^k}$ – це кількість способів, якими можна розкласти *k* різних предметів у *n* ящиків.

Приклад 1.1. Скласти всі двозначні числа з трьох цифр 3, 5, 7 (числа з однаковими цифрами не враховувати).

Розв'язання. Маємо 35, 37, 53, 57, 73, 75. Загальну кількість різних чисел можна підрахувати за формулою (1.2): $A_3^2 = 3 \cdot 2 = 6.$

Приклад 1.2. У ліфт шістнадцятиповерхового будинку зайшло на першому поверсі 10 осіб. Скількома способами вони можуть вийти з ліфта, починаючи з другого поверху?

Розв'язання. Задача зводиться до розкладання 10 предметів у 15 ящиків. Кількість таких способів дорівнює $\overline{A_{15}^{10}} = 15^{10}$.

Означення. Розміщення з *n* елементів по *n* називають **перестановками.**

Різні перестановки відрізняються лише порядком розташування елементів. Кількість перестановок з *n* елементів дорівнює добутку всіх натуральних чисел від 1 до *n*

$$P_n = 1 \cdot 2 \cdot \ldots \cdot n = n! \qquad (1.4)$$

Означення. **Перестановками з повтореннями** називають розміщення з *n* елементів, які мають однотипні елементи.

Кількість різних перестановок з повторенням, які можна утворити з *n* елементів, серед яких є k_1 елементів першого типу, k_2 елементів другого типу, …, k_m елементів *m*-го типу,

$$P_n(k_1, k_2, \ldots, k_m) = \frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_m!}. \qquad (1.5)$$

Приклад 1.3. Скількома способами можна скласти список з п'яти студентів?

Розв'язання. За формулою (1.4) отримаємо:

$$P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120.$$

Приклад 1.4. Скільки різних «слів» можна утворити перестановкою літер у слові «математика»?

Розв'язання. Слово «математика» містить *n* = 10 літер. Літер одного типу: «м» – дві, «а» – три, «т» – дві, «е» – одна, «и» – одна, «к» – одна. За формулою (1.5) отримаємо:

$$P_{10}(2,\ 3,\ 2,\ 1,\ 1,\ 1) = \frac{10!}{2! \cdot 3! \cdot 2! \cdot 1! \cdot 1! \cdot 1!} = 151\,200.$$

Означення. **Комбінацією** (сполукою) з *n* елементів по *k* називають довільну підмножину з *k* елементів із множини *M*, яка відрізняється одна від одної принаймні одним елементом.

Порядок розташування елементів у комбінаціях не має значення. Кількість комбінацій з *n* елементів по *k* $(k \le n)$ позначають C_n^k та обчислюють за формулою

$$C_n^k = \frac{n!}{k!(n-k)!}. \tag{1.6}$$

***Означення*. Комбінацією з повтореннями** з n елементів по k називають довільну множину з k елементів із множини M (елементи не обов'язково різні).

Кількість комбінацій з повтореннями з n елементів по k обчислюють за формулою

$$\overline{C_n^k} = C_{n+k-1}^k = \frac{(n+k-1)!}{k!(n-1)!}. \tag{1.7}$$

***Зауваження*.** $\overline{C_n^k}$ – це кількість способів, якими можна розкласти k однакових предметів по n ящиках.

Приклад 1.5. На конференції присутні 30 студентів. Скількома способами можна обрати президію у складі трьох осіб?

Розв'язання. Кількість способів дорівнює кількості комбінацій із 30 елементів по 3:

$$C_{30}^3 = \frac{30!}{3!\cdot 27!} = \frac{30\cdot 29\cdot 28}{1\cdot 2\cdot 3} = 4060.$$

Приклад 1.6. У кондитерській є 6 різних сортів тістечок. Скількома способами можна купити 8 тістечок?

Розв'язання. Шукане число

$$\overline{C_6^8} = C_{13}^8 = C_{13}^5 = \frac{13!}{5!\cdot 8!} = \frac{13\cdot 12\cdot 11\cdot 10\cdot 9}{1\cdot 2\cdot 3\cdot 4\cdot 5} = 1287.$$

Для обчислення цих сполук в пакеті Excel вбудовані наступні функції: **ФАКТР, ЧИСЛОКОМБ, ПЕРЕСТ,** перші дві з яких знаходяться в категорії функцій *математичні,* а третя − *статистичні.*

Приклади виконання завдань із використанням пакету Excel

Приклад 1.7. Скільки перестановок можна утворити із трьох букв А, Б та В?

Розв'язання. Можна утворити шість перестановок: АБВ, АВБ, ВАБ, ВБА, БАВ, ВБА, якщо букви не повторюються, тоді маємо 3!=6.

Факторіал можна обчислити, використовуючи функцію Excel **ФАКТР**(число), яка активізується за допомогою команд

Вставка⇒Функція⇒Математические⇒ФАКТР.

Відкриється діалогове вікно (рис. 1.1), де в поле «число» — потрібно вказати невід'ємне ціле число (n), факторіал якого обчислюється.

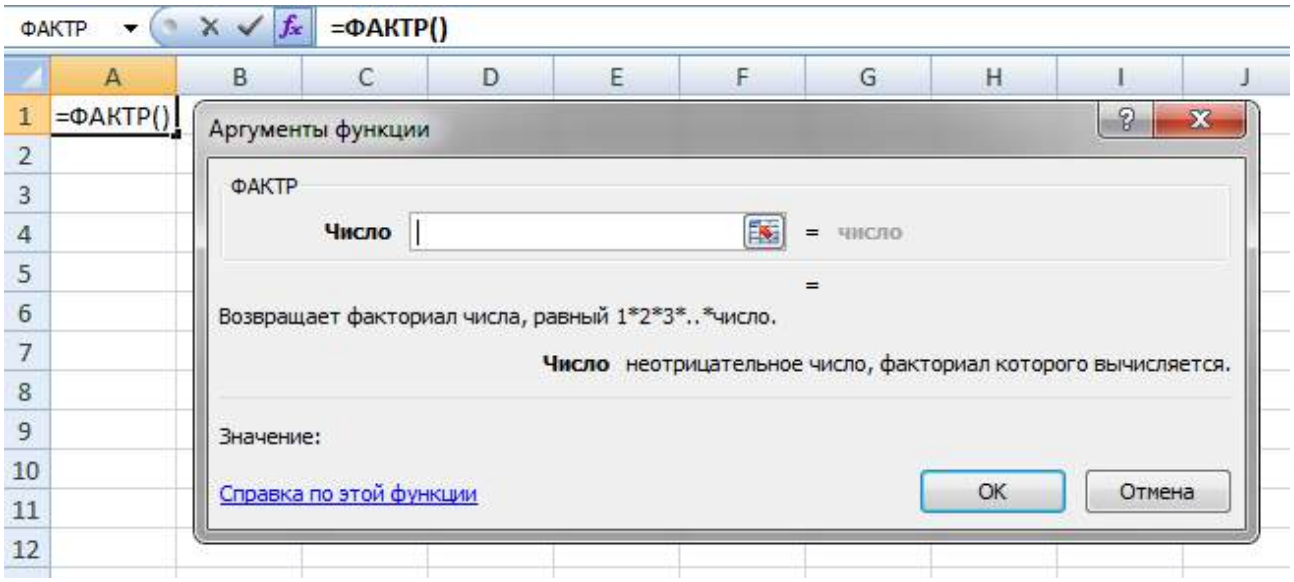


Рис. 1.1. Обчислення кількості перестановок

Приклад 1.8. Скільки перестановок по дві букви можна скласти із трьох букв А, Б і В?

Розв'язання. Із трьох букв А, Б і В можна скласти шість перестановок по дві букви: АБ, БА, АВ, ВА, БВ, ВБ. За формулою (1.1) знаходимо $A_n^k = 3 \cdot (3-1) = 6.$

Обчислення числа розміщень можна здійснити за допомогою функції **ПЕРЕС**Т(число; число_выбранных), яку викликають за допомогою команд **Вставка⇒ Функція⇒Статистические⇒ПЕРЕСТ.**

Відкриється діалогове вікно (рис. 1.2), де в поле «число» — необхідно вказати невід'ємне ціле число, яке задає кількість елементів (n), а в поле «число_выбранных» − ціле число, що задає кількість елемент у кожному розміщенні (k).

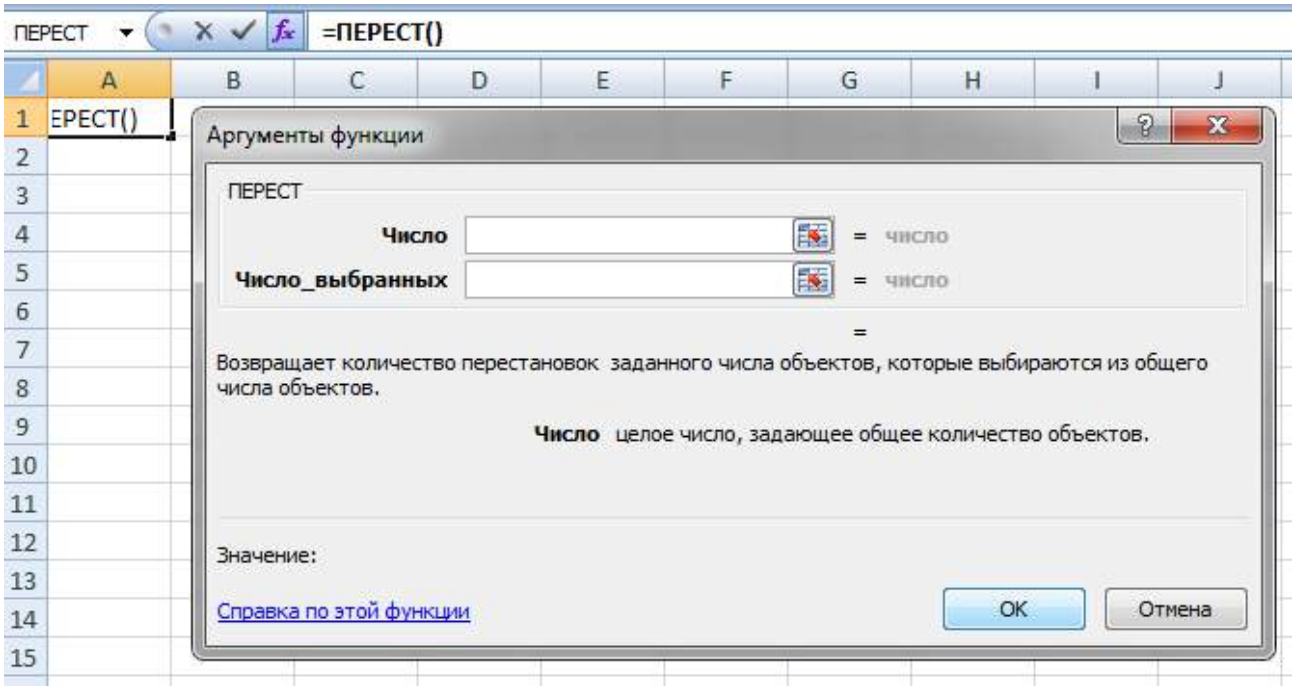


Рис. 1.2. Обчислення розміщень

Приклад 1.9. Скільки комбінацій по дві букви можна скласти із трьох букв А, Б і В?

Розв'язання. Із трьох букв А, Б і В по дві букви, без врахування порядку розташування, можна скласти всього три комбінації: АБ, АВ, БВ. За формулою (1.5) знаходимо: $C_3^2 = \frac{3!}{2!(3-2)!} = 3$.

Обчислення кількості комбінацій можна обчислити за допомогою функції **ЧИСЛОКОМБ** (число; число_выбранных) (рис. 1.3), яку викликають послідовністю команд

Вставка⇒ Функція⇒ Математические⇒ЧИСЛОКОМБ.

В діалоговому вікні «Аргументы функции» в поле «число» − вписують невід'ємне ціле число, яке задає кількість елементів (n), в поле «число_выбранных» − ціле число, що задає кількість елементів у кожній комбінації (k).

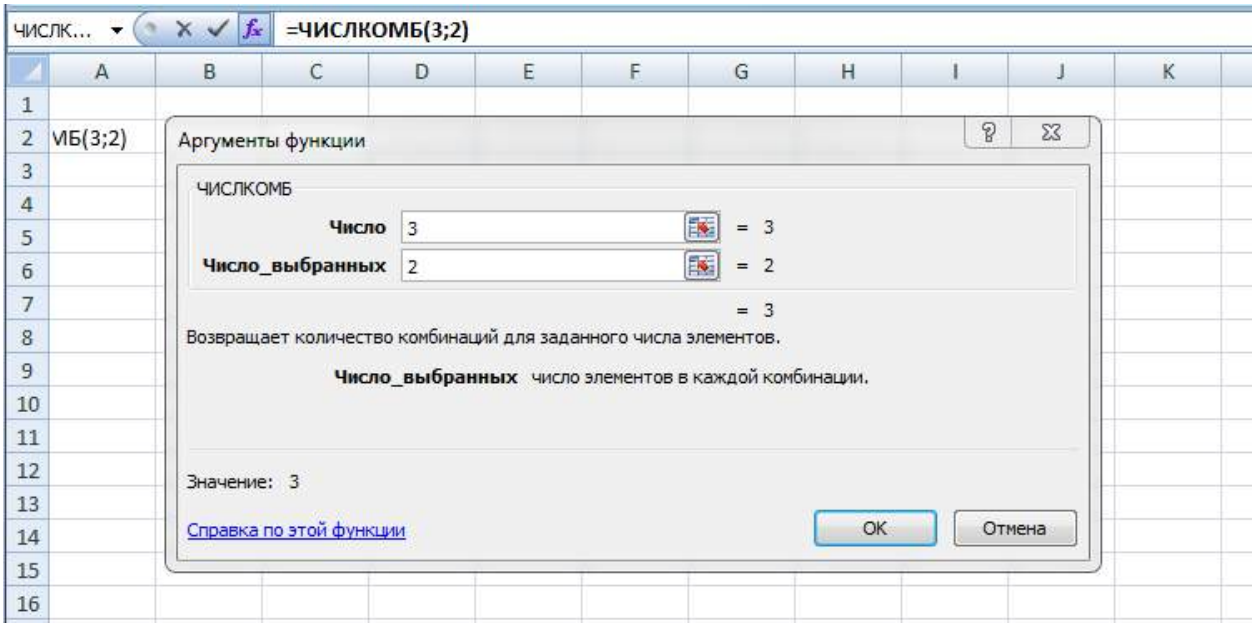


Рис. 1.3. Обчислення кількості комбінацій

Зауваження. Розв'язуючи комбінаторну задачу, перш за все потрібно відповісти на запитання − з яким із основних понять в даній ситуації ми маємо справу? Для цього відповідаємо на два питання:

1. Усі елементи множини використовуються чи ні? Якщо використовуються всі, то це перестановка.

2. Важливий порядок розміщення елементів або ні? Якщо порядок важливий, то це розміщення. У протилежному випадку − комбінація.

1.1.3. Простір елементарних подій. Випадкові події та операції над ними

За допомогою теорії ймовірностей та математичної статистики в науці та техніці досліджують випадкові явища і події та їх закономірності.

1.1.3.1. Простір елементарних подій

Означення. Експерименти, які в незмінних умовах можна повторити будь-яку кількість разів, але результати яких непередбачувані, у теорії ймовірностей називають **випробуваннями** (стохастичними експериментами).

Означення. **Елементарною подією** називають найпростіший результат випробування в певному експерименті, позначають ω.

Означення. Сукупність усіх можливих елементарних подій випробування називають **простором елементарних** подій і позначають Ω.

Означення. Підмножина A простору елементарних подій називається **випадковою подією**.

Отже, з кожним стохастичним експериментом пов'язують простір елементарних подій Ω (сукупність всіх можливих результатів випробування) та підмножини цього простору – випадкові події, які належать цьому простору.

1.1.3.2. Випадкові події та операції над ними

Перш ніж продовжити викладання понять теорії ймовірностей, потрібно розглянути деякі важливі співвідношення між випадковими подіями [10]. Для випадкових подій вводять операції, які можна подати мовою операцій із множинами.

Означення. **Сумою** *(об'єднанням)* подій A і B називають подію $A+B$ $(A \cup B)$, яка настає тоді, коли настає принаймні одна з подій A чи B (елементарні події, сприятливі або для події A, або для події B, або для обох подій A і B). На рис. 1.4 за допомогою діаграми Венна зображено суму подій.

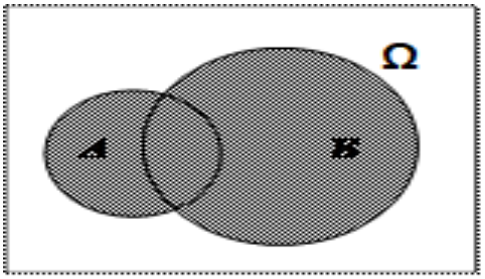


Рис. 1.4. Сума подій

Означення. **Добутком** *(перетином)* A і B (рис. 1.5) називають подію AB $(A \cap B)$, яка настає тоді, коли настають обидві події A і B (елементарні події, які сприятливі і для події A, і для події B).

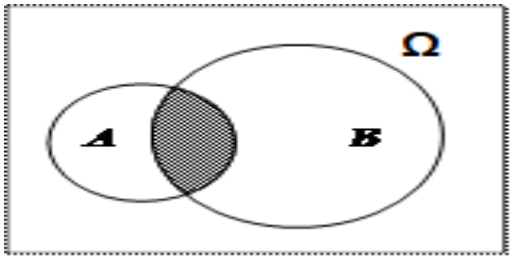


Рис. 1.5. Перетин подій

Означення. Події A і B називають **несумісними**, якщо $A \cap B = \varnothing$. Несумісні події не можуть відбуватися одночасно.

Означення. **Різницею** подій A і B (рис. 1.6) називають подію $A \setminus B$, яка настає тоді, коли настає подія A і не настає подія B.

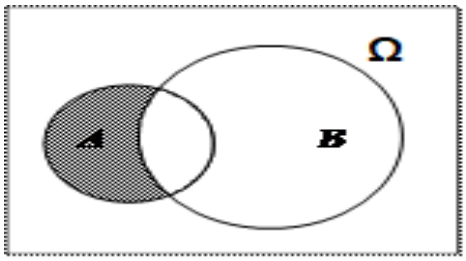


Рис. 1.6. Різниця подій

Означення. Подію $\overline{A}$ називають **протилежною** до події A (рис. 1.7), якщо $\overline{A}$ настає тоді, коли не настає A (сприятливими для $\overline{A}$ є всі елементарні події, які не входять в A).

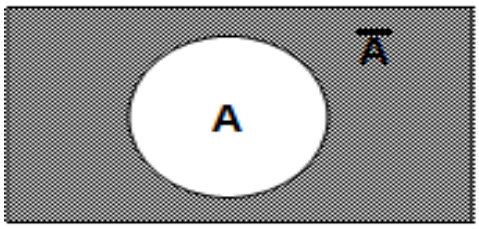


Рис. 1.7. Протилежні події

Властивості операцій над подіями

1. $A \cup \overline{A} = \Omega,\ A \cap \overline{A} = \varnothing,\ \overline{A} = \Omega \setminus A.$
2. $A \cup B = B \cup A,\ A \cap B = B \cap A.$
3. $(A \cup B) \cup C = A \cup (B \cup C),\ (A \cap B) \cap C = A \cap (B \cap C).$
4. $(A \cup B) \cap C = (A \cap C) \cup (B \cap C).$
5. $A \cap B \cup C = (A \cup C) \cap (B \cup C).$
6. $\overline{A \cup B} = \overline{A} \cap \overline{B},\ \overline{A \cap B} = \overline{A} \cup \overline{B}$ (закони де Моргана).

Доведення 5.

Нехай $\omega \in A \cap B \cup C$. Покажемо, що $\omega \in (A \cup C) \cap (B \cup C)$. Якщо $\omega \in A \cap B \cup C$, то $\omega \in A \cap B$ або $\omega \in C$. У першому випадку $\omega \in A$ і $\omega \in A$, отже, $\omega \in A \cup C, \omega \in B \cup C,$ а тому $\omega \in (A \cup C) \cap (B \cup C)$. У другому випадку $\omega \in C$, отже, $\omega \in A \cup C$ і $\omega \in B \cup C,$ і знов $\omega \in (A \cup C) \cap (B \cup C)$.

Нехай тепер $\omega \in (A \cup C) \cap (B \cup C)$. Покажемо, що $\omega \in A \cap B \cup C$. Якщо $\omega \in (A \cup C) \cap (B \cup C)$, то $\omega \in A \cup C$ і $\omega \in A \cup B$. За умови $\omega \in C$ маємо $\omega \in A \cap B \cup C$.

Якщо $\omega \notin C$, то $\omega \in A$, $\omega \in B$, тому $\omega \in A \cap B$ і, отже, $\omega \in A \cap B \cup C$. Властивість доведено.

Основні поняття теорії ймовірностей можна представити на мові теорії множин у вигляді таблиці.

Теорія множин	Теорія ймовірності
Множина Ω	Ω–простір елементарних подій
Множина Ω	Ω – достовірна подія – подія, яка настає в кожному проведенні експеримента
$\varnothing$ – порожня множина	$\varnothing$ – неможлива подія – подія, яка не настає в будь-якому проведенні експеримента
$A \subset B$	Подія A сприяє події B
$A \cup B$ (сума множин A та B)	$A \cup B$ – подія, яка полягає в тому, що настає по крайній мірі одна із подій A або B
$A \cap B$	$A \cap B$ – подія, яка полягає в тому, що настає і подія A, і B
$\bar{A} = \Omega \setminus A$	$\bar{A}$ – протилежна до A подія, яка полягає в тому, що подія A не настане
$A \cap B = \varnothing$ (A та B – множини, які не перетинаються)	$A \cap B = \varnothing$, A і B – несумісні події
$A \setminus B$ (різниця множин A і B)	$A \setminus B$ – різниця подій A і B, настане подія A, але не настане подія B

1.1.4. Ймовірності подій

Ще до того, як було введене поняття ймовірності, що ґрунтувалось на відносній частоті настання події, під час дослідження азартних ігор використовувалось інше поняття ймовірності, яке називають зараз класичним. Розглянемо класичне означення ймовірності, а також статистичне та геометричне поняття ймовірності.

1.1.4.1. Класичне означення ймовірності

Класичне означення ймовірності ввів Лаплас (1749–1827).

Означення. **Ймовірністю** події A називають відношення кількості результатів випробування, сприятливих для A, до кількості всіх рівноможливих і попарно несумісних наслідків випробування:

$$P(A) = \frac{m}{n}. \tag{1.8}$$

Побудова логічно повноцінної теорії ймовірностей ґрунтується на аксіоматичному визначенні випадкової події і її ймовірності. У системі аксіом, запропонованій О. М. Колмогоровим, невизначуваними поняттями є елементарна подія і ймовірність. Приведемо аксіоми, що визначають імовірність:

1. Кожній події A поставлено у відповідність невід'ємне дійсне число $P(A)$. Це число називається ймовірністю події A.

2. Ймовірність достовірної події дорівнює одиниці:

$$P(\Omega) = 1.$$

3. Ймовірність настання хоча б однієї із попарно несумісних подій дорівнює сумі ймовірностей цих подій.

Якщо події $A_1, A_2, \ldots$ попарно несумісні, то ймовірність суми подій дорівнює сумі ймовірностей цих подій:

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

Аксіому 3 називають аксіомою адитивності для подій A і B, що є несумісні:

$$P(A \cup B) = P(A) + P(B).$$

Аксіому додавання ймовірностей інколи називають «теоремою додавання» (для дослідів, що зводяться до схеми випадків, її можна довести).

Виходячи з цих аксіом, властивості ймовірностей і залежності між ними виводять як теореми.

Приклад 1.10. Знайти ймовірність того, що кількість очок, яка випадає на гральному кубику за одного підкидання, буде непарною (подія *A*).

Розв'язання. Простір елементарних подій $\Omega = \{1, 2, 3, 4, 5, 6\}$ містить $n = 6$ елементарних подій, подія $A = \{1, 3, 5\}$ містить $m = 3$ сприятливі події. Тоді

$$P(A) = \frac{m}{n} = \frac{3}{6} = \frac{1}{2}.$$

Приклад 1.11. З урни, в якій 12 білих та 8 чорних кульок, виймають навмання 3 кульки. Яка ймовірність того, що:

1) всі три кульки будуть чорними (подія *A*);

2) дві кульки будуть чорними та одна білою (подія *B*).

Розв'язання. 1) Кількість усіх елементарних подій дорівнює кількості комбінацій з 20 по 3, тобто

$$n = C_{20}^3 = \frac{20!}{3!(20-3)!} = \frac{18 \cdot 19 \cdot 20}{1 \cdot 2 \cdot 3} = 1140.$$

Кількість випадків, які сприяють події *A*, становить $m = C_8^3 = \frac{8!}{3! \cdot 5!} = \frac{6 \cdot 7 \cdot 8}{1 \cdot 2 \cdot 3} = 56.$

Тоді ймовірність виймання трьох чорних кульок

$$P(A) = \frac{C_8^3}{C_{20}^3} = \frac{56}{1140} = \frac{14}{285} \approx 0,049.$$

2) Як і в п. 1 $n = C_{20}^3 = 1140.$ Знаходимо кількість сприятливих подій: дві чорні кульки можна взяти з восьми чорних $C_8^2 = \frac{8!}{2! \cdot 6!} = \frac{7 \cdot 8}{2} = 28$ способами, а одну білу з 12 кульок $C_{12}^1 = 12$ способами. За правилом множення (принцип добутку) $m = C_8^2 \cdot C_{12}^1$. Отже,

$$P(B)=\frac{m}{n}=\frac{C_8^2\cdot C_{12}^1}{C_{20}^3}=\frac{28\cdot 12}{1140}=\frac{84}{285}\approx 0,29.$$

1.1.4.2. Статистичне означення ймовірності

***Означення*. Відносною частотою** події називають відношення кількості випробувань, в яких подія настала m разів, до загальної кількості *n* фактично проведених випробувань:

$$W(A)=\frac{m}{n}.$$

Відносна частота $\frac{m}{n}$ прямує за ймовірністю до числа *p*, якщо для довільного $\varepsilon>0$ ймовірність того, що $\left|\frac{m}{n}-p\right|<\varepsilon$ прямує до одиниці за $n\to\infty$, тобто

$$\lim_{n\to\infty}P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=1.$$

Відносна частота може бути обчислена після того, коли проведено серію експериментів, і, загалом кажучи, вона змінюється, якщо здійснити іншу серію з *n* експериментів або якщо змінити *n*. Однак практика показує, що за достатньо великих *n* для більшості таких серій експериментів відносна частота має властивість стійкості.

***Означення*.** Число p називають **статистичною ймовірністю** статистично стійкої події, якщо відносна частота цієї події прямує за ймовірністю до числа p за $n\to\infty$, тобто $\frac{m}{n}\xrightarrow{p}p, n\to\infty.$

Зауважимо, що теорія ймовірностей має справу тільки зі статистично стійкими експериментами, і коли кажуть, що ймовірність деякої випадкової події дорівнює, наприклад 0,28, то практично це означає, що у середньому в кожних 100 дослідах ця подія відбувається 28 разів.

На практиці використовують статистичне означення ймовірності, вважаючи за ймовірність події відносну частоту або число, яке є близьким до неї.

Наприклад, якщо в серії достатньо великої кількості випробувань виявилось, що відносна частота дуже близька до 0,28, тоді це число і приймають за статистичну ймовірність події.

1.1.4.3. Геометричні ймовірності

Нехай Ω – деяка область на прямій, площині або в просторі, A – деяка частина області Ω. В області Ω навмання вибирають точку, вважаючи, що вибір точок області рівноможливий. Імовірність того, що вибрана точка належить A, визначається рівністю

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

де $\text{mes } A$, $\text{mes } \Omega$ – міра (довжина, площа, об'єм) A, Ω.

Зауваження. У випадку класичного означення ймовірності не тільки $P(\varnothing) = 0$, а й з того, що $P(A) = 0$ випливає, що $A = \varnothing$. Для геометричної ймовірності ця обставина не має місця. Нехай, наприклад, Ω − плоска область, A − лінія нульової площі. Тоді за формулою обчислення геометричної ймовірності $P(A) = 0$, хоча подія A не є неможливою (точка може потрапити в область).

Приклад 1.12. У коло радіусом R вписано правильний трикутник. Визначити ймовірність того, що навмання поставлена в коло точка потрапить у трикутник.

Розв'язання. Нехай подія A – точка потрапляє у трикутник. Тоді

$$\text{mes } A = S_{\Delta} = \frac{3\sqrt{3}}{4} R^2, \ \text{mes } \Omega = \pi R^2.$$

За формулою геометричної ймовірності

$$P(A) = \frac{\text{mes } A}{\text{mes } \Omega} = \frac{3\sqrt{3}}{4\pi}.$$

Приклад 1.13. Відрізок довжиною l розділили на три частини, вибираючи дві точки поділу навмання. Знайти ймовірність того, що з утворених відрізків можна скласти трикутник.

Розв'язання. Позначимо довжини утворених частин через $x, y, l-(x+y), \quad x>0, y>0, \quad 0<x+y<l.$ Для того, щоб із утворених відрізків можна було скласти трикутник, необхідно і достатньо виконання умов:

$$x<y+(l-x-y), \quad y<x+(l-x-y), \quad (l-x-y)<x+y,$$

з яких $x<\frac{l}{2}, y<\frac{l}{2}, \quad x+y>\frac{l}{2}.$

Простором елементарних подій є прямокутний трикутник з катетами довжиною l (кут O – прямий), mes $\Omega=\frac{1}{2}l^2.$ Із отриманих обмежень нерівностями для множини точок, які дають можливість складання трикутника, маємо також mes $A=S_{\Delta}$, трикутник також є прямокутний, але з катетами $\frac{l}{2}.$

Отже, $P(A)=\frac{\text{mes } A}{\text{mes } \Omega}=\frac{\frac{1}{8}l^2}{\frac{1}{4}l^2}=\frac{1}{4}.$

1.1.5. Теореми додавання ймовірностей

Теореми додавання і множення ймовірностей використовують для визначення ймовірностей складних подій, які можна записати через інші події. Розглянемо теореми додавання для сумісних та несумісних подій.

1.1.5.1. Теорема додавання для несумісних подій

Означення. Дві події називають **несумісними**, якщо вони не можуть одночасно настати в одному досліді, іншими словами, настання однієї події виключає можливість настання другої.

Означення. Події $A_1, A_2, \ldots, A_k$ утворюють **повну групу** несумісних подій, якщо $A_i \cap A_j=\varnothing$ за $i\neq j$ $(i,\ j=1,\ 2,\ \ldots,k)$, $\bigcup\limits_{i=1}^{k}A_i=\Omega.$

Теорема додавання для несумісних подій

Якщо події A і B несумісні ($A\cap B=\varnothing$), причому відомі їх імовірності P(A) та P(B), то ймовірність суми цих подій дорівнює сумі їх імовірностей:

$$P(A \cup B) = P(A) + P(B). \qquad (1.9)$$

Доведення. Дійсно, нехай n – кількість усіх елементарних подій у певному досліді; m_1 – кількість елементарних подій, сприятливих події A; m_2 – кількість елементарних подій, сприятливих події B. Тоді настанню події $A \cup B$ сприяють $m_1 + m_2$ елементарних подій. Отже, за класичним означенням імовірності

$$P(A \cup B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B).$$

Наслідок 1. Якщо події A_1, A_2, …, A_k попарно несумісні, то ймовірність суми подій дорівнює сумі ймовірностей цих подій:

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i).$$

Наслідок 2. Ймовірність протилежної до A події $\overline{A}$:

$$P(\overline{A}) = 1 - P(A). \qquad (1.10)$$

Дійсно, оскільки $A \cup \overline{A} = \Omega$, то $P(A \cup \overline{A}) = P(\Omega) = 1$. З другого боку, $P(A \cup \overline{A}) = P(A) + P(\overline{A})$.

Отже, $P(A) + P(\overline{A}) = 1$, звідки $P(\overline{A}) = 1 - P(A)$.

Наслідок 3. Сума ймовірностей подій A_1, A_2, …, A_k, що утворюють повну групу подій, дорівнює одиниці:

$$P(A_1) + P(A_2) + \ldots + P(A_k) = 1.$$

Доведення. Так як настання однієї із подій повної групи подій є достовірною подією, а ймовірність достовірної події дорівнює 1, то

$$P(A_1 + A_2 + \ldots + A_k) = 1.$$

Будь-які дві події повної групи несумісні, тому можна використати теорему додавання для несумісних подій:

$$P(A_1 + A_2 + \ldots + A_k) = P(A_1) + P(A_2) + \ldots + P(A_k) = 1.$$

Приклад 1.14. У партії з 20 деталей є 16 стандартних. Знайти ймовірність того, що серед навмання взятих трьох деталей виявиться принаймні одна стандартна.

Розв'язання. Нехай подія A_1: виявиться одна стандартна; подія A_2: виявиться дві стандартні; подія A_3: виявиться три стандартні деталі. Ці події попарно несумісні.

Нехай подія A: серед навмання взятих трьох деталей виявиться принаймні одна стандартна. Отже, $A = A_1 \cup A_2 \cup A_3$, і за формулою (1.9) маємо:

$$P(A) = P(A_1) + P(A_2) + P(A_3).$$

Обчислимо ймовірності подій A_1, A_2, A_3:

$$P(A_1) = \frac{C_{16}^1 \cdot C_4^2}{C_{20}^3} = \frac{8}{95};\ P(A_2) = \frac{C_{16}^2 \cdot C_4^1}{C_{20}^3} = \frac{40}{95};\ P(A_3) = \frac{C_{16}^3}{C_{20}^3} = \frac{28}{57}.$$

Отже, $P(A) = \frac{8}{95} + \frac{40}{95} + \frac{28}{57} = \frac{284}{285}$.

На рис. 1.8. показано обчислення за допомогою пакету Excel, де в комірках D_2–D_4 та E_2–E_4 обчислено число комбінацій, а для використання формули (1.9) за допомогою функції **СУММПРОИЗВ** категорії «математичні» обчислено значення кількості наслідків, що сприяють настанню події. Задавши вручну в Excel формулою дію ділення, отримано шукану ймовірність 0,9965.

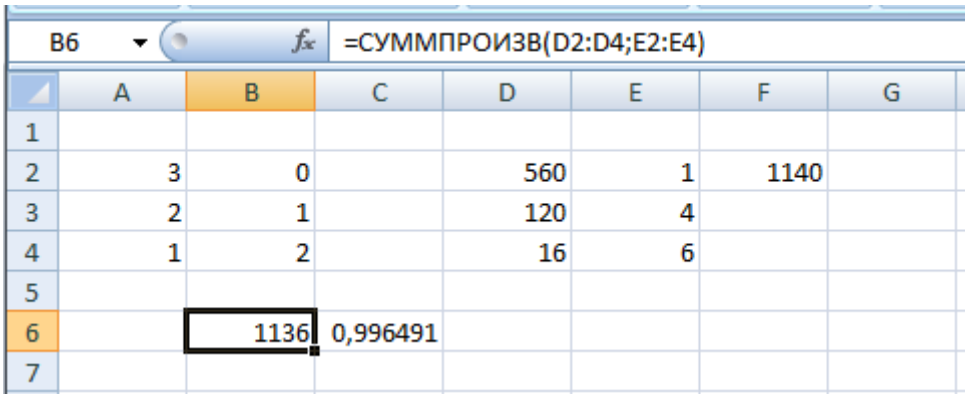


Рис. 1.8. Вигляд обчислень на екрані

Інший спосіб. Під час розв'язування задач часто буває зручно переходити до протилежної події. Так, якщо подія A_0 – не виявиться жодної стандартної деталі, тоді подія A_0 протилежна до події A, і події A_0, A_1, A_2, A_3 утворюють повну групу попарно несумісних подій. Отже, будемо використовувати формулу (1.10). Обчисливши ймовірність $P(A_0) = \frac{C_4^3}{C_{20}^3} = \frac{1}{285}$, отримаємо $P(A) = 1 - \frac{1}{285} = \frac{284}{285}$.

Зауваження. Наведений приклад є частинним для стандартної задачі: В партії з n деталей є m стандартних. Знайти ймовірність того, що серед навмання взятих k деталей виявиться принаймні одна стандартна.

Розв'язання. Події «серед навмання взятих деталей принаймні одна нестандартна» і «серед навмання взятих деталей немає жодної стандартної» − протилежні. Позначимо першу з них через A, а другу, відповідно, $\overline{A}$.

Обчислимо $P(\overline{A})$. Загальна кількість способів, якими можна взяти k деталей із n, дорівнює C_n^k. Кількість нестандартних деталей дорівнює $n-m$, тому із цієї кількості деталей можна C_{n-m}^k способами взяти навмання k нестандартні деталі. Тому ймовірність того, що серед k взятих немає стандартних, дорівнює $P(\overline{A}) = \frac{C_{n-m}^k}{C_n^k}$.

Отже, шукана ймовірність (за формулою (1.10)):

$$p(A) = 1 - P(\overline{A}) = 1 - \frac{C_{n-m}^k}{C_n^k}.$$

1.1.5.2. Теорема додавання для сумісних подій

Нехай дві події A і B сумісні, причому відомі ймовірності цих подій $P(A)$, $P(B)$ та ймовірність їх сумісного настання $P(A \cap B)$.

Теорема додавання (для сумісних подій). *Імовірність настання принаймні однієї з двох сумісних подій A і B дорівнює сумі ймовірностей цих подій без імовірності їх спільного настання*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \qquad (1.11)$$

Доведення. Дійсно, оскільки події A і B сумісні, то подія $A \cup B$ настане, якщо настане одна з трьох несумісних подій $A \cap \overline{B}$, $\overline{A} \cap B$ або $A \cap B$:

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

Подія A настане, якщо настане одна з двох несумісних подій $A \cap \overline{B}$ або $A \cap B$. Аналогічно, подія B настане, якщо настане одна з двох несумісних подій $\overline{A} \cap B$ або $A \cap B$.

Отже, $A=(A\cap\overline{B})\cup(A\cap B);\ B=(\overline{A}\cap B)\cup(A\cap B),$

і за теоремою додавання ймовірностей несумісних подій (1.9) маємо:

$$P(A)=P(A\cap\overline{B})+P(A\cap B)\ \Rightarrow P(A\cap\overline{B})=P(A)-P(A\cap B);$$

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

Таким чином,

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

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

Зауваження. Теорема може бути узагальнена на довільну скінченну кількість сумісних подій. Наприклад, для трьох сумісних подій

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

$$+P(A\cap B\cap C).$$

Зауваження. Якщо дві події A і B несумісні, то їх суміщення є неможливою подією, тому $P(A\cap B)=0.$ Формула (1.11) має тоді вигляд $P(A\cup B)=P(A)+P(B)$, як теорема додавання для несумісних подій. Отже, формула (1.11) справедлива, як для сумісних, так і для несумісних подій.

Завдання для самоконтролю

1. Що називається сполуками елементів певної множини?

2. Які сполуки елементів певної множини називаються перестановками, розміщеннями та комбінаціями?

3. За якими формулами обчислюються перестановки, розміщення та комбінації?

4. За якими функціями програми Excel обчислюються перестановки, розміщення та комбінації?

5. Які є два основні правила комбінаторики? Навести приклади.

6. Як визначити, з яким із основних понять комбінаторики в даній ситуації необхідно користуватись?

7. Дати означення простору елементарних подій, випадкової події.

8. Які події називають достовірними, неможливими, випадковими? Навести приклади.

9. Яка подія називається елементарною; складеною елементарною подією? Навести приклади.

10. Які події називаються сумісними, несумісними, рівноможливими?

11. Дати класичне означення ймовірності випадкової події.

12. Що називається простором елементарних випадкових подій? Навести приклади.

13. Дати означення класичної ймовірності, сформулювати аксіоми ймовірності.

14. Означити статистичну ймовірність, її зв'язок з відносною частотою події.

15. Означити геометричну ймовірність, навести приклади його використання.

16. Сформулювати теореми додавання ймовірностей для сумісних та несумісних подій.

17. Скільки існує п’ятизначних чисел, які діляться на: 1) 5? 2) 2?

Відповідь. а)180000; б) 450000.

18. Скільки всього шестизначних парних чисел можна скласти з цифр 1, 2, 3 ,5, 7, 9, якщо в кожному з цих чисел ні одна цифра не повторюється?

Відповідь. $P_5 = 120$.

19. Скількома способами можна розмістити групу студентів із 20 осіб у 5 вагонів метро? *Відповідь.* 5^{20}.

20. Скільки трицифрових чисел можна записати за допомогою цифр: а) 1, 5 та 9; б) 0, 3 та 5, якщо цифри в числі не повторюються? *Відповідь.* а) 6; б) 4.

21. 1) Скількома способами можна обрати на конференцію 3 студентів із групи 25 студентів? 2) Скількома способами можна обрати президію із 3 студентів, визначаючи головуючого та його першого та другого заступників, якщо в групі 20 студентів?

22. Двічі підкидають монету. Описати простір елементарних подій Ω, події: A – при першому підкиданні випав герб, B – принаймні один раз випаде герб. Знайти: $A \cup B$, $A \cap B$, $A \setminus B$, $B \setminus A$, $\bar{A}$.

Відповідь. $\Omega = \{гг, гц, цг, цц\}$, $A = \{гг, гц\}$, $B = \{гг, гц, цг\}$, $A \cup B = \{гг, гц, цг\}$, $A \cap B = \{гг, гц\}$, $A \setminus B = \varnothing$, $B \setminus A = \{цг\}$, $\bar{A} = \{цг, цц\}$.

23. Тричі підкидають монету. Описати простір елементарних подій Ω та подію A – герб випаде принаймні два рази.

Відповідь. $\Omega = \{ггг, ггц, гцг, цгг, гцц, цгц, ццг, ццц\}$, $A = \{ггг, ггц, гцг, цгг\}$.

24. Вісім різних підручників навмання покладено на поличку. Знайти ймовірність того, що студент візьме дві потрібні для нього книжки і вони розташовані поруч? *Відповідь.* $p = \dfrac{7 \cdot 2! \cdot 6!}{8!} = \dfrac{1}{4}$.

25. Кинули два гральних кубики. Описати простір елементарних подій. Знайти ймовірність того, що:

а) сума очок на гранях, які випали, дорівнює 6;

б) сума очок на гранях, які випали, дорівнює 8, а різниця 4;

в) сума очок на гранях, які випали, дорівнює 5, а добуток 4.

26. Пристрій складається з 5 елементів, з яких 2 зношені. Коли пристрій включають, 2 елементи підключаються довільно, випадковим чином. Знайти ймовірність того, що підключаться незношені елементи.

27. Партія з 30 виробів містить 10% браку. Знайти ймовірність того, що серед 7 виробів, узятих випадково: а) тільки 2 бракованих; б) жодного бракованого. *Відповідь.* а) 0,119; б) 0,436.

28. Комплект містить 12 виробів, 5 з яких коштують по 3 гривні кожний, інші – по 1 гривні. Знайти ймовірність того, що взяті навмання 4 вироби коштують разом 10 гривень.

29. В урні 5 кульок з номерами від 1 до 5. Навмання по одній беруть 3 кульки. Яка ймовірність того, що послідовно з'являться кульки з номерами 3, 1, 5? *Відповідь.* $\dfrac{1}{120}$

30. Група з 24 студентів, серед яких 5 відмінників, довільно розбивається на дві підгрупи. Знайти ймовірність того, що три відмінники будуть у першій підгрупі (подія A). *Відповідь.* $P(A)=\frac{m}{n}=\frac{C_5^3\cdot C_{19}^9}{C_{24}^{12}}\approx 0,34.$

31. Користуючись таблицею смертності, складеною для страхових компаній на основі вибірки із 10 млн. чоловік, визначити статистичну ймовірність таких подій: 1) подія A – рік життя народженої людини складе 60–65 років; 2) подія B – людина, яка дожила до 60 років, проживе принаймні ще 5 років.

Вік	Кількість живих	Кількість померлих
0	10 000 000	70 800
20	9 664 994	17 300
21	9 647 694	17 655
60	7 698 698	156 592
61	7 542 106	167 736
65	6 800 531	215 917
95	6 415	6 415

Відповідь. 1) 0,09; 2) 0,88.

32. В партії із 1000 деталей відділом технічного контролю виявлено 10 нестандартних. Якою є відносна частота виготовлення нестандартних деталей?

33. При стрільбі в мішень отримана відносна частота влучень 0,8. Скільки зроблено пострілів, якщо промахів виявилось 15. *Відповідь.* 75.

34. При випробуванні партії виробів відносна частота якісних виробів виявилась рівною 0,9. Знайти кількість якісних виробів, якщо всього перевірено 200.

35. До авіакаси у випадковий час протягом 10 хв звернулись 2 пасажири. Обслуговування одного пасажира триває 2 хв. Знайти ймовірність того, що пасажир, який звернувся другим, буде вимушений зачекати. *Відповідь.* 0,36.

36. Кожне з двох дійсних додатних чисел не більше 4. Знайти ймовірність того, що їх добуток також буде не більше 4. *Відповідь.* $\frac{1+\ln 4}{4}$.

Варіанти завдань для індивідуальної роботи

Користуючись елементами комбінаторики та програмою Excel розв'язати наступні задачі двома способами:

− за допомогою формул комбінаторики;

− із використанням пакету Excel.

Завдання 1. Серед *n* лотерейних квитків *k* виграшних. Навмання купили *m* квитків. Визначити ймовірність того, що серед них *l* виграшних.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
n	10	10	10	10	11	11	11	12	12	12	9	9	9	8	8
l	2	2	3	3	2	3	3	3	2	2	2	3	2	2	2
m	4	3	5	5	5	4	5	8	8	5	4	5	3	4	5
k	6	6	7	6	7	8	7	5	3	4	6	6	7	5	4

Завдання 2. У ліфт *k*-поверхового будинку сіли *n* пасажирів ($n<k$). Кожний незалежно від інших із однаковою ймовірністю може вийти на будь-якому (починаючи з другого) поверсі. Визначити ймовірність того, що:

1) усі вийшли на різних поверхах;
2) хоча б двоє вийшли на одному поверсі.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
k	6	7	8	9	10	11	12	13	14	13	12	11	10	9	8
n	4	4	5	5	6	4	4	3	3	4	3	3	4	3	4

Зауваження. Завдання 2 виконати із використанням функції пакету Excel **ГИПЕРГЕОМЕТ** *(Число_успехов_в_выборке; Размер_выборки; Число_успехов_в_совокупности; Размер_совокупности).* Порівняти отримані результати.

1.2. Основні теореми теорії ймовірностей. Послідовності випробувань

1.2.1. Умовні ймовірності та незалежні події

Для теорем множення ймовірностей часто використовують поняття умовної ймовірності та залежних і незалежних подій.

1.2.1.1. Теореми множення ймовірностей

Як зазначалось, в основі означення ймовірності випадкової події лежить сукупність певних умов. Якщо ж ніяких інших обмежень, окрім цих умов, у разі обчислення ймовірності $P(A)$ не накладається, така ймовірність називається ***безумовною***. Якщо ж подія A відбувається за умови, що відбулась інша подія B, причому $P(B) > 0$, то ймовірність настання події A називають ***умовною*** і обчислюють за формулою

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \ (P(B) > 0). \tag{1.12}$$

Означення. Ймовірність настання події A за умови, що подія B вже відбулася, називають **умовною** і позначають $P(A|B)$.

Тоді $P(A|B) = \dfrac{N(A \cap B)}{N(B)} = \dfrac{\dfrac{N(A \cap B)}{n}}{\dfrac{N(B)}{n}} = \dfrac{P(A \cap B)}{P(B)}, \ P(B) > 0,$

де n – кількість наслідків стохастичного експерименту, $N(B)$, $N(A \cap B)$ – кількості наслідків стохастичного експерименту, що сприяють відповідно події B та перетину події A з подією B.

Означення. **Добутком** двох подій A і B називають подію C, яка полягає в сумісному настанні цих подій.

Означення. Подію A називають **незалежною** від події B, якщо ймовірність події A не залежить від того, чи відбулася подія B і справедлива рівність

$$P(A|B) = P(A). \tag{1.13}$$

Означення. Подію A називають **залежною** від події B, якщо ймовірність події A залежить від того, чи відбулася подія B, тобто $P(A|B) \neq P(A)$.

Теорема множення ймовірностей залежних подій. Розглянемо дві залежні події A і B, причому відомі ймовірності $P(A)$ і $P(B|A)$. Ймовірність суміщення цих подій обчислюють за теоремою:

Імовірність сумісного настання двох залежних подій дорівнює добутку ймовірності однієї з них на умовну ймовірність іншої, обчисленої за умови, що перша відбулася:

$$P(AB) = P(A) \cdot P(B|A)$$

або

$$P(AB) = P(B) \cdot P(A|B). \qquad (1.14)$$

Наслідок. Якщо події A_i $(i = \overline{1,n})$ залежні, то

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|(A_1 \cap A_2)) \cdot \ldots \cdot P(A_n|(A_1 \cap A_2 \cap \cdots \cap A_{n-1})),$$

тобто ймовірність сумісного настання декількох залежних подій дорівнює добутку ймовірності однієї з них на умовні ймовірності решти, причому ймовірність кожної наступної події обчислюється за припущення, що всі попередні події відбулися.

Зокрема, для трьох залежних подій *A, B, C* маємо:

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B). \qquad (1.15)$$

Теорема множення ймовірностей незалежних подій. *Для двох незалежних подій A і B*

$$P(A \cap B) = P(A) \cdot P(B). \qquad (1.16)$$

Означення. Події називають **незалежними в сукупності**, якщо ймовірність будь-якої події не залежить від настання довільної групи інших.

Події $A_1, A_2, \ldots A_n$ будуть незалежними в сукупності тоді і тільки тоді, коли імовірність добутку довільної їх комбінації дорівнює добутку відповідних імовірностей, тобто

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \ldots \cdot P(A_{i_k})$$

для будь-яких $k = 2, 3, \ldots n, \ 1 \le i_1 < i_2 < \ldots < i_k \le n.$

Якщо ця рівність виконується за $k = 2$, то події A_1, A_2, … A_n називають **попарно незалежними**. Попарно незалежні події можуть не бути незалежними в сукупності.

Задача Бернштейна. На трьох гранях правильного тетраедра, зробленого з однорідного матеріалу, написано цифри 1, 2, 3, а на четвертій грані – всі три цифри 1, 2, 3. Позначимо A_i (i = 1, 2, 3) подію, яка полягає в тому, що підкинутий тетраедр впаде на грань, на якій є цифра i. Показати, що події A_1, A_2, A_3 попарно незалежні, але не незалежні в сукупності.

Розв'язання. Очевидно, що $P(A_i) = \frac{2}{4} = \frac{1}{2}$ ($i = 1, 2, 3$). Покажемо, що події A_1, A_2, A_3 попарно незалежні. Оскільки $P(A_1 \cap A_2) = \frac{1}{4}$ (подія A_1, A_2 настає тоді і тільки тоді, коли тетраедр падає на четверту грань), то $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$, тобто події A_1 та A_2 незалежні. Аналогічно незалежні події A_1 та A_3, A_2 та A_3. Але події A_1, A_2, A_3 не незалежні в сукупності, оскільки

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{1}{8}.$$

Приклад 1.15. Студент прийшов на екзамен, підготувавши лише 20 з 25 питань програми. Екзаменатор задав йому три запитання. Знайти ймовірність того, що студент знає відповіді на них.

Розв'язання. Нехай подія A – студент знає відповіді на всі три запитання; подія A_i (i = 1, 2, 3) – студент знає відповідь на i-те запитання. Тоді $A = A_1 \cap A_2 \cap A_3$ і за формулою умовної ймовірності

$$P(A) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2).$$

Очевидно, що $P(A_1) = \frac{20}{25}$, $P(A_2 \mid A_1) = \frac{19}{24}$ (оскільки залишилось запитань 24, з яких студент знає 19), аналогічно $P(A_3 \mid A_1 \cap A_2) = \frac{18}{23}$. Підставивши в формулу, отримуємо:

$$P(A)=\frac{20}{25}\cdot\frac{19}{24}\cdot\frac{18}{23}=\frac{57}{115}=0{,}496.$$

Відповідь. 0,496.

Приклад 1.16. Знайти:

1) імовірність влучити в мішень один раз при трьох пострілах − подія A, якщо ймовірність влучити в мішень при першому пострілі (подія B) становить 0,7, при другому (подія C) – 0,8, а при третьому (подія D) – 0,85.

2) імовірність влучити в мішень принаймні один раз при вказаних трьох пострілах.

Розв'язання.

1) Події A – одне влучення за трьох пострілів відповідає один із варіантів: один із стрільців влучає, а два інші не влучають, тому за формулами (1.12) – теорема додавання для несумісних подій та (1.16) – теорема множення ймовірностей для незалежних подій, маємо

$$P(A)=P(B)\cdot P(\bar{C})\cdot P(\bar{D})+P(\bar{B})\cdot P(C)\cdot P(\bar{D})+P(\bar{B})\cdot P(\bar{C})\cdot P(D),$$

тобто

$$P(A)=0{,}7\cdot 0{,}2\cdot 0{,}15+0{,}3\cdot 0{,}8\cdot 0{,}15+0{,}3\cdot 0{,}2\cdot 0{,}85=0{,}108.$$

Обчислення за допомогою програмного пакету Microsoft Excel виконують, ввівши відповідні значення ймовірності події та їй протилежної у відповідні комірки та задавши аналог функції за допомогою арифметичних дій множення та додавання (рис. 1.9).

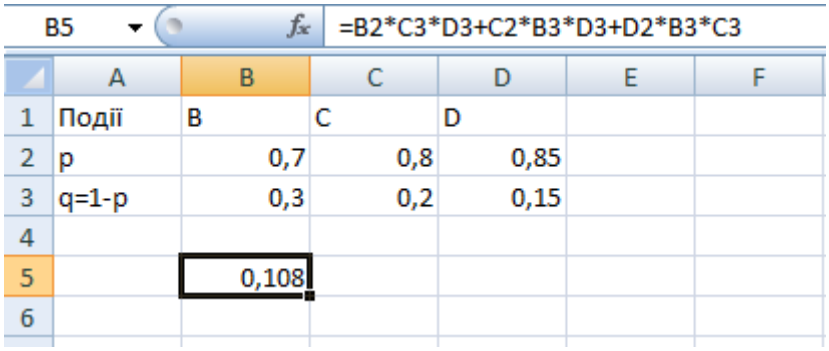


Рис. 1.9. Вигляд побудованої функції до задачі

2) Перейдемо до протилежної події – стрілець не влучив жодного разу в мішень із трьох пострілів (подія $\overline{A}$). Тоді $\overline{A} = \overline{B}\overline{C}\overline{D}$. Оскільки події *B, C, D*, а разом з ними відповідні протилежні події – незалежні, тому

$$P(\overline{A}) = P(\overline{B}) \cdot P(\overline{C}) \cdot P(\overline{D}) = 0,3 \cdot 0,2 \cdot 0,15 = 0,009,$$

звідки і знаходимо ймовірність події *A*:

$$P(A) = 1 - P(\overline{A}) = 1 - 0,009 = 0,991.$$

Відповідь. 1) 0,108; 2) 0,991.

1.2.1.2. Імовірність настання принаймні однієї події

Нехай у результаті *n* послідовних випробувань може настати *n* подій $A_1, A_2, \ldots, A_n$, незалежних в сукупності, ймовірності яких відомі $p_1, p_2, \ldots, p_n$. Тоді ймовірності протилежних подій $\overline{A_1}, \overline{A_2}, \ldots, \overline{A_n}$ дорівнюють

$$q_1 = 1 - p_1,\ q_2 = 1 - p_2, \ldots,\ q_n = 1 - p_n.$$

Теорема. *Ймовірність настання принаймні однієї з подій* $A_1, A_2, \ldots, A_n$, *незалежних в сукупності, обчислюють за формулою*

$$P(A) = 1 - q_1 q_2 \ldots q_n. \qquad (1.17)$$

Наслідок. Якщо події $A_1, A_2, \ldots, A_n$ мають однакову ймовірність *p*, то ймовірність настання принаймні однієї з цих подій $P(A) = 1 - q^n$.

Приклад 1.17. Ймовірність принаймні одного влучення в ціль за чотирьох пострілів дорівнює 0,9984. Знайти ймовірність влучення в ціль за одного пострілу.

Розв'язання. Нехай подія *A* – принаймні одне влучення в ціль за чотирьох пострілів, *p* – ймовірність влучення за одного пострілу, $q = 1 - p$ – ймовірність промаху за одного пострілу. Тоді $P(A) = 1 - q^4 = 0,9984$, тобто $q^4 = 0,0016$, $q = 0,2$. Отже, $p = 1 - q = 0,8$.

1.2.1.3. Надійність системи

***Означення*. Надійністю** системи називають імовірність її безвідмовної роботи протягом певного часу *t* (гарантійний термін).

Системи складаються з *n* елементів, з'єднаних послідовно (рис. 1.10) або паралельно (рис. 1.11).

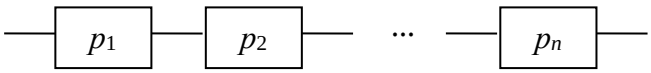


Рис. 1.10. Схема послідовного з’єднання елементів

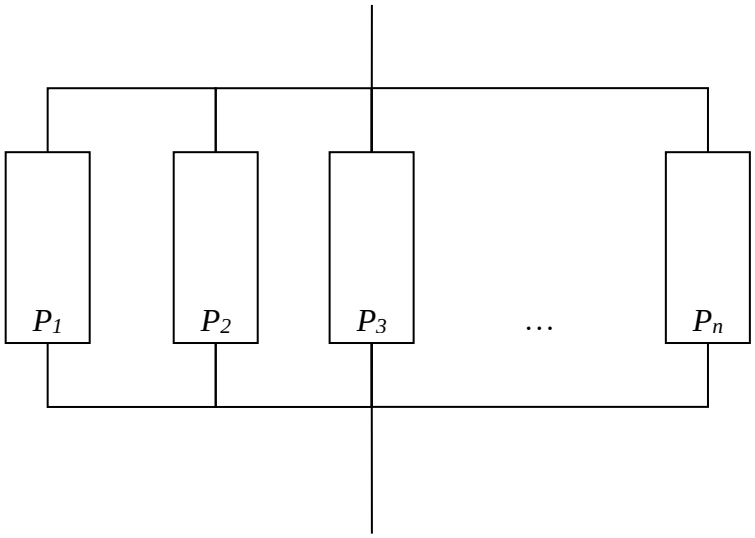


Рис. 1.11. Схема паралельного з’єднання елементів

У разі обчислення надійності систем потрібно виразити їх надійність через надійність елементів та блоків. Надійність елементів вважать відомою, оскільки вона пов’язана з технологією їх виготовлення та умовами експлуатації.

Позначимо p_k надійність *k*-го елемента, q_k – ймовірність виходу з ладу за час t *k*-го елемента, P – надійність блока.

Послідовне з’єднання незалежних елементів (рис. 1.10). Такий блок працюватиме безвідмовно лише той час, коли усі елементи працюватимуть безвідмовно. За теоремою множення ймовірностей незалежних подій імовірність безвідмовної роботи такого блока

$$P = p_1 \cdot p_2 \cdot \ldots \cdot p_n .$$

Паралельне з’єднання (рис. 1.11). Такий блок працюватиме безвідмовно, якщо принаймні один елемент не вийде з ладу, тому ймовірність P безвідмовної роботи

$$P = 1 - q_1 \cdot q_2 \cdot \ldots \cdot q_n .$$

Приклад 1.18. Прилад складено з двох елементів, з'єднаних послідовно, які працюють незалежно. Ймовірність відмови елементів дорівнює 0,05 та 0,08. Знайти ймовірність відмови приладу.

Розв'язання. Відмова приладу – подія, протилежна до його безвідмовної роботи. Ймовірності безвідмовної роботи елементів

$$p_1 = 1 - 0,05 = 0,95;\ \ p_2 = 1 - 0,08 = 0,92.$$

Ймовірність безвідмовної роботи такого блока (послідовне з'єднання)

$$P = p_1 \cdot p_2 = 0,95 \cdot 0,92 = 0,874.$$

Ймовірність відмови приладу

$$P = 1 - p_1 \cdot p_2 = 1 - 0,874 = 0,126$$

або $P = q_1 + q_2 - q_1 \cdot q_2 = 0,05 + 0,08 - 0,05 \cdot 0,08 = 0,126.$

Відповідь. 0,126.

1.2.2. Формули повної ймовірності та Байєса

Якщо подія A має складну структуру, то пряме визначення її ймовірності може виявитись надто важким завданням (наприклад, за класичним означенням з використанням комбінаторного аналізу). Виникає питання про можливість розглянути подію A з іншими подіями, ймовірності яких відомі, і на основі розглянутих подій знайти ймовірність події A. Виявляється, що в багатьох випадках це можливо, і відповідь на поставлене запитання дає формула повної ймовірності.

Теорема (формула повної ймовірності). *Нехай випадкова подія A може відбуватись лише сумісно з однією із несумісних між собою подій* $H_1,\ H_2,\ \ldots,H_n,$ *що утворюють повну групу* $\left(\bigcup_{i=1}^{n} H_i = \Omega,\ H_i \cap H_j = \varnothing \text{ çà } i \neq j\right)$, *які називають гіпотезами,* $P(H_i) > 0, \quad (i = 1, 2, \ldots, n)$. *Тоді ймовірність події A обчислюють за формулою*

$$P(A) = \sum_{i=1}^{n} P(H_i) P\left(A \mid H_i\right). \tag{1.18}$$

Отже, ймовірність події A дорівнює сумі добутків імовірностей гіпотез та умовних імовірностей події A за цих гіпотез.

Доведення. Подію A можна подати у такому вигляді:

$$A = \Omega \cap A = \left(\bigcup_{i=1}^{n} H_i \right) \cap A = \bigcup_{i=1}^{n} (H_i \cap A).$$

Оскільки події H_i несумісні, то події $H_i \cap A$ також несумісні. За теоремою додавання для несумісних подій

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

і за теоремою множення $P(H_i \cap A) = P(H_i) P(A \mid H_i)$. Отже,

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

Якщо до випробування відомі апріорні ймовірності гіпотез $P(H_1)$, $P(H_2)$, … , $P(H_n)$, а в результаті випробування настала подія A, і $P(A) > 0$, то, з урахуванням настання цієї події, умовні ймовірності гіпотез обчислюють за **формулою Байєса**:

$$P(H_i \mid A) = \frac{P(H_i) P(A \mid H_i)}{P(A)}, \; i = 1, \; 2, \; \ldots, n. \qquad (1.19)$$

Приклад 1.19. Два працівники сфери бізнесу заповнюють документи з власними прогнозами щодо курсу певних коштовних паперів, після чого їх складають у спільну папку. Ймовірність відхилення прогнозованого значення курсу від фактичного понад допустиму похибку для першого працівника дорівнює 0,1; для другого – 0,2. Перший працівник заповнив 40 документів, другий – 60. Під час перевірки навмання взятий із папки документ виявився з суттєвою похибкою. Визначити ймовірність того, що його склав перший працівник.

Розв'язання. Позначимо події H_1, H_2 – документ заповнив відповідно перший, другий працівник, A – навмання взятий із папки документ виявився з суттєвою похибкою. Тоді за умовою:

$$P(H_1) = \frac{40}{100} = 0,4; \; P(H_2) = \frac{60}{100} = 0,6; \; p(A \mid H_1) = 0,1; \; P(A \mid H_2) = 0,2.$$

За формулою повної ймовірності (1.18):

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = 0,4 \cdot 0,1 + 0,6 \cdot 0,2 = 0,16.$$

За формулою Байєса отримаємо

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(A)} = \frac{0,4 \cdot 0,1}{0,16} = 0,25.$$

Отже, ймовірність заповнення документа з суттєвою похибкою першим працівником

$$P(H_1|A) = 0,25.$$

Відповідь. 0,16; 0,25.

Приклад 1.20. Імовірність поразки команди групи технологів у матчі з командою групи економістів при дощовій погоді становить 0,5, а при відсутності дощу – 0,6. Імовірність дощу у день матчу становить 0,2. 1) Знайти ймовірність уникнення поразки командою групи технологів. 2) Команда групи технологів зазнала поразки. Яка ймовірність того, що матч відбувався при дощовій погоді?

Розв'язання. 1) Нехай подія A – поразка команди групи технологів. Утворимо дві гіпотези: H_1 – під час матчу дощова погода; H_2 – під час матчу не буде дощу. За умовою задачі

$$P(H_1) = 0,2; \quad P(H_2) = 0,8; \quad P(A|H_1) = 0,5; \quad P(A|H_2) = 0,6.$$

За формулою повної ймовірності (2.21), яка для даного прикладу має вигляд:

$$P(A) = P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2)$$

знаходимо ймовірність поразки команди групи технологів

$$P(A) = 0,2 \cdot 0,5 + 0,8 \cdot 0,6 = 0,58.$$

Тоді ймовірність уникнення поразки становить:

$$P(\overline{A}) = 1 - P(A) = 1 - 0,58 = 0,42.$$

Зауваження. Обчислення ймовірності за формулою (1.18) можна виконати за допомогою вбудованої математичної функції **СУММПРОИЗВ** (Массив1, Масив 2, Масив 3) (рис. 1.12), задаючи для даної задачі тільки 2 масиви даних, один з них є ймовірностями прийнятих гіпотез, а другий – відповідними умовними ймовірностями (рис. 1.13).

2) За формулою Байєса знаходимо ймовірність того, що йшов дощ під час матчу, в якому команда групи технологів зазнала поразки:

$$P(H_1 \mid A) = \frac{P(H_1) \cdot P(A \mid H_1)}{P(A)} = \frac{0,2 \cdot 0,5}{0,58} = 0,17.$$

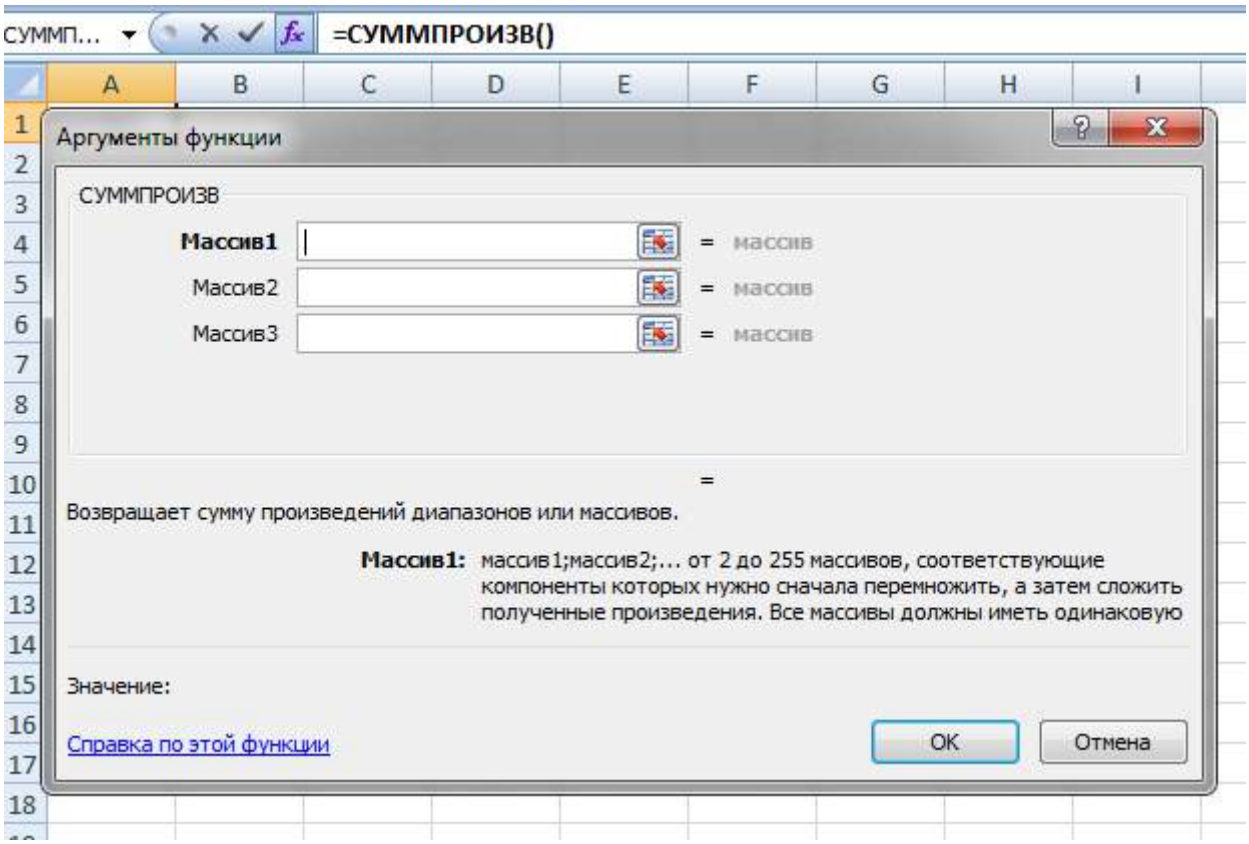


Рис. 1.12. Вигляд функції на екрані

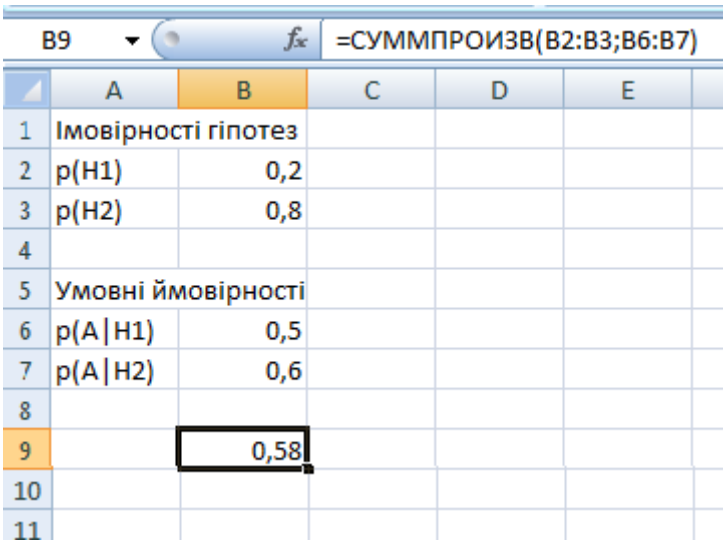


Рис. 1.13. Обчислення за формулою повної ймовірності

Обчислення за допомогою описання функції в Excel, тобто задання арифметичних дій відповідно до формули Баєса, подано на рис. 1.14.

C9 =B2*B6/B9

	A	B	C	D
1	Імовірності гіпотез			
2	p(H1)	0,2		
3	p(H2)	0,8		
4				
5	Умовні ймовірності			
6	p(A|H1)	0,5		
7	p(A|H2)	0,6		
8				
9		0,58	0,172414	

Рис. 1.14. Обчислення за формулою Байєса

Відповідь. 1) 0,58; 2) 0,17.

1.2.3. Послідовні незалежні випробування. Граничні теореми формули Бернуллі

1.2.3.1. Послідовні незалежні випробування. Формула Бернуллі

Означення. Якщо n випробувань проводити за однакових умов і ймовірність настання події A в усіх випробуваннях однакова та не залежить від настання або ненастання A в інших випробуваннях, таку послідовність незалежних випробувань називають ***схемою Бернуллі.***

Нехай проводиться скінченна кількість спроб n, в результаті яких може з'явитися подія A з певною ймовірністю p, причому ця подія не залежить від наслідків інших спроб. Такі спроби назвемо *незалежними відносно події* A.

Обчислимо ймовірність того, що в результаті проведення n незалежних спроб подія A настане рівно k разів, якщо в кожній із спроб вона настає із сталою ймовірністю $P(A) = p$ або не настає з ймовірністю $P(\overline{A}) = q$ $(p+q=1)$. Позначимо шукану ймовірність $P_n(k)$, це означає, що в n спробах подія A з'явиться k разів. Зауважимо, що тут не вимагається, щоб подія A повторилася k разів у певній послідовності. Для розв'язання поставленої задачі за великих

значень n і k безпосереднє застосування теорем додавання і множення ймовірностей приводить до громіздких розрахунків, тому зручніше користуватися формулою Бернуллі, виведення якої ми розпочнемо.

Імовірність того, що подія A в n спробах настане рівно k разів, а в решті $(n-k)$ спроб настане протилежна подія $\overline{A}$, за теоремою множення ймовірностей незалежних подій дорівнює $p^k \cdot q^{n-k}$. При цьому подія A в n спробах може настати рівно k разів у різних комбінаціях, кількість яких C_n^k. Оскільки всі комбінації подій – події несумісні і не важливо, в якій послідовності настане подія A або подія $\overline{A}$, то, застосовуючи теорему додавання ймовірностей несумісних подій, отримаємо ***формулу Бернуллі:***

$$P_n(k) = C_n^k p^k q^{n-k}. \qquad (1.20)$$

Отже, ймовірність $P_n(k)$ того, що подія A настане k разів у n випробуваннях $(0 \le k \le n)$ обчислюють за **формулою Бернуллі**:

$$P_n(k) = C_n^k p^k q^{n-k}.$$

Число k_0, за якого ймовірність $P_n(k_0)$ найбільша, називають ***найімовірнішим*** числом настання події A. Його визначають співвідношенням

$np - q \le k_0 \le np + p$ або

$$(n+1)p - 1 \le k_0 \le (n+1)p\,. \qquad (1.21)$$

Число k_0 має бути цілим. Якщо $(n+1)p-$ ціле число, то найбільшого значення ймовірність $P_n(k)$ досягає у двох випадках:

$$k_1 = (n+1)p - 1 \text{ та } k_2 = (n+1)p.$$

Зауваження. Ймовірність настання події A у n випробуваннях схеми Бернуллі *менше ніж m разів* знаходять за формулою

$$P_n(k < m) = P_n(0) + P_n(1) + \cdots + P_n(m-1).$$

Імовірність настання події *А не менше ніж* m *разів* знаходять за формулою $P_n(k \ge m) = P_n(m) + P_n(m+1) + \cdots + P_n(n)$ або за такою формулою:

$$P_n(k \ge m) = 1 - \sum_{k=0}^{m-1} P_n(k)\,.$$

Імовірність настання події *A принаймні один раз у n випробуваннях* доцільно обчислювати за формулою $P_n(1 \le m \le n) = 1 - q^n$.

1.2.3.2. Граничні теореми формули Бернуллі

Якщо проводять випробування за схемою Бернуллі [4] та числа n і k великі, то обчислення ймовірності за формулою Бернуллі викликає певні труднощі. У таких випадках для обчислення цих імовірностей застосовують асимптотичні (наближені) формули, які випливають із локальної та інтегральної теорем Муавра-Лапласа та граничної тереми Пуассона. Назва «гранична» в обох випадках пов'язана з тим, що ці теореми встановлюють поведінку ймовірності $P_n(k)$ або $P_n(k_1 \le k \le k_2)$ за певних умов, до яких обов'язково входить умова $n \to \infty$.

З огляду на це за досить великих значень n та за малих p або q замість формули Бернуллі часто використовують наближені асимптотичні формули.

1. Формула Пуассона

Теорема (теорема Пуассона). *Якщо* $n \to \infty$ *і* $p \to 0$ *так, що* $np \to \lambda$, $0 < \lambda < \infty$, то

$$P_n(k) = C_n^k p^k q^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda} \tag{1.22}$$

для будь-якого сталого $k = 0,\ 1,\ 2, \ldots$.

Наслідок. Імовірність настання події A k разів у n випробуваннях схеми Бернуллі знаходять за наближеною **формулою Пуассона**

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \tag{1.23}$$

де $\lambda = np$.

Ця формула дає досить точне наближення за значень p, близьких до нуля $(p < 0{,}1)$, тобто для подій, що рідко трапляються і для достатньо великих n $(npq \le 9)$.

2. Локальна теорема Муавра–Лапласа

Означення. Локальною функцією Лапласа називають функцію вигляду

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Основні властивості локальної функції Лапласа

1. Парність функції: $\varphi(-x) = \varphi(x)$;

2. $\varphi(x)$ визначена для всіх дійсних значень змінної;

3. $\varphi(x) \to 0$ за $x \to \pm\infty$;

4. $\varphi_{\max} = \varphi(0) = \frac{1}{\sqrt{2\pi}}$.

Теорема (локальна теорема Муавра–Лапласа). *Якщо ймовірність p настання події A в кожній спробі стала і така, що $0 < p < 1$, то ймовірність числа k настання події в n спробах обчислюють за наближеною рівністю*

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{k - np}{\sqrt{npq}}\right), \text{ де } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \qquad (1.24)$$

Формула (1.24) дає добре наближення, якщо n достатньо велике, p та q не дуже близькі до нуля, $npq > 9$.

3. Інтегральна теорема Муавра–Лапласа

Означення. Інтегральною функцією Лапласа називають функцію вигляду

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Як правило, табулюється функція $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$, яка пов'язана із функцією $F(x)$ співвідношенням $F(x) = \frac{1}{2} + \Phi(x)$.

Основні властивості функції Лапласа $\Phi(x)$

1. Непарність функції: $\Phi(-x) = -\Phi(x)$;

2. $\Phi(0) = 0$;

3. $\Phi(x) \approx 0{,}5$ для $x \geq 5$.

Теорема (інтегральна теорема Муавра–Лапласа). *Ймовірність того, що в n незалежних випробуваннях схеми Бернуллі, в яких подія A може відбутись з*

імовірністю p, подія A відбудеться не менше k_1 *та не більше* k_2 *разів, наближено рівна*

$$P_n(k_1 \le k \le k_2) \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right), \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt. \qquad (1.25)$$

Формула (1.25) дає добре наближення, якщо n достатньо велике, p та q не дуже близькі до нуля, $npq > 9$. Для всіх значень $x \ge 5$ можна вважати $\Phi(x) \approx 0{,}5$.

4. Імовірність відхилення відносної частоти від сталої ймовірності в незалежних випробуваннях

Теорема (Я. Бернуллі). *Якщо в n незалежних випробуваннях імовірність p настання події A однакова і подія A настала m разів, то для будь-якого додатного числа* α *справедлива рівність*

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| \ge \alpha\right) = 0,$$

тобто границя ймовірності відхилення відносної частоти $\frac{m}{n}$ *події A від її ймовірності на величину, що більша або дорівнює* α*, рівна нулю*.

Зауваження. З використанням інтегральної функції Лапласа формулу записують у вигляді

$$P\left(\left|\frac{m}{n} - p\right| \le \alpha\right) \approx 2\Phi\left(\alpha\sqrt{\frac{n}{pq}}\right). \qquad (1.26)$$

Зауваження. Послідовність випробувань із різними ймовірностями

У практичній діяльності не завжди в n незалежних випробуваннях імовірності настання події A рівні, наприклад, вони дорівнюють $p_1,\ p_2,\ \ldots,\ p_n$, відповідно, й імовірності ненастання події A: $q_1 = 1 - p_1,\ q_2 = 1 - p_2,\ \ldots,\ q_n = 1 - p_n$. У цьому випадку використовують твірну функцію

$$\varphi_n(z) = \prod_{k=1}^{n} (q_k + p_k z).$$

Потрібна ймовірність $P_n(m)$ рівна коефіцієнту при z^m.

Проста течія подій

Означення. Течією подій називають послідовність таких подій, які настають у випадкові моменти часу.

Означення. Течію подій називають ***пуассонівською***, якщо вона:

1. *Стаціонарна* – залежить від кількості k настання події та часу t і не залежить від моменту свого початку.

2. Має *властивість відсутності післядії* – ймовірність настання події не залежить від настання або ненастання події раніше та не впливає на найближче майбутнє.

3. *Ординарна* – ймовірність настання більше однієї події в малий проміжок часу є величина нескінченно мала порівняно з імовірністю настання події один раз у цей проміжок часу.

Означення. Середнє число λ настання події A в одиницю часу називають ***інтенсивністю течії***.

Теорема (математична модель простої течії подій). *Для пуассонівської течії подій імовірність настання A k разів за час t знаходять за формулою*

$$P_t(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

де λ – інтенсивність течії.

Приклад 1.21. Монету підкинули 6 разів. Знайти ймовірність того, що герб випаде 4 рази.

Розв'язання. За умовою задачі $n = 6,\ p = q = \frac{1}{2}$ та $k = 4.$ За формулою Бернуллі $P_6(4) = C_6^4 \cdot \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^2 = \frac{6!}{2! \cdot 4!}\left(\frac{1}{2}\right)^6 = \frac{5 \cdot 6}{2} \cdot \frac{1}{64} = \frac{15}{64}.$

Приклад 1.22. Ймовірність народження хлопчика дорівнює 0,515, а дівчинки 0,485. У певній сім'ї шестеро дітей. Знайти ймовірність того, що серед них не більше двох дівчаток.

Розв'язання. За умовою $n = 6,\ p = 0{,}485,\ q = 0{,}515.$ Шукана ймовірність дорівнює

$$P_6(0)+P_6(1)+P_6(2)=(0,515)^6+C_6^1\cdot 0,485\cdot(0,515)^5+C_6^2\cdot(0,485)^2\cdot(0,515)^4=$$
$$=0,018657+0,10542+0,2482=0,3723.$$

Приклад 1.23. Підручник надруковано тиражем 10 000 примірників. Ймовірність бракованого брошурування підручника дорівнює 0,0001. Знайти ймовірність того, що тираж має 5 бракованих підручників.

Розв'язання. Згідно з умовою задачі $n=10\,000$ досить велике число, а ймовірність $p=0,0001$ – мала, $\lambda=np=1.$ Застосовуючи формулу Пуассона (1.23), отримаємо

$$P_{10000}(5)\approx\frac{1^5}{5!}e^{-1}=\frac{1}{120\cdot e}\approx 0,0031.$$

Приклад 1.24. За нового технологічного процесу 80 % виготовленої продукції має найвищу якість. Знайти найбільш імовірну кількість виготовлених виробів найвищої якості серед 250 виготовлених виробів.

Розв'язання. Потрібно використати співвідношення – формулу (1.21)

$$np-q\le m_0\le np+p,$$

де $n=250,\ p=0,8,\ q=0,2,$ тоді $199,8\le m_0\le 200,8,$ але m_0 має бути цілим числом, тому $m_0 = 200.$

Приклад 1.25. Імовірність успіху в кожному випробуванні дорівнює 0,25. Яка ймовірність того, що за 300 випробувань успішними будуть рівно 75 випробувань.

Розв'язання. За локальною формулою Муавра–Лапласа $P_n(k)\approx$ $\approx\frac{1}{\sqrt{npq}}\varphi\left(\frac{k-np}{\sqrt{npq}}\right),$ де $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$ тоді за $n=300,\ k=75,\ p=0,25$ отримуємо:

$$\frac{k-np}{\sqrt{npq}}=\frac{75-300\cdot 0,25}{\sqrt{300\cdot 0,25\cdot 0,75}}=0,\ P_{300}(75)\approx\frac{1}{7,5}\varphi(0)=\frac{0,3989}{7,5}\approx 0,0532.$$

Приклад 1.26. Імовірність виходу з ладу за час t одного приладу рівна 0,1. Визначити ймовірність того, що за час t із 100 приладів вийде з ладу від 6 до 18 приладів.

Розв'язання. Використовуємо інтегральну формулу Муавра-Лапласа:

$$P_n(k_1 \le k \le k_2) \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right),\ \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}} dt.$$

За умовою задачі $n = 100,\ k_2 = 18,\ k_1 = 6,\ p = 0{,}1;\ q = 1 - p = 0{,}9,$ тоді

$$P_{100}\left(6 \le k \le 18\right) \approx \Phi\left(\frac{18 - 100 \cdot 0{,}1}{\sqrt{100 \cdot 0{,}1 \cdot 0{,}9}}\right) - \Phi\left(\frac{6 - 100 \cdot 0{,}1}{\sqrt{100 \cdot 0{,}1 \cdot 0{,}9}}\right) = \Phi\left(\frac{8}{3}\right) - \Phi\left(-\frac{4}{3}\right) \approx$$

$$\approx \Phi\left(2{,}66\right) - \Phi\left(-1{,}33\right) = \Phi\left(2{,}66\right) + \Phi\left(1{,}33\right) = 0{,}49609 + 0{,}40824 = 0{,}90433.$$

Приклад 1.27. Імовірність того, що деталь нестандартна, дорівнює 0,2. Знайти ймовірність того, що серед випадково відібраних 300 деталей відносна частота натрапляння на нестандартні деталі відхилиться від імовірності 0,2 за абсолютною величиною не більше ніж на 0,01.

Розв'язання. За умовою задачі $n = 300,\ p = 0{,}2;\ q = 1 - p = 0{,}8;\ \alpha = 0{,}01,$ тоді потрібно знайти ймовірність $P\left(\left|\frac{m}{300} - 0{,}2\right| \le 0{,}01\right)$. За теоремою Бернуллі

$$P\left(\left|\frac{m}{300} - 0{,}2\right| \le 0{,}01\right) \approx 2\Phi\left(0{,}01\sqrt{\frac{300}{0{,}2 \cdot 0{,}8}}\right) = 2\Phi(0{,}43) = 0{,}328.$$

Зміст отриманого результату: якщо згідно з умовою задачі взяли достатньо велику кількість спроб, тобто по 300 деталей у кожній, то приблизно для 33 % цих спроб відхилення відносної частоти від сталої ймовірності 0,2 за абсолютною величиною не перевищить 0,01.

Приклад 1.28. Працівник обслуговує три верстати, що працюють незалежно один від одного. Ймовірність того, що протягом години перший верстат не потребуватиме уваги працівника, дорівнює 0,9, а для другого та третього верстата – 0,8 та 0,85 відповідно. Яка ймовірність того, що протягом години:

а) ні один верстат не потребуватиме уваги працівника;

б) усі три верстати потребуватимуть уваги працівника;

в) принаймні один верстат потребуватиме уваги працівника.

Розв'язання. Цей приклад можна розв'язати, використовуючи теореми множення та додавання ймовірностей. Розв'яжемо задачу з використанням твірної функції, яка в цьому разі матиме вигляд

$$\varphi_n(z) = \prod_{k=1}^{3}(q_k + p_k z) = (0{,}1 + 0{,}9z)(0{,}2 + 0{,}8z)(0{,}15 + 0{,}85z) =$$

$$= 0{,}003 + 0{,}056z + 0{,}329z^2 + 0{,}612z^3.$$

Тоді коефіцієнт при z^k $(k = 0, 1, 2, 3)$ рівний імовірності того, що протягом години уваги працівника не потребують k верстатів. Тож відповіді на запитання такі:

а) імовірність того, що усі три верстати не потребують уваги працівника, дорівнює коефіцієнту при z^3, тобто $P_3(3) = 0{,}612$;

б) $P_3(0) = 0{,}003$;

в) $P_3(1 \le m \le 3) = 1 - P_3(0) = 1 - 0{,}003 = 0{,}997.$

Приклад 1.29. Середня кількість замовлень, що надходять до комбінату побутового обслуговування кожної години, дорівнює 3. Знайти ймовірність того, що протягом двох годин надійде 5 замовлень.

Розв'язання. Маємо просту течію подій з інтенсивністю $\lambda = 3$. За формулою $P_t(k) = \dfrac{(\lambda t)^k}{k!}e^{-\lambda t}$ отримуємо $P_2(5) = \dfrac{(3\cdot 2)^5}{5!}e^{-3\cdot 2} = \dfrac{6^5}{120}e^{-6} \approx 0{,}16.$

Приклад 1.30. Скільки разів потрібно підкинути монету, щоб з імовірністю 0,6 можна було сподіватись, що відхилення відносної частоти випадання герба від імовірності p = 0,5 буде за модулем не більше 0,01?

Розв'язання. За умовою α = 0,01; p = 0,5, отже q = 0,5, тому

$$P\left(\left|\frac{m}{n} - 0{,}5\right| \le 0{,}01\right) = 0{,}6$$

і за формулою (2.15): $2\Phi\left(\alpha\sqrt{\dfrac{n}{pq}}\right) = 2\Phi\left(0{,}01\sqrt{\dfrac{n}{0{,}5\cdot 0{,}5}}\right) = 0{,}6,$ тобто $\Phi\left(0{,}02\sqrt{n}\right) = 0{,}03,$ звідси $0{,}02\sqrt{n} = 0{,}84,$ тобто n =1764.

Користуючись формулою Бернуллі, її граничними теоремами та програмою Excel розв'язати наступні задачі:

Приклад 1.31. Обчислити ймовірність народження 6-ти дівчаток в багатодітній сім'ї, яка налічує 10 дітей, якщо ймовірність народження дівчинки дорівнює 0,48.

Розв'язання. Імовірність народжених дівчаток із 10 дітей обчислюємо за формулою Бернуллі

$$P_{10}(6) = C_{10}^{6} \cdot 0,48^{6} \cdot 0,52^{4} \approx 0,18.$$

Використаємо функцією Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*): *число_успехов*= 6*; число_испытаний* = 10*; вероятность_успеха* = 0,48*; интегральная* − 0 (рис. 1.15).

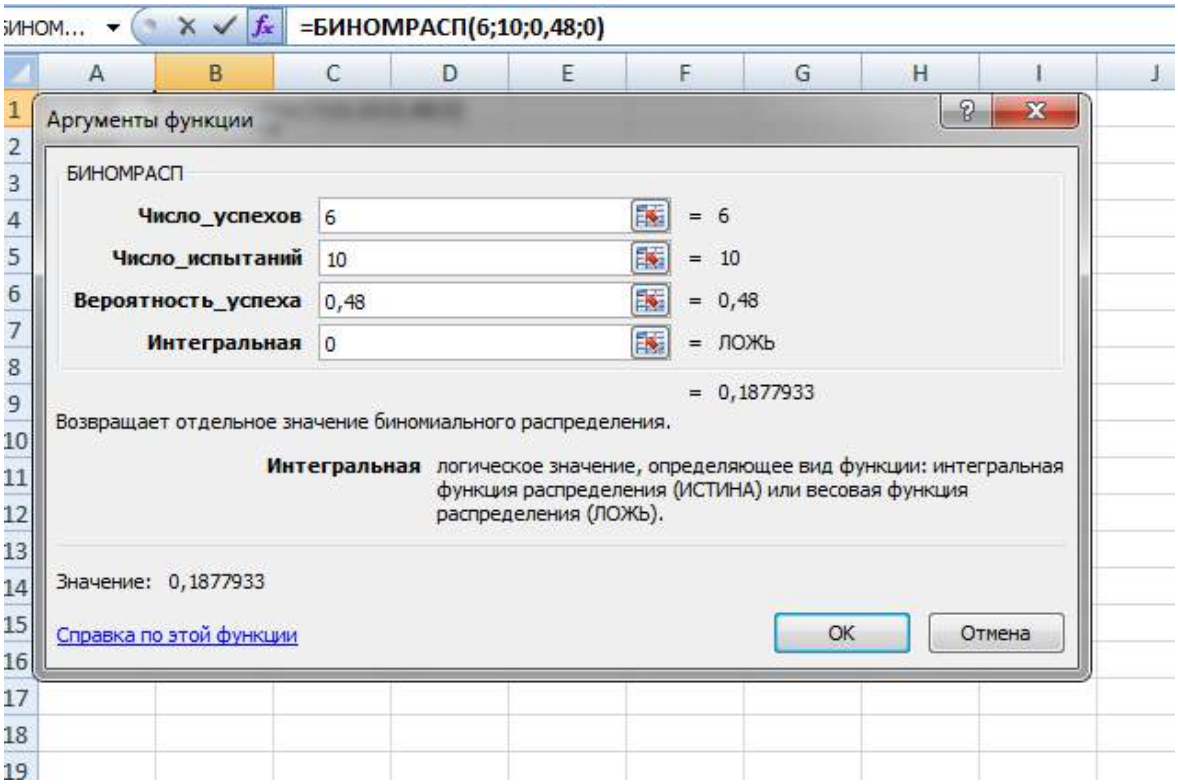


Рис. 1.15. Використання формули Бернуллі

Відповідь. $P_{10}(6) = 0,1878.$

Приклад 1.32. Імовірність влучити в мішень при одному пострілі дорівнює 0,8. Знайти найімовірніше число влучень із 6 пострілів та відповідну ймовірність.

Розв'язання. Знайдемо величину виразу $np + p = 0{,}8 \cdot 6 + 0{,}8 = 5{,}6$ – не ціле число. Тоді найбільше ціле число, яке не перевищує 5,6, дорівнює 5. Таким чином, найбільш імовірне ціле число влучень $k_0 = 5$. Імовірність п'яти влучень із шести пострілів обчислюємо за формулою Бернуллі

$$P_6(5) = C_6^5 \cdot 0{,}8^5 \cdot 0{,}2 = 0{,}39.$$

Зауваження. Використання функції Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*): *число_успехов* = 5*; число_испытаний* = 6*; вероятность_успеха* = 0,8*; интегральная* −0 дасть такий же результат.

Відповідь. $k_0 = 5$, $P_6(5) = 0{,}39$.

Приклад 1.33. Імовірність влучити в мішень при одному пострілі дорівнює 0,8. Знайти ймовірність того, що із 100 пострілів число влучень буде: а) точно 90; б) належить проміжку від 75 до 85.

Розв'язання. Згідно з умовою задачі $np = 80 > 5$ і $nq = 20 > 5$, тому можна скористатися формулами Муавра-Лапласа.

а) За локальною формулою Муавра-Лапласа (1.24) знаходимо:

$$x = \frac{90 - 0{,}8 \cdot 100}{\sqrt{100 \cdot 0{,}8 \cdot 0{,}2}} = \frac{10}{4} = 2{,}5.$$

Для $x = 2{,}5$ знаходимо (за таблицею значень, додаток: $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$) $\varphi(x) = 0{,}0175$ та, відповідно,

$$P_{100}(90) = \frac{0{,}0175}{4} = 0{,}004.$$

Знаючи функцію Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*), обчислення проводять набагато швидше (рис. 1.16).

б) За інтегральною формулою Муавра-Лапласа (1.25) знаходимо:

$$x_1 = \frac{75 - 80}{4} = -1{,}25; \quad x_2 = \frac{85 - 80}{4} = 1{,}25.$$

Для $x=1{,}25$ знаходимо за таблицею значень функції Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$:

$$\Phi(1{,}25)=0{,}39435.$$

Тоді $\Phi(-1{,}25)=-0{,}39435$. Шукану ймовірність знаходимо як різницю

$$P_{100}(75<k<85)=\Phi(1{,}25)-\Phi(-1{,}25)=0{,}39435-(-0{,}39435)=0{,}79.$$

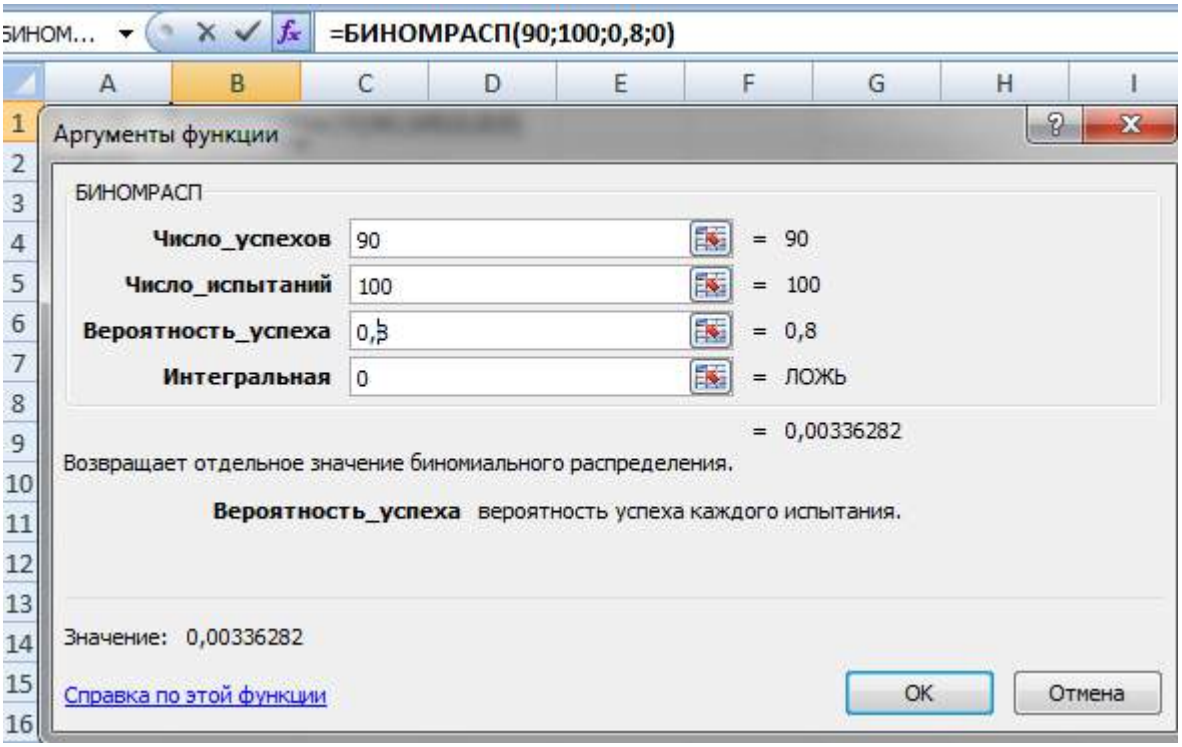


Рис. 1.16. Обчислення за локальною формулою Муавра-Лапласа

Зауваження. Можна також провести обчислення із використанням тієї ж функції Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*), де для пункту а) брали *интегральная* – 0, але для пункту б) потрібно взяти значення *интегральная* – 1 (повертає значення: кількість успішних випробувань не менше значення аргументу *число_испытаний* , тобто обчислюють два значення функції для $k=75, k=85$ та знаходять їх різницю (рис. 1.17).

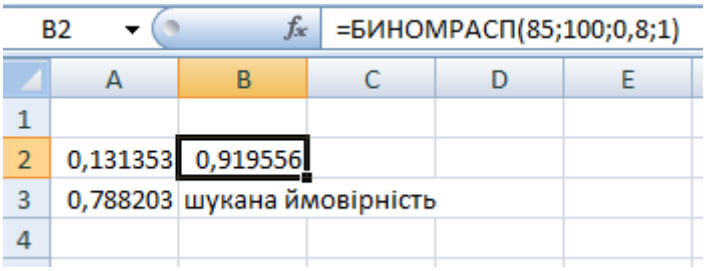


Рис. 1.17. Обчислення для інтегральної формули Муавра-Лапласа

Відповідь. а) 0,004; б) 0,79.

Приклад 1.34. Енциклопедія має 1500 сторінок. Імовірність друкарської помилки на одній сторінці дорівнює 0,001. Знайти ймовірність того, що в енциклопедії:

а) буде точно 3 помилки;

б) не буде жодної помилки;

в) буде хоча б одна помилка.

Розв'язання. а) За умовою ймовірність друкарської помилки на одній сторінці $p = 0{,}001 < 0{,}01$, а добуток

$$npq = 1500 \cdot 0{,}001 \cdot 0{,}999 = 1{,}49 < 5 \; (np = 1500 \cdot 0{,}001 = 1{,}5 < 20).$$

Тоді за формулою Пуассона (1.23) та за таблицею значень (додаток) обчислюємо

$$P_{1500}(3) = \frac{1{,}5^3}{3!} \cdot e^{-1{,}5} \approx 0{,}125.$$

Скористаємось функцією Excel категорії «Статистические» **ПУАССОН** (x; *среднее; интегральная);* для таких параметрів: x – кількість сприятливих подій; *среднее* – середнє значення $\lambda = n \cdot p$; *интегральная* – 0 або 1. За умовою задачі $x = 3$; *среднее* $\lambda = 1{,}5$; *интегральная* – 0, тобто обчислення точної кількості подій – 3 помилки для книги із 1500 сторінками (рис. 1.18).

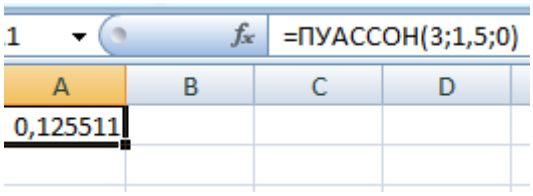


Рис. 1.18. Обчислення за допомогою функції **ПУАССОН**

б) Імовірність того, що в словнику не буде жодної помилки, тобто $k=0$, обчислюємо за тією ж формулою Пуассона

$$P_{1500}(0)=\frac{1{,}5^{0}}{0!}\cdot e^{-1{,}5}=\frac{1}{1\cdot\sqrt{e^{3}}}\approx\frac{1}{4{,}48}\approx 0{,}223.$$

в) Подія A – у словнику буде хоча б одна помилка, є протилежною до події – у словнику немає жодної помилки. Тому

$P(A)=1-P_{1500}(0)=1-0{,}223=0{,}777$.

Зауваження. Якщо аргумент *«интегральная»* дорівнює 1 («істина»), тоді функція **ПУАССОН** повертає ймовірність того, що кількість винадкових подій була в діапазоні від 0 до x включно.

Відповідь. а) 0,125; б) 0,223; в) 0,777.

Завдання для самоконтролю

1. Пояснити на прикладах залежні та незалежні події та обчислення їх ймовірностей.
2. Навести приклади повної групи несумісних подій.
3. Сформулювати теореми додавання ймовірностей для сумісних та несумісних подій. Навести приклади їх використаня.
4. Сформулювати теореми множення ймовірностей для залежних та незалежних подій. Навести приклади їх використаня.
5. Сформулювати теорему обчислення повної ймовірності, пояснити на прикладі.
6. Навести приклади використання із програмного документу Excel функцій категорії «математичні» для обчислень за формулою повної ймовірності.
7. Записати формулу Байєса та пояснити її використання.
8. Яку послідовність незалежних випробувань називають схемою Бернуллі?
9. Написати і пояснити формулу Бернуллі.
10. Як визначають найімовірніше число настання події в серії незалежних випробувань?

11. У яких випадках використовують формулу Пуассона?

12. За яких умов локальна формула Муавра-Лапласа дає добре наближення?

13. Коли використовують інтегральну формулу Муавра-Лапласа?

14. Навести приклад розв'язання задачі за формулами Бернуллі, Пуассона та локальною формулою Муавра-Лапласа, пояснити їх раціональність у використанні та найкраще наближення.

15. Скласти приклад математичної моделі простої течії подій.

16. Є 8 кандидатів на отримання роботи. Серед них є люди з відповідною кваліфікацією (подія A) і люди, що закінчили другий курс КПІ ім. Ігоря Сікорського (подія B), а також інші. Їх кількості подані таблицею.

	A	$\overline{A}$	Усього
B	1	3	4
$\overline{B}$	2	2	4
Усього	3	5	8

Усі кандидати мають рівні шанси на отримання роботи. Знайти ймовірність того, що роботу отримає людина без кваліфікації, або яка ще подовжує навчатись. *Відповідь.* 0,75.

17. Фірма має можливість отримати два контракти. Ймовірність отримання першого контракту дорівнює 0,9, а другого – 0,8. Вважаючи ці події незалежними, знайти ймовірності подій: а) фірма отримає обидва контракти; б) фірма отримає принаймні один контракт. *Відповідь.* а) 0,72; б) 0,98.

18. Імовірність хоча б одного влучення в ціль при чотирьох пострілах дорівнює 0,9919. Знайти ймовірність влучення в ціль при одному пострілі. *Відповідь.* 0,3.

19. Знайти ймовірність того, що навмання взятий виріб виявиться першосортним, якщо 4% усіх виробів є браком, а першосортні вироби складають 75% від усіх небракованих.

20. Пристрій містить два елементи, які працюють незалежно. Імовірності відмов елементів дорівнюють 0,05 та 0,08. Знайти ймовірність відмови пристрою, якщо для цього досить, щоб відмовив хоча б один елемент.

21. З опитаних бізнесменів 80 % віддають перевагу зберіганню грошей у банку (подія A), 60 % вкладає гроші в цінні папери (подія B), 50 % одночасно тримають гроші в банку та вкладають у цінні папери. Результати опитування подано в таблиці

	A	$\overline{A}$	Разом
B	0,5		0,5
$\overline{B}$			
Разом	0,8		1

Заповнити таблицю до кінця. Знайти ймовірність того, що навмання обраний бізнесмен тримає гроші в банку або у вигляді цінних паперів.

Відповідь. 0,9.

22. Економічний факультет провів обстеження працевлаштування своїх випускників. Імовірність того, що людина, яка працює в сфері бізнесу та має прибуток вище N грн., становить 0,9, а поза сферою бізнесу – 0,3. З'ясовано, що 80 % випускників працюють в сфері бізнесу. Знайти: а) ймовірність того, що навмання обраний випускник має прибуток вище N грн.; б) за умови отримання прибутку випускником більше N гривень яка ймовірність того, що він працює в сфері бізнесу? *Відповідь.* а) 0,78; б) 0,923.

23. Задачу розв'язують самостійно 2 відмінники, 3 посередні студенти та 5 студентів, що навчаються добре. Ймовірність розв'язання задачі відмінником дорівнює 0,9, студентом, який навчається добре, – 0,8 та посереднім – 0,5. До дошки навмання викликають одного зі студентів. 1) Знайти ймовірність того, що студент розв'язав задачу. 2) Викликаний студент розв'язав задачу. Знайти ймовірність того, що він: а) відмінником; б) студент із середнім рівнем успішності. *Відповідь.* 1) 0,73; 2) а) 0,247; б) 0,205.

24. Підприємець тримає свої гроші на валютному та гривневому рахунках у пропорції 3:2. Ймовірність падіння курсу валют в наступному місяці складає 0,02, а гривні – 0,06. Знайти: а) ймовірність того, що гроші підприємця девальвувалися; б) за умови девальвуації грошей підприємця яка ймовірність того, що це трапилося через падіння курсу гривні?

25. Фабрика виготовляє однотипну продукцію на трьох конвеєрних лініях, які мають однакову продуктивність. На першій лінії виробляється продукція тільки першого сорту, на другій лінії продукція першого сорту становить 90 %, а на третій – 85 %. а) Знайти ймовірність того, що навмання взятий виріб буде першосортним; б) за умови отримання першосортного виробу знайти ймовірність того, що він виготовлений на третій лінії.

26. В першій урні 3 білі та 1 чорна куля, в другій – 2 білі і 2 чорні кулі. З першої урни в другу навмання переклали одну кулю, а потім з другої урни взяли одну кулю. Яка ймовірність того, що взята куля біла? *Відповідь.* $\frac{11}{20}$.

Зауваження. Задачі розв'язувати, користуючись теоретичними знаннями та, по-можливості, засобами програмного забезпечення Excel.

Варіанти завдань для індивідуальної роботи

Завдання 1. У двох партіях k_1 та k_2% (відсотків) доброякісних виробів відповідно. Навмання вибирають по одному виробу з кожної партії. Яка ймовірність виявити серед них:

1) хоча б один бракований виріб;
2) два браковані вироби;
3) один доброякісний та один бракований виріб?

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
k_1	71	78	87	72	79	86	73	81	85	74	82	84	75	83	75
k_2	47	39	31	46	38	32	45	37	33	44	36	34	43	35	42

Завдання 2. Імовірність виграшу в лотерею на один квиток дорівнює p. Куплено n квитків. Знайти: 1) найбільш імовірне число виграшних квитків та відповідну ймовірність; 2) імовірність виграшу на k квитків.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
p	0.7	0.6	0.2	0.7	0.9	0.8	0.4	0.3	0.1	0.5	0.4	0.3	0.5	0.3	0.2
n	7	9	10	6	8	9	5	7	30	14	6	4	13	25	20
k	3	5	4	3	6	4	2	5	10	5	5	2	4	11	5

Розділ 2. Випадкові величини

У теорії ймовірностей поряд з поняттям «випадкова подія» і «ймовірність» одним з основних є поняття «випадкова величина». Наприклад, час безвідмовної роботи деякого приладу, кількість випадань герба за багаторазового підкидання монети тощо.

В теорії ймовірностей і математичній статистиці часто доцільно пов'язувати різноманітні випадкові події з дійсними числами і проводити відповідні обчислення, тому виникає нове поняття випадкової величини, про яке йтиметься далі.

2.1. Дискретні та неперервні випадкові величини

2.1.1. Види випадкових величин та способи їх задання

***Означення*. Випадковою величиною** називають таку величину, яка внаслідок випробування може набути лише одного числового значення, яке зумовлене результатом експерименту.

Отже, ***випадковою величиною***, пов'язаною з певним дослідом, називають величину, яка під час кожного проведення досліду може набувати того чи іншого числового значення, залежно від випадку.

Між випадковими подіями і випадковими величинами є тісний зв'язок. Випадкова подія – це якісна характеристика випадкового результату досліду, а випадкова величина – його кількісна характеристика. Випадкові величини за певними властивостями поділяються на ***дискретні*** та ***неперервні***.

***Означення*. Дискретною випадковою величиною** (ДВВ) називають таку величину, яка внаслідок випробування може набути відокремлених, ізольованих одне від одного числових значень з відповідними ймовірностями.

Інакше кажучи, вона має таку властивість, що кожне з її можливих значень має окіл, який вже не містить жодного з інших значень цієї самої величини. Всі

можливі значення дискретної випадкової величини можуть бути перенумеровані:

$$x_1, x_2, ..., x_n, ...$$

Означення. **Неперервною випадковою величиною** (НВВ) називають величину, яка може набувати будь-якого числового значення з певного обмеженого інтервалу (a, b) або необмеженого інтервалу $(-\infty, +\infty)$. Наприклад, випадкова величина X – час безвідмовної роботи приладу, неперервна, оскільки її можливе значення $t > 0$.

Означення. Співвідношення, яке встановлює зв'язок між можливими значеннями випадкової величини і ймовірностями, з якими приймають ці значення, називають **законом розподілу ймовірностей випадкової величини**.

Для дискретної випадкової величини X закон розподілу може бути заданий таблично або графічно.

У першому випадку закон розподілу називають ***рядом розподілу ймовірностей*** випадкової величини X:

X	0	1	2	3
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

У першому рядку таблиці записують усі можливі значення випадкової величини, а в другому − відповідні їм імовірності. Оскільки події $\{X = x_1\}$, $\{X = x_2\}$, ..., $\{X = x_n\}$ становлять повну групу несумісних подій, то за теоремою додавання ймовірностей маємо:

$$\sum_{k=1}^{\infty} p_k = 1,$$

тобто сума ймовірностей усіх можливих значень випадкової величини дорівнює одиниці.

Графічна форма закону розподілу називається ***багатокутником розподілу***: по осі абсцис відкладаємо можливі значення x_k випадкової величини X, а по осі

ординат – імовірності p_k цих значень; точки (x_k, p_k) послідовно з'єднуємо відрізками прямих.

Закон розподілу неперервної випадкової величини може бути заданий графічно або аналітично $p_k = f(x_k)$ (за допомогою формули). Табличне задання неможливе, оскільки ймовірність отримати будь-яке значення неперервної величини дорівнює нулю, що пов'язано не з неможливістю самої події (потрапляння в певну точку на числовій осі), а з нескінченно великою кількістю можливих випадків.

З огляду на це, для неперервних випадкових величин (як, зрештою, і для дискретних) визначають імовірність потрапляння в деякий інтервал числової осі.

Імовірність потрапляння випадкової величини X в інтервал $[a, b]$ визначають як імовірність події $P(a \le X < b)$.

Для кількісного оцінювання закону розподілу випадкової величини (дискретної або неперервної) задають ***функцію розподілу ймовірностей випадкової величини,*** котру визначають як імовірність того, що випадкова величина X набуде значення, меншого від певного фіксованого числа x і позначають $F(x) = P(X < x)$ або $F(x) = P(-\infty < X < x)$.

Функцію розподілу $F(x)$ називають ***інтегральною функцією*** розподілу ймовірностей випадкової величини.

Знаючи функцію розподілу $F(x)$, можна обчислити ймовірність потрапляння випадкової величини у деякий інтервал $[a, b)$:

$$P(a < X < b) = F(b) - F(a). \qquad (2.1)$$

Справді, випадкова подія $(X < b)$ – об'єднання двох несумісних подій $(X < a)$ та $(a \le X < b)$.

Отже, за теоремою додавання ймовірностей несумісних подій маємо:

$$P(X < b) = P(X < a) + P(a \le X < b).$$

Звідси, $P(a \le X < b) = P(X < b) - P(X < a)$, або, враховуючи позначення,

$$P\left(a \le X < b\right) = F(b) - F(a).$$

Зауваження. Для дискретної випадкової величини X

$$F(x) = \sum_{x_k < x} p_k . \tag{2.2}$$

Властивості функцій розподілу

1. $0 \le F(x) \le 1.$

2. Функція розподілу неспадна: якщо $x_1 < x_2$, то $F(x_1) \le F(x_2).$

3. Функція розподілу неперервна зліва: $\lim\limits_{x \to x_1 - 0} F(x) = F(x_1).$

4. Імовірність потрапляння випадкової величини X у проміжок $\left[a,\ b\right)$:

$$P\left(a \le X < b\right) = F(b) - F(a). \tag{2.3}$$

5. $P\left(X \ge x\right) = 1 - F(x).$

6. $F(-\infty) = 0, \quad F(+\infty) = 1.$

Зауваження. Неперервна випадкова величина X, що набуває значення в інтервалі (a, b), має незліченну кількість можливих значень, тому набуття X певних значень, наприклад $X = a$ або $X = b$ – події з нульовою ймовірністю. Це означає, що $P\left(X = a\right) = 0$ та $P\left(X = b\right) = 0$. З огляду на це справедливі рівності:

$$P\left(a < X < b\right) = P\left(a \le X < b\right) = P\left(a < X \le b\right) = P\left(a \le X \le b\right).$$

Тоді, згідно із зауваженням, для такої *неперервної випадкової величини* $F(x) = 0$ за $x \le a$; $F(x) = 1$ за $x \ge b.$

Графік функції розподілу зображено на рис. 2.1.

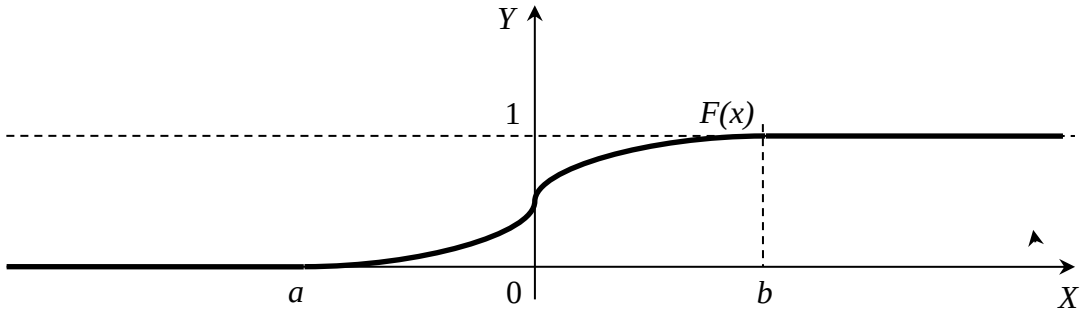


Рис. 2.1. Неперервна функція розподілу

Приклад 2.1. Побудувати закон розподілу випадкової величини X – кількості випадань герба за трьох підкидань однієї монети.

Розв'язання. За трьох підкидань монети випадкова величина X – кількість випадань герба – може набувати значення $x_1 = 0,\ x_2 = 1,\ x_3 = 2,\ x_4 = 3$ із відповідними ймовірностями, які обчислимо за формулою Бернуллі:

$$p_1 = P_3(0) = C_3^0\left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^3 = \frac{1}{8},\ \ p_2 = P_3(1) = C_3^1\left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right)^2 = \frac{3}{8};$$

$$p_3 = P_3(2) = C_3^2\left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right) = \frac{3}{8},\ \ p_4 = P_3(3) = C_3^3\left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^0 = \frac{1}{8}.$$

Ряд розподілу має такий вигляд:

X	0	1	2	3
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Побудуємо функцію розподілу випадкової величини X, заданої цим рядом розподілу:

– якщо $x \le 0$, $\quad F(x) = P\left(X < 0\right) = 0;$

– якщо $0 < x \le 1$, $\quad F(x) = P\left(X < 1\right) = P\left(X = 0\right) = \frac{1}{8};$

– якщо $1 < x \le 2$, $\quad F(x) = P\left(X < 2\right) = P\left(X = 0\right) + P\left(X = 1\right) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2};$

– якщо $2 < x \le 3$,

$$F(x) = P\left(X < 3\right) = P\left(X = 0\right) + P\left(X = 1\right) + P\left(X = 2\right) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8};$$

– якщо $x > 3$,

$$F(x) = P\left(X < 4\right) = P\left(X = 0\right) + P\left(X = 1\right) + P\left(X = 2\right) + P\left(X = 3\right) = 1.$$

Приклад 2.2. Нехай функцію розподілу деякої неперервної випадкової величини X задано у вигляді $F(x) = \begin{cases} 0, & x \le 0; \\ a(1 - \cos x), & 0 < x \le \pi; \\ 1, & x > \pi. \end{cases}$

Визначити значення коефіцієнта a.

Розв'язання. Значення коефіцієнта обчислимо, користуючись властивостями функції розподілу.

Оскільки функція неперервна зліва, то за $x=\pi$ маємо $a(1-\cos\pi)=1,$ звідки $a=\frac{1}{2}$. З другого боку, $F(\pi)-F(0)=1,$ тобто $a(1-\cos\pi)-a(1-\cos 0)=1$ і, відповідно, $2a=1,\ a=\frac{1}{2}.$

Закон розподілу ймовірностей неперервних випадкових величин може бути заданий також і ***щільністю розподілу.***

Нехай неперервна випадкова величина X задана неперервною і диференційованою функцією розподілу $F(x)$. Імовірність потрапляння цієї випадкової величини в деякий інтервал $(x,\ x+\Delta x)$ знайдемо на підставі співвідношення (2.3):

$$P\left(x<X<x+\Delta x\right)=F(x+\Delta x)-F(x),$$

тобто як приріст функції розподілу на цьому інтервалі.

Відношення

$$\frac{P\left(x<X<x+\Delta x\right)}{\Delta x}$$

виражає середню ймовірність, яка припадає на одиницю довжини інтервалу.

Перейшовши до границі за $\Delta x\to 0,$ отримаємо:

$$\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=F'(x).$$

***Означення*. Диференціальною функцією розподілу** або **щільністю розподілу ймовірностей** неперервної випадкової величини називають

$$f(x)=F'(x). \qquad (2.4)$$

Назва *щільність ймовірностей* випливає з рівності

$$f(x)=\lim_{\Delta x\to 0}\frac{P\left(X<x+\Delta x\right)-P\left(X<x\right)}{\Delta x}.$$

Властивості диференціальних функцій розподілу

1. $f(x) \geq 0, x \in R.$

2. $\int_{-\infty}^{+\infty} f(x)dx = 1.$

3. $P(a \leq X < b) = \int_{a}^{b} f(x)dx.$

4. Якщо всі можливі значення випадкової величини належать інтервалу $(a,\ b)$, то $\int_{a}^{b} f(x)dx = 1.$

За відомою диференціальною функцією розподілу $f(x)$ знаходять інтегральну функцію розподілу $F(x) = \int_{-\infty}^{x} f(t)dt.$

Геометричне тлумачення щільності розподілу випливає із формули $P(a \leq X < b) = \int_{a}^{b} f(x)dx$: імовірність потрапляння випадкової величини X на проміжок $[a, b)$ обчислюють як площу криволінійної трапеції, обмеженої зверху графіком функції $y = f(x)$, знизу – відрізком $[a, b]$ осі абсцис, зліва і справа – відрізками прямих $x = a, \quad x = b.$

2.1.2. Числові характеристики випадкових величин

Під час розв'язування практичних задач досить важко, а іноді й неможливо визначити функцію розподілу випадкової величини *X (диференціальну чи інтегральну)*, тому потрібно вміти характеризувати її розподіл через параметри, найважливіші з яких – математичне сподівання, дисперсія та середнє квадратичне відхилення. Розглянемо вказані параметри розподілу випадкових величин, як дискретних, так і неперервних.

2.1.2.1. Математичне сподівання

Означення. **Математичним сподіванням** *дискретної випадкової величини* X називають суму добутків усіх можливих значень x_k випадкової величини X та відповідних імовірностей p_k

$$M(X) = \sum_k x_k p_k. \tag{2.5}$$

Якщо при цьому множина можливих значень X нескінченна, то накладається умова абсолютної збіжності ряду.

Математичне сподівання називають центром розсіювання випадкової величини і воно відповідає середньому значенню випадкової величини.

Означення. **Математичним сподіванням** *неперервної випадкової величини* X, яка задана щільністю розподілу $f(x)$, називають

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx, \tag{2.6}$$

якщо цей інтеграл абсолютно збіжний. Якщо можливі значення неперервної випадкової величини X належать проміжку $[a, b]$, то

$$M(X) = \int_a^b xf(x)dx. \tag{2.7}$$

Означення, потрібні для розуміння властивостей числових характеристик випадкових величин

1. *Добутком сталої величини* C та дискретної випадкової величини X називають дискретну випадкову величину CX, можливі значення якої рівні добутку сталої C та можливих значень X, а ймовірності цих значень дорівнюють імовірностям відповідних значень X.

2. Дві випадкові величини називають *незалежними*, якщо закон розподілу однієї з них не залежить від того, яких можливих значень набуває інша величина. Кілька випадкових величин називають взаємно незалежними, якщо закони розподілу будь-якої кількості цих величин не залежать від того, яких можливих значень набули інші величини.

3. *Добутком незалежних дискретних випадкових величин* X та Y називають випадкову величину XY, можливі значення якої дорівнюють добуткам кожного можливого значення X та кожного можливого значення Y, а ймовірності цих значень дорівнюють добуткам імовірностей можливих значень співмножників. Якщо деякі добутки $x_i y_i$ рівні між собою, то їх імовірності додаються.

4. *Сумою дискретних випадкових величин* X та Y називають випадкову величину $X + Y$, можливі значення якої дорівнюють сумам кожного можливого значення X та кожного можливого значення Y, а імовірності цих значень для незалежних величин X і Y дорівнюють добуткам імовірностей доданків; для залежних величин – добуткам імовірностей одного доданку та умовних імовірностей другого. Якщо деякі суми $x_i + y_j$ рівні між собою, то їх імовірності додаються.

Властивості математичного сподівання

1. Математичне сподівання сталої величини – стала $M(C) = C$.

2. $M(CX) = CM(X)$.

3. $M(X + Y) = M(X) + M(Y)$.

Наслідки: *а*) $M(X_1 + X_2 + \ldots + X_n) = M(X_1) + M(X_2) + \ldots + M(X_n)$;

б) $M(X - Y) = M(X) - M(Y)$.

4. Математичне сподівання добутку двох незалежних випадкових величин:

$$M(XY) = M(X)M(Y).$$

2.1.2.2. Дисперсія. Середнє квадратичне відхилення

Означення. **Дисперсією** випадкової величини називають математичне сподівання квадрата різниці випадкової величини і її математичного сподівання:

$$D(X) = M(X - M(X))^2. \qquad (2.8)$$

Теорема. *Дисперсія випадкової величини дорівнює різниці між математичним сподіванням квадрата цієї величини і квадратом її математичного сподівання:*

$$D(X) = M(X^2) - \left(M(X)\right)^2. \tag{2.9}$$

Доведення. Згідно з означенням дисперсії та властивостями математичного сподівання маємо:

$$D(X) = M(X - M(X))^2 = M(X^2 - 2M(X)X + (M(X))^2) =$$

$$= M(X^2) - 2M(X)M(X) + (M(X))^2 = M(X^2) - 2(M(X))^2 + (M(X))^2 =$$

$$= M(X^2) - (M(X))^2.$$

За теоремою зручно для обчислення дисперсії використовувати формули:

1) для дискретної випадкової величини

$$D(X) = \sum_k x_k^2 p_k - (M(X))^2; \tag{2.10}$$

2) для неперервної випадкової величини

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - (M(X))^2; \tag{2.11}$$

якщо її можливі значення належать відрізку $[a, b]$, тоді

$$D(X) = \int_a^b x^2 f(x)dx - (M(X))^2. \tag{2.12}$$

Розмірність дисперсії не збігається з розмірністю випадкової величини. Для того, щоб розмірності були однаковими, вводять поняття ***середнього квадратичного відхилення (стандартного відхилення)***.

***Означення*. Середнім квадратичним відхиленням** випадкової величини називають

$$\sigma_x = \sigma(X) = \sqrt{D(X)}. \tag{2.13}$$

Дисперсія і середнє квадратичне відхилення – міра розсіювання значень випадкової величини навколо її математичного сподівання. В економічних задачах дисперсію часто беруть за міру ризику, оскільки вона характеризує

мінливість, розсіювання випадкової величини прибутку, зумовленої, наприклад, випадковими коливаннями цін на енергоносії.

Середнє квадратичне відхилення σ обчислюють, на відміну від дисперсії, у тій самій розмірності, що й випадкову величину. Саме це є причиною його широкого застосування для характеристики відхилень та ймовірної оцінки поведінки випадкової величини. Зокрема, середнє квадратичне відхилення має надзвичайно важливе значення для критеріальної характеристики так званого *принципу практичної впевненості*. ***Принцип практичної впевненості*** можна сформулювати так: *якщо ймовірність деякої події в певному досліді досить мала, то можна бути практично впевненим у тому, що за одноразового проведення досліду подія А не настане*. Що стосується підприємницької діяльності, *принцип практичної впевненості*, мабуть, можна сформулювати так: *можна бути майже впевненим, що запланований захід, дія чи прийняте рішення будуть здійснені, якщо ймовірність їх нездійснення та ризик досить малі.*

Властивості дисперсії

1. $D(C) = 0,\ C$ – стала величина.

2. $D(CX) = C^2 D(X).$

3. Дисперсія суми незалежних випадкових величин дорівнює сумі їх дисперсій:

$$D(X + Y) = D(X) + D(Y).$$

Наслідок. Дисперсія різниці двох незалежних випадкових величин дорівнює сумі їх дисперсій: $D(X - Y) = D(X) + D(Y).$

Приклад 2.3. За законом розподілу дискретної випадкової величини

X	1	2	3
P	0,4	0,5	0,1

знайти $D(X),\ \sigma(X).$

Розв'язання. Знаходимо $M(X)=1\cdot 0{,}4+2\cdot 0{,}5+3\cdot 0{,}1=1{,}7.$ Також $M(X^2)=1\cdot 0{,}4+4\cdot 0{,}5+9\cdot 0.1=3{,}3$. Тоді за формулою (2.9)

$$D(X)=M(X^2)-(M(X))^2=3{,}3-1{,}7^2=0{,}41,$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{0{,}41}=0{,}64.$$

Приклад 2.4. По мішені проводяться чотири незалежні постріли. Ймовірність влучення при одному пострілі дорівнює 0,25. Скласти ряд розподілу випадкової величини X − кількості влучень в мішень та обчислити його основні числові характеристики. Визначити функцію розподілу $F(x)$ та побудувати її графік.

Розв'язання. Випадкова величина X − кількість попадань в мішень може приймати значення 0, 1, 2, 3, 4. Оскільки розглядувані випробування задовольняють схемі Бернуллі, то для запису закону розподілу використовують формулу Бернуллі. У даному випадку $n=4,\ p=0{,}25,\ q=1-p=0{,}75.$

Також для складання ряду розподілу використаємо функцію Excel категорії «Статистические» **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*) із такими параметрами: *Число_успехов*: k − змінна величина, яка приймає значення: 0, 1, 2, 3, 4; *Число_испытаний*: 4 − кількість незалежних випробувань; *Вероятность_успеха:* 0,25 − ймовірність успіху у кожному випробуванні; *Интегральная:*0 − для знаходження ймовірності випадкової події $P(X=k)$.

Відповідні ймовірності, знайдені за допомогою даної функції **БИНОМРАСП**, запишемо у вигляді ряду розподілу (табл. 2.1).

Табл. 2.1

k	0	1	2	3	4	Сума
P	0,3164	0,4219	0,2109	0,0469	0,0039	1

Знайдемо основні числові характеристики розподілу даної випадкової величини: математичне сподівання, дисперсію, середнє квадратичне відхилення.

Математичне сподівання дискретної випадкової величини обчислюється за формулою (2.5):

$$M(X)=\sum_{k=0}^{4} x_k p_k = 0\cdot 0{,}3164+1\cdot 0{,}4219+2\cdot 0{,}2109+3\cdot 0{,}0469+4\cdot 0{,}0039=1.$$

Для обчислення дисперсії знайдемо $M(X^2)$

$$M(X^2)=\sum_{k=0}^{4} (x_k)^2 p_k = 0\cdot 0{,}3164+1\cdot 0{,}4219+4\cdot 0{,}2109+9\cdot 0{,}0469+16\cdot 0{,}0039=1{,}75$$

Звідси

$$D(X)=M(X^2)-(M(X))^2=1{,}75-1^2=0{,}75.$$

Середнє квадратичне відхилення обчислимо за формулою

$$\sigma(X)=\sqrt{D(X)}=\sqrt{0{,}75}=0{,}866025.$$

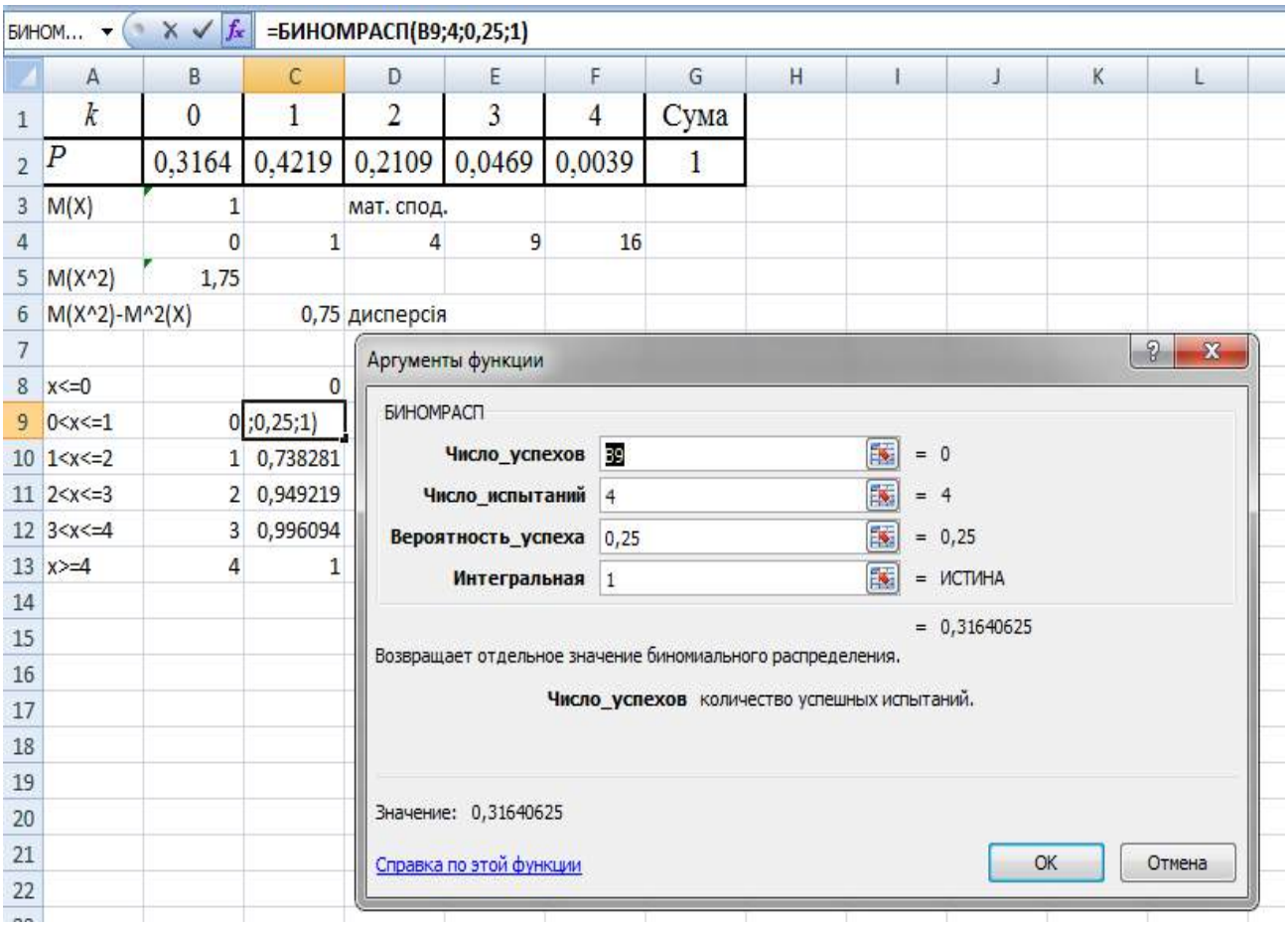
Зауваження. Всі основні числові характеристики зручно обчислювати в Excel, власне для математичного сподівання використати функцію категорії «Математические» **СУММПРОИЗВ** (*массив 1, массив 2, массив 3*), для дисперсії обчислити за цією ж функцією $M(X^2)$ і задати формулу (3.13), для середнього квадратичного відхилення використати функцію **КОРЕНЬ** (*число*).

Для функції розподілу випадкової величини скористаємось функцією **БИНОМРАСП** (*число_успехов; число_испытаний; вероятность_успеха; интегральная*) з параметрами: *Число_успехов*: k − змінна величина, яка приймає значення 0, 1, 2, 3, 4; *Число_испытаний*: 4 − кількість незалежних випробувань; *Вероятность_успеха:* 0,25 − ймовірність успіху кожного випробування; *Интегральная:* 1 − для знаходження функції розподілу $F(x)$. При використанні функції **БИНОМРАСП** для побудови функції розподілу $F(x)$ потрібно врахувати, що **БИНОМРАСП** повертає значення $P(X \le x)$, а не $P(X < x)$. Результат застосування табличного процесора Microsoft Excel для розв'язування даної задачі показано на рис. 2.2.

Відповідні значення, знайдені за допомогою даної функції, наведено в табл. 2.1.

Функція розподілу $F(x)$ має вигляд

$$F(x)=\begin{cases} 0, & \text{якщо} \quad x \le 0; \\ 0,3164, & \text{якщо} \quad 0 < x \le 1; \\ 0,7383, & \text{якщо} \quad 1 < x \le 2; \\ 0,9492, & \text{якщо} \quad 2 < x \le 3; \\ 0,9961, & \text{якщо} \quad 3 < x \le 4; \\ 1, & \text{якщо} \quad x > 4. \end{cases}$$



	A	B	C	D	E	F	G
1	k	0	1	2	3	4	Сума
2	P	0,3164	0,4219	0,2109	0,0469	0,0039	1
3	M(X)	1		мат. спод.			
4		0	1	4	9	16	
5	M(X^2)	1,75					
6	M(X^2)-M^2(X)		0,75	дисперсія			
7							
8	x<=0		0				
9	0<x<=1	0	;0,25;1)				
10	1<x<=2	1	0,738281				
11	2<x<=3	2	0,949219				
12	3<x<=4	3	0,996094				
13	x>=4	4	1				

Рис. 2.2. Вигляд обчислень на екрані

Її можна зобразити графічно в Excel (рис. 2.3).

	A	B
1	X	F(x)
2	-2	0
3	-1	0
4	0	0
5	0	0,3164
6	0,999999	0,3164
7	1	0,7383
8	1,999999	0,7383
9	2	0,9492
10	2,999999	0,9492
11	3	0,9961
12	3,999999	0,9961
13	4	1
14	5	1
15	6	1

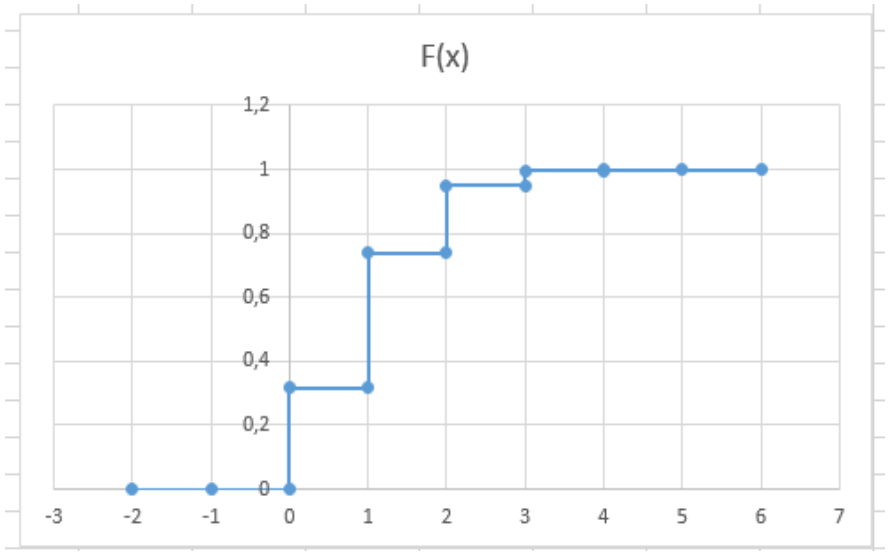


Рис. 2.3. Вигляд функції розподілу на екрані

Приклад 2.5. Середня кількість відвідувачів магазину протягом 10-ти хвилинного інтервалу дорівнює 2. Поява відвідувачів у магазині відбувається випадково і незалежно один від одного. Потрібно:

1. Скласти ряд розподілу випадкової величини – кількості відвідувачів магазину протягом 10 хвилин і побудувати його графік.

2. Знайти числові характеристики цього розподілу.

3. Обчислити функцію розподілу кількості відвідувачів магазину протягом 10 хвилин.

4. Обчислити ймовірність того, що протягом 10 хвилин кількість відвідувачів магазину виявиться менше 3 і не менше 3.

Розв'язання. Нехай випадкова величина X – кількість відвідувачів магазину протягом 10 хвилин. Дана дискретна випадкова величина може набувати значення X: $0,1,2,\ldots,n$, середнє значення дорівнює 2. Отже, для даної величини можна використовувати формулу Пуассона з $\lambda = 2$ і $X = k = 0,1,2,\ldots,n.$

1. Для складання ряду розподілу скористаємось функцією Excel **ПУАССОН** (*x*; *среднее; интегральная);* за таких параметрів: $x = k$ – змінна

величина, яка набуває значення: $0,1,2,\ldots,n$; *среднее:* 2 − середнє значення $\lambda = n \cdot p$; *интегральная:* 0 − для знаходження ймовірності випадкової події $X = k$.

Відповідні ймовірності, знайдені за допомогою даної функції, наведено в табл. 2.2, а відповідний графік − на рис. 2.4.

Табл. 2.2

k	0	1	2	3	4	5	6	7	8	9
p_k	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361	0,0120	0,0034	0,0009	0,0002

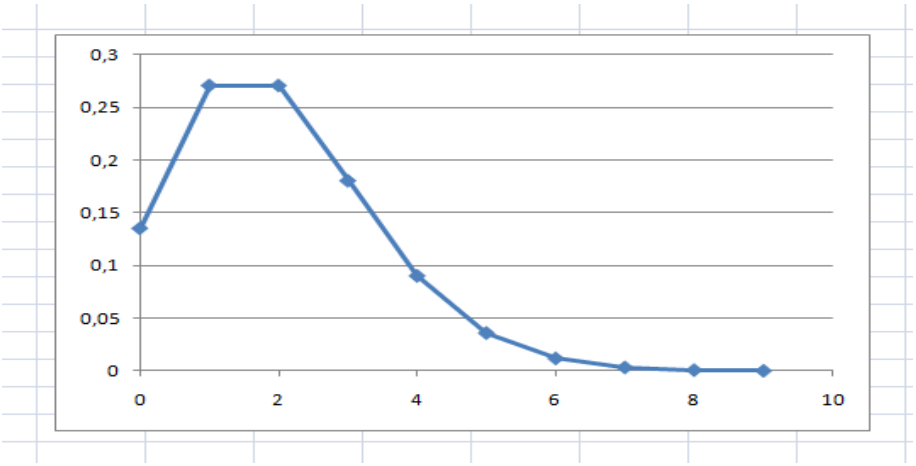


Рис. 2.4. Багатокутник розподілу

Для побудови ймовірнісного багатокутника (багатокутника розподілу) використано команду «**Диаграмма**» ***(Вставка ⇒ Диаграмма ⇒ Точечная**...)* з відповідними параметрами.

2. В даному прикладі числові характеристики обчислюються за відповідними формулами або, як на рис. 2.5, в програмному документі, де комірка E6 показана на рис. 2.6.

$M(X) \approx 2, \quad D(X) \approx 2, \quad \sigma(X) = 1,4142.$

F6 =E6-E5*E5

	A	B	C	D	E	F	G	H	I	J	K
1		0	1	4	9	16	25	36	49	64	81
2	k	0	1	2	3	4	5	6	7	8	9
3	p_k	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361	0,012	0,0034	0,0009	0,0002
4											
5		Математичне сподівання			1,9994						
6		Дисперсія			5,9952	1,9976					
7		Середнє квадр. Відхилення				1,413365					

Рис. 2.5. Обчислення числових характеристик

=СУММПРОИЗВ(B1:K1;B3:K3)

Рис. 2.6. Обчислення для комірки E6

3. Функцію розподілу зручно обчислити за допомогою функції **ПУАССОН** (*x; среднее; интегральная*) для параметрів «*x*» та «среднее» за умовою «*итегральная*» =1. Результати обчислень наведені в табл. 2.3, графік функції зображено на рис. 2.7.

Табл. 2.3

k	0	1	2	3	4	5	6	7	8	9
$F(x)$	0,1353	0,4060	0,6767	0,8571	0,9473	0,9834	0,9955	0,9989	0,9998	1,0000

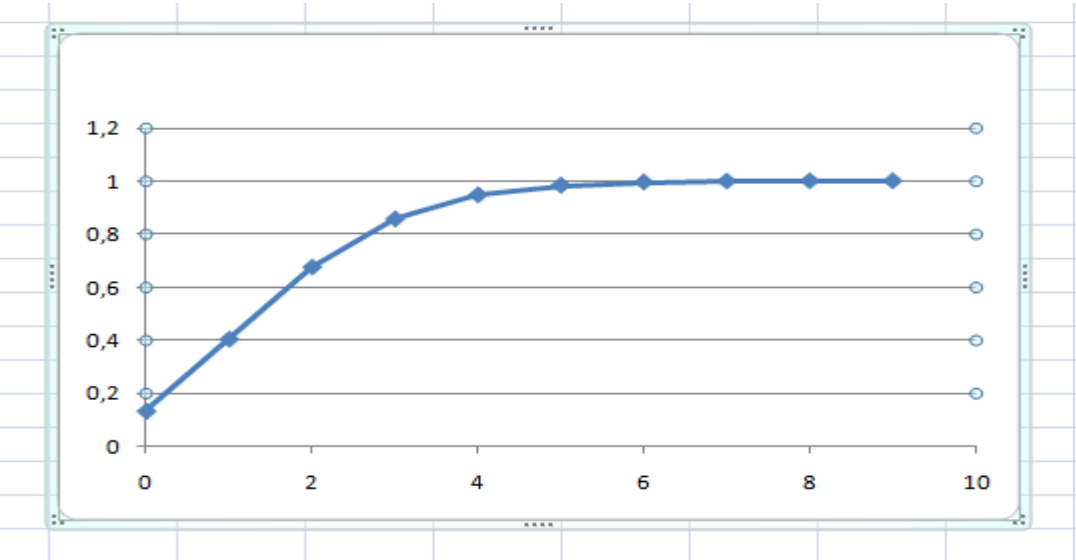


Рис. 2.7. Вигляд функції розподілу

4. Імовірність того, що протягом 10 хвилин кількість відвідувачів магазину виявиться менше 3, можна обчислити за формулою

$$P(X<3)=P(X=0)+P(X=1)+P(X=2)=0{,}1353+0{,}2707+0{,}2707=0{,}6767,$$

де числові дані взяті з таблиці (додаток). Такий самий результат можна одержати зразу, із використанням функції **ПУАССОН** за $X\le 2$ (рис. 2.8).

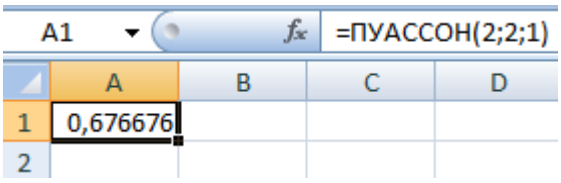


Рис. 2.8. Використання функції **ПУАССОН**

Імовірність того, що протягом 10 хвилин кількість відвідувачів магазину виявиться не менше 3, тобто 3 і більше, можна обчислити за формулою переходу до протилежної події:

$$P(X\ge 3)=1-P(X<3)=1-0{,}6767=0{,}3133.$$

Приклад 2.6. Закон розподілу неперервної випадкової величини X задано формулою

$$F(x)=\begin{cases}0, & \text{якщо} \quad x\le -1;\\ \dfrac{(x+1)^3}{64}, & \text{якщо} \quad -1<x\le 3;\\ 1, & \text{якщо} \quad x>3.\end{cases}$$

Знайти $f(x)$ та побудувати графіки функцій $f(x)$ і $F(x)$. Обчислити $P(0<X<2)$.

Розв'язання. За означенням диференціальної функції (3.4) маємо

$$f(x)=F'(x)=\begin{cases}0, & \text{якщо} \quad x\le -1;\\ \dfrac{3\cdot(x+1)^2}{64}, & \text{якщо} \quad -1<x\le 3;\\ 0, & \text{якщо} \quad x>3.\end{cases}$$

Графіки функцій $f(x)$ і $F(x)$ будують за їх табличними значеннями:

x	-1	0	1	2	3	4
$F(x)$	0	0,016	0,016	0,016	0,016	1
$f(x)$	0	0,047	0,188	0,422	0,750	0

При виконанні обчислень в Excel інтегральну та диференціальну функцію задають вручну, вибравши за її виглядом відповідні значення незалежної змінної (рис. 2.9).

Імовірність події $0 < X < 2$ обчислимо за формулою (2.3):

$$P(0 < x < 2) = F(2) - F(0) = \frac{27}{64} - \frac{1}{64} = \frac{13}{32}.$$

Такий же результат із використанням диференціальної функції:

$$P(0 < x < 2) = \int_0^2 \frac{3 \cdot (x+1)^2}{64} dx = \frac{(x+1)^3}{64}\Big|_0^2 = \frac{27}{64} - \frac{1}{64} = \frac{13}{32}.$$

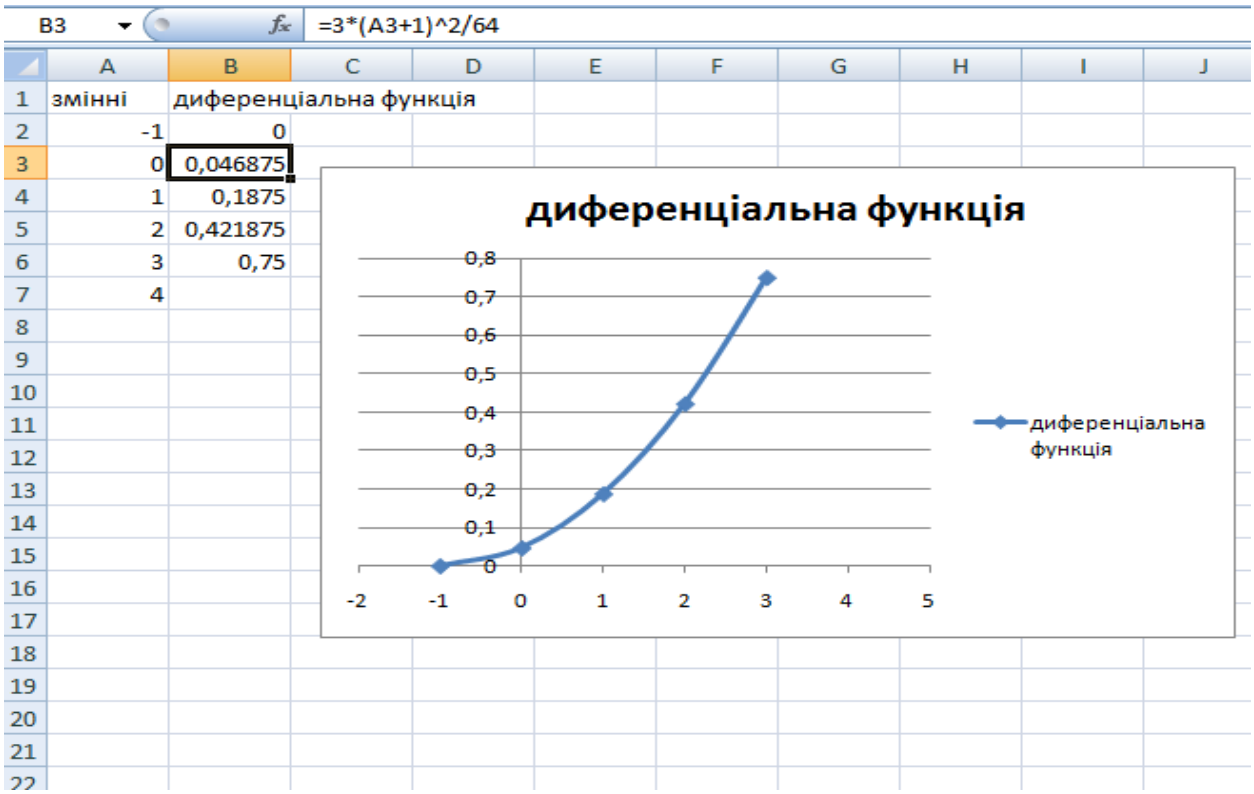


Рис. 2.9. Вигляд обчислень на екрані для побудови графіка

Приклад 2.7. Диференціальна функція розподілу випадкової величини X має вигляд: $f(x)=\begin{cases} 0, & \text{якщо} \quad x \le 0; \\ \frac{1}{2}\sin x, & \text{якщо} \quad 0 < x \le \pi; \\ 0, & \text{якщо} \quad x > \pi. \end{cases}$

Знайти $F(x)$ та побудувати графіки функцій $f(x)$ і $F(x)$. Обчислити $P\left(\frac{\pi}{6} < X < \frac{\pi}{2}\right)$.

Розв'язання. Згідно з формулою (2.12) маємо:

$$F(x)=\int_0^x \frac{1}{2}\sin t \cdot dt = \frac{1}{2}(-\cos t)\Big|_0^x = \frac{1}{2}(1-\cos x).$$

Таким чином функція розподілу ймовірності має вигляд

$$F(x)=\begin{cases} 0, & \text{якщо} \quad x \le 0; \\ \frac{1}{2}(1-\cos x), & \text{якщо} \quad 0 < x \le \pi; \\ 1, & \text{якщо} \quad x > \pi. \end{cases}$$

Графіки функцій $f(x)$ та $F(x)$, побудовані за їх табличними значеннями, зображено відповідно на рисунках 2.10 та 2.11.

x	0,00	0,52	1,05	1,57	2,09	2,64	3,14
$f(x)$	0,00	0,25	0,43	0,50	0,43	0,24	0,00
$F(x)$	0,00	0,07	0,25	0,50	0,75	0,94	1,00

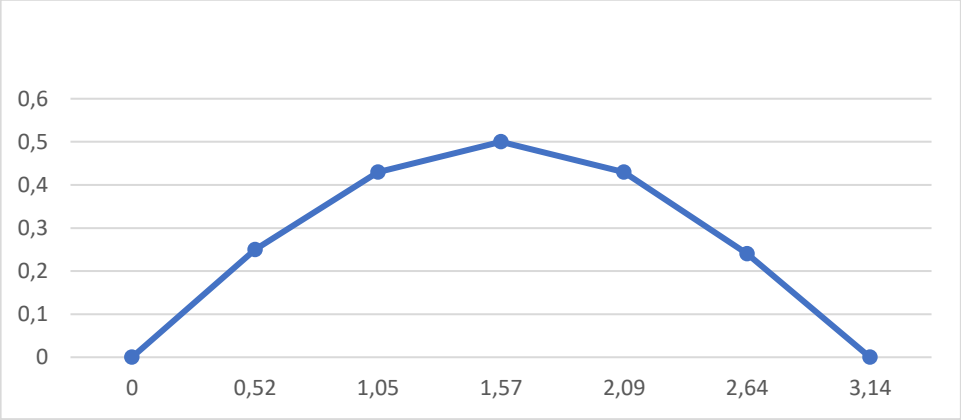


Рис. 2.10. Диференціальна функція

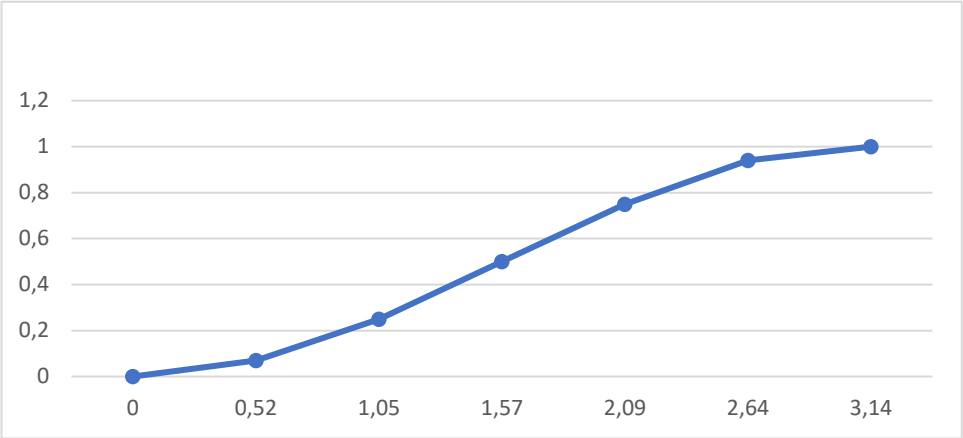


Рис. 2.11. Інтегральна функція

Імовірність події $P\left(\frac{\pi}{6} < x < \frac{\pi}{2}\right)$ можна обчислити за формулою

$$P\left(\frac{\pi}{6} < x < \frac{\pi}{2}\right) = \frac{1}{2}\int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \sin t dt = \frac{1}{2}\left(-\cos t\right)\Big|_{\frac{\pi}{6}}^{\frac{\pi}{2}} = -\frac{1}{2}\left(0 - \frac{\sqrt{3}}{2}\right) = \frac{\sqrt{3}}{4},$$

також: $P\left(\frac{\pi}{6} < x < \frac{\pi}{2}\right) = F\left(\frac{\pi}{2}\right) - F\left(\frac{\pi}{6}\right) = \frac{\sqrt{3}}{4}.$

Приклад 2.8. Знайти числові характеристики випадкової величини X, яка задана функцією розподілу $F(x) = \begin{cases} 0, & x \le 0; \\ \frac{x^2}{25}, & 0 < x \le 5; \\ 1, & x > 5. \end{cases}$

Розв'язання. Спочатку знайдемо диференціальну функцію розподілу, тобто щільність розподілу ймовірності за формулою $f(x) = F'(x)$.

$$f(x) = \begin{cases} \frac{2}{25}x, & 0 < x \le 5, \\ 0, & x \notin (0, 5]. \end{cases}$$

За формулою $M(X) = \int_a^b xf(x)dx$ знаходимо математичне сподівання:

$$M(X)=\int_0^5 x\cdot\frac{2}{25}xdx=\frac{2}{25}\cdot\frac{x^3}{3}\Big|_0^5=\frac{2}{75}\left(5^3-0\right)=\frac{10}{3}.$$

Для дисперсії використовуємо формулу $D(X)=\int_a^b x^2 f(x)dx-(M(X))^2$:

$$D(X)=\int_0^5 x^2\cdot\frac{2}{25}xdx-\left(\frac{10}{3}\right)^2=\frac{2}{25}\cdot\frac{x^4}{4}\Big|_0^5-\frac{100}{9}=\frac{25}{18}.$$

Середнє квадратичне відхилення

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{25}{18}}=\frac{5}{3\sqrt{2}}=\frac{5\sqrt{2}}{6}\approx 1{,}17.$$

2.1.2.3. Однаково розподілені взаємно незалежні випадкові величини

Нехай X_1, X_2, …, X_n – n взаємно незалежних однаково розподілених випадкових величин, які мають однакові характеристики. Позначимо через a і D відповідно їх математичне сподівання і дисперсію, $\sigma=\sqrt{D}$. Нехай $\overline{X}=\frac{X_1+X_2+\ldots+X_n}{n}$ – це середнє арифметичне цих величин.

Тоді числові характеристики $\overline{X}$ такі:

1. $M(\overline{X})=a.$

2. $D(\overline{X})=\frac{D}{n}.$ 3. $\sigma(\overline{X})=\frac{\sigma}{\sqrt{n}}.$

З огляду на ці результати середнє арифметичне досить великої кількості взаємно незалежних випадкових величин має менше розсіювання, ніж кожна окрема величина. Тому середнє арифметичне кількох вимірювань беруть за наближене значення вимірюваної величини.

2.1.2.4. Початкові та центральні моменти, інші числові характеристики

***Означення*. Початковим моментом порядку** k (k – натуральне число) випадкової величини X називають математичне сподівання величини X^k:

$$\nu_k=M(X^k). \qquad (2.14)$$

Зокрема, $\nu_1 = M(X)$, $\nu_2 = M(X^2)$, тоді формулу (2.9) для обчислення дисперсії можна записати так: $D(X) = \nu_2 - \nu_1^2$.

***Означення*. Центральним моментом порядку** k **(**k – натуральне число**)** випадкової величини X називають математичне сподівання величини $(X - M(X))^k$:

$$\mu_k = M(X - M(X))^k. \tag{2.15}$$

Зокрема, $\mu_1 = M(X - M(X)) = 0$, $\mu_2 = M(X - M(X))^2 = D(X)$, відповідно порівнюючи формули дисперсії маємо: $\mu_2 = \nu_2 - \nu_1^2$.

Справедливі формули:

$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3;$$

$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4.$$

Для симетричного розподілу кожен центральний момент непарного порядку дорівнює нулю, тому кожен відмінний від нуля центральний момент непарного порядку характеризує ступінь асиметрії розподілу.

***Означення*.** Величину, яка є безрозмірною і має вигляд

$$A_s = \frac{\mu_3}{\sigma^3} \tag{2.16}$$

називають **асиметрією розподілу** (характеристика зсуву – вершини графіка щільності розподілу).

***Означення*. Ексцесом** називають нормований центральний момент четвертого порядку

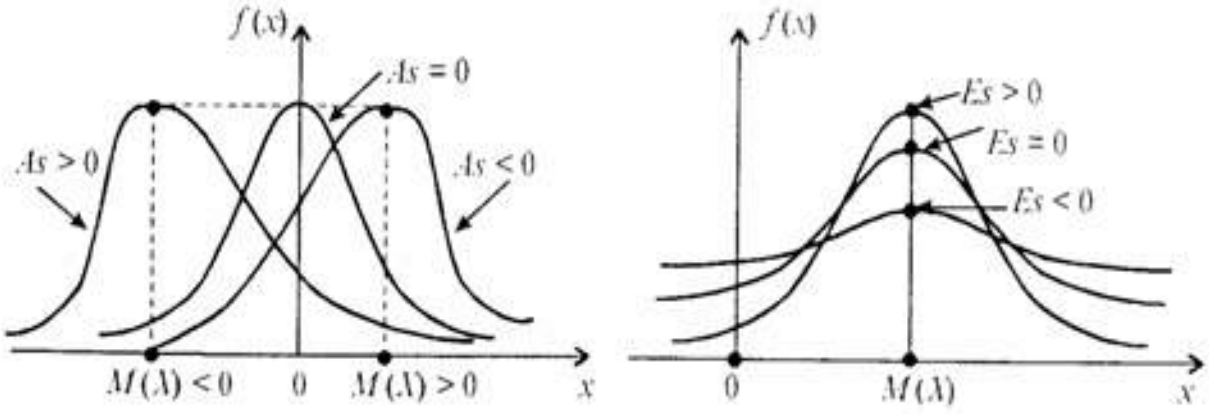
$$E_s = \frac{\mu_4}{\sigma^4} - 3, \tag{2.17}$$

що є характеристикою гостроверхості (плосковерхості) вершини графіка щільності розподілу.

Для наочності графіки щільності розподілу ймовірностей неперервної випадкової величини $f(x)$ за різних значень асиметрії A_s та ексцесу E_s наведено на рис. 2.12.

***Означення*. Модою** $M_o(X)$ розподілу неперервної випадкової величини X із щільністю розподілу $f(x)$ називають кожне значення x, за якого $f(x)$ має максимум. Розподіли, які мають одну моду, називають *унімодальними.*

***Означення*. Медіаною** $M_e(X)$ розподілу неперервної випадкової величини X називають можливе значення x, за якого пряма $x = M_e(X)$ ділить криволінійну трапецію, обмежену кривою розподілу та віссю Ox, на частини рівної площі.



а *б*

Рис. 2.12. Графіки $f(x)$ за різних значень: *а* – асиметрії; *б* – ексцесу

Приклад 2.9. Випадкова величина X задана щільністю розподілу

$$f(x) = \begin{cases} -\frac{3}{4}x^2 + 6x - \frac{45}{4}, & x \in (3,\ 5); \\ 0, & x \notin (3,\ 5). \end{cases}$$

Знайти моду, математичне сподівання і медіану випадкової величини X.

Розв'язання. Запишемо щільність розподілу ймовірностей у вигляді: $f(x) = -\frac{3}{4}(x-4)^2 + \frac{3}{4}$. Звідси видно, що за $x = 4$ щільність розподілу $f(x)$ досягає максимуму, отже, і $M_o(X) = 4$. Крива симетрична відносно прямої $x = 4$, тому $M(X) = 4,\ M_e(X) = 4$.

Зауваження. В середовищі Excel функція **СКОС** повертає вибірковий коефіцієнт асиметрії, функція **ЭКСЦЕСС** повертає вибірковий коефіцієнт

ексцесу, функція **МОДА** повертає моду (значення, що найчастіше зустрічається) набору даних, **МЕДІАНА** − величина, що міститься всередині варіаційного ряду.

Завдання для самоконтролю

1. Означити випадкову величину та навести приклади.
2. Пояснити відмінність ДВВ від НВВ.
3. Означити закон розподілу ДВВ та пояснити побудову многокутника розподілу.
4. Пояснити на прикладі побудову функції розподілу ДВВ.
5. Записати основні числові характеристики дискретної випадкової величини, їх обчислення та практичне обґрунтування.
6. Пояснити на прикладі обчислення математичного сподівання за формулою та за допомогою функцій Excel.
7. Означення дисперсії випадкової величини, основні формули її обчислення для дискретних та неперервних випадкових величин.
8. Пояснити на прикладі обчислення дисперсії за формулою та за допомогою функцій Excel.
8. Навести приклади використання функції **БИНОМРАСП.**
9. Навести приклади використання функції **ПУАССОН.**
10. Пояснити використання «*интегральная*» 0 чи 1 для обчислень у функціях **БИНОМРАСП, ПУАССОН.**
11. За даним законом розподілу дискретної випадкової величини X побудувати багатокутник розподілу. Знайти $M(X), D(X)$ і $\sigma(X)$.

X	-2	-1	0	4	5	7
p	0,12	0,18	0,2	0,3	0,17	0,03

12. Завод випускає 96% виробів першого сорту та 4% − виробів другого сорту. Навмання відібрали партію з 100 виробів. Побудувати закон розподілу ймовірностей

дискретної випадкової величини X – кількості виробів другого сорту в цій вибірці, зобразити багатокутник розподілу, знайти функцію розподілу ймовірностей $F(x)$; обчислити $M(X)$, $D(X)$, $\sigma(X)$.

13. Телефонна станція обслуговує 1000 абонентів. Імовірність того, що протягом години абонент розмовлятиме по телефону дорівнює, в середньому, 0,002. Знайти $M(X)$, $D(X)$, $\sigma(X)$ дискретної випадкової величини X – кількості абонентів, що розмовляють протягом години; функцію розподілу ймовірностей та побудувати багатокутник розподілу.

14. Серед 15 однакових мобільних телефонів 14 відповідають вимогам стандарту. Побудувати закон розподілу дискретної випадкової величини X – кількості телефонів, що відповідають вимогам стандарту серед 2 навмання взятих. Обчислити $M(X)$, $D(X)$ та побудувати функцію розподілу.

15. Імовірність того, що футболіст реалізує одинадцятиметровий штрафний удар дорівнює 0,9. Футболіст виконав три такі удари. Побудувати закон розподілу ймовірностей дискретної випадкової величини X – кількості реалізованих штрафних. Обчислити $M(X)$, $D(X)$, $\sigma(X)$.

16. Неперервна випадкова величина задана щільністю розподілу $y = f(x)$. Записати інтегральну функцію розподілу $y = F(x)$; обчислити числові характеристики $M(X), D(X)$, а також обчислити $P(\alpha \le X \le \beta)$. Зробити креслення графіків диференціальної та інтегральної функцій розподілу.

16.1. $f(x) = \begin{cases} 0, & x \le 0,\ x > \pi/2, \\ \cos x, & 0 < x \le \pi/2; \end{cases}$, $\alpha = -\pi/4;\ \beta = \pi/4$.

16.2. $f(x) = \begin{cases} 0, & x \le 1, \\ x - \dfrac{1}{2}, & 1 < x \le 2, \\ 0, & x > 2; \end{cases}$ $\alpha = 1;\ \beta = 1{,}5$.

17. Задана функція розподілу неперервної випадкової величини X. Знайти коефіцієнт A; записати щільність розподілу $f(x)$; обчислити числові характеристики $M(X), D(X),$ а також $P(\alpha \le X \le \beta)$. Зробити креслення графіків функції розподілу та щільності розподілу.

17.1. $F(x) = \begin{cases} 0, & x \le 0, \\ A(-x^2 + 4x), & 0 < x \le 2, \\ 1, & x > 2; \end{cases} \quad \alpha = 0{,}5;\ \beta = 1{,}5.$

17. 2. $F(x) = \begin{cases} 0, & x \le \dfrac{\pi}{6}, \\ A\cos 3x, & \dfrac{\pi}{6} < x \le \dfrac{\pi}{3}, \\ 1, & x > \dfrac{\pi}{3}; \end{cases} \quad \alpha = \dfrac{\pi}{6}, \quad \beta = \dfrac{\pi}{4}.$

Варіанти завдань для індивідуальної роботи

Завдання 1. За даним законом розподілу дискретної випадкової величини X побудувати багатокутник розподілу. Знайти $M(X), D(X)$ і $\sigma(X)$.

Вказівка. Виконувати обчислення за формулами та за допомогою програмного документа.

1.

X	-2	-1	0	4	5	7
p	0,12	0,18	0,2	0,3	0,17	0,03

2.

X	-3	-1	0	1	3	5
p	0,13	0,17	0,2	0,3	0,18	0,02

3.

X	-2	-1	0	1	3	4
p	0,15	0,2	0,25	0,2	0,15	0,05

4.

X	2	4	7	9	12	15
p	0,05	0,15	0,35	0,2	0,15	0,1

5.

X	3	4	7	9	12	14
p	0,1	0,3	0,2	0,05	0,15	0,2

6.

X	1	4	8	9	12	13
p	0,12	0,18	0,2	0,3	0,18	0,02

7.

X	-5	-4	0	1	2	4
p	0,15	0,2	0,25	0,2	0,15	0,05

8.

X	-2	-1	0	1	4	6
p	0,12	0,28	0,22	0,18	0,12	0,08

9.

X	-7	-5	-2	1	5	9
p	0,13	0,17	0,2	0,3	0,18	0,02

10.

X	-8	-6	-2	1	5	6
p	0,12	0,28	0,22	0,18	0,12	0,08

11.

X	-3	-2	0	1	2	4
p	0,13	0,17	0,2	0,3	0,18	0,02

12.

X	-6	-5	-2	3	5	7
p	0,13	0,17	0,2	0,3	0,18	0,02

13.

X	-8	-4	-2	2	5	8
p	0,12	0,18	0,2	0,3	0,17	0,03

14.

X	-10	-5	-2	1	5	10
p	0,12	0,28	0,22	0,18	0,12	0,08

15.

X	-9	-5	-1	1	5	7
p	0,05	0,15	0,35	0,2	0,15	0,1

Завдання 2.

Проводять k незалежних випробувань, в яких імовірність успіху для кожного з них дорівнює p. Скласти ряд розподілу випадкової величини X – кількості успішних випробувань та обчислити основні числові характеристики.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
p	0.2	0.55	0.3	0.4	0.6	0.2	0.45	0.3	0.25	0.6	0.4	0.3	0.35	0.3	0.45
k	3	2	4	3	5	4	3	5	3	2	4	5	4	2	6

Завдання 3. Задана функція розподілу неперервної випадкової величини X. Знайти коефіцієнт A; записати щільність розподілу $f(x)$; обчислити числові характеристики MX, DX, а також $P\{\alpha \le X \le \beta\}$. Зробити креслення інтегральної функції розподілу та щільності розподілу.

1. $F(x)=\begin{cases}0, & x\le 0,\\ A\left(2x-\dfrac{x^2}{3}\right), & 0<x\le 3,\\ 1, & x>3\end{cases}$ $\alpha=1;\ \beta=2$.

2. $F(x)=\begin{cases}0, & x\le 0,\\ Ax^2, & 0<x\le 1,\\ 1, & x>1\end{cases}$ $\alpha=0{,}3;\ \beta=0{,}8$.

3. $F(x)=\begin{cases}0, & x\le 2,\\ A\left(x^2-4x+4\right), & 2<x\le 4,\\ 1, & x>4\end{cases}$ $\alpha=0{,}2;\ \beta=3{,}5$.

4. $F(x)=\begin{cases}0, & x\le 0,\\ Ax(2-x), & 0<x\le 1,\\ 1, & x>1\end{cases}$ $\alpha=0{,}1;\ \beta=0{,}4$.

5. $F(x)=\begin{cases}0, & x\le 2,\\ A(x-2), & 2<x\le 8,\\ 1, & x>8;\end{cases}$ $\alpha=3;\ \beta=5$.

6. $F(x)=\begin{cases}0, & x\le -1,\\ A(x+1)^2, & -1<x\le 2,\\ 1, & x>2;\end{cases}$ $\alpha=-0{,}5;\ \beta=1{,}5$.

7. $F(x)=\begin{cases}0, & x\le 1,\\ A(x-1)^2, & 1<x\le 2,\\ 1, & x>2;\end{cases}$ $\alpha=1;\ \beta=\dfrac{4}{3}$.

8. $F(x)=\begin{cases}0, & x\le 0,\\ Ax^3, & 0<x\le 2,\\ 1, & x>2;\end{cases}$ $\alpha=0{,}2;\ \beta=1{,}7$.

9. $F(x)=\begin{cases}0, & x\le 0,\\ A(3x^2-x^3), & 0<x\le 2,\\ 1, & x>2;\end{cases}$ $\alpha=0{,}3;\ \beta=1{,}2$.

10. $F(x)=\begin{cases}0, & x\le 2,\\ A(9x^2-24x-x^3+20), & 2<x\le 4,\\ 1, & x>4;\end{cases}$ $\alpha=2{,}5;\ \beta=3{,}5$.

11. $F(x)=\begin{cases}0, & x\le 0,\\ A(4x-x^2), & 0<x\le 2,\\ 1, & x>2;\end{cases}$ $\alpha=0{,}5;\ \beta=1{,}5$.

12. $F(x)=\begin{cases}0, & x\le -1,\\ A(2\arcsin x+\pi), & -1<x\le 1,\\ 1, & x>1;\end{cases}$ $\alpha=0;\ \beta=0{,}5$.

13. $F(x)=\begin{cases}0, & x\le 0,\\ A(1-\cos x), & 0<x\le \pi,\\ 1, & x>\pi;\end{cases}$ $\alpha=\dfrac{\pi}{4};\ \beta=\dfrac{\pi}{2}$.

14. $F(x)=\begin{cases}0, & x\le 0,\\ A(1-e^{-\lambda x}), & x>0,\ \lambda>0;\end{cases}$ $\alpha=-\infty;\ \beta=\dfrac{1}{\lambda}$.

15. $F(x)=\begin{cases}0, & x\le -1,\\ \dfrac{(1+x)^2}{2}, & -1<x\le 0,\\ A\left(2-(1-x)^2\right), & 0<x\le 1,\\ 1, & x>1\end{cases}$ $\alpha=-0{,}5;\ \beta=0{,}5$.

2.2. Основні закони розподілу випадкових величин

2.2.1. Основні закони розподілу дискретних випадкових величин, їх основні числові характеристики

2.2.1.1. Біномний розподіл

Проводять n однакових незалежних випробувань, у кожному з яких може відбутись подія A з імовірністю p $(0 < p < 1)$. Випадкова величина X – кількість разів настання події A. Ряд **біномного розподілу**

X	0	1	2	…	k	…	n
P	$p_n(0)$	$p_n(1)$	$p_n(2)$	⋯	$p_n(k)$	⋯	$p_n(n)$

де $p_n(k) = C_n^k p^k q^{n-k}$, $q = 1 - p$, $(k = 0, 1, 2, \ldots, n)$.

Для біномного розподілу

$$M(X) = np;$$
$$D(X) = npq.$$

Доведення. Нехай випадкова величина X_k набуває двох значень: 1, якщо подія A настала в k-му випробуванні, і 0, якщо подія A не настала в k-му випробуванні $(k = 1, 2, \ldots, n)$. Ряд розподілу для випадкової величини X_k має такий вигляд:

X_k	0	1
P	$1 - p$	p

Тоді $X = X_1 + X_2 + \ldots + X_n$. Визначимо числові характеристики для випадкової величини X_k: $M(X_k) = 0 \cdot (1 - p) + 1 \cdot p = p$, $D(X_k) = M(X_k^2) - (M(X_k))^2 =$ $= 0^2 \cdot (1 - p) + 1^2 \cdot p - p^2 = p - p^2 = p(1 - p) = pq.$

Тоді $M(X) = M(X_1) + M(X_2) + \ldots + M(X_n) = np;$

$$D(X) = D(X_1) + D(X_2) + \ldots + D(X_n) = npq.$$

Отже,

$$M(X)=np,\ D(X)=npq. \qquad (2.18)$$

Приклад 2.10. В партії 10 % нестандартних деталей. Навмання вибрали 4 деталі. Знайти закон розподілу дискретної випадкової величини X – кількості нестандартних деталей серед чотирьох відібраних та обчислити $M(X)$, $D(X)$, $\sigma(X)$.

Розв'язання. Випадкова величина X може набувати значення $x_1=0$, $x_2=1$, $x_3=2$, $x_4=3$, $x_5=4$. Очевидно, $n=4$, $p=0,1$; $q=1-p=0,9$. За формулою Бернуллі обчислимо ймовірності значень випадкової величини:

$$p_4(0)=q^4=0,9^4=0,6561;$$

$$p_4(1)=C_4^1pq^3=4\cdot 0,1\cdot 0,9^3=0,2916;$$

$$p_4(2)=C_4^2p^2q^2=\frac{4\cdot 3}{1\cdot 2}\cdot 0,1^2\cdot 0,9^2=6\cdot 0,01\cdot 0,81=0,0486;$$

$$p_4(3)=C_4^3p^3q=4\cdot 0,1^3\cdot 0,9=0,0036;$$

$$p_4(4)=p^4=0,1^4=0,0001.$$

Контроль обчислень: $0,6561+0,2916+0,0486+0,0036+0,0001=1$.

Ряд розподілу для випадкової величини X має такий вигляд:

X	0	1	2	3	4
P	0,6561	0,2916	0,0486	0,0036	0,0001

За формулами (2.18):

$$M(X)=np=4\cdot 0,1=0,4;$$

$D(X)=npq=4\cdot 0,1\cdot 0,9=0,36$, відповідно, $\sigma(X)=\sqrt{D(X)}=\sqrt{0,36}=0,6$.

Зауваження. Контроль проведених обчислень стосовно ряду розподілу зручно виконати в програмному середовищі із використанням функції **БИНОМРАСП** (рис. 2.13).

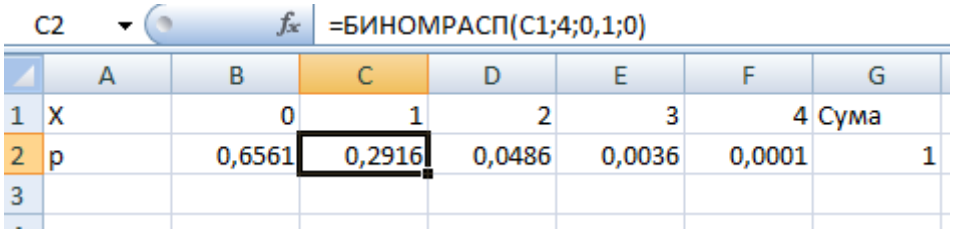


Рис. 2.13. Закон розподілу

2.2.1.2. Розподіл Пуассона

Якщо в схемі незалежних повторних випробувань n досить велике, а p або $(1-p)$ прямує до нуля, тоді біномний закон розподілу апроксимує **розподіл Пуассона**, параметр якого $\lambda = np$, причому $p \le 0{,}1$ або $p \ge 0{,}9$:

X	0	1	2	…	m	…
P	p_0	p_1	p_2	…	p_m	…

Відповідні ймовірності обчислюють за формулою Пуассона:

$$P(X=m)=\frac{\lambda^m}{m!}e^{-\lambda},\ \lambda>0\ \ (m=0,\ 1,\ 2,\ \ldots),$$

причому $\sum\limits_{m=0}^{\infty}\frac{\lambda^m}{m!}e^{-\lambda}=e^{-\lambda}\sum\limits_{m=0}^{\infty}\frac{\lambda^m}{m!}=e^{-\lambda}\cdot e^{\lambda}=1.$

Знаходимо математичне сподівання та дисперсію:

$$M(X)=\sum_{m=0}^{\infty}m\frac{\lambda^m e^{-\lambda}}{m!}=\sum_{m=1}^{\infty}m\frac{\lambda^m e^{-\lambda}}{m(m-1)!}=\lambda e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!}=\lambda e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}=\left|\begin{matrix}\text{заміна}\\ n=m-1\end{matrix}\right|=$$

$$=\lambda e^{-\lambda}\cdot e^{\lambda}=\lambda,\ \text{тобто}\ M(X)=\lambda.$$

$$M(X^2)=\sum_{m=0}^{\infty}m^2\frac{\lambda^m e^{-\lambda}}{m!}=\sum_{m=1}^{\infty}m\frac{m\lambda^m e^{-\lambda}}{m(m-1)!}=\lambda\sum_{m=1}^{\infty}m\frac{\lambda^{m-1}e^{-\lambda}}{(m-1)!}=$$

$$=\lambda\sum_{m=1}^{\infty}\big((m-1)+1\big)\frac{\lambda^{m-1}e^{-\lambda}}{(m-1)!}=\lambda\left[\sum_{m=1}^{\infty}(m-1)\frac{\lambda^{m-1}e^{-\lambda}}{(m-1)!}+\sum_{m=1}^{\infty}\frac{\lambda^{m-1}e^{-\lambda}}{(m-1)!}\right]=\left|\begin{matrix}\text{заміна}\\ n=m-1\end{matrix}\right|=$$

$$=\lambda\left[\sum_{n=0}^{\infty}n\frac{\lambda^n e^{-\lambda}}{n!}+\sum_{n=0}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}\right]=\lambda(\lambda+e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!})=\lambda(\lambda+e^{-\lambda}e^{\lambda})=\lambda(\lambda+1)=\lambda^2+\lambda,$$

$$D(X)=\lambda^2+\lambda-\lambda^2=\lambda.$$

Отже,

$$M(X)=\lambda,\ D(X)=\lambda,\ \sigma(X)=\sqrt{\lambda}. \qquad (2.19)$$

Розподіл Пуассона використовують у задачах статистичного контролю якості, в теорії надійності, теорії масового обслуговування, для прогнозування кількості вимог на виплату страхових компенсацій за рік, кількості дефектів в однакових виробах.

Приклад 2.11. На складі знаходиться 15000 підручників для школярів. Імовірність неправильного брошурування підручника дорівнює $p=0,00035$. Записати закон розподілу випадкової величини X – кількості бракованих підручників та обчислити його основні характеристики.

Розв'язання. За допомогою функції **ПУАССОН** побудовано ряд розподілу (можливе також значення випадкової величини і 13 з досить маленькою ймовірністю). Показано (рис. 2.14) також обчислення математичного сподівання.

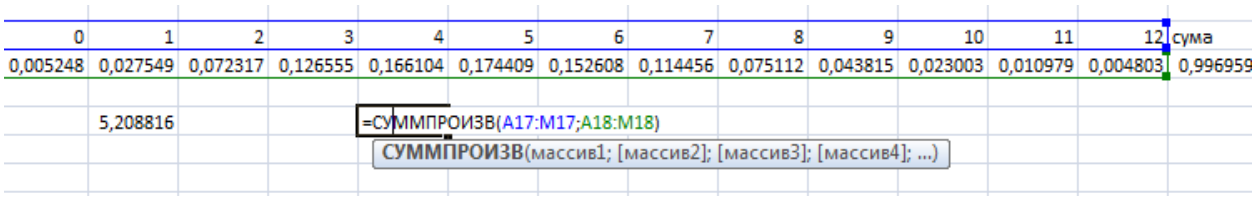


Рис.2.14. Вигляд обчислень на екрані

За формулами 2.19 маємо

$$M(X)=\lambda=0,00035\cdot 15000=5,25;\quad D(X)=5,25.$$

2.2.1.3. Геометричний розподіл

Проводять незалежні випробування, в кожному з яких може відбутись подія A з імовірністю p $(0<p<1)$. Дискретна випадкова величина X – кількість проведених випробувань до першого настання події A.

Ряд імовірностей цього розподілу – нескінченно спадна геометрична прогресія зі знаменником $q=1-p$, сума всіх членів якої дорівнює одиниці.

X	1	2	3	…	n	…
P	p	pq	pq^2	…	pq^{n-1}	…

Для даного розподілу

$$M(X)=\frac{1}{p},\ D(X)=\frac{q}{p^2}.$$

Обчислимо математичне сподівання:

$$M(X)=1p+2pq+3pq^2+\ldots+npq^{n-1}+\ldots+=p(1+2q+3q^2+\ldots+nq^{n-1}+\ldots)=$$

$$=p(q+q^2+q^3+\ldots+q^n+\ldots)'=p\left(\frac{q}{1-q}\right)'=p\frac{1}{(1-q)^2}=\frac{p}{p^2}=\frac{1}{p}.$$

Отже, $M(X)=\frac{1}{p}$.

Для $D(X)=M(X^2)-(M(X))^2=M(X^2)-\frac{1}{p^2}$.

X	0	1	2	…	n	…
P	p	pq	pq^2	…	pq^n	…

$$M(X^2)=1p+4pq+9pq^2+\ldots+n^2pq^{n-1}+\ldots=p(1+4q+9q^2+\ldots+n^2q^{n-1}+\ldots).$$

Для обчислення суми в дужках розглянемо суму

$$1+2q+3q^2+\ldots+nq^{n-1}+\ldots=\frac{1}{(1-q)^2},$$

звідси

$$q(1+2q+3q^2+\ldots+nq^{n-1}+\ldots)=\frac{q}{(1-q)^2}$$

або

$$q+2q^2+3q^3+\ldots+nq^n+\ldots=\frac{q}{(1-q)^2}.$$

Диференціюючи ліву та праву частини цієї рівності, отримуємо:

$$1+4q+9q^2+\ldots+n^2q^{n-1}+\ldots=\frac{1+q}{(1-q)^3}.$$

Отже, $M(X^2)=p\dfrac{1+q}{(1-q)^3}=\dfrac{1+q}{p^2}$,

$$D(X)=M(X^2)-\frac{1}{p^2}=\frac{1+q}{p^2}-\frac{1}{p^2}=\frac{q}{p^2}.$$

Геометричний розподіл застосовують у різноманітних задачах статистичного контролю якості виробів, у теорії надійності та страхових розрахунках.

Зауваження. Можливий також випадок, коли перше значення випадкової величини – 0. Тоді ряд розподілу має вигляд

X	0	1	2	…	n	…
P	p	pq	pq^2	…	pq^n	…

Можна довести, що в даному випадку $M(X)=\dfrac{q}{p}$, $D(X)=\dfrac{q}{p^2}$.

Наприклад, розглянемо таку задачу.

Задача. В урні міститься 5 білих і 7 чорних куль. Випробування полягає у вийманні по одній кулі з поверненням. Розглянемо дві випадкові величини: X – кількість вийнятих чорних куль до першого виймання білої кулі; Y – кількість виймань, які закінчуються з першим вийманням білої кулі. Скласти ряди розподілу для цих випадкових величин. Обчислити математичне сподівання та дисперсію.

Розв'язання. За умовою задачі та класичним означенням імовірності $p=\dfrac{5}{12}$ – ймовірність виймання білої кулі в разі одного виймання, тоді $q=1-p=\dfrac{7}{12}$.

Випадкова величина X набуде значення 0, якщо першою вийнята біла куля, тому $P(X=0)=p=\frac{5}{12}$. Також $X=1$, якщо першою вийнята чорна куля, а другою – біла. Відповідно,

$$P(X=1)=pq=\frac{5}{12}\cdot\frac{7}{12}=\frac{35}{144}.$$

Аналогічно, для $X=2$ обчислюємо

$$P(X=2)=pq^2=\frac{5}{12}\cdot\left(\frac{7}{12}\right)^2=\frac{245}{1728}$$

і так далі. Маємо ряд розподілу для випадкової величини X:

X	0	1	2	…	n	…
P	p	pq	pq^2	…	pq^n	…
	$\frac{5}{12}$	$\frac{5}{12}\cdot\frac{7}{12}$	$\frac{5}{12}\cdot\left(\frac{7}{12}\right)^2$	…	$\frac{5}{12}\cdot\left(\frac{7}{12}\right)^n$	…

За формулами обчислюємо числові характеристики:

$$M(X)=\frac{1}{p}=\frac{12}{5},\quad D(X)=\frac{q}{p^2}=\frac{84}{25}.$$

Випадкова величина Y набуває найменшого значення 1, якщо першою вийняли білу кулю, оскільки виймання припиняється. Тоді $P(Y=1)=p=\frac{5}{12}$.

Ряд розподілу для випадкової величини Y буде таким:

Y	1	2	3	…	n	…
P	p	pq	pq^2	…	pq^{n-1}	…
	$\frac{5}{12}$	$\frac{5}{12}\cdot\frac{7}{12}$	$\frac{5}{12}\cdot\left(\frac{7}{12}\right)^2$	…	$\frac{5}{12}\cdot\left(\frac{7}{12}\right)^{n-1}$	…

Тоді за відповідними формулами

$$M(Y)=\frac{q}{p}=\frac{7}{5},\ D(Y)=\frac{q}{p^2}=\frac{35}{12}.$$

Зауваження. Геометричний розподіл характеризує кількість *m* випробувань (проведених за схемою Бернуллі з імовірністю *p* настання події в кожному випробуванні) до першого настання події; ***розподіл Паскаля*** − до *k*-го настання випадку. Геометричний розподіл для $k=1$ є окремим випадком розподілу Паскаля, для якого

$$P(X=m)=C_{m-1}^{k-1}p^k q^{m-k}, m=k,k+1,...$$

Числові характеристики розподілу Паскаля:

$$M(X)=\frac{k}{p},\ D(X)=\frac{k\cdot q}{p^2}.$$

На відміну від закону Паскаля, не настання події за біномним розподілом характеризує розподіл кількості не появи подій до *k*-го випадку настання події. Його функція ймовірності

$$P(X=m)=C_{m+k-1}^{m}p^k q^m, m=0,1,....$$

та числові характеристики: $M(X)=\frac{kq}{p},\ D(X)=\frac{k\cdot q}{p^2}.$

Приклад 2.12. Імовірність влучення у ціль дорівнює 0,05. Проводиться стрільба в ціль до третього влучення. Потрібно:

1. Скласти закон розподілу кількості зроблених пострілів.

2. Знайти математичне сподівання і дисперсію цієї випадкової величини.

Розв'язання. Випадкова величина *X* – кількість пострілів до 3 влучення, $p=0,05,\quad k=3,\quad q=0,95.$

1. Побудуємо таблицю розподілу, обчисливши відповідні ймовірності для значень випадкової величини.

$P(X=3)=C_{3-1}^{2}\cdot 0,05^3\cdot 0,95^{3-3}=0,00125;$

$P(X=4)=C_{4-1}^{2}\cdot 0,05\cdot 0,95^{4-3}=3\cdot 0,000125\cdot 0,95=0,00035625;$

$P(X=5)=C_{5-1}^{2}\cdot 0,05^3\cdot 0,95^{5-3}=6\cdot 0,000125\cdot 0,9025=0,000676875;$

$$P(X=m)=C_{m-1}^{2}\cdot 0{,}05^{3}\cdot 0{,}95^{m-3}.$$

x_i	3	4	5	…	m	…
p_i	0,00125	0,00035625	0,000676875	…	$C_{m-1}^{2}\cdot 0{,}05^{3}\cdot 0{,}95^{m-3}$	…

2. Числові характеристики:

$$M(X)=\frac{k}{p}=\frac{3}{0{,}05}=60;\ D(X)=\frac{k\cdot q}{p^{2}}=\frac{3\cdot 0{,}95}{0{,}05^{2}}=1140.$$

2.2.1.4. Гіпергеометричний розподіл

Такий розподіл має вигляд:

$$P(X=m)=\frac{C_{k}^{m}\cdot C_{N-k}^{n-m}}{C_{N}^{n}},\ m=0,\ 1,\ 2,\ \ldots,\ n,\ k\geq n.$$

Він визначає ймовірність отримання m елементів з певною властивістю серед n елементів, взятих із сукупності N елементів, яка містить k елементів саме з такою властивістю.

Цей розподіл використовують у багатьох задачах статистичного контролю якості.

Зауваження. Якщо обсяг вибірки n малий порівняно з обсягом N сукупності, тобто $\frac{n}{N}\leq 0{,}1;\ \frac{n}{k}\leq 0{,}1;\ \frac{n}{N-k}\leq 0{,}1,$ то ймовірності у гіпергеометричному розподілі будуть близькими до відповідних імовірностей біномного розподілу з $p=\frac{k}{N}$.

У статистиці це означає, що розрахунки ймовірностей для вибірки без повторення мало відрізнятимуться від розрахунків імовірностей для повторної вибірки.

Для гіпергеометричного розподілу

$$M(X)=\frac{kn}{N},\ D(X)=\frac{nk(N-k)}{N^{2}}\cdot\frac{(N-n)}{(N-1)}.\qquad(2.20)$$

Приклад 2.13. Серед дев'яти однотипних виробів п'ять відповідають стандарту, а решта − ні. Навмання береться s виробів. Визначити закон розподілу цілочислової випадкової величини X – кількості появи числа виробів, що відповідають стандарту і обчислити для цієї величини числові характеристики, якщо їх максимальна кількість $s = 4$.

Розв'язання. У даному випадку випадкова величина X задовольняє гіпергеометричному закону розподілу при $N = 9,\ M = 5,\ N - M = 4$ та s, що може приймати різні значення від 0 до 4.

Використовують формулу обчислення відповідної ймовірності, яка матиме такий вигляд:

$$p_k = P(X = k) = \frac{C_5^k C_4^{s-k}}{C_9^s}.\ (s = 4,\ p_k = \frac{C_5^k C_4^{4-k}}{C_9^s},\ \text{де}$$

$$k = s - (N - M), \ldots, s = 0,1,2,3,4).$$

За функцією пакету Excel **ГИПЕРГЕОМЕТ**

(Число_успехов_в_выборке; Размер_выборки; Число_успехов в_совокупности; Размер_совокупности), де *Число_успехов_в_выборке*: k − змінна величина, яка приймає значення: 0, 1, 2, 3, 4; *Размер_выборки:* s − кількість навмання взятих виробів; *Число_успехов_в_совокупности*: M − кількість виробів, які сприяють події; *Размер_совокупности*: N −загальна кількість виробів) обчислення спрощуються (рис. 2.15).

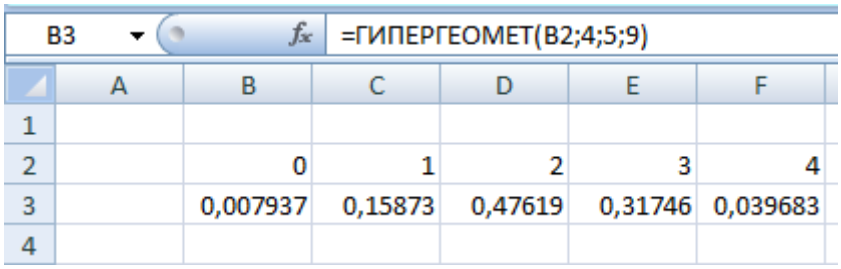


Рис. 2.15. Вигляд обчислень на екрані

В табл. 2.4 наведені гіпергеометричний закон та числові характеристики закону, отримані в Excel.

Табл. 2.4

X	0	1	2	3	4	Сума
p_k	0,007937	0,15873	0,47619	0,31746	0,039683	1
$k \cdot p_k$	0	0,15873	0,952381	0,952381	0,15873	$M(X)$ =2,2222
$(k - M(X))^2 \cdot p_k$	0,039193	0,237115	0,023516	0,192044	0,125416	$D(X)$ =0,6172
						$\sigma(X)$ =0,7856

Аналогічні результати можна одержати, використовуючи формули (2.20):

$$M(X) = \frac{Ms}{N} = \frac{5 \cdot 4}{9} = 2,2222;$$

$$D(X) = \frac{Ms}{N}\left(1 - \frac{M}{N}\right)\left(1 - \frac{s-1}{N-1}\right) = \frac{5 \cdot 4}{9}\left(1 - \frac{5}{9}\right)\left(1 - \frac{4-1}{9-1}\right) = 0,6173;$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0,61728} = 0,7857.$$

Приклад 2.14. Партія містить 200 виробів, серед яких 25 бракованих. Для перевірки якості з партії відібрали 10 виробів. Якщо кількість бракованих виробів не перевищує одиниці, то партію приймають. Знайти ймовірність того, що партію буде прийнято. Визначити вказану ймовірність, якщо апроксимувати гіпергеометричний розподіл біномним розподілом і законом розподілу Пуассона.

Розв'язання. Застосуємо формулу гіпергеометричного закону розподілу. Партію буде прийнято, якщо кількість бракованих серед відібраних десяти дорівнюватиме нулю або одиниці.

$$P(X \le 1) = P(X = 0) + P(X = 1) = \frac{C_{175}^{10}}{C_{200}^{10}} + \frac{C_{25}^{1} \cdot C_{175}^{9}}{C_{200}^{10}} \approx 0,638.$$

Обчислимо ймовірність за допомогою формули біномного закону розподілу,

$$p = \frac{25}{200} = \frac{1}{8};$$

$$P(X \le 1) = P(X = 0) + P(X = 1) = \left(\frac{7}{8}\right)^{10} + 10 \cdot \frac{1}{8} \cdot \left(\frac{7}{9}\right)^{9} \approx 0,639.$$

Обчислимо ту ж ймовірність за допомогою закону розподілу Пуассона:

$$\lambda = np = 10 \cdot \frac{1}{8} = 1,25.$$

$$P(X \le 1) = P(X = 0) + P(X = 1) = e^{-1,25} + 1,25e^{-1,25} \approx 0,644.$$

На рис. 2.16 показано обчислення за описаними законами, для зручності для біномного закону та закону Пуассона використано інтегральну функцію 1.

B9 =ПУАССОН(1;1,25;1)

	A	B	C	D
1		Гіпергеометричний закон		
2		0	1	
3		0,254465329	0,383230917	0,637696
4				
5		Біномний закон		
6		0,638897828		
7				
8		Закон Пуассона		
9		0,644635793		

Рис. 2.16. Обчислення за різними законами

Як бачимо, похибки обчислення в разі апроксимації гіпергеометричного розподілу порівняно невеликі.

2.2.1.5. Поліномний розподіл

Цей розподіл має вигляд

$$P_n(X_1 = m_1;\ X_2 = m_2;\ \ldots,\ X_s = m_s) = \frac{n!}{m_1! \cdot m_2! \cdot \ldots \cdot m_s!} p_1^{m_1} \cdot p_2^{m_2} \cdot \ldots \cdot p_s^{m_s}.$$

Його застосовують тоді, коли внаслідок кожного із здійснених повторних випробувань може з'явитися s різних подій A_i з імовірностями p_i, причому $\sum_{i=1}^{s} p_i = 1$. Його числові характеристики такі:

$$M(X_i) = np_i,\ D(X_i) = np_iq_i,\ q_i = 1 - p_i, i = 1,\ \ldots,\ s.$$

Отже, можна зробити висновоек, що поліномний розподіл є узагальненням біномного.

2.2.1.6. Дискретний рівномірний розподіл

Дискретна випадкова величина рівномірно розподілена, якщо вона набуває n значень з однаковими ймовірностями. Згідно з умовою нормування

$$\sum_{i=1}^{n} p_i = 1, \;\; p_i = \frac{1}{n}, \;\; i = 1, 2, ..., n.$$

Прикладом рівномірно розподіленої випадкової величини є кількість очок, що випадає за одного підкидання грального кубика. Її закон розподілу має вигляд

x	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Легко показати, що для рівномірно розподіленої випадкової величини

$$M(X) = \frac{n+1}{2}, \; D(X) = \frac{n^2 - 1}{12}. \qquad (2.21)$$

Подамо у вигляді таблиці залежності основних числових характеристик дискретних розподілів від параметрів цих розподілів.

№	Розподіл	$M(X)$	$D(X)$
1	Біномний	np	npq
2	Пуассонівський	λ	λ
3	Геометричний: *а*) $P(X = n) = pq^n, \; n = 1, 2, ...$ *б*) $P(X = n) = pq^n, \; n = 0, 1, 2, ...$	$\frac{1}{p}$ $\frac{q}{p}$	$\frac{q}{p^2}$ $\frac{q}{p^2}$
4	Гіпергеометричний	$\frac{kn}{N}$	$\frac{nk(N-k)}{N^2} \cdot \frac{(N-n)}{(N-1)}$
5	Рівномірний	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$

Зауваження. В Excel для обчислення ймовірності окремого значення біномного розподілу або значення випадкової величини за заданою ймовірністю використовують функції **БИНОМРАСП** та **КРИТБИНОМ.**

Функція **КРИТБИНОМ** обчислює найменше значення кількості успішних випробувань випадкової величини, для якого інтегральний біномний розподіл більший від заданої величини (критерію) або дорівнює їй. Цю функцію використовують у задачах, пов'язаних з контролем якості.

2.2.2. Основні закони розподілу неперервних випадкових величин, їх основні числові характеристики

2.2.2.1. Рівномірний розподіл

Означення. Величина X ***розподілена рівномірно*** на проміжку $[a,\ b]$, якщо усі її можливі значення належать цьому проміжку та щільність розподілу ймовірностей має такий вигляд:

$$f(x)=\begin{cases}\dfrac{1}{b-a},\ x\in[a,\ b];\\ 0,\ x\notin[a,\ b].\end{cases} \tag{2.22}$$

Для рівномірно розподіленої випадкової величини на проміжку $[a,b]$:

$$P(x_1<X<x_2)=\frac{x_2-x_1}{b-a},\ a\le x_1\le x_2\le b,$$

тобто ймовірність потрапляння X в інтервал $(x_1,\ x_2)$ дорівнює відношенню довжини цього інтервалу до довжини всього проміжку $[a,\ b]$.

Функція розподілу рівномірного закону розподілу має такий вигляд:

$$F(x)=\begin{cases}0,\ x<a;\\ \dfrac{x-a}{b-a},\ x\in[a,\ b];\\ 1,\ x>b.\end{cases}$$

Рівномірний закон розподілу легко моделювати. За допомогою функціональних перетворень із величин, розподілених рівномірно, можна отримувати величини з довільним законом розподілу. Графіки щільності ймовірності та функції розподілу показано на рис. 2.17 та рис. 2.18. Цей розподіл задовольняють похибки округлення різноманітних розрахунків.

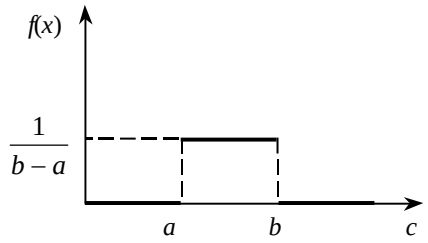


Рис. 2.17. Щільність розподілу ймовірності

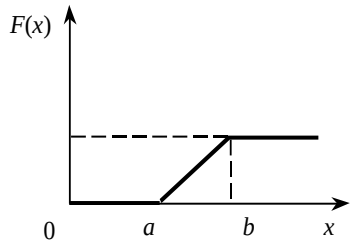


Рис. 2.18. Функція розподілу

Знайдемо числові характеристики цього розподілу:

$$M(X)=\int_a^b xf(x)dx=\int_a^b \frac{x}{b-a}dx=\frac{1}{b-a}\frac{x^2}{2}\Big|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{(b+a)(b-a)}{2(b-a)}=\frac{a+b}{2}.$$

Отже, $M(X)=\dfrac{a+b}{2}$, тобто $M(X)$ – середина відрізка $[a,b]$.

Для дисперсії $D(X)=\int_a^b x^2 f(x)dx-(M(X))^2=\int_a^b \frac{x^2}{b-a}dx-\left(\frac{a+b}{2}\right)^2=$

$$=\frac{1}{b-a}\frac{x^3}{3}\Big|_a^b-\frac{a^2+2ab+b^2}{4}=\frac{b^3-a^3}{3(b-a)}-\frac{a^2+2ab+b^2}{4}=$$

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

Таким чином,

$$M(X)=\frac{a+b}{2},\ D(X)=\frac{(b-a)^2}{12}. \qquad (2.23)$$

Приклад 2.15. Рівномірно розподілена випадкова величина X задана щільністю розподілу $f(x)=\begin{cases}\dfrac{1}{2l},\ x\in(a-l,\ a+l);\\ 0,\ x\notin(a-l,\ a+l).\end{cases}$

Знайти $M(X),\ D(X)$.

Розв'язання. За формулами (2.23)

$$M(X)=\frac{a-l+a+l}{2}=a,\ D(X)=\frac{(a+l-a+l)^2}{12}=\frac{4l^2}{12}=\frac{l^2}{3}.$$

2.2.2.2. Показниковий розподіл

Означення. Випадкову величина X називають розподіленою за ***показниковим розподілом***, якщо її щільність розподілу ймовірностей має такий вигляд:

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & x\le 0,\end{cases}$$

де $\lambda>0$ – параметр розподілу.

Показниковий розподіл задовольняють: час телефонної розмови, час ремонту техніки, час безвідмовної роботи комп'ютерної мережі.

Якщо випадкова величина X має показниковий розподіл з параметром λ, тоді

$$M(X)=\frac{1}{\lambda},\ D(X)=\frac{1}{\lambda^2}.$$

Дійсно, $M(X)=\lambda\int\limits_0^{+\infty}xe^{-\lambda x}dx=\lambda\lim\limits_{b\to\infty}\int\limits_0^{b}xe^{-\lambda x}dx=\left|\begin{matrix}u=x & du=dx\\ dv=e^{-\lambda x}dx & v=-\frac{1}{\lambda}e^{-\lambda x}\end{matrix}\right|=$

$$=\lambda\lim_{b\to\infty}\left(-\frac{1}{\lambda}xe^{-\lambda x}\Big|_0^b+\frac{1}{\lambda}\int_0^b e^{-\lambda x}dx\right)=-\lim_{b\to\infty}be^{-\lambda b}-\lim_{b\to\infty}\frac{1}{\lambda}e^{-\lambda x}\Big|_0^b=$$

$$=-\lim_{b\to\infty}\frac{1}{\lambda e^{\lambda b}}+\frac{1}{\lambda}=\frac{1}{\lambda}.$$

Для визначення дисперсії обчислимо

$$M(X^2)=\lambda\int\limits_0^{+\infty}x^2e^{-\lambda x}dx=\lambda\lim\limits_{b\to\infty}\int\limits_0^{b}x^2e^{-\lambda x}dx=$$

$$=\left|\begin{matrix}u=x^2 & du=2xdx\\ dv=e^{-\lambda x}dx & v=-\frac{1}{\lambda}e^{-\lambda x}\end{matrix}\right|=\lambda\lim_{b\to\infty}\left(-\frac{1}{\lambda}x^2e^{-\lambda x}\Big|_0^b+\frac{2}{\lambda}\int_0^b xe^{-\lambda x}dx\right)=$$

$$=-\lim_{b\to\infty}\frac{b^2}{e^{\lambda b}}+\frac{2}{\lambda}\cdot\frac{1}{\lambda}=\frac{2}{\lambda^2}.$$

$$D(X)=M(X)-(M(X))^2=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2},$$

тобто $D(X)=\frac{1}{\lambda^2}$.

Отже, для показникового розподілу

$$M(X)=\frac{1}{\lambda},\ D(X)=\frac{1}{\lambda^2}. \tag{2.24}$$

Зауваження. Якщо випадкова величина X розподілена за показниковим розподілом, то її інтегральна функція розподілу має такий вигляд:

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

відповідно,

$$P\left(a<X<b\right)=\begin{cases}e^{-a\lambda}-e^{-b\lambda},\ a\geq 0;\\ 1-e^{-b\lambda},\ a<0,\ b>0;\\ 0,\ b<0.\end{cases}$$

Графіки диференціальної та інтегральної функцій показникового розподілу зображено на рис. 2.19 *а*, *б*.

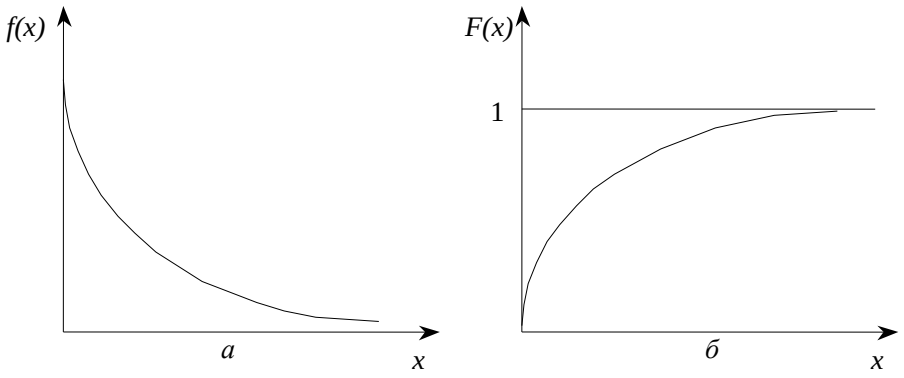


Рис. 2.19. Диференціальна та інтегральна функції показникового розподілу:

а – $f(x)$; *б* – $F(x)$

Показниковий розподіл часто трапляється в теорії масового обслуговування, теорії надійності. Нехай t – час безвідмовної роботи деякого елемента, λ – інтенсивність відмов (середня кількість відмов за одиницю часу).

Тоді час роботи елемента *t* можна вважати неперервною випадковою величиною, розподіленою за показниковим законом із функцією розподілу

$$F(x) = p(t < x) = 1 - e^{-\lambda x} \quad (\lambda > 0),$$

яка визначає ймовірність відмови елемента за час, менший від *x*.

Імовірність

$$R(x) = p(t \geq x) = e^{-\lambda x}$$

називають ***функцією надійності***, яка визначає ймовірність безвідмовної роботи елемента за час *x*.

Приклад 2.16. Час *t* роботи дитячого іграшкового телефона має показниковий розподіл. Знайти ймовірність його безвідмовної роботи протягом 30 годин, якщо середній час роботи без підзарядки 20 годин.

Розв'язання. За умовою задачі математичне сподівання величини *t* дорівнює 20, тому $\lambda = \frac{1}{m} = \frac{1}{20}$. За формулою $R(x) = p(t \geq x) = e^{-\lambda x}$, маємо:

$$R(x = 30) = e^{-\frac{30}{20}} = e^{-1,5} = 0,2231.$$

Отже, ймовірність того, що іграшковий телефон працюватиме не менше 20 годин, дорівнює 22 %.

Приклад 2.17. Знайти числові характеристики випадкової величини, розподіленої за законом

$$f(x) = \begin{cases} 4e^{-4x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Розв'язання. Для показникового розподілу випадкової величини з параметром $\lambda = 4$ (2.24) маємо:

$$M(X) = \frac{1}{\lambda} = \frac{1}{4} = 0,25; \; D(X) = \frac{1}{\lambda^2} = \frac{1}{16} = 0,0625.$$

Приклад 2.18. Величина *X* розподілена за показниковим розподілом

$$f(x) = \begin{cases} 3e^{-3x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Знайти ймовірність того, що X потрапить в інтервал $(0,4;1)$.

Розв'язання. Параметр розподілу $\lambda = 3,$ тому

$$P(0,4 < X < 1) = e^{-0,4\cdot 3} - e^{-1\cdot 3} = e^{-1,2} - e^{-3} = 0,302 - 0,050 = 0,252.$$

Приклад 2.19. Час безвідмовної роботи в годинах деякого приладу є неперервна випадкова величина X, яка задана щільністю розподілу

$$f(x) = 0,08 \cdot e^{-0,08t}, t \geq 0.$$

Знайти наступні характеристики:

1. Середній час безвідмовної роботи приладу.

2. Імовірність безвідмовної роботи приладу протягом 4 год.

3. Імовірність відмови приладу в інтервалі часу від 4 до 20 год.

4. Надійність (ймовірність безвідмовної роботи) системи протягом 4 год., якщо вона складається з п'яти послідовно підключених однакових приладів. Вважати відмови приладів незалежними.

Розв'язання.

1. Середній час безвідмовної роботи приладу є математичним сподіванням випадкової величини X – часу безвідмовної роботи. З вигляду щільності розподілу робимо висновок, що випадкова величина X розподілена за показниковим законом з параметром $\lambda = 0,08$. Тому

$$M(X) = \frac{1}{\lambda} = 12,5 \text{ год.}$$

2. Імовірність безвідмовної роботи приладу протягом часу t знайдемо за формулою $P\{X \geq t\} = 1 - F(t) = e^{-\lambda t}$.

За умовою задачі $\lambda = 0,08; \quad t = 4$ год, тому

$$P\{X \geq 4\} = e^{-0,32} \approx 0,726.$$

3. Спочатку знайдемо ймовірність безвідмовної роботи приладу в інтервалі часу від 4 до 20 год:

$$P\{4 \leq X \leq 20\} = F(20) - F(4) = e^{-0,32} - e^{-1,6} \approx 0,524.$$

Тоді ймовірність відмови приладу в цьому інтервалі дорівнює

$$p = 1 - P\{4 \le X \le 20\} = 1 - 0{,}524 = 0{,}476.$$

4. Надійність системи послідовно підключених приладів за теоремою множення ймовірностей дорівнює:

$$P_{системи}(t) = \prod_{i=1}^{n} P_i(t) = \prod_{i=1}^{n} e^{-\lambda_i t} = e^{-t\sum_{i=1}^{n}\lambda_i},$$

де λ_i – параметри показникового закону розподілу безвідмовної роботи для кожного приладу. За умовою задачі прилади однакові, тому $\lambda_1 = \lambda_2 = \ldots = \lambda_5 = 0{,}08$, тому $\sum_{i=1}^{5}\lambda_i = 5 \cdot 0{,}08 = 0{,}4.$ Отже, надійність роботи системи протягом чотирьох годин дорівнює

$$P_{системи}(4) = e^{-4 \cdot 0{,}4} = e^{-1{,}6} \approx 0{,}202.$$

Відповідь. 1) 12,5 год.; 2) $p \approx 0{,}726$; 3) 0,476; 4) 0,202.

Завдання для самоконтролю

1. Біномний закон розподілу, доведення формул обчислення його основних числових характеристик.

2. Закон Пуассона, його числові характеристики. Практичне застосування закону Пуассона.

3. Геометричний закон розподілу, можливі варіанти прийняття перших його числових значень, навести приклад.

4. Навести приклад побудови гіпергеометричного закону розподілу, його апроксимації біномним законом та законом Пуассона.

5. Вивести формули обчислення числових характеристик рівномірно розподіленої неперервної випадкової величини.

6. Записати основні формули для показникового розподілу, отримати для нього інтегральну функцію розподілу.

7. Пояснити використання функції надійності на практиці.

8. Корпорація вкладає гроші у купівлю земельної ділянки вартістю 100 000 ум. о., сподіваючись, що через два роки ціна на земельну ділянку зросте на $q\%$. Є чотири прогнози зміни $q\%$ через два роки

$q\%$	50%	45%	40%	30%
Імовірність	0,2	0,3	0,3	0,2

Побудувати ряд розподілу випадкової величини X – вартості земельної ділянки через два роки та знайти її числові характеристики.

9. У рекламних цілях торгова фірма вкладає в кожну десяту одиницю товару приз вартістю 100 грн. Побудувати ряд розподілу випадкової величини X – розміру виграшу при 5 покупках та знайти її числові характеристики.

10. Скласти закон розподілу випадкової величини X – кількості пакетів трьох акцій, за якими власником буде отримано прибуток, якщо ймовірність отримання прибутку за кожним пакетом дорівнює 0,5; 0,6 та 0,7. Обчислити числові характеристики випадкової величини X.

11. В білеті є 4 задачі. Ймовірність правильного розв'язання задач дорівнює 0,7. Побудувати ряд розподілу випадкової величини X – кількості правильно розв'язаних задач в білеті та знайти її числові характеристики.

12. Торговий агент має 5 телефонних номерів потенційних покупців і телефонує доти, доки не отримає замовлення на покупку товару. Ймовірність того, що потенційний покупець зробить замовлення дорівнює 0,4. Побудувати ряд розподілу випадкової величини X – кількості телефонних розмов, які проведе торговий агент та знайти її числові характеристики.

13. Ціна поділки шкали вимірного приладу дорівнює 2. Припускаючи, що похибка при округленні розподілена рівномірно, знайти числові характеристики похибки округлення та ймовірність того, що похибка округлення не перевищує 0,4.

14. Випадкова величина X задає час безвідмовної роботи системи і має розподіл $f(x) = Ae^{-0,1t}, t \geq 0$. Знайти коефіцієнт A та надійність (імовірність безвідмовної роботи системи) протягом часу τ. Яка ймовірність того, що час безвідмовної роботи системи буде меншим від математичного сподівання?

15. Час відновлення каналу зв'язку має показниковий розподіл. Середній час відновлення дорівнює 10 хв. Яка ймовірність того, що час відновлення буде знаходитись в межах від 5 до 25 хв?

16. Час прийому та обробки одного повідомлення є випадкова величина X, яка розподілена за показниковим законом. В середньому за хвилину приймається 6 повідомлень. Яка ймовірність того, що повідомлення буде прийнято та оброблено протягом 4 хв.?

17. Середнє число замовлень таксі, які поступають в диспетчерський пункт за одну хвилину, дорівнює 3. Знайти ймовірність того, що за 2 хвилини поступить: а) чотири замовлення; б) менше чотирьох замовлень; в) не менше двох замовлень.

Варіанти завдань для індивідуальної роботи

Завдання 1.

Проводять k незалежних випробувань, в яких імовірність успіху для кожного з них дорівнює p. Скласти ряд розподілу випадкової величини X – кількості успішних випробувань та обчислити основні числові характеристики.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
p	0.3	0.55	0.3	0.4	0.65	0.7	0.45	0.3	0.8	0.6	0.4	0.3	0.35	0.3	0.5
k	4	5	3	4	6	4	3	5	4	3	4	5	4	5	6

Завдання 2. Отримано партію k підручників з математики для однієї шкіл м. Києва. Ймовірність неправильного брошурування підручника дорівнює p.

Записати закон розподілу випадкової величини X – кількості бракованих підручників та обчислити його основні характеристики.

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
p	0.0002	0.00055	0.0003	0.0004	0.0006	0.00012	0.0005	0.00013	0.00025
k	10000	20000	40000	30000	25000	14000	30000	50000	13000

10	**11**	**12**	**13**	**14**	**15**
0.0006	0.00024	0.00035	0.00045	0.0008	0.000075
20000	40000	15000	45000	20000	60000

Вказівка. В завданнях 1, 2 бажано будувати ряд розподілу в програмному середовищі Excel, а обчислення числових характеристик за відомими формулами відповідного розподілу.

Завдання 3. Випадкова величина X розподілена рівномірно на інтервалі $(a;b)$. Записати диференційну $f(x)$ та інтегральну $F(x)$ функції, знайти $M(X)$, $D(X)$, $\sigma(X)$, якщо

1) $a=1, b=6$; 2) $a=2, b=8$; 3) $a=3, b=8$;

4) $a=3, b=6$; 5) $a=3, b=7$; 6) $a=4, b=10$;

7) $a=-1, b=4$; 8) $a=2, b=9$; 9) $a=2, b=7$;

10) $a=3, b=9$; 11) $a=3, b=10$; 12) $a=-2, b=4$;

13) $a=2, b=12$; 14) $a=12, b=16$; 15) $a=5, b=10$.

Завдання 4. Час t безперервної роботи електроприладу має показниковий розподіл. Знайти ймовірність його безвідмовної роботи протягом N годин, якщо середній час його роботи – T годин, записавши відповідну диференціальну функцію розподілу.

1) $N=200, \quad T=100$; 2) $N=250, \quad T=100$;

3) $N=300, \quad T=200$; 4) $N=300, \quad T=150$;

5) $N=300, \quad T=290$; 6) $N=200, \quad T=180$;

7) $N=400, \quad T=200$; 8) $N=400, \quad T=250$;

9) $N = 400, \quad T = 300;$

10) $N = 400, \quad T = 380;$

11) $N = 400, \quad T = 350;$

12) $N = 500, \quad T = 300;$

13) $N = 500, \quad T = 400;$

14) $N = 500, \quad T = 450;$

15) $N = 500, \quad T = 150.$

2.3. Нормальний закон розподілу та його значення у теорії ймовірностей. Граничні теореми теорії ймовірностей

У значній частині задач випадкова величина – це сума великої кількості малих доданків, що визначаються факторами, які діють незалежно один від одного. За таких обставин слід очікувати, що випадкова величина має нормальний розподіл або близький до нього.

2.3.1. Нормальний закон розподілу, його основні характеристики

Означення. Випадкову величину X називають **розподіленою нормально,** якщо її диференціальна функція розподілу ймовірностей має вигляд

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

де $a,\ \sigma$ – параметри розподілу.

Графік функції $f(x)$ називають ***нормальною кривою*** або ***кривою Гаусса.***

За $a=0$ та $\sigma=1$ нормальну криву називають *нормованою*, функція розподілу щільності має вигляд

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

тобто це функція Лапласа, табульована, зокрема в електронній формі.

Нормальний закон розподілу повністю визначається своїм математичним сподіванням та дисперсією (середнім квадратичним відхиленням):

$$M(X)=a,\ D(X)=\sigma^2,\ \sigma(X)=\sigma. \qquad (2.25)$$

Справді, $M(X)=\int\limits_{-\infty}^{+\infty}xf(x)dx=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}xe^{-\frac{(x-a)^2}{2\sigma^2}}dx=\left|\begin{array}{l}t=\frac{x-a}{\sqrt{2}\sigma},\\ x=\sigma\sqrt{2}t+a\\ dx=\sigma\sqrt{2}dt\end{array}\right|=$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma\sqrt{2}t + a) e^{-t^2} \sigma\sqrt{2} dt = \frac{\sigma^2 2}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} te^{-t^2} dt + \frac{\sigma\sqrt{2}a}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt =$$

$$= \frac{a}{\sqrt{\pi}} \sqrt{\pi} = a,$$

оскільки $\int_{-\infty}^{+\infty} te^{-t^2} dt = 0,$ як інтеграл від непарної функції з симетричними межами (за умови збіжності інтеграла $\int_{0}^{+\infty} te^{-t^2} dt$) і $\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi},$ як інтеграл Пуассона.

Обчислимо дисперсію

$$D(X) = \int_{-\infty}^{+\infty} \left(x - M(X)\right)^2 dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left| \begin{array}{l} t = \dfrac{x-a}{\sqrt{2}\sigma} \\ x = \sigma\sqrt{2}t + a \\ dx = \sigma\sqrt{2}dt \end{array} \right| =$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^2 e^{-t^2} dt = \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t \cdot 2te^{-t^2} dt = \left| \begin{array}{l} u = t \quad du = dt \\ dv = 2te^{-t^2} dt \quad v = -e^{-t^2} \end{array} \right| = -\frac{\sigma^2}{\sqrt{\pi}} t e^{-t^2} \Big|_{-\infty}^{+\infty} +$$

$$+ \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \frac{\sigma^2}{\sqrt{\pi}} \sqrt{\pi} = \sigma^2.$$

Отже, $M(X) = a,\ D(X) = \sigma^2.$

Для нормального розподілу ймовірність потрапляння випадкової величини в інтервал $(\alpha,\ \beta)$ обчислюють за формулою

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right), \tag{2.26}$$

де $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt$ – функція Лапласа.

Інтегральна функція розподілу має вигляд

$$F(x) = \int_{-\infty}^{x} f(t) dt = \Phi\left(\frac{x-a}{\sigma}\right) - \Phi(-\infty) = \Phi\left(\frac{x-a}{\sigma}\right) + \frac{1}{2},$$

тобто

$$F(x)=\frac{1}{2}+\Phi\left(\frac{x-a}{\sigma}\right). \qquad (2.27)$$

Графіки щільності розподілу ймовірностей та функції розподілу наведено відповідно на рис. 2.20 та рис. 2.21.

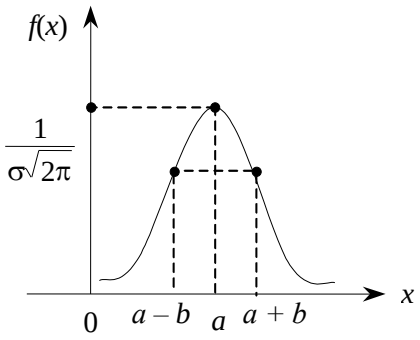


Рис. 2.20. Диференціальна функція нормального розподілу

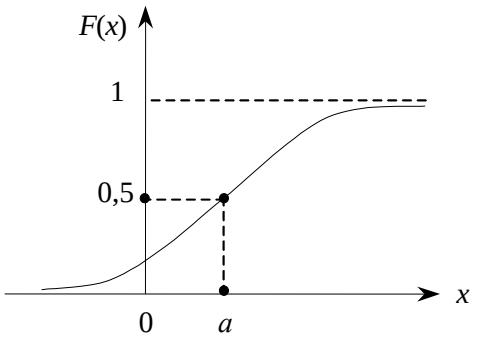


Рис. 2.21. Інтегральна функція нормального розподілу

Зауваження. Нормально розподілену випадкову величину із параметрами розподілу a, σ часто позначають $N(a,\ \sigma)$.

2.3.2. Правило трьох сигм

Із рівності (2.26) випливає:

$$P\left(|X-a|\le\delta\right)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Зокрема, $P\left(|X-a|\le t\sigma\right)=2\Phi(t),\ t>0$.

Підставляючи послідовно в останню нерівність $t=1,\ 2,\ 3,$ за допомогою таблиці функції Лапласа отримують:

$$P\left(|X-a|\le\sigma\right)=2\Phi(1)=0,68;$$

$$P\left(|X-a|\le 2\sigma\right)=2\Phi(2)=0,95;$$

$$P\left(|X-a|\le 3\sigma\right)=2\Phi(3)=0,997.$$

Число 0,997 мало відрізняється від одиниці, тому подію з імовірністю 0,997 можна вважати майже достовірною.

Рівність

$$P\left(\left|X-a\right|\le 3\sigma\right)=2\Phi(3)=0,997$$

називають ***«правилом*** 3σ***»***: *майже достовірно, що нормально розподілена випадкова величина може відхилятись від свого математичного сподівання не більше, ніж на потроєне середнє квадратичне відхилення.*

На практиці його використовують так: *якщо закон розподілу випадкової величини X невідомий, але* $\left|X-a\right|<3\sigma$, *тоді можна припустити, що X розподілена нормально.*

Нормальний закон розподілу широко застосовують у математичній статистиці. Головна особливість нормального закону полягає в тому, що він є граничним законом, до якого наближуються інші закони розподілу за типових умов.

Приклад 2.20. Для нормально розподіленої випадкової величини X знайти $P\left(\alpha<X<\beta\right)$, якщо $a=2,\ \sigma=2,\ \alpha=1,\ \beta=6.$

Розв'язання. За формулою $P\left(\alpha<X<\beta\right)=\Phi\left(\dfrac{\beta-a}{\sigma}\right)-\Phi\left(\dfrac{\alpha-a}{\sigma}\right)$

маємо

$$P(1<X<6)=\Phi\left(\frac{6-2}{2}\right)-\Phi\left(\frac{1-2}{2}\right)=\Phi(2)-\Phi(-0,5)=\Phi(2)+\Phi(0,5)=$$

$$=0,47725+0,19146=0,66871.$$

Приклад 2.21. Зріст студентів однієї із академічних груп розподілено за нормальним законом, причому $a=175$ см, а $\sigma=6$ см. Визначити ймовірність того, що принаймні один із 5 викликаних навмання студентів матиме зріст від 170 до 180 см.

Розв'язання. Позначимо подію A – із 5 викликаних студентів зріст принаймні одного належить проміжку (170, 180); тоді $\overline{A}$ – зріст усіх 5 викликаних студентів не належить проміжку.

Обчислення проведемо за формулою (2.26).

$$P(170 < X < 180) = \Phi\left(\frac{180-175}{6}\right) - \Phi\left(\frac{170-175}{6}\right) = \Phi\left(\frac{5}{6}\right) - \Phi\left(-\frac{5}{6}\right) =$$

$$= 2\Phi\left(\frac{5}{6}\right) = 2\Phi(0,833) = 2 \cdot 0,2967 = 0,5934.$$

Отже, ймовірність того, що зріст одного викликаного студента не належить проміжку (170, 180)

$$p = 1 - P\left(170 < X < 180\right) = 1 - 0,5934 = 0,4066.$$

За теоремою множення ймовірностей незалежних подій знаходимо ймовірність події $\overline{A}$:

$$P(\overline{A}) = (0,4066)^5 = 0,0111.$$

Отже, тоді ймовірність події A становить:

$$P(A) = 1 - P(\overline{A});$$

$$P(A) = 1 - 0,0111 = 0,9889.$$

Зауваження.

Для обчислень, пов'язаних із нормальним законом розподілу, засобами Excel використовують функцію **НОРМРАСП** (*x*; *среднее*; *стандартное_отклонение*; *интегральная*), де *x* – значення, для якого будується розподіл; *«среднее»* − середнє арифметичне розподілу (математичне сподівання); «*стандартное_отклонение*» − стандартне відхилення розподілу; *«интегральная»* − логічне значення, яке визначає форму функції. Якщо *«интегральная»* = ЛОЖЬ, то функція **НОРМРАСП** використовує функцію щільності ймовірності, а в разі *«интегральная»* = ИСТИНА – інтегральну функцію і значення функції з геометричного погляду, дає площу криволінійної фігури, обмеженої кривою щільності розподілу ймовірностей, вертикальною прямою, яка проходить через точку *x* (ліворуч від цієї прямої) та віссю абсцис. Функцію **НОРМСТРАСП** використовують для стандартного нормального розподілу, для якого математичне сподівання дорівнює 0 і стандартне відхилення $\sigma = 1$.

2.3.3. Розподіли χ^2 та Стьюдента

2.3.3.1. Розподіл χ^2

Нехай $X_i (i = 1,\ 2, \ldots, n)$ – нормально розподілені нормовані незалежні величини (їх математичне сподівання дорівнює 0, середні квадратичні відхилення дорівнюють 1). Тоді сума квадратів цих величин

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \tag{2.28}$$

розподілена ***за законом*** χ^2 з $k = n$ степенями вільності.

Якщо величини X_i зв'язані одним лінійним співвідношенням, наприклад

$$\sum_{i=1}^{n} X_i = n\overline{X},$$

то кількість степенів вільності буде $k = n - 1.$

Диференціальна функція розподілу χ^2 має вигляд

$$f(x) = \begin{cases} 0, \ x \le 0; \\ \dfrac{1}{2^{\frac{k}{2}} \cdot \Gamma\left(\dfrac{k}{2}\right)} \cdot e^{-\frac{x}{2}} \cdot x^{\frac{k}{2}-1}, \ x > 0, \end{cases}$$

де $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ – гамма-функція, $\Gamma(n+1) = n!.$

Розподіл χ^2 визначається параметром – кількість степенів вільності *k*. Але важливо, що розподіл χ^2 залежить лише від одного параметра – кількості степенів вільності *k* , що дорівнює кількості незалежних доданків у сумі (2.28). У разі наявності *l* зв'язків між величинами X_i, які не ведуть до порушення нормального закону розподілу, кількість незалежних величин зменшується, а отже, зменшується й кількість степенів вільності суми квадратів: вона дорівнює $k = n - l.$ Можна довести, що математичне сподівання і дисперсія випадкової величини χ^2 дорівнюють відповідно:

$$\mathrm{M}\left(\chi^2(k)\right) = k, \quad D\left(\chi^2(k)\right) = 2k.$$

Диференціальну функцію розподілу «хі-квадрат» інтегрувати складно. Тому складено таблиці, де вказано критичні точки розподілу випадкової величини. Критичні точки «хі-квадрат» розподілу – це такі граничні (найменші) значення $\chi^2_{k,\alpha}$, які випадкова величина, що має χ^2-розподіл із відомою кількістю степенів вільності *k*, перевищить з імовірністю α. Отже,

$$\alpha = P\left(\chi^2 > \chi^2_{k,\alpha}\right).$$

Імовірність α в цьому разі іноді називають ***рівнем значущості.***

Якщо $k \to \infty$, то розподіл «хі-квадрат» прямує до нормального розподілу з параметрами $a = k, \quad \sigma = \sqrt{2k}\, a.$ Якщо $k \geq 30$, то цей розподіл уже можна апроксимувати нормальним розподілом.

2.3.3.2. Розподіл Стьюдента

Розподіл Стьюдента було введено 1908 року англійським статистом В. Госсетом, який працював на фабриці з виробництва пива. Ймовірнісно-статистичні методи використовувалися для ухвалення економічних і технічних рішень на цій фабриці, тому її керівництво забороняло В. Госсету публікувати наукові статті під своїм ім'ям. Проте він мав можливість публікуватися під псевдонімом Стьюдент. Історія Госсета-Стьюдента показує, що вже сто років тому менеджерам Великобританії була очевидна велика економічна ефективність ймовірносно-статистичних методів.

Нині розподіл Стьюдента − один з найбільш відомих розподілів серед використовуваних під час аналізу реальних даних. Його застосовують для оцінювання математичного сподівання, прогнозного значення та інших характеристик за допомогою надійних інтервалів, під час перевірки гіпотез про значення математичних сподівань, коефіцієнтів регресійної залежності, гіпотез однорідності вибірок.

Якщо випадкова величина Y має стандартний розподіл $(N(0;\ 1))$, а випадкова величина X − розподіл χ^2 із k степенями вільності, то величина $Z_k = Y\sqrt{\dfrac{k}{\chi^2_k}}$

характеризується ***t-розподілом або розподілом Стьюдента*** з k степенями вільності та диференціальною функцією розподілу

$$f(z)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{z^2}{k}\right)^{-\frac{k+1}{2}},\qquad -\infty<z<\infty.$$

Розподіл Стьюдента, як і розподіл Пірсона, залежить від єдиного параметра – кількості степенів вільності k. Для розподілу випадкової величини t складено таблиці.

Математичне сподівання і дисперсія випадкової величини t дорівнюють відповідно

$$M(t_k)=0,\ D(t_k)=\frac{k}{k-2},\quad k>2.$$

У разі зростання k ***розподіл Стьюдента*** швидко наближується до нормального розподілу. Загалом крива щільності розподілу подібна за формою до кривої щільності нормального розподілу з математичним сподіванням 0 і дисперсією 1 з тією відмінністю, що у ***t-розподілу*** вона трохи нижча і ширша. За кількості степенів вільності, що прямує до нескінченості, ***t-розподіл*** прямує до нормального розподілу з математичним сподіванням 0 і дисперсією 1. Отже, Якщо $k\to\infty$, то розподіл Стьюдента прямує до стандартного нормального розподілу, і вже для $k\geq 25$, його замінюють цим розподілом.

2.3.4. Закон великих чисел та центральна гранична теорема

У теорії ймовірностей, особливо в разі її застосування на практиці, важливу роль відіграють події з імовірностями, близькими до нуля або до одиниці, тому одна з основних задач теорії ймовірностей – встановлення закономірностей, що можуть виникати з імовірністю, близькою до одиниці, й особливо таких закономірностей, які виникають в результаті спільної дії великої кількості незалежних випадкових факторів. Закон великих чисел і є одним з найважливіших тверджень такого типу.

Під законом великих чисел у теорії ймовірностей розуміють групу теорем, кожна з яких встановлює умови, за яких середнє арифметичне випадкових величин має властивість стійкості. Тобто, середній результат дії великої кількості випадкових явищ практично перестає бути випадковим. У масі однорідних явищ випадкові особливості компенсуються і виявляються якісь закономірності, що служать предметом дослідження. Саме у виявленні загальних умов, що забезпечують статистичну стійкість середніх, полягає неминуча наукова цінність закону великих чисел.

Граничні теореми, які встановлюють відповідність між теоретичними та дослідними характеристиками випадкових подій, називають ***законом великих чисел.***

2.3.4.1. Нерівність і теорема Чебишова

Розглянемо спочатку нерівність Чебишова. Нехай випадкова величина X має скінченне математичне сподівання $M(X)$ і скінченну дисперсію $D(X)$. Тоді для будь-якого $\varepsilon > 0$ виконується ***нерівність Чебишова***

$$P\left(\left|X - M(X)\right| < \varepsilon\right) \geq 1 - \frac{D(X)}{\varepsilon^2}. \tag{2.29}$$

Перейшовши до протилежної події, маємо:

$$P\left(\left|X - M(X)\right| \geq \varepsilon\right) \leq \frac{D(X)}{\varepsilon^2}. \tag{2.30}$$

Наслідок (***правило*** 3σ). Для будь-якої випадкової величини X, яка має скінченне математичне сподівання та скінченну дисперсію, виконується нерівність

$$P\left(\left|X - M(X)\right| < 3\sigma\right) \geq 1 - \frac{\sigma^2}{(3\sigma)^2} = 1 - \frac{1}{9} = \frac{8}{9}. \tag{2.31}$$

Зауваження. Нерівність Чебишова на практиці має обмежене значення, оскільки часто дає грубу оцінку. Нехай, наприклад, $\varepsilon = \frac{\sigma}{2}$, тоді за нерівністю (2.30) матимемо:

$$P\left(\left|X - M(X)\right| \geq \frac{\sigma}{2}\right) \leq \frac{\sigma^2}{\sigma^2/4} = 4.$$

Але і так зрозуміло, що ймовірність не може бути більшою за одиницю. Якщо, тепер $\varepsilon = 10\cdot\sigma$, то

$$P\left(\left|X - M(X)\right| \geq 10\sigma\right) \leq \frac{\sigma^2}{100\sigma^2} = 0,01.$$

Це вже непогана оцінка для ймовірності.

Попри те, що на практиці цей варіант закону великих чисел застосовують нечасто, теоретичне значення нерівності Чебишова дуже велике.

Однією з фундаментальних теорем теорії ймовірностей є **теорема Чебишова.** *Ймовірність того, що абсолютне відхилення середнього арифметичного попарно незалежних випадкових величин від середнього арифметичного їх математичних сподівань не перевищить як завгодно малого додатного числа, буде як завгодно близька до одиниці, якщо кількість випадкових величин достатньо велика.*

Теорема Чебишова дає змогу обґрунтувати правило середнього, яким широко користуються в практиці вимірювань: за достатньо великої кількості вимірювань з імовірністю, як завгодно близькою до одиниці, можна вважати, що середнє арифметичне результатів вимірювань буде як завгодно мало відрізнятись від вимірюваної величини, тобто точно характеризує вимірювану величину.

Теорема Чебишова має велике теоретичне, практичне та методологічне значення. На цій теоремі ґрунтується вибірковий метод, який застосовують у статистиці. Його суть полягає в тому, що за порівняно невеликою випадковою вибіркою роблять висновок про всю сукупність досліджуваних об'єктів.

Означення. Послідовність випадкових величин $X_1, X_2, \ldots, X_n, \ldots$ називається *збіжною за ймовірністю* до числа γ, якщо для довільного $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|X_n - \gamma\right| < \varepsilon\right) = 1.$$

Позначення збіжності за ймовірністю $X_n \xrightarrow{p} \gamma,\ n \to \infty.$

Теорема Чебишова (закон великих чисел). *Нехай* $X_1, X_2, \ldots, X_n, \ldots$ – *послідовність попарно незалежних випадкових величин зі скінченними*

математичними сподіваннями $M(X_i) = a_i < \infty\, i$ *дисперсіями, обмеженими однією і тією ж сталою* $D(X_i) \le C, i \in N$. *Тоді*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{a_1 + a_2 + \cdots + a_n}{n} \xrightarrow{p} 0,\ n \to \infty. \qquad (2.32)$$

Наслідок. Нехай X_1, X_2, …, X_n, … – послідовність попарно незалежних випадкових величин з однаковими математичними сподіваннями $M(X_i) = a$ і дисперсіями, обмеженими однією і тією ж сталою $D(X_i) \le C,\ i \in N$. Тоді

$$\frac{X_1 + X_2 + \cdots + X_n}{n} - a \xrightarrow{p} 0,\ n \to \infty.$$

Отже, це твердження обґрунтовує правило середнього арифметичного. За виконаних *n* незалежних вимірювань, позбавлених систематичних помилок та здійснених з деякою гарантованою точністю, *за достатньо великої кількості вимірювань з імовірністю, як завгодно близькою до одиниці, середнє арифметичне результатів вимірювань буде як завгодно мало відрізнятися від вимірюваної величини.*

Теорема Бернуллі. Закон великих чисел у формі Бернуллі – безпосередній наслідок теореми Чебишова. Теорема Бернуллі відображає в математичній формі об’єктивно притаманну багатьом випадковим явищам властивість стійкості відносних частот. Ця теорема дає теоретичне обґрунтування заміни невідомої ймовірності події її частотою або статистичною ймовірністю, одержаною в *n* послідовних незалежних випробуваннях, що проводять за одного і того самого комплексу умов. Ця теорема поклала початок теорії ймовірностей як науки.

Теорема. *Нехай* μ_n *– кількість настання події A за n послідовних незалежних випробувань, у кожному з яких імовірність настання події A дорівнює p. Тоді*

$$\frac{\mu_n}{n} \xrightarrow{p} p,\ n \to \infty. \qquad (2.33)$$

Теорема Маркова. *Нехай* $X_1, X_2, \ldots, X_n, \ldots$ *є послідовністю випадкових величин, які мають скінченні математичні сподівання* $(M(X_i) = a_i < \infty)$, *для яких виконується умова* $\lim\limits_{n\to\infty} \dfrac{D(X_1 + X_2 + \cdots + X_n)}{n^2} = 0.$

Тоді

$$\frac{X_1 + X_2 + \cdots X_n}{n} - \frac{a_1 + a_2 + \cdots + a_n}{n} \xrightarrow{p} 0, \quad n \to \infty. \tag{2.34}$$

Твердження Маркова справедливе для будь-якого закону розподілу.

Зауваження. Велике практичне значення має теорема Пуассона, яка є узагальненням теореми Бернуллі на випадок, коли незалежні випробування відбуваються не обов'язково за однакових умов. Наприклад, часто доводиться перевіряти відповідність теоретично обчисленої ймовірності події фактичній відносній частоті цієї події. Для визначення відносної частоти потрібне багаторазове відтворення експерименту, та далеко не завжди це можна зробити в тих самих умовах. Однак перевірку можна здійснити, якщо вдається обчислити ймовірності для різних умов. У цьому випадку можна порівняти спостережену в досліді відносну частоту із середнім арифметичним обчислених імовірностей.

Закон великих чисел лежить в основі різних видів страхування (страхування життя людини на різні терміни, страхування майна тощо). У разі планування асортименту товарів широкого ужитку враховують попит на них у населення. У цьому попиті виявляється дія закону великих чисел.

2.3.4.2. Центральна гранична теорема

Граничні теореми, що встановлюють граничні закони розподілу випадкових величин, об'єднують загальною назвою – ***центральна гранична теорема.***

У групі теорем, що носять загальну назву *«центральна гранична теорема»,* йдеться про закономірності розподілу суми досить великої кількості випадкових величин. Різні форми центральної граничної теореми відрізняються одна від одної тими чи тими умовами, за яких виникає нормальний закон. Досить загальні умови має **теорема О. М. Ляпунова:** *якщо випадкова величина є сумою взаємно*

незалежних випадкових величин $X_1, X_2, \ldots, X_n$, *кожна з яких мало впливає на суму, то за наявності досить великої кількості доданків* n *закон розподілу їх суми стає як завгодно близьким до нормального, незважаючи на закон розподілу доданків.*

За теоремою Ляпунова вплив кожного окремого доданка на суму за великих *n* дуже малий, і в разі необмеженого збільшення кількості доданків закон розподілу їх суми необмежено наближується до нормального з математичним сподіванням і дисперсією, які дорівнюють сумам відповідних числових характеристик доданків.

Центральна гранична теорема пояснює велике поширення нормального закону розподілу і є теоретичною основою застосування нормального розподілу для багатьох практичних задач: *за широких припущень сума великої (але скінченної) кількості незалежних випадкових величин розподілена згідно із законом, близьким до нормального.* Наприклад, на відлагодженому виробництві якість продукції змінюється за нормальним законом унаслідок того, що виробнича похибка є результатом сумарної дії великої кількості випадкових величин. Окремим випадком центральної граничної теореми є інтегральна формула Лапласа.

Отже, і закон великих чисел, і центральна гранична теорема становлять теоретичну основу розв'язання багатьох задач, що виникають у практичній діяльності людини.

Отже, сформулюємо граничні теореми.

Теорема Ляпунова. *Якщо для послідовності попарно незалежних випадкових величин* $X_1, X_2, \ldots, X_n, \ldots$ *можна знайти таке число* $\delta > 0$, *що*

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

то

$$\lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} M(X_k)}{\sqrt{\sum_{k=1}^{n} D(X_k)}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Зауваження. На практиці зазвичай найлегше перевірити умову Ляпунова для δ = 1. Якщо послідовність випадкових величин задовольняє умову Ляпунова, то вона задовольняє також умову Ліндеберга. Обернене твердження неправильне.

Теорема Ліндеберга-Леві. *Якщо попарно незалежні випадкові величини* $X_1, X_2, \ldots, X_n, \ldots$ *однаково розподілені і мають математичне сподівання a та дисперсію* σ^2, *то*

$$\lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - na}{\sigma\sqrt{n}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt\,.$$

Центральна гранична теорема пояснює достатньо широке застосування нормального закону розподілу: *якщо випадкова величина формується під впливом багатьох незалежних факторів, кожен з яких здійснює на неї незначний вплив, то розподіл цієї величини близький до нормального.*

Приклад 2.22. Нормально розподілена випадкової величина X має математичне сподівання $M(X) = 4$ та середньоквадратичне відхилення $\sigma(X) = 2.$ Ставиться завдання:

1. Записати вигляд диференціальної функції розподілу $f(x)$, побудувати її графік та графік інтегральної функції розподілу $F(x)$.

2. Знайти ймовірність того, що внаслідок випробування випадкова величина набуде значення з інтервалів: $(2, 6)$, $(a-3\sigma, a+3\sigma)$ (тобто знайти $P(2 < X < 6)$ і $P(a-3\sigma < X < a+3\sigma)$).

Розв'язання. 1. Згідно умови задачі $a = 4,\ \sigma = 2,$ тому за формулою щільності розподілу

$$f(x)=\frac{1}{2\sqrt{2\pi}}e^{-\frac{(x-4)^2}{2\cdot 2^2}}=\frac{1}{2\sqrt{2\pi}}e^{-\frac{(x-4)^2}{8}}.$$

Для побудови графіка функції $f(x)$ знайдемо значення у точках $x=-2,\ 0,\ \ldots,\ 12$, використовуючи функцією пакету Excel, яка має вигляд: **НОРМРАСП**(*x*; среднее; стандартное_откл; интегральная), де *x*– значення, для якого будується розподіл; *среднее* – математичне сподівання $M(X)$; *стандартное_откл* – середнє квадратичне відхилення $\sigma(X)$; *интегральная* – логічне значення 0 або 1. (Якщо «*интегральная*»: – 0, отримаємо значення диференціальної функції розподілу $f(x)$ в точці *x*, якщо «*интегральная*»: 1 – функція повертає значення інтегральної функції розподілу $F(x)$ в точці *x*).

Для побудови графіка функції $F(x)$ знайдемо її значення у тих же точках $x=-2,\ 0,\ \ldots,\ 12,$ використовуючи функцію **НОРМРАСП**(*x*; $M(X)$; $\sigma(X)$;1).

2. Використовуючи формулу (2.26), одержимо:

$$P(2<X<6)=\int_2^6 f(t)dt=\Phi\left(\frac{6-4}{2}\right)-\Phi\left(\frac{2-4}{2}\right)=2\Phi(1)=2\cdot 0,34134\approx 0,683,$$

де $\Phi(1)$ знаходимо за таблицею значень функції Лапласа.

Цю ймовірність легко обчислити, якщо скористатись функцією **НОРМРАСП,** використовуючи різницю виразів **НОРМРАСП**(6;4;2;1) і **НОРМРАСП**(2;4;2;1), що дорівнює 0,683. Результат застосування функції **НОРМРАСП** та графіки функції $f(x)$ і $F(x)$ показано на рис. 2.22.

Для інтервалу $(a-3\sigma, a+3\sigma)=(-2,10)$ шукана ймовірність обчислюється аналогічно за допомогою різниці виразів

НОРМРАСП(10;4;2;1), **НОРМРАСП**(-2;4;2;1)

і дорівнює наближено 0,9973, що і співпадає з теоретичними міркуваннями.

Зауваження. В пакеті Excel, в залежності від встановленого офісу, використовують ті ж описані функції, але запис їх англомовний. Для виокремлення потрібної бажано дивитись на довідку стосовно потрібної функції.

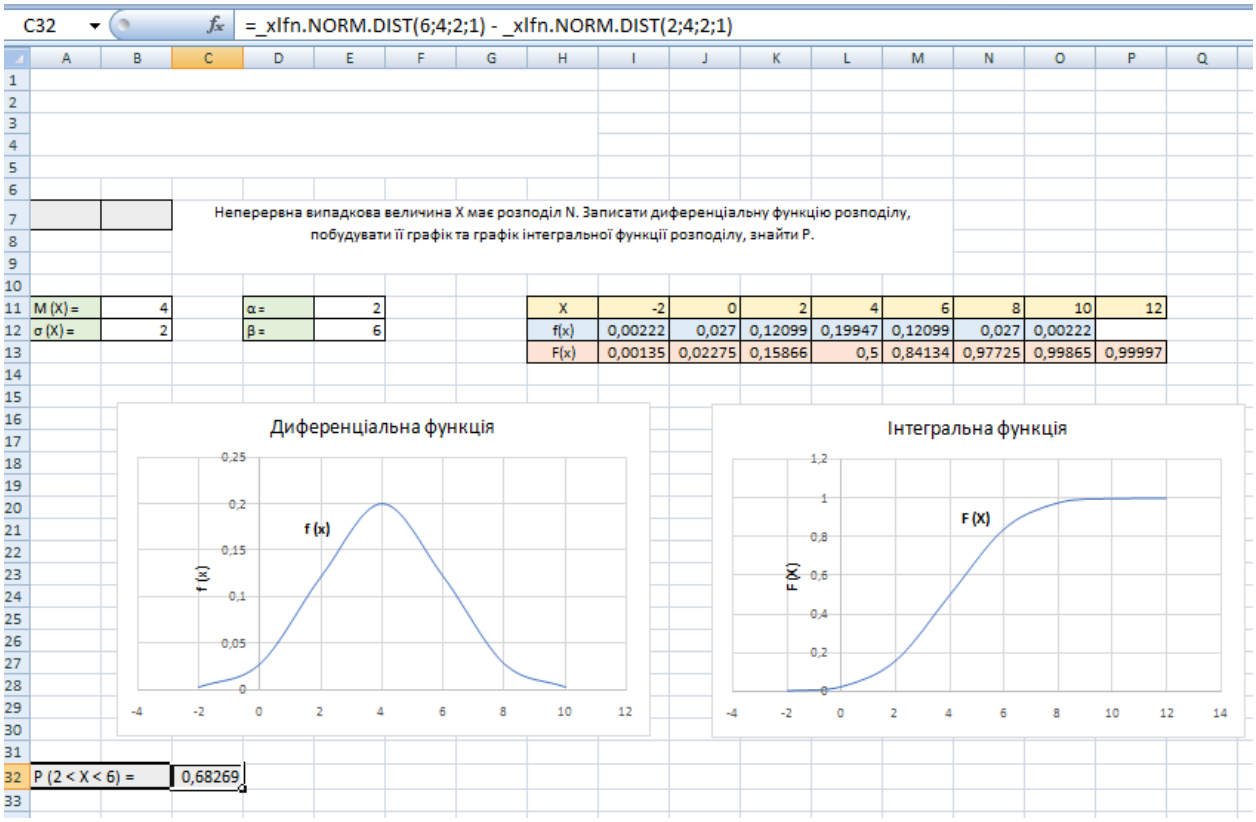


Рис. 2.22. Обчислення за допомогою **НОРМРАСП**

Приклад 2.23. Прилад складається із 100 незалежно працюючих елементів. Імовірність відмови кожного елемента за час T дорівнює 0,03. Оцінити ймовірність того, що абсолютна величина різниці між кількістю (математичним сподіванням) відмов за час T буде: а) менше двох; б) не менше двох.

Розв'язання.

а) Позначимо через X – кількість елементів, які відмовили за час T. Тоді $M(X) = np = 100 \cdot 0,03 = 3,\ D(X) = npq = 100 \cdot 0,03 \cdot 0,97 = 2,91.$

Використовуємо нерівність Чебишова (2.29), підставивши дані $M(X) = 3,\quad D(X) = 2,91$ та $\varepsilon = 2$:

$$P\left(|X - 3| < 2\right) \geq 1 - \frac{2,91}{4} \approx 0,27.$$

б) Події $|X - 3| < 2$ і $|X - 3| \geq 2$ є протилежними, тому

$$P\left(|X - 3| > 2\right) = 1 - P\left(|X - 3| < 2\right) = 0,73.$$

Приклад 2.24. Імовірність того, що деталь пройшла перевірку ВТК, дорівнює 0,1. Знайти ймовірність того, що серед 200 навмання обраних деталей виявиться неперевірених від 10 до 30 деталей.

Розв'язання.

За умовою задачі маємо

$M(X) = np = 200 \cdot 0{,}1 = 20,\ D(X) = npq = 200 \cdot 0{,}1 \cdot 0{,}9 = 18.$ Тоді

$$|X - 20| \le \varepsilon \Leftrightarrow 20 - \varepsilon \le X \le 20 + \varepsilon,$$

звідки $\varepsilon = 10.$ Відповідно, за нерівністю Чебишова тоді

$$P(|X - 20| \le 10) \ge 1 - \frac{18}{100} = 0{,}82.$$

Завдання для самоконтролю

1. Означити нормальний закон розподілу випадкової величини, побудувати графік її щільності розподілу.

2. Інтегральна функція нормального розподілу та інтегральна функція Лапласа.

3. Основні числові характеристики нормального розподілу, їх отримання.

4. Правило «трьох сигм», його тлумачення.

5. Розподіл χ^2.

6. Розподіл Стьюдента, практичне застосування та зв'язок із нормальним законом розподілу.

7. Закон великих чисел (формулювання граничних теорем).

8. Нерівність та теорема Чебишова.

9. Закон великих чисел у формі Бернуллі та Пуассона, їх практичне значення.

10. Центральна гранична теорема, формулювання та пояснення змісту.

11. Випадкова величина X розподілена нормально з математичним сподіванням $M(X)=10$. Імовірність потрапляння X в інтервал (10; 20) дорівнює 0,2. Знайти ймовірність потрапляння X в інтервал (20; 30). *Відповідь.* 0,15.

12. Автомат штампує деталі. Контролюється довжина деталі X, яка розподілена нормально з математичним сподіванням 50 мм. Фактична довжина виготовлених деталей не менша 32 мм і не більша 68 мм. Знайти ймовірність того, що довжина навмання взятої деталі більша 55 мм. *Відповідь.* 0,0823.

13. Математичне сподівання і середнє квадратичне відхилення нормальної розподіленої випадкової величини X відповідно дорівнюють 20 і 5. Знайти ймовірність того, що в результаті випробування випадкова величина X набуде значення що належить інтервалу (15;25). *Відповідь.* 0,68268.

14. Ціна поділки шкали вимірювального приладу дорівнює 0,1. Покази приладу заокруглюють до найближчої цілої поділки. Знайти ймовірність того, що при відліку буде допущена похибка більша від 0,02. *Відповідь.* 0,6.

15. Деталі, які випускає цех, за розмірами мають нормальний закон розподілу з математичним сподіванням $M(X)=5$ і середнім квадратичним відхиленням $\sigma(X)=0,9$. Знайти ймовірність того, що розмір діаметра навмання взятої деталі відрізняється від математичного сподівання не більше, ніж на 2 см. *Відповідь.* 0,9736.

16. Пристрій складається з 10 елементів, що працюють незалежно один від одного. Імовірність відмови кожного елементу за час t дорівнює 0,05. За нерівністю Чебишова оцініть імовірність того, що абсолютна величина різниці між кількістю елементів, що відмовили за час t, і математичним сподіванням відмов, буде меншою за 2.

17. Довжина виробів є випадковою величиною із середнім значенням 90см і дисперсією 0,0225. Оцініть ймовірність того, що: а) довжина взятого на контроль виробу відхилиться за абсолютною величиною від середнього значення

не більше ніж на 0,4 см; б) довжина виробу буде в межах від 89,5 до 90,5см. *Відповідь.* а) 0,859; б) 0,91.

18. Середня кількість пасажирів, що проходять через станцію метрополітену становить 800 осіб. Оцінити ймовірність того, що в один день через цю станцію пройде не більше 1200 осіб. *Відповідь*. 0,333.

Варіанти завдань для індивідуальної роботи

Завдання. Неперервна випадкова величина X має розподіл $N(a;\sigma)$. Записати диференціальну функцію розподілу $f(x)$, побудувати її графік та графік інтегральної функції розподілу, знайти $P(\alpha < X < \beta)$, використовуючи функцією пакету Excel **НОРМРАСП**, якщо

1) $a=3,\ \sigma=4,\ \alpha=0{,}5,\ \beta=9;$
2) $a=3,\ \sigma=4,\ \alpha=2,\ \beta=9;$
3) $a=0,\ \sigma=12,\ \alpha=3,\ \beta=6;$
4) $a=2,\ \sigma=8,\ \alpha=5,\ \beta=9;$
5) $a=2,\ \sigma=1,\ \alpha=4,\ \beta=11;$
6) $a=5,\ \sigma=1,\ \alpha=1,\ \beta=5;$
7) $a=3,\ \sigma=2,\ \alpha=2,\ \beta=5;$
8) $a=1,\ \sigma=25,\ \alpha=1,\ \beta=3;$
9) $a=2,\ \sigma=25,\ \alpha=3,\ \beta=8;$
10) $a=3,\ \sigma=9,\ \alpha=0{,}25,\ \beta=10;$
11) $a=-1,\ \sigma=-1,\ \alpha=0,\ \beta=9;$
12) $a=-1,\ \sigma=5,\ \alpha=2,\ \beta=7;$
13) $a=-2,\ \sigma=9,\ \alpha=0,\ \beta=7;$
14) $a=-2,\ \sigma=4,\ \alpha=1,\ \beta=5{,}25;$
15) $a=12,\ \sigma=4,\ \alpha=5,\ \beta=9;$
16) $a=13,\ \sigma=4,\ \alpha=6,\ \beta=9;$
17) $a=11,\ \sigma=6,\ \alpha=6,\ \beta=12;$
18) $a=11,\ \sigma=6,\ \alpha=5,\ \beta=8;$
19) $a=-1,\ \sigma=4,\ \alpha=7,\ \beta=9{,}25;$
20) $a=1,\ \sigma=4,\ \alpha=0,\ \beta=14.$

2.4. Багатовимірні випадкові величини

В прикладних задачах зазвичай доводиться розглядати не одну, а декілька випадкових величин, які одночасно вимірюються (спостерігаються) в одному й тому ж експерименті. При цьому з кожною елементарною подією пов'язана система (сукупність) двох чи більше випадкових величин. Наприклад, успішність студента характеризується системою n випадкових величин, $X_1, X_2, ..., X_n$ – оцінки, отримані ним у семестрі. Такі питання, а також функції випадкових аргументів розглянемо далі.

2.4.1. Функція одного випадкового аргумента, її розподіл та числові характеристики

Означення. Якщо кожному можливому значенню випадкової величини X відповідає одне можливе значення випадкової величини Y, то Y називають **функцією випадкового аргумента X:**

$$Y = \varphi(X).$$

Такі функції розглядають як дискретні, так і неперервні.

Нехай X – *дискретна випадкова величина*. Якщо різним можливим значенням випадкової величини X відповідають різні значення функції Y, то ймовірності відповідних значень X та Y різні. Якщо різним можливим значенням випадкової величини X відповідають значення Y, серед яких є рівні між собою, то ймовірності рівних значень Y додаються.

Нехай дискретна випадкова величина X задана рядом розподілу

X	x_1	x_2	…	x_n	…
P	p_1	p_2	…	p_n	…

і $Y = \varphi(X)$.

Тоді справедливі такі формули:

$$M\left(\varphi(X)\right) = \sum_k \varphi(x_k)\, p_k; \tag{2.35}$$

$$D\left(\varphi(X)\right) = \sum_k \varphi^2(x_k)\, p_k - \left(M\left(\varphi(X)\right)\right)^2. \tag{2.36}$$

Нехай X – *неперервна випадкова величина* з щільністю розподілу $f(x)$, а $Y=\varphi(X)$. Якщо $y=\varphi(x)$ – диференційована, строго зростаюча або строго спадна функція, обернена до якої має вигляд $x=\psi(y)$, то щільність розподілу $g(y)$ випадкової величини $Y=\varphi(X)$ обчислюють за формулою:

$$g(y)=f(\psi(y))\cdot|\psi'(y)|. \qquad (2.37)$$

Доведемо формулу (2.37). Припустимо, що функція $\varphi(x)$ *монотонно зростає*, неперервна і диференційована на (a,b). Функція розподілу $g(y)$ випадкової величини Y визначається за формулою

$$g(y)=P(Y<y).$$

Якщо функція $y=\varphi(x)$ монотонно зростає, то подія $(Y<y)$ еквівалентна події ($X<\varphi^{-1}(y)$), де $x=\varphi^{-1}(y)=\psi(y)$ − функція, обернена до функції $y=\varphi(x)$, яка теж монотонно зростаюча, неперервна і диференційована.

Отже, $g(y)=P(Y<y)=P(X<\psi(y))=\int\limits_{a}^{\psi(y)} f(x)dx$.

Диференціюючи цей вираз за змінною y, отримаємо щільність розподілу випадкової величини Y:

$$g(y)=g'(y)=f\left(\psi(y)\right)|\psi'(y)|.$$

Якщо функція $\varphi(x)$ на (a,b) *монотонно спадає*, то подія $(Y<y)$ еквівалентна події ($X>\psi(y)$). Отже,

$$g(y)=\int\limits_{\psi(y)}^{b} f(x)dx,\ g(y)=-f\left(\psi(y)\right)|\psi'(y)|.$$

Оскільки щільність розподілу не може бути від‘ємною, то отримані формули можна об‘єднати в одну формулу (2.37):

$$g(y)=f\left(\psi(y)\right)|\psi'(y)|.$$

Зауваження. Якщо функція $y=\varphi(x)$ в інтервалі можливих значень X не монотонна, то потрібно розбити цей інтервал на такі інтервали, на яких $\varphi(x)$

монотонна, знайти щільності розподілів $g_i(y)$ для кожного з інтервалів монотонності, а потім подати $g(y)$ у вигляді суми: $g(y)=\sum_i g_i(y)$.

Числові характеристики випадкової величини *Y* обчислюють за формулами:

$$M\left((\varphi(X)\right)=\int_{-\infty}^{+\infty}\varphi(x)\,f(x)dx;$$

$$D\left(\varphi(X)\right)=\int_{-\infty}^{+\infty}\varphi^2(x)\,f(x)dx-\left(M(\varphi(X))\right)^2.$$

Зокрема, якщо можливі значення *X* належать інтервалу (a,b), то

$$M\left(\varphi(X)\right)=\int_{a}^{b}\varphi(x)\,f(x)dx;$$

$$D\left(\varphi(X)\right)=\int_{a}^{b}\varphi^2(x)\,f(x)dx-\left(M(\varphi(X))\right)^2.$$

Приклад 2.25. Дана дискретна випадкова величина *X*

X	0	$\frac{\pi}{6}$	$\frac{\pi}{2}$
P	0,3	0,5	0,2

та $Y=\sin X$. Знайти $M(Y),\ D(Y),\ \sigma(Y)$.

Розв'язання. Знаходимо розподіл випадкової величини Y:

$$y_1=\sin 0=0,\ y_2=\sin\frac{\pi}{6}=\frac{1}{2},\ y_3=\sin\frac{\pi}{2}=1.$$

Y	0	$\frac{1}{2}$	1
P	0,3	0,5	0,2

$$M(Y)=0\cdot 0,3+0,5\cdot 0,5+1\cdot 0,2=0,45;$$

$$M(Y^2)=0\cdot 0,3+0,25\cdot 0,5+1\cdot 0,2=0,325;$$

$$D(Y)=M(Y^2)-(M(Y))^2=0,325-(0,45)^2=0,1225;$$

$$\sigma(Y)=\sqrt{D(Y)}=\sqrt{0,1225}=0,35.$$

Приклад 2.26. Випадкова величина *X* задана щільністю розподілу

$$f(x)=\begin{cases}0,\ x\le 0\ \ або\ \ x>1;\\ 2x,\ 0<x\le 1;\end{cases}\ Y=X^2.$$

Знайти $M(Y), D(Y)$.

Розв'язання.

$$M(Y)=\int_0^1 x^2\cdot 2xdx=2\int_0^1 x^3dx=\frac{x^4}{2}\Big|_0^1=\frac{1}{2};$$

$$M(Y^2)=\int_0^1 x^4\cdot 2xdx=2\int_0^1 x^5dx=\frac{x^6}{3}\Big|_0^1=\frac{1}{3};$$

$$D(Y)=M(Y^2)-(M(Y))^2=\frac{1}{3}-\left(\frac{1}{2}\right)^2=\frac{1}{12}.$$

Приклад 2.27. Задана щільність розподілу $f(x)$ випадкової величини *X*, можливі значення якої належать інтервалу $(0,+\infty)$. Знайти щільність розподілу випадкової величини $Y=\ln X$.

Розв'язання. Функція $y=\ln x$ монотонна в інтервалі $(0,+\infty)$, отже має обернену функцію $x=e^y$, звідки $x'=e^y$. Тоді за формулою (2.37) отримаємо:

$$g(y)=e^y f(e^y),\ y\in(-\infty,\ +\infty).$$

Приклад 2.28. Дискретна випадкова величина *X* задана рядом розподілу

X	–1	1	2
P	0,1	0,5	0,4

Знайти розподіл функції $Y=X^2$.

Розв'язання. Можливі значення *Y*: $y_1=1,\ y_2=1,\ y_3=4$. Значення $y_1=y_2$ рівні між собою. Ймовірність цього значення $p_1=0,1+0,5=0,6$.

Одержимо розподіл *Y*:

Y	1	4
P	0,6	0,4

Приклад 2.29. Випадкова величина X розподілена нормально з параметрами $a=0,\ \sigma_x=1,$ тобто відома щільність розподілу $f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Знайти закон розподілу функції $Y=X^2$.

Розв'язання. Обернена функція $x=\varphi^{-1}(y)$ неоднозначна: $x_1=-\sqrt{y}$ для $x<0$ і $x_2=\sqrt{y}$ для $x\geq 0$.

Отже, $g(y)=\frac{1}{\sqrt{2\pi}}\cdot e^{-y/2}\left|-\frac{1}{2\sqrt{y}}\right|+\frac{1}{\sqrt{2\pi}}\cdot e^{-y/2}\left|\frac{1}{2\sqrt{y}}\right|=\frac{1}{\sqrt{2\pi y}}e^{-y/2},\ y>0$.

Розглянемо задачу побудови закону розподілу лінійної функції.

Нехай $Y=\alpha X+\beta$, де α,β – невипадкові величини. Оскільки $y=\alpha x+\beta$ монотонна функція, то обернена функція $x=\frac{y-\beta}{\alpha}$ теж монотонна. Маємо $x'_y=\frac{1}{\alpha}$ і за формулою

$$g(y)=f\left(\frac{y-\beta}{\alpha}\right)\cdot\frac{1}{|\alpha|}.$$

Вираз показує, що лінійне перетворення випадкової величини X тотожне зміні масштабу зображення кривої $f(x)$ і переносу початку координат в нову точку. Вигляд кривої $f(x)$ при такому перетворенні не змінюється.

Покажемо, що лінійна функція $Y=\alpha X+\beta$ розподілена нормально, якщо аргумент X – нормально розподілена випадкова величина із щільністю розподілу $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$. Припустивши $\alpha>0$, знайдемо похідну: $x'_y=\frac{1}{\alpha}$. За формулою запишемо щільність розподілу функції

$$g(y)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\left(\frac{y-\beta}{\alpha}-a\right)^2}{2\sigma^2}},\ \text{або}\ g(y)=\frac{1}{(\sigma\alpha)\sqrt{2\pi}}e^{-\frac{(y-(\beta+\alpha a))^2}{2(\alpha\sigma)^2}}\ e^{-\frac{(y-(\beta+\alpha a))^2}{2(\alpha\sigma)^2}}.$$

Таким чином, лінійна функція теж розподілена нормально з параметрами $\sigma_y = \sigma\alpha$ та $a_y = \beta + \alpha a$.

2.4.2. Функція двох випадкових аргументів

Якщо кожній парі можливих значень випадкових величин X і Y відповідає одне можливе значення випадкової величини Z, то Z називають ***функцією двох випадкових величин*** X та Y:

$$Z = \varphi(X,\ Y).$$

Задача визначення закону розподілу функції декількох випадкових аргументів є складнішою. Тому розглянемо випадок функції двох аргументів. Нехай задана система двох неперервних випадкових величин (X,Y), щільність розподілу якої $f(x,y)$ відома, і випадкова величина $Z = \varphi(X,Y)$, де $\varphi(X,Y)$ − функція. Потрібно знайти закон розподілу величини Z. Геометрично функція $z = \varphi(x,y)$ зображається деякою поверхнею в просторі. Знайдемо функцію розподілу величини Z:

$$G(z) = P(Z < z) = P(\ \varphi(X,Y) < z).$$

Проведемо площину, паралельну площині Oxy, на відстані z від неї. Ця площина перетне поверхню $z = \varphi(x,y)$ по деякій кривій L. Спроектуємо криву L на площину Oxy. Ця проекція, рівняння якої $z = \varphi(x,y)$, розділить площину Oxy на дві області − для однієї висота поверхні над площиною Oxy буде менше z, це область D; а для іншої – більше z. Щоб виконувалась рівність для функції розподілу, випадкова точка (X,Y) повинна попасти в область D, яка визначається нерівністю $\varphi(x,y) < z$. Отже,

$$G(z) = P(\ (X,Y) \subset D) = \iint_D f(x,y)dxdy.$$

Тут z входить неявно. Диференціюючи вираз за z, отримаємо щільність розподілу величини Z:

$$g(z) = G'(z).$$

На практиці достатньо побудувати на площині Oxy криву, рівняння якої $z=\varphi(x,y)$, і вияснити, по який бік цієї кривої $Z<z$, а по який $Z>z$, тоді інтегрувати по області D, для якої $Z<z$.

Закон розподілу суми $Z=X+Y$ за відомими розподілами доданків

1. Нехай X та Y – дискретні незалежні випадкові величини. Можливі значення $Z=X+Y$ дорівнюють сумам кожного можливого значення X та кожного можливого значення Y; імовірності можливих значень дорівнюють добуткам імовірностей доданків.

2. Якщо X та Y – неперервні незалежні випадкові величини, які мають відповідно щільності розподілу $f_1(x)$ і $f_2(y)$, то щільність розподілу $g(z)$ суми $Z=X+Y$ за умови, що щільність розподілу принаймні одного з аргументів задана в інтервалі $(-\infty,+\infty)$, можна знайти за формулою

$$g(z)=\int_{-\infty}^{+\infty} f_1(x)\,f_2(z-x)dx$$

або за рівносильною формулою

$$g(z)=\int_{-\infty}^{+\infty} f_1(z-y)\,f_2(y)dy.$$

Функція $g(z)$, утворена з функцій $f_1(x)$ і $f_2(y)$ за наведеними формулами, називається ***згорткою*** або ***композицією*** цих функцій.

Якщо можливі значення аргументів невід'ємні, то щільність розподілу суми $Z=X+Y$ обчислюють за формулою

$$g(z)=\int_{0}^{z} f_1(x)\,f_2(z-x)dx$$

або за рівносильною формулою

$$g(z)=\int_{0}^{z} f_1(z-y)\,f_2(y)dy.$$

Зауваження. Сума незалежних нормально розподілених випадкових величин також має нормальний розподіл. Математичне сподівання і дисперсія цієї

суми відповідно дорівнюють сумам математичних сподівань і дисперсій доданків.

Приклад 2.30. Для незалежних випадкових величин X та Y відомі щільності їх розподілів

$$f_1(x)=\frac{1}{3}e^{-\frac{x}{3}}\ \ (0\le x<\infty),\ \ f_2(y)=\frac{1}{5}e^{-\frac{y}{5}}\ \ (0\le y<\infty).$$

Знайти щільність розподілу випадкової величини $Z=X+Y.$

Розв'язання. За формулою $g(z)=\int\limits_0^z f_1(x)\,f_2(z-x)dx$ обчислимо:

$$g(z)=\int\limits_0^z\frac{1}{3}e^{-\frac{x}{3}}\cdot\frac{1}{5}e^{-\frac{(z-x)}{5}}dx=\frac{1}{15}e^{-\frac{z}{5}}\int\limits_0^z e^{-\frac{2x}{15}}dx=\frac{1}{15}e^{-\frac{z}{5}}\left(-\frac{15}{2}e^{-\frac{2x}{15}}\right)\Bigg|_0^z=\frac{1}{2}e^{-\frac{z}{5}}\left(1-e^{-\frac{2z}{15}}\right).$$

Оскільки можливі значення X та Y невід'ємні і $Z=X+Y$, то $z\ge 0$. Отже,

$$g(z)=\begin{cases}\dfrac{1}{2}e^{-\frac{z}{5}}\left(1-e^{-\frac{2z}{15}}\right),\ 0\le z<+\infty;\\ 0,\ -\infty<z<0.\end{cases}$$

Приклади композицій законів розподілу

1. Скласти композицію *нормального закону* розподілу із щільністю

$$f_1(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$$

та рівномірного

$$f_2(y)=\begin{cases}\dfrac{1}{\beta-\alpha},\quad \alpha<y<\beta,\\ 0,\quad y\le\alpha,\quad y\ge\beta.\end{cases}$$

Розв'язання. Застосувавши формулу згортки, отримаємо

$$g(z)=\int\limits_\alpha^\beta\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(z-y-a)^2}{2\sigma^2}}\cdot\frac{1}{\beta-\alpha}dy=\frac{1}{\beta-\alpha}\frac{1}{\sigma\sqrt{2\pi}}\int\limits_\alpha^\beta e^{-\frac{(y-(z-a))^2}{2\sigma^2}}dy.$$

Вираз $\frac{1}{\sigma\sqrt{2\pi}}\int_{\alpha}^{\beta} e^{-\frac{(y-(z-a))^2}{2\sigma^2}}\,dy$ – це ймовірність потрапляння в інтервал (α,β) нормально розподіленої величини з центром $z-a$.

Отже,

$$g(z)=\frac{1}{\beta-\alpha}\left(\Phi\left(\frac{\beta-(z-a)}{\sigma}\right)-\Phi\left(\frac{\alpha-(z-a)}{\sigma}\right)\right).$$

2. Скласти композицію *двох нормальних законів розподілу*

$$f_1(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\ f_1(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}.$$

Розв'язання. Знайдемо щільність розподілу суми

$$g(z)=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(z-y)^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-\frac{(z-y)^2+y^2}{2}}dy.$$

Перетворивши показник степеня в суму квадратів

$$(z-y)^2+y^2=z^2-2zy+y^2+y^2=\left(y\sqrt{2}-\frac{z}{\sqrt{2}}\right)^2+\frac{z^2}{2}$$

та зробивши заміну змінної $t=y\sqrt{2}-\frac{z}{\sqrt{2}}$ $\left(dy=\frac{1}{\sqrt{2}}dt\right)$, отримаємо

$$g(z)=\frac{1}{2\pi}\frac{1}{\sqrt{2}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}e^{-\frac{z^2}{4}}dt=\frac{1}{2\pi}\frac{1}{\sqrt{2}}e^{-\frac{z^2}{4}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt.$$

Оскільки $\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}$, то остаточно отримаємо:

$$g(z)=\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2}}e^{-\frac{z^2}{2(\sqrt{2})^2}}.$$

Таким чином, закон розподілу суми двох незалежних нормальних випадкових величин з параметрами $a_1=0$, $\sigma_1=1$ та $a_2=0$, $\sigma_2=1$ теж є нормальним законом з параметрами $\sigma_z=\sqrt{2}$, $a_z=0.$

3. Скласти композицію *двох показникових законів розподілу* з параметрами λ_1 і λ_2, тобто $f_1(x)=\lambda_1 e^{-\lambda_1 x}\ (x>0)$; $f_2(y)=\lambda_2 e^{-\lambda_2 y}\ (y>0)$.

Розв'язання. Отримуємо

$$g(z)=\int_0^z \lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2(z-x)}\,dx=\lambda_1\lambda_2 e^{-\lambda_2 z}\int_0^z e^{(\lambda_2-\lambda_1)x}\,dx=\frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}\left(e^{-\lambda_1 z}-e^{-\lambda_2 z}\right),\ (z>0).$$

Композиція двох показникових законів розподілу з параметрами λ_1 і λ_2 називається узагальненим законом Ерланга другого порядку.

4. Розглянемо на прикладі, як знаходити композицію законів розподілу двох дискретних випадкових величин, а саме композицію *двох законів розподілу Пуассона* з параметрами λ_1 та λ_2.

Розв'язання. Імовірність події $(X+Y=m)$, $(m=0,1,2,....)$ знайдемо за формулою:

$$P(X+Y=m)=\sum_{k=0}^{m}P(X=k)\cdot P(Y=m-k)=\sum_{k=0}^{m}\frac{\lambda_1^k}{k!}e^{-\lambda_1}\cdot\frac{\lambda_2^{m-k}}{(m-k)!}e^{-\lambda_2}=$$

$$==\frac{e^{-(\lambda_1+\lambda_2)}}{m!}\cdot\sum_{k=0}^{m}\frac{m!\lambda_1^k}{k!}\frac{\lambda_2^{m-k}}{(m-k)!}=\frac{(\lambda_1+\lambda_2)^m}{m!}\cdot e^{-(\lambda_1+\lambda_2)}.$$

Таким чином, сума двох незалежних випадкових величин, розподілених за законом Пуассона, теж розподілена за законом Пуассона з параметром $\lambda_1+\lambda_2$.

2.4.3. Двовимірні випадкові величини

Випадкові величини, які входять в систему (X_1, X_2, …, X_n), можуть бути як дискретними, так і недискретними (неперервними або мішаними).

Розглянемо стохастичний експеримент, яким є підкидання двох монет. Елементарною подією буде випадання герба і положення монет одна відносно одної. Нехай випадкова величина X – кількість гербів, які випали, а випадкова величина Y – відстань між монетами. У цьому разі X – дискретна випадкова величина, яка набуває значень 0, 1, 2, а Y – неперервна випадкова величина із значеннями в R.

Для наочності дослідження *n*-вимірних випадкових величин зручно користуватися геометричною інтерпретацією. Так, систему двох випадкових величин (*X, Y*) можна зобразити *випадковою точкою* на площині *Oxy* з коорди-

натами X та Y, або, що рівносильно, *випадковим вектором*, напрямленим з початку координат в точку (X, Y). Аналогічно, систему трьох випадкових величин (X, Y, Z) можна зобразити в тривимірному просторі як випадковий вектор, напрямлений з початку координат в точку (X, Y, Z).

Якщо вимірність величини більше від 3, то геометрична інтерпретація втрачає наочність, але користуватися геометричною термінологією зручно. Так, про систему випадкових величин (X_1, X_2, …, X_n) будемо казати як про випадкову точку в n-вимірному просторі R^n або як про випадковий вектор, напрямлений з початку координат у точку (X_1, X_2, …, X_n): $\overrightarrow{X}$ = (X_1, X_2, …, X_n).

На n-вимірні випадкові величини поширюються майже без змін основні означення, які було розглянуто для одновимірної випадкової величини.

Очевидно, що властивості n-вимірної випадкової величини не обмежуються властивостями окремих складових випадкових величин; суттєві також зв'язки між складовими величинами.

Повна характеристика n-вимірної випадкової величини – її закон розподілу, як і для окремих випадкових величин, може мати різні форми: функція розподілу, щільність розподілу, таблиця ймовірностей окремих значень випадкового вектора тощо.

Для практики важливий випадок $n = 2$.

Означення. Функцією розподілу ймовірностей (або сумісною функцією розподілу) двовимірної випадкової величини (X, Y) називається така функція двох аргументів x, y, яка визначає ймовірність сумісного настання подій

$$F(x, y) = P\left(X < x,\ Y < y\right).$$

Подія в дужках означає добуток подій: $\left(X < x, Y < y\right) = \left(X < x\right) \bigcap \left(Y < y\right)$.

Користуючись геометричною інтерпретацією двовимірної випадкової величини (X, Y) як випадкової точки на площині, можна дати геометричне тлумачення функції розподілу: це ймовірність влучення в нескінченний прямокутник із вершиною в точці (X, Y) (рис. 2.23), який лежить лівіше і нижче від неї.

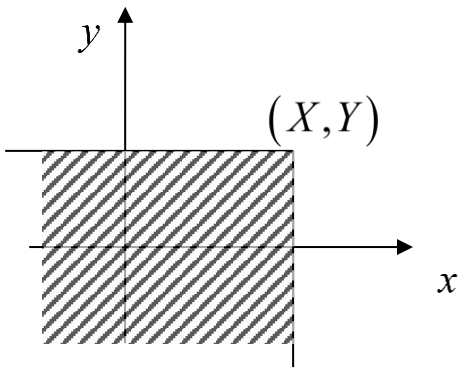


Рис. 2.23. Нескінченний прямокутник із вершиною в точці (*X, Y*)

Властивості функції розподілу двовимірної випадкової величини (*X, Y*) аналогічні властивостям одновимірної випадкової величини:

1. $0 \le F(x,y) \le 1.$

2. Неперервність зліва за кожним аргументом:

$$\lim_{x \to x_1 - 0} F(x,y) = F(x_1,y); \quad \lim_{y \to y_1 - 0} F(x,y) = F(x,y_1).$$

3. Функція $F(x,y)$ – неспадна функція за кожним аргументом:

$$F(x_1,y) \le F(x_2,y), \text{ якщо } x_1 < x_2;$$

$$F(x,y_1) \le F(x,y_2), \text{ якщо } y_1 < y_2.$$

4. $F(x, +\infty) = F_1(x)$ – функція розподілу компонента *X,*

$F(+\infty, y) = F_2(y)$ – функція розподілу компонента Y.

Властивість встановлює природний зв'язок між функцією розподілу двовимірної випадкової величини і функціями розподілу $F_1(x)$, $F_2(y)$, які також називають частинними функціями одновимірних випадкових величин X та *Y*.

5. $F(-\infty, y) = 0,\ F(x, -\infty) = 0,\ F(-\infty, -\infty) = 0,\ F(+\infty, +\infty) = 1.$

6. Імовірність потрапляння випадкової точки до прямокутника $(x_1 \le X \le x_2;\ y_1 \le Y \le y_2)$ обчислюють за формулою

$$P(x_1 < X < x_2;\ y_1 < Y < y_2) = (F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1)).$$

2.4.3.1. Дискретні двовимірні випадкові величини

Означення. Двовимірну випадкову величину називають **дискретною,** якщо множина значень, яких вона може набути, скінченна або зліченна.

Означення. **Законом розподілу дискретної двовимірної випадкової величини** називають перелік можливих значень цієї величини (x_i, y_k) та їх імовірностей $p(x_i, y_k),\ i = 1,\ 2,\ \ldots,\ m;\ k = 1,\ 2,\ \ldots,\ n.$

Закон розподілу такої випадкової величини задають таблицею:

Y \ X	x_1	x_2	…	x_m	$\sum$
y_1	p_{11}	p_{21}	…	p_{m1}	q_1
y_2	p_{12}	p_{22}	…	p_{m2}	q_2
…	…	…	…	…	…
y_n	p_{1n}	p_{2n}	…	p_{mn}	q_n
$\sum$	p_1	p_2	…	p_m	1

причому $q_1 = \sum_{i=1}^{m} p_{i1},\ q_2 = \sum_{i=1}^{m} p_{i2},\ \ldots, q_n = \sum_{i=1}^{m} p_{in};\ p_1 = \sum_{i=1}^{n} p_{1i},\ p_2 = \sum_{i=1}^{n} p_{2i},\ \ldots,\ p_m = \sum_{i=1}^{n} p_{mi}.$

Події $(X = x_i,\ Y = y_k),\ i = 1,\ 2,\ \ldots, m;\ k = 1,\ 2,\ \ldots, n$ утворюють повну групу, тому сума ймовірностей наведеної таблиці дорівнює 1. Додаючи ймовірності стовпчиків і рядків, одержують закони розподілу компонентів X та Y:

X	x_1	x_2	...	x_m
P	p_1	p_2	...	p_m

Y	y_1	y_2	...	y_n
P	q_1	q_2	...	q_n

2.4.3.2. Неперервні двовимірні випадкові величини

Означення. Двовимірну випадкову величину (X, Y) називають **неперервною,** якщо вона набуває значення в усіх точках деякої області площини

Oxy, що рівносильно існуванню такої функції $f(x, y)$, за якої функцію розподілу $F(x, y)$ можна подати у вигляді

$$F(x,\ y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,\ y)dxdy.$$

Означення. Функцію $f(x, y)$ називають **щільністю розподілу ймовірностей** *двовимірної випадкової величини* (X, Y).

Якщо функція $f(x, y)$ неперервна, то

$$f(x,\ y) = \frac{\partial^2 F(x,\ y)}{\partial x \partial y}. \qquad (2.38)$$

Властивості функції щільності розподілу ймовірностей

1. $f(x,\ y) \geq 0$ за всіх x, y.

2. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,\ y)dxdy = 1.$

3. Якщо D – будь-яка область у площині Oxy, то

$$P\left((X,\ Y) \in D\right) = \iint_{D} f(x,\ y)dxdy.$$

Знаючи щільність розподілу $f(x, y)$ двовимірної випадкової величини (X, Y), знаходять щільності розподілу $f_1(x)$ і $f_2(x)$ для її компонентів X та Y. Справді, функція розподілу випадкової величини X

$$F_1(x) = F(x,\ +\infty) = \int_{-\infty}^{x} dx \int_{-\infty}^{+\infty} f(x,\ y)dy.$$

Диференціюючи обидві частини рівності за x, отримаємо:

$$f_1(x) = \frac{\partial F_1(x)}{\partial x} = \int_{-\infty}^{+\infty} f(x,\ y)dy.$$

Аналогічно, $F_2(y) = F(+\infty, y) = \int_{-\infty}^{y} dy \int_{-\infty}^{+\infty} f(x,\ y)dx$ та

$$f_2(y) = \int_{-\infty}^{+\infty} f(x,\ y)dx.$$

Теорема. *Випадкові величини X та Y незалежні тоді і тільки тоді, коли функція розподілу пари* (X, Y) *дорівнює добутку функцій розподілу компонентів:*

$$F(x, y) = F_1(x) \cdot F_2(y).$$

Наслідок. Неперервні випадкові величини X та Y незалежні тоді і тільки тоді, коли щільність розподілу ймовірностей пари (X, Y) дорівнює добутку щільностей розподілу ймовірностей компонентів:

$$f(x, y) = f_1(x) \cdot f_2(y).$$

Приклад 2.31. Задано дискретну двовимірну випадкову величину:

Y \ X	19	21	23	25
19	0,032	0,118	0,102	0,078
24	0,098	0,072	0,088	0,042
29	0,062	0,048	0,122	0,138

Знайти закони розподілу компонент X таY.

Розв'язання. Додаючи стовпчики і рядки таблиці ймовірностей, отримують відповідні закони розподілу випадкових величин X та Y (рис. 2.23):

	X				
Y	19	21	23	25	**сума**
19	0,032	0,118	0,102	0,078	**0,33**
24	0,098	0,072	0,088	0,042	**0,3**
29	0,062	0,048	0,122	0,138	**0,37**
сума	**0,192**	**0,238**	**0,312**	**0,258**	**1**

Рис. 2.23. Побудова законів розподілу

X	19	21	23	25
P	0,192	0,238	0,312	0,258

Y	19	24	29
P	0,33	0,3	0,37

Приклад 2.32. Функцію розподілу двовимірної випадкової величини (X, Y) задано формулою:

$$F(x, y) = \begin{cases} (1 - e^{-4x})(1 - e^{-2y}), & x \geq 0, \ y \geq 0; \\ 0, & x < 0 \ \text{або} \ y < 0. \end{cases}$$

Знайти щільність розподілу ймовірностей $f(x, y)$.

Розв'язання. Використовуємо формулу (2.38): $f(x, y) = \dfrac{\partial^2 F(x, y)}{\partial x \partial y}$, тому обчислюємо

$$\frac{\partial F(x)}{\partial x} = \begin{cases} 4e^{-4x}(1-e^{-2y}),\ x \geq 0,\ y \geq 0; \\ 0,\ x < 0 \quad або \quad y < 0, \end{cases}$$

тоді

$$\frac{\partial^2 F(x,\ y)}{\partial x \partial y} = f(x,\ y) = \begin{cases} 8e^{-4x}e^{-2y},\ x \geq 0,\ y \geq 0; \\ 0,\ x < 0 \quad або \quad y < 0. \end{cases}$$

Приклад 2.33. За відомою щільністю розподілу двовимірної випадкової величини $(X,\ Y)$, яка задана формулою $f(x,\ y) = \dfrac{20}{\pi^2(16+x^2)(25+y^2)}$, знайти її функцію розподілу $F(x, y)$.

Розв'язання. Обчислення проводимо згідно з формулою

$$F(x,\ y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x,\ y)dxdy.$$

Отже, $F(x,\ y) = \displaystyle\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} \frac{20}{\pi^2(16+x^2)(25+y^2)}dxdy = \frac{20}{\pi^2}\int\limits_{-\infty}^{x}\frac{dx}{4^2+x^2}\int\limits_{-\infty}^{y}\frac{dy}{5^2+y^2} =$

$$= \frac{20}{\pi^2}\lim_{a\to-\infty}\left(\frac{1}{4}\operatorname{arctg}\frac{x}{4}\right)\Bigg|_a^x \cdot \lim_{b\to-\infty}\left(\frac{1}{5}\operatorname{arctg}\frac{y}{5}\right)\Bigg|_b^y =$$

$$= \frac{1}{\pi^2}\left(\operatorname{arctg}\frac{x}{4} - \lim_{a\to-\infty}\operatorname{arctg}\frac{a}{4}\right)\left(\operatorname{arctg}\frac{y}{5} - \lim_{b\to-\infty}\operatorname{arctg}\frac{b}{5}\right) =$$

$$= \frac{1}{\pi^2}\left(\operatorname{arctg}\frac{x}{4} + \frac{\pi}{2}\right)\left(\operatorname{arctg}\frac{y}{5} + \frac{\pi}{2}\right) = \left(\frac{1}{\pi}\operatorname{arctg}\frac{x}{4} + \frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg}\frac{y}{5} + \frac{1}{2}\right).$$

Таким чином, функція розподілу має вигляд

$$F(x, y) = \left(\frac{1}{\pi}\operatorname{arctg}\frac{x}{4} + \frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg}\frac{y}{5} + \frac{1}{2}\right).$$

2.4.3.3. Числові характеристики двовимірної випадкової величини. Коефіцієнт кореляції та його властивості

Нехай (X, Y) – дискретна двовимірна випадкова величина, для якої закон розподілу задано таблицею розподілу. Тоді, додаючи стовпчики і рядки таблиці

ймовірностей, знаходять закони розподілу компонентів X та Y. Відповідно обчислюють їх числові характеристики.

Згідно з даними прикладу 2.31 можемо визначити числові характеристики кожного компонента за відомими формулами для одновимірної дискретної випадкової величини (рис. 2.24):

C17 f_x =(C10)^2*C11+(D10)^2*D11+(E10)^2*E11+(F10)^2*F11-(C16)^2

	A	B	C	D	E	F	G	H	I
1									
2			X						
3		Y	19	21	23	25	сума		
4		19	0,032	0,118	0,102	0,078	0,33		
5		24	0,098	0,072	0,088	0,042	0,3		
6		29	0,062	0,048	0,122	0,138	0,37		
7		сума	0,192	0,238	0,312	0,258	1		
8									
9									
10		**X**	**19**	**21**	**23**	**25**			
11		**P**	**0,192**	**0,238**	**0,312**	**0,258**			
12									
13		**Y**	**19**	**24**	**29**				
14		**P**	**0,33**	**0,3**	**0,37**				
15									
16		M(X)	22,272						
17		D(X)	4,526016						
18		σ(X)	2,127444						
19									
20		M(Y)	24,2						
21		D(Y)	17,46						
22		σ(Y)	4,178516						
23									

Рис. 2.24. Обчислення числових характеристик компонент розподілу

Зауваження. На рис. 2.24. показано побудову формули для обчислення дисперсії згідно з її аналітичним виглядом.

Нехай (X, Y) – неперервна двовимірна випадкова величина з щільністю розподілу $f(x, y)$. Якщо $f_1(x)$ і $f_2(y)$ – щільності розподілу її компонентів X та Y, то за означенням математичного сподівання одновимірної випадкової величини

$$M(X) = \int_{-\infty}^{+\infty} x f_1(x) dx, \quad M(Y) = \int_{-\infty}^{+\infty} y f_2(y) dy .$$

Підставивши $f_1(x)=\int\limits_{-\infty}^{+\infty} f(x,y)dy, \quad f_2(y)=\int\limits_{-\infty}^{+\infty} f(x,y)dx$ отримаємо:

$$M(X)=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xf(x,y)dxdy, \quad M(Y)=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} yf(x,y)dxdy.$$

Якщо $f(x,y)$ набуває значення в деякій області D площини Oxy, то

$$M(X)=\iint\limits_{D} xf(x,y)dxdy, \quad M(Y)=\iint\limits_{D} yf(x,y)dxdy.$$

***Означення*. Коваріацією** *(кореляційним моментом)* випадкових величин X та Y називається математичне сподівання добутку різниць випадкових величин і їх математичних сподівань:

$$K(X,Y)=\operatorname{cov}(X,Y)=M\left(\left(X-M(X)\right)\left(Y-M(Y)\right)\right).$$

Властивості коваріації

1. $K(X,Y)=K(Y,X)$.
2. $K(X,Y)=D(X)$.
3. Для довільних випадкових величин $X, \quad Y$ виконується рівність

$$D(X+Y)=D(X)+D(Y)+2K(X,Y).$$

Доведення.

$$D(X+Y)=M\left(X+Y-M(X+Y)\right)^2=M\left(\left(X-M(X)\right)+\left(Y-M(Y)\right)\right)^2=$$

$$=M\left(X-M(X)^2\right)+2M(X-M(X))(Y-M(Y))+M\left(Y-M(Y)^2\right)=$$

$$=D(X)+2K(X,Y)+D(Y).$$

4) Коваріація випадкових величин дорівнює різниці математичного сподівання добутку цих випадкових величин і добутку їхніх математичних сподівань

$$K(X,Y)=M(XY)-M(X)M(Y).$$

Доведення.

$$K(X,Y)=M\left(\left(X-M(X)\right)\left(Y-M(Y)\right)\right)=$$

$$=M\left(X\cdot Y-X\cdot M(Y)-Y\cdot M(X)+M(X)\cdot M(Y)\right)=$$

$$= M\left(X \cdot Y\right) - M(X) \cdot M(Y) - M(Y) \cdot M(X) + M(X) \cdot M(Y) = M(X \cdot Y) - M(X)M(Y).$$

Отже, після розкриття дужок отримали рівність, яку використовують для обчислення коваріації.

$$K(X, Y) = M(XY) - M(X)M(Y).$$

З останньої формули випливає, що для дискретної двовимірної випадкової величини (X, Y)

$$K(X, Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_{ij} - \sum_{i=1}^{m} x_i p_i \sum_{j=1}^{n} y_j q_j$$

та для неперервної двовимірної випадкової величини $\left(X, Y\right)$:

$$K(X, Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x, y)dxdy - \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x, y)dxdy \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x, y)dxdy.$$

Якщо $f(x, y)$ набуває значення в деякій області D площини Oxy, то

$$K(X,\ Y) = \iint_{D} xyf(x, y)dxdy - \iint_{D} xf(x, y)dxdy \iint_{D} yf(x, y)dxdy.$$

Якщо випадкові величини X, Y незалежні, то їхня коваріація дорівнює нулю.

Це твердження слідує з формули обчислення та з того, що математичне сподівання добутку незалежних випадкових величин дорівнює добутку їхніх математичних сподівань.

Означення. **Дисперсією** *(дисперсійною матрицею) двовимірної випадкової величини* (X, Y) *називається сукупність чотирьох чисел* d_{ik}, $(d_{ik} = d_{ki})$, яка є матрицею другого порядку:

$$D\left(X, Y\right) = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}.$$

Тоді, у разі двовимірної неперервної випадкової величини для елементів матриці

$$d_{11} = D(X) = \int_{-\infty}^{+\infty} \left(x - M(X)\right)^2 f_1(x)dx = \int_{-\infty}^{+\infty} x^2 f_1(x)dx - \left(M(X)\right)^2,$$

$$d_{22} = D(Y) = \int_{-\infty}^{+\infty} \left(y - M(Y)\right)^2 f_1(y)dy = \int_{-\infty}^{+\infty} y^2 f_1(y)dy - \left(M(Y)\right)^2,$$

$$d_{12} = d_{21} = K(X, Y),$$

або

$$D(X)=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\big(x-M(X)\big)^2 f(x,y)dxdy=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}x^2 f(x,y)dxdy-\big(M(X)\big)^2,$$

$$D(Y)=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\big(y-M(Y)\big)^2 f(x,y)dxdy=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}y^2 f(x,y)dxdy-\big(M(Y)\big)^2.$$

Якщо $f(x, y)$ набуває значення в деякій області D площини O_{xy}, то

$$D(X)=\iint\limits_{D}\big(x-M(X)\big)^2 f(x, y)dxdy=\iint\limits_{D}x^2 f(x, y)dxdy-\big(M(X)\big)^2,$$

$$D(Y)=\iint\limits_{D}\big(y-M(Y)\big)^2 f(x, y)dxdy=\iint\limits_{D}y^2 f(x, y)dxdy-\big(M(Y)\big)^2.$$

Точка $\big(M(X), M(Y)\big)$ називається ***центром розсіювання*** двовимірної випадкової величини.

Якщо випадкові величини X та Y незалежні, то $K(X,Y)=0$, дисперсійна матриця є діагональною матрицею:

$$D\big(X,Y\big)=\begin{pmatrix} D(X) & 0 \\ 0 & D(Y) \end{pmatrix}.$$

Звідси випливає, що коваріація є характеристикою *зв'язку* між величинами системи $\big(X,Y\big)$. Коваріація описує не тільки зв'язок між X та Y, а також і їх розсіювання. Дійсно, якщо принаймні одна з величин мало відрізняється від свого математичного сподівання, то коваріація буде малою, як би не були зв'язані між собою випадкові величини X та Y.

Істотний недолік коваріації – те, що її розмірність збігається з добутком розмірностей випадкових величин, тому замість коваріації використовують безрозмірну величину – *коефіцієнт кореляції*, який характеризує *тісноту зв'язку* між величинами X та Y.

Означення. **Коефіцієнтом кореляції** між випадковими величинами X і Y називається число

$$\rho=\rho(X,Y)=\frac{K(X,Y)}{\sigma(X)\sigma(Y)}.$$

Властивості коефіцієнта кореляції

Теорема 1. *Для будь-яких випадкових величин* $-1 \le \rho(X,Y) \le 1.$

Теорема 2. *Коефіцієнт кореляції між випадковими величинами* X *і* Y *дорівнює* ± 1 *тоді і тільки тоді, коли* X *і* Y *зв'язані лінійною залежністю* $Y = aX + b,$ *причому* $\rho(X,Y) = 1$ *за* $a > 0,$ $\rho(X,Y) = -1$ *за* $a < 0.$

Теорема 3. *Якщо випадкові величини* X *та* Y *незалежні, то* $\rho(X,Y) = 0.$

Наслідок. Якщо $\rho(X,Y) \neq 0,$ то X та Y залежні випадкові величини.

***Означення*.** Випадкові величини X та Y, для яких $\rho(X,Y) = 0,$ називаються **некорельованими,** для яких $\rho(X,Y) \neq 0$ називаються **корельованими**.

Зауваження. З некорельованості двох випадкових величин, взагалі кажучи, не випливає їх незалежності. Для нормально розподілених випадкових величин некорельованість рівносильна незалежності.

Коефіцієнт кореляції служить для оцінювання тісноти лінійного зв'язку між випадковими величинами X та Y: що ближче модуль коефіцієнта кореляції до одиниці, то зв'язок сильніший, що ближче модуль коефіцієнта кореляції до нуля, то зв'язок слабший.

Приклад 2.34. Неперервна двовимірна випадкова величина (X, Y) задана щільністю розподілу $f(x,y) = \begin{cases} Axy, (x,y) \in D; \\ 0, (x,y) \notin D, \end{cases}$ де область D обмежена прямими $x + y - 1 = 0,\ x = 0,\ y = 0$ (рис. 2.25).

Знайти: $A,\ M(X),\ M(Y),\ D(X),\ D(Y),\ K(X,Y),\ \rho(X,Y).$

Розв'язання.

Оскільки $f(x,y)$ набуває значення в області D, то

$$\iint_D f(x,y)dxdy = 1.$$

Одержимо $\iint_D Axydxdy = A\int_0^1 xdx \int_0^{1-x} ydy = A\int_0^1 x\frac{y^2}{2}\Big|_0^{1-x} dx = \frac{A}{2}\int_0^1 x(1-x)^2 dx =$

$$= \frac{A}{2}\int_0^1 (x - 2x^2 + x^3)dx = \frac{A}{2}\left(\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4} \right)\Bigg|_0^1 = \frac{A}{24};\ \frac{A}{24} = 1 \text{ і звідси } A = 24.$$

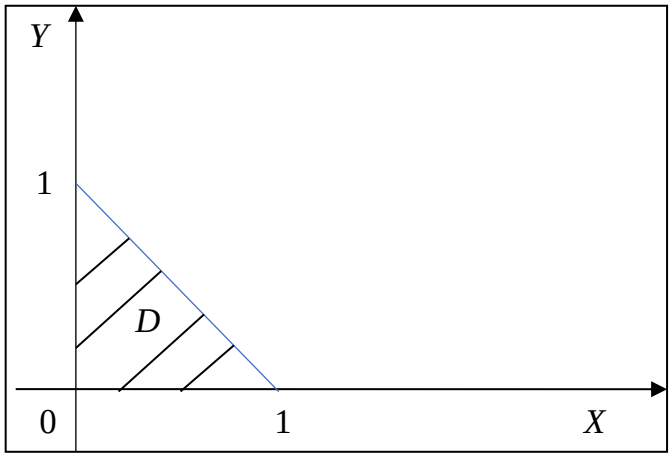


Рис. 2.25. Задання області D

$$f(x, y) = \begin{cases} 24\,xy,\ (x, y) \in D; \\ 0,\ (x, y) \notin D. \end{cases}$$

Визначимо числові характеристики:

$$M(X) = \iint_D xf(x, y)dxdy = 24\int_0^1 x^2 dx \int_0^{1-x} ydy = 24\int_0^1 x^2 \frac{y^2}{2}\Bigg|_0^{1-x} = 12\int_0^1 x^2(1-x)^2 dx =$$

$$= 12\int_0^1 (x^2 - 2x^3 + x^4)dx = 12\left(\frac{x^3}{3} - \frac{x^4}{2} + \frac{x^5}{5} \right)\Bigg|_0^1 = \frac{2}{5};$$

$$M(Y) = \iint_D yf(x, y)dxdy = 24\int_0^x y^2 dy \int_0^{1-y} xdx = \frac{2}{5};$$

$$M(X^2) = \iint_D x^2 f(x, y)dxdy = 24\int_0^1 x^3 dx \int_0^{1-x} ydy = 24\int_0^1 x^3 \frac{y^2}{2}\Bigg|_0^{1-x} dx = 12\int_0^1 x^3(1-x)^2 dx =$$

$$= 12\int_0^1 (x^3 - 2x^4 + x^5)dx = 12\left(\frac{x^4}{4} - \frac{2x^5}{5} + \frac{x^6}{6} \right)\Bigg|_0^1 = \frac{1}{5};$$

$$M(Y^2) = \iint_D y^2 f(x, y)dxdy = 24\int_0^1 y^3 dy \int_0^{1-y} xdx = \frac{1}{5};$$

$$D(X) = M(X^2) - (M(X))^2 = \frac{1}{5} - \frac{4}{25} = \frac{1}{25};\ \sigma(X) = \frac{1}{5};$$

$$D(Y)=M(Y^2)-(M(Y))^2=\frac{1}{5}-\frac{4}{25}=\frac{1}{25};\ \sigma(Y)=\frac{1}{5};$$

$$M(XY)=\iint\limits_D xyf(x,y)dxdy=24\int\limits_0^1 x^2dx\int\limits_0^{1-x}y^2dy=24\int\limits_0^1 x^2\frac{y^3}{3}\Bigg|_0^{1-x}dx=8\int\limits_0^1 x^2(1-x)^3dx=$$

$$=8\int\limits_0^1(x^2-3x^3+3x^4-x^5)dx=8\left(\frac{x^3}{3}-\frac{3x^4}{4}+\frac{3x^5}{5}-\frac{x^6}{6}\right)\Bigg|_0^1=\frac{2}{15};$$

$$K(X,Y)=M(XY)-M(X)M(Y)=\frac{2}{15}-\frac{2}{5}\cdot\frac{2}{5}=-\frac{2}{75};$$

$$\rho(X,Y)=\frac{K(X,Y)}{\sigma(X)\sigma(Y)}=-\frac{2}{3}.$$

2.4.3.4. Лінійна регресія

Означення. Пряма $y=\alpha_1x+\alpha_0$ називається **прямою регресії** (або ***прямою середньоквадратичної регресії***) Y на X, якщо коефіцієнти α_1 і α_0 вибрано так, щоб середнє квадратичне відхилення $\alpha_1X+\alpha_0$ від Y було мінімальним:

$$M(Y-\alpha_1X-\alpha_0)^2=\min.$$

Аналогічно визначається пряма регресії X на Y.

Тоді коефіцієнти α_1 і α_0 набувають таких значень:

$$\alpha_1=\rho\frac{\sigma_y}{\sigma_x},\ \alpha_0=M(Y)-\alpha_1M(X).$$

Коефіцієнт α_1 називається коефіцієнтом регресії Y на X. Рівняння прямої регресії Y на X має вигляд

$$y=\rho\frac{\sigma_y}{\sigma_x}x+M(Y)-\rho\frac{\sigma_y}{\sigma_x}M(X)$$

або

$$y-M(Y)=\rho\frac{\sigma_y}{\sigma_x}(x-M(X)).$$

Аналогічно пряма регресії X на Y має вигляд

$$x-M(X)=\rho\frac{\sigma_x}{\sigma_y}(y-M(Y)).$$

Обидві прямі регресії проходять через точку $(M(X),M(Y))$ – центр розсіювання двовимірної випадкової величини. Прямі регресії збігаються тоді і тільки тоді, коли $\rho=\pm1$.

Величина $(1-\rho^2)\sigma_y^2$ називається залишковою регресією Y на X. Вона характеризує величину похибки у разі заміни Y на $\alpha_1 X+\alpha_0$.

Приклад 2.35. За даними прикладу 31 знайти рівняння прямих регресії Y на X ,також X на Y.

Розв'язання. Відомо, що $M(X)=22,27;\quad D(X)=4,53;\quad \sigma(X)=2,13;$

$$M(Y)=24,2;\quad D(Y)=17,46;\quad \sigma(Y)=4,179.$$

Обчислимо коефіцієнт коваріації та відповідно, коефіцієнт кореляції (рис. 2.26).

J3 =B4*(C3*C4+D3*D4+E3*E4+F3*F4)+B5*(C3*C5+D3*D5+E3*E5+F3*F5)+B6*(C3*C6+D3*D6+E3*E6+F3*F6)

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2			X									
3		Y	19	21	23	25	сума		M(XY)	539,828		
4		19	0,032	0,118	0,102	0,078	0,33		K(X,Y)=M(X)M(Y)		0,8456	
5		24	0,098	0,072	0,088	0,042	0,3		ρ(X,Y)		0,095123	
6		29	0,062	0,048	0,122	0,138	0,37					
7		сума	0,192	0,238	0,312	0,258	1					
8												

Рис. 2.26. Вигляд обчислень з побудовою формули

$$K(X,Y)=M(XY)-M(X)M(Y)=0,8456;$$

$$\rho(X,Y)=\frac{K(X,Y)}{\sigma_x\sigma_y}=0,095;$$

$$\rho\cdot\frac{\sigma_y}{\sigma_x}=0,095\cdot\frac{4,179}{2,13}=0,1864;\quad \rho\frac{\sigma_x}{\sigma_y}=0,095\cdot\frac{2,13}{4,179}=0,0484.$$

Рівняння прямої регресії Y на X:

$$y-24,2=0,1864\,(x-22,27),\quad y=0,1864x+20,0396.$$

Рівняння прямої регресії X на Y:

$$x-22,27=0,0484\,(y-24,2)$$

або для її побудови (рис. 2.27) зручніше подати у вигляді залежності від змінної x.

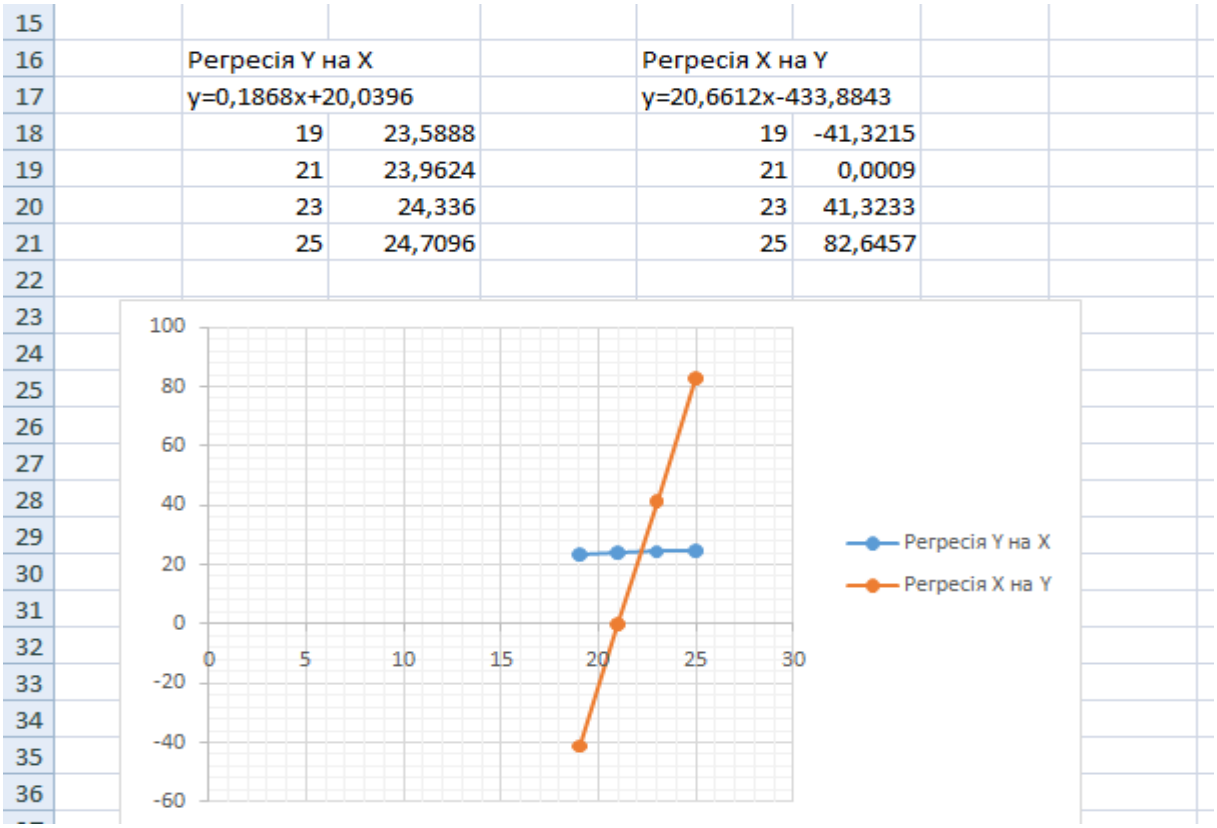


Рис. 2.27. Прямі регресії

2.4.4. Умовні закони розподілу

Поняття умовної ймовірності події розглядали в розділі 1. Імовірність події A за умови, що відбулась подія B, обчислюється за формулою (1.12), якщо $P(B) \neq 0$.

2.4.4.1. Умовні закони розподілу компонентів дискретної двовимірної випадкової величини

Нехай (X, Y) – дискретна двовимірна випадкова величина, закон розподілу якої задають таблицею. Припустимо, що в результаті випробування величина Y набула значення $Y = y_j$, при цьому X набуде одного із своїх можливих значень $x_1, x_2, ..., x_m$. Позначимо умовну ймовірність того, що X набуде значення x_i за умови, що $Y = y_j$ через $p(x_i \mid y_j)$.

Y \ X	x_1	x_2	...	x_m	$\sum$
y_1	p_{11}	p_{21}	…	p_{m1}	q_1
Y_2	p_{12}	p_{22}	…	p_{m2}	q_2
…	…	…	…	…	…
y_n	p_{1n}	p_{2n}	…	$p_{mò}$	q_n
$\sum$	p_1	p_2	…	p_m	1

***Означення*. Умовним законом розподілу** дискретної випадкової величини X за фіксованого значення випадкової величини $Y = y_j$ називають множину можливих значень випадкової величини X та відповідних їм умовних імовірностей, обчислених за фіксованого значення $Y = y_j$, тобто розподіл

X	x_1	x_2	…	x_m
$p(X \mid y_j)$	$p(x_2 \mid y_j)$	$p(x_2 \mid y_j)$	⋯	$p(x_m \mid y_j)$

Умовні ймовірності $p(x_i \mid y_j)$ обчислюють за формулою

$$p(x_i \mid y_j) = \frac{p_{ij}}{q_j} \ (i = 1,\ 2,\ \ldots,\ m,\ \ j = 1,\ 2,\ \ldots,\ n).$$

Аналогічно визначається умовний закон розподілу компоненти Y за $X = x_i$

Y	y_1	y_2	…	y_n
$p(Y \mid x_i)$	$p(y_1 \mid x_i)$	$p(y_2 \mid x_i)$	…	$p(y_n \mid x_i)$

$$p(y_j \mid x_i) = \frac{p_{ij}}{p_i} \ (i = 1,\ 2,\ \ldots,\ m,\ \ j = 1,\ 2,\ \ldots,\ n).$$

2.4.4.2. Умовні закони розподілу компонентів неперервної двовимірної випадкової величини

Нехай задана двовимірна неперервна випадкова величина (X, Y). Як і для двовимірної дискретної випадкової величини, для двовимірної неперервної випадкової величини розглядають умовні закони розподілу. Умовні закони

розподілу для неперервних випадкових величин X, Y, що утворюють вектор (X, Y), визначаються умовними щільностями ймовірностей.

***Означення*. Умовною щільністю** $\varphi(x \mid y)$ *розподілу компоненти* X *за даного значення* $Y = y$ називається відношення щільності розподілу двовимірної випадкової величини $f(x, y)$ до щільності розподілу $f_2(y)$ компоненти Y:

$$\varphi(x \mid y) = \frac{f(x, y)}{f_2(y)}.$$

Аналогічно визначається *умовна щільність компоненти* Y *за* $X = x$:

$$\psi(y \mid x) = \frac{f(x, y)}{f_1(x)}.$$

Приклад 2.36. Підприємство виконує три великих замовлення щодня. Ймовірність появи браку через технічні неполадки під час виконання першого замовлення та зупинення процесу виробництва становить 0,15, другого – 0,1 та третього – 0,05. Потрібно:

1. Скласти закон розподілу системи (X, Y), де X – кількість зупинок виробництва під час виконання першого замовлення, Y – сумарна кількість зупинок.

2. Записати одновимірні закони розподілів кожного компонента.

3. Записати основні числові характеристики системи випадкових величин (X, Y).

4. Записати умовні закони розподілів $\{X \mid Y = 1\}, \{Y \mid X = 1\}$.

5. Обчислити відповідні умовні математичні сподівання для компонент системи (X, Y).

Розв'язання.

1. Випадкова величина X набуває можливі значення 0 і 1, а для Y – 0, 1, 2 та 3. Знайдемо ймовірність p_{ij} спільної появи всіх можливих значень x_i та y_j. Позначимо p_1 – ймовірність появи браку під час виконання першого замовлення

та зупинки процесу, p_2- другого, відповідно, p_3- третього Тоді ймовірності протилежних подій будуть відповідно, q_1, q_2, q_3.

За умовою задачі $p_1 = 0{,}15;\ p_2 = 0{,}1;\ p_3 = 0{,}05$ і, відповідно, $q_1 = 1 - 0{,}15 = 0{,}85;\ q_2 = 0{,}9;\ q_3 = 0{,}95$.

Знаходимо ймовірності подій:

$$p_{11} = p\{X = 0, Y = 0\} = q_1 q_2 q_3 \approx 0{,}727;$$

$$p_{12} = p\{X = 0, Y = 1\} = q_1(1 - q_2 q_3) \approx 0{,}123;$$

$$p_{13} = p\{X = 0, Y = 2\} = q_1 p_2 p_3 \approx 0{,}004;$$

$$p_{14} = p\{X = 0, Y = 3\} = 0- \text{ неможлива подія;}$$

$$p_{21} = p\{X = 1, Y = 0\} = 0- \text{ неможлива подія;}$$

$$p_{22} = p\{X = 1, Y = 1\} = p_1 q_2 q_3 \approx 0{,}128;$$

$$p_{23} = p\{X = 1, Y = 2\} = p_1(1 - q_2 q_3) \approx 0{,}028;$$

$$p_{24} = p\{X = 1, Y = 3\} = p_1 p_2 p_3 \approx 0{,}001.$$

Таблиця розподілу системи матиме вигляд:

Y	X	
	0	1
0	0,727	0
1	0,123	0,128
2	0,004	0,028
3	0	0,001

Умова нормування системи виконана.

2. Ряди розподілу компонент X та Y знайдемо як суму ймовірностей відповідно за рядками й стовпчиками:

X	0	1
p	0,854	0,157

Y	0	1	2	3
q	0,727	0,252	0,032	0,001

3. Обчислюємо основні числові характеристики випадкових величин X та Y, користуючись рядами розподілу цих величин за відомими формулами або із використанням програмного документа (рис. 2.28).

$$M(X) = 0 \cdot 0{,}854 + 1 \cdot 0{,}157 = 0{,}157;$$

$$D(X) = M(X^2) - (M(X))^2 = 0 \cdot 0{,}854 + 1 \cdot 0{,}157 - 0{,}157^2 = 0{,}157 - 0{,}246 = 0{,}132;$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0{,}132} = 0{,}364.$$

Отже, отримані результати:

$$M(X) = 0{,}157;\ D(X) = 0{,}132;\ \sigma(X) = 0{,}364.$$

Аналогічно проведені обчислення для компонента Y:

$$M(Y) = 0{,}318;\ D(Y) = 0{,}288;\ \sigma(Y) = 0{,}535.$$

M10 fx =СУММ(B2:C5)

	A	B	C	D	E	F	G	H	I
1	Y\X	0	1	q	yq	y^2q	M(Y)	D(Y)	σ(Y)
2	0	0,72675	0	0,72675	0	0	0,31825	0,285967	0,534759
3	1	0,12325	0,12825	0,2515	0,2515	0,2515			
4	2	0,00425	0,028	0,03225	0,0645	0,129			
5	3	0	0,00075	0,00075	0,00225	0,00675			
6	p	0,85425	0,157						
7	xp	0	0,157						
8	x^2p	0	0,157						
9	M(X)	0,157							
10	D(X)	0,132351							
11	σ(X)	0,363801							

Рис. 2.28. Обчислення числових характеристик

Обчислюємо кореляційний момент випадкових величин X та Y за формулою $K(X,Y) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j p_{ij} - M(X) \cdot M(Y)$:

$$K(X,Y) = 0{,}128 + 2 \cdot 0{,}028 + 3 \cdot 0{,}001 - 0{,}157 \cdot 0{,}318 = 0{,}1371.$$

Тоді коефіцієнт кореляції $\rho(X,Y)$ знаходимо за формулою

$$\rho(X,Y) = \frac{K(X,Y)}{\sigma_x \cdot \sigma_y} = \frac{0{,}1371}{0{,}364 \cdot 0{,}535} \approx 0{,}7.$$

Одержаний результат свідчить про достатньо щільну, яка наближається до лінійної, додатну кореляційну залежність між випадковими величинами X та Y.

4. Знайдемо умовні закони розподілів. У пункті 2 прикладу обчислені ймовірності значень $X=1, \quad Y=1$:

$$P(X=1)=0{,}157; \quad P(Y=1)=0{,}252.$$

Отже, відповідні умовні ймовірності знаходимо так:

$$P\{X=0\,|\,Y=1\}=\frac{P\{X=0,Y=1\}}{P\{Y=1\}}=\frac{p_{12}}{q_2}=\frac{0{,}123}{0{,}252}=0{,}488;$$

$$P\{X=1\,|\,Y=1\}=\frac{P\{X=1,Y=1\}}{P\{Y=1\}}=\frac{p_{22}}{q_2}=\frac{0{,}128}{0{,}252}=0{,}508.$$

Отримані результати записуємо у вигляді таблиці:

X	0	1
$P\{X\,\|\,Y=1\}$	0,488	0,508

Аналогічно обчислюємо умовні ймовірності можливих значень $y_1=0,\ y_2=1,\ y_3=2$ для компоненти Y:

$$P\{Y=0\,|\,X=1\}=\frac{P\{X=1,Y=0\}}{P\{X=1\}}=\frac{p_{21}}{p_2}=\frac{0}{0{,}157}=0;$$

$$P\{Y=1\,|\,X=1\}=\frac{P\{X=1,Y=1\}}{P\{X=1\}}=\frac{p_{22}}{p_2}=\frac{0{,}128}{0{,}157}=0{,}815;$$

$$P\{Y=2\,|\,X=1\}=\frac{P\{X=1,Y=2\}}{P\{X=1\}}=\frac{p_{23}}{p_2}=\frac{0{,}028}{0{,}157}=0{,}178;$$

$$P\{Y=3\,|\,X=1\}=\frac{P\{X=1,Y=3\}}{P\{X=1\}}=\frac{p_{24}}{p_2}=\frac{0{,}001}{0{,}157}=0{,}005;$$

Отримані результати записуємо у вигляді таблиці:

Y	0	1	2	3
$P\{Y\,\|\,X=1\}$	0	0,815	0,178	0,005

5. Обчислюємо умовні математичні сподівання, користуючись одержаними рядами розподілу:

$$M\{X \mid Y=1\} = \sum_{i=1}^{n} x_i P\{X = x_i \mid Y=1\} = 0 \cdot 0,488 + 1 \cdot 0,508 = 0,508;$$

$$M\{Y \mid X=1\} = \sum_{j=1}^{n} y_j P\{Y = y_j \mid X=1\} = 0 \cdot 0 + 1 \cdot 0,815 + 2 \cdot 0,178 + 3 \cdot 0,005 = 1,186.$$

2.4.5. Приклади деяких важливих для практики двовимірних розподілів випадкових величин

Двовимірний біномний розподіл. Двовимірна випадкова величина (X_1, X_2) має *двовимірний біномний розподіл* з параметрами (n, p_1, p_2), $0 < p_i < 1,\ p_1 + p_2 = 1,$ якщо

$$P(X_1 = k_1,\ X_2 = k_2) = \frac{n!}{k_1! \cdot k_2!} p_1^{k_1} \cdot p_2^{k_2},\ k_1 + k_2 = n.$$

Двовимірний біномний розподіл – окремий випадок поліномного розподілу.

Поліномний розподіл – багатовимірний аналог біномного розподілу.

Двовимірний рівномірний розподіл в області *D*. *Двовимірний рівномірний розподіл* в області *D* визначається щільністю

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{1}{S_D}, & (x, y) \in D; \\ 0, & (x, y) \notin D, \end{cases}$$

де S_D – площа області *D*, тобто щільність сумісного розподілу зберігає стале значення в області *D*, якій належать усі можливі значення випадкового вектора (X, Y).

Двовимірний експоненціальний розподіл. *Означення*. Двовимірна випадкова величина (X, Y) називається *експоненціально розподіленою*, якщо її функція розподілу визначається формулою:

$$F_{X,Y}(x, y) = \begin{cases} 0,\ x < 0 \ \ або \ \ y < 0; \\ 1 - e^{-(\lambda_1 + \lambda_{12})x} - e^{-(\lambda_2 + \lambda_{12})y} + e^{-\lambda_1 x - \lambda_2 y - \lambda_{12} \max\{x, y\}},\ x \geq 0,\ y \geq 0. \end{cases}$$

Двовимірний нормальний розподіл. Нехай X та Y − нормально розподілені і незалежні випадкові величини, задані щільностями розподілу ймовірностей

$$f_1(x)=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{(x-a_1)^2}{2\sigma_1^2}};\qquad f_2(y)=\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{(y-a_2)^2}{2\sigma_2^2}}.$$

Тоді на основі теореми множення щільностей маємо:

$$f(x,y)=\frac{1}{\sigma_1\sigma_2\cdot 2\pi}e^{-\frac{1}{2}\left[\frac{(x-a_1)^2}{\sigma_1^2}+\frac{(y-a_2)^2}{\sigma_2^2}\right]}.$$

Якщо центр розсіювання збігається з початком координат, тобто $a_1=a_2=0,$ тоді маємо канонічну форму двовимірного нормального розподілу:

$$f(x,y)=\frac{1}{\sigma_1\sigma_2\cdot 2\pi}e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_1^2}+\frac{y^2}{\sigma_2^2}\right)}.$$

Фіксованому значенню $f(x,y)=C$ відповідає деяке стале значення показника степеня $\frac{x^2}{\sigma_1^2}+\frac{y^2}{\sigma_2^2}=k^2,$ де k − const, звідки отримаємо рівняння

$$\frac{x^2}{(k\sigma_1)^2}+\frac{y^2}{(k\sigma_2)^2}=1,$$

яке називають рівнянням еліпса однакової щільності або еліпсом розсіювання. Осі симетрії називають головними осями розсіювання.

Нехай X та Y − *залежні нормально розподілені* випадкові величини. Тоді

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\;e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a_1)^2}{\sigma_1^2}-2\rho\frac{(x-a_1)(y-a_2)}{\sigma_1\sigma_2}+\frac{(y-a_2)^2}{\sigma_2^2}\right]},$$

де $\rho=\rho(x,y)$ − коефіцієнт кореляції.

Рівняння

$$\frac{(x-a_1)^2}{\sigma_1^2}-2\rho\frac{(x-a_1)(y-a_2)}{\sigma_1\sigma_2}+\frac{(y-a_2)^2}{\sigma_2^2}=k^2$$

є рівнянням еліпса з центром у точці $(a_1,a_2),$ а осі симетрії утворюють з віссю Ox кути, які визначаються з рівняння $\operatorname{tg}2\alpha=\frac{2\rho\sigma_1\sigma_2}{\sigma_1^2-\sigma_2^2}.$

Орієнтація еліпсів розсіювання відносно координатних осей перебуває у прямій залежності від значення коефіцієнта кореляції. Якщо X та Y некорельовані $(\rho = 0)$, то головні осі розсіювання паралельні осям координат.

Щільність розподілу для некорельованих нормальних випадкових величин

$$f(x, y) = \frac{1}{\sigma_1 \sigma_2 \cdot 2\pi} \; e^{-\frac{1}{2}\left[\frac{(x-a_1)^2}{\sigma_1^2} + \frac{(y-a_2)^2}{\sigma_2^2}\right]},$$

тобто така, як і для незалежних нормально розподілених випадкових величин. Це означає, що для них поняття некорельованості та незалежності еквівалентні.

Умовні закони нормального розподілу

$$f(y \mid x) = \frac{f(x, y)}{\int\limits_{-\infty}^{\infty} f(x, y)dy} = \frac{1}{\sigma_2 \cdot \sqrt{2\pi}\sqrt{1-\rho^2}} \; e^{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[y-a_2-\rho\frac{\sigma_2}{\sigma_1}(x-a_1)\right]^2};$$

$$f(x \mid y) = \frac{f(x, y)}{\int\limits_{-\infty}^{\infty} f(x, y)dx} = \frac{1}{\sigma_1 \cdot \sqrt{2\pi}\sqrt{1-\rho^2}} \; e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x-a_1-\rho\frac{\sigma_1}{\sigma_2}(y-a_2)\right]^2}.$$

Легко бачити, що ці вирази є щільностями нормального розподілу з математичними сподіваннями

$$M[Y \mid X = x] = a_2 + \rho\frac{\sigma_2}{\sigma_1}(x - a_1) = y(x);$$

$$M[X \mid Y = y] = a_1 + \rho\frac{\sigma_1}{\sigma_2}(y - a_2) = x(y),$$

і середніми квадратичними відхиленнями

$$\sigma(y \mid x) = \sigma_2\sqrt{1-\rho^2}, \;\; \sigma(x \mid y) = \sigma_1\sqrt{1-\rho^2}.$$

Завдання для самоконтролю

1. Функція одного випадкового аргумента, її розподіл та числові характеристики, навести приклади.
2. Пояснити закон розподілу лінійної функції.
3. Функція двох випадкових аргументів, поняття згортки та його

застосування до обчислення щільності розподілу суми $Z = X + Y$.

4. Скласти композицію законів розподілу двох дискретних випадкових величин.

5. Записати закон розподілу суми незалежних нормально розподілених випадкових величин.

6. Багатовимірні випадкові величини, приклади.

7. Властивості функції розподілу двовимірної випадкової величини, незалежність випадкових величин.

8. Закон розподілу дискретної двовимірної випадкової величини, закони розподілу його компонентів.

9. Неперервні двовимірні випадкові величини, властивості функції щільності розподілу ймовірностей.

10. Обчислення основних числових характеристик законів розподілу компонентів дискретної та неперервної випадкової величини.

11. Поняття коваріації випадкових величин, її властивості.

12. Коефіцієнт кореляції, його властивості та формули обчислення.

13. Означення прямої регресії, її побудова.

14. Умовні закони розподілу компонентів дискретної двовимірної випадкової величини, формули обчислення його основних числових характеристик.

15. Умовні закони розподілу компонентів неперервної двовимірної випадкової величини, умовна щільність.

16. Випадкова величина X рівномірно розподілена в інтервалі $(-4, 4)$. Знайти щільність розподілу імовірності випадкової величини $Y = X^2$.

17. За законом розподілу дискретного двовимірного вектора знайти закон розподілу випадкових величин $Z = X + Y$, $W = X \cdot Y$.

Y \ X	0	1	3	$\sum$
-2	0,1	0	0,2	0,3
1	0,2	0,1	0,25	0,55
5	0	0,05	0,1	0,15
$\sum$	0,3	0,15	0,55	1

18. Двовимірна дискретна випадкова величина (X, Y) задана таблицею розподілу

Y \ X	-1	2	3	$\sum$
0	0,1	0,1	0,1	0,3
2	0,05	0,2	0,25	0,6
5	0,15	0	0,05	0,2
$\sum$	0,3	0,3	0,4	1

Знайти рівняння прямих регресії Y на X та X на Y.

19. Авіакомпанія виконує два рейси на добу. Ймовірність затримки першого рейсу через технічні причини дорівнює 0,1, а другого – 0,05. Потрібно:

1. Скласти закон розподілу системи (X, Y), де X – кількість затримок першого рейсу, Y – сумарна кількість затримок двох рейсів.

2. Записати одномірні закони розподілів кожної компоненти;

3. Записати основні числові характеристики системи випадкових величин (X, Y).

4. Записати умовні закони розподілів $\{X/_{Y=1}\}, \{Y/_{X=1}\}$.

5. Обчислити відповідні умовні математичні сподівання для компонент системи (X, Y).

20. На поліграфічному виробництві за одну зміну друкується по одному комплекту шкільних підручників на двох друкарських машинах. Імовірність

того, що перший комплект матиме чималу кількість браку та потребуватиме передруковування дорівнює 0,2, а для другого – 0,06. Потрібно:

1. Скласти закон розподілу системи (X, Y), де X – кількість комплектів для передруковування для першої друкарської машини (ймовірність передруковування дорівнює 0,2), Y – сумарна кількість комплектів для передруковування.

2. Написати одномірні закони розподілів кожного компонента.

3. Написати основні числові характеристики системи випадкових величин (X, Y).

4. Написати умовні закони розподілів $\{X|Y=1\}$, $\{Y|X=1\}$ та обчислити відповідні умовні математичні сподівання для компонент.

Зауваження. Випадкова величина X набуває можливі значення 0 і 1, а Y, відповідно, 0, 1, 2.

Варіанти завдань для індивідуальної роботи

Завдання1. Фірма планує відкрити ще одне кафе. З метою оптимального планування можливої кількості відвідувачів кафе протягом деякого часу аналітики фірми провели статистичне дослідження кількості відвідувачів і прибутку, який був отриманий. У таблиці представлений закон розподілу системи двох дискретних випадкових величин: X – кількість відвідувачів кафе за деякий час і Y – прибуток, який отримала фірма в ум. од. Вважати, що k – номер варіанту.

Y	X			
	k	$k+2$	$k+4$	$k+6$
k	$0{,}002(35-k)$	$0{,}002(k+40)$	$0{,}002(70-k)$	$0{,}002(k+20)$
$k+5$	$0{,}002(k+30)$	$0{,}002(55-k)$	$0{,}002(k+25)$	$0{,}002(40-k)$
$k+10$	$0{,}002(50-k)$	$0{,}002(k+5)$	$0{,}002(80-k)$	$0{,}002(k+50)$

Виконати наступні завдання:

1. Скласти закони розподілу одномірних випадкових величин X та Y.

2. Знайти математичне сподівання, дисперсію та середнє квадратичне відхилення випадкових величин X та Y.

3. Обчислити кореляційний момент $K(X,Y)$ і коефіцієнт кореляції $\rho(X,Y)$.

4. Побудувати умовний закон розподілу випадкової величини X за умови, що випадкова величина Y набуває значення $k+10$ та умовний закон розподілу випадкової величини Y за умови, що випадкова величина X набуває значення $k+2$.

5. Знайти умовні математичні сподівання $M\left(Y \mid X = k+2\right)$, $M\left(X \mid Y = k+10\right)$.

Завдання 2. Скласти закон розподілу дискретної двовимірної випадкової величини (навести приклад самостійно), описати всі його характеристики, зробити креслення ліній регресії. Побудувати для даного закону умовний закон розподілу однієї компоненти при зафіксованому значенні іншої та обчислити можливі числові характеристики.

Зауваження. Зручно будувати лінії регресії в Excel, виразивши для кожної прямої рівняння прямої у вигляді залежності y від x, задавши відповідну кількість значень незалежної змінної та обчисливши значення функції в цих точках.

Частина ІІ. Математична статистика

Розділ 3. Елементи математичної статистики

3.1. Елементи математичної статистики. Вибірковий метод

У теорії ймовірностей вивчають питання, пов’язані з випадковими подіями, їх імовірностями та числовими характеристиками. Але в більшості випадків, які трапляються на практиці, точне значення цих характеристик невідоме. Метою кожного наукового дослідження власне і є встановлення закономірностей явищ, які спостерігають, та використання їх у повсякденній практичній діяльності. З огляду на це постає питання про їх експериментальне визначення й узагальнення висновку у вигляді закону. Предметом математичної статистики є розроблення методів збору та оброблення статистичних даних для одержання наукових та практичних висновків. Звичайним є використання статистичних методів у техніці, медицині, економіці, соціології, політології.

3.1.1. Основні задачі математичної статистики

Математичною статистикою називається наука, яка займається розробкою методів відбору, опису та аналізу дослідних даних з метою вивчення закономірностей випадкових масових явищ.

У свою чергу, встановлення цих закономірностей ґрунтується на вивченні методами теорії ймовірностей статистичних даних – результатів досліду або спостережень.

Отже, математична статистика вивчає методи, які дають змогу за результатами випробувань робити певні ймовірнісні висновки.

Основні задачі математичної статистики:

1) вказати способи збору та групування (якщо даних дуже багато) статистичної інформації;

2) визначити закон розподілу випадкової величини або системи випадкових величин за статистичними даними;

3) оцінити невідомі параметри розподілу;

4) перевірити правдоподібність припущень про закон розподілу випадкової величини, про форму зв'язку між випадковими величинами або про значення оцінюваного параметра.

Можна сказати, що ***основне завдання математичної статистики*** – розроблення методів аналізу статистичних даних залежно від мети дослідження.

Методи математичної статистики ефективно використовують під час розв'язування багатьох наукових задач, задач щодо організації технологічного процесу, планування, управління та ціноутворення.

Методи математичної статистики дають можливість представити велику кількість результатів спостереження у компактному, зручному для означення вигляді. Вони дають можливість виділити суттєву інформацію із множинності спостережень, представивши її у вигляді невеликої кількості зведених показників. Якщо виявляється, що наявних даних недостатньо для розуміння суті явища та потрібне проведення додаткового експерименту, то методи математичної статистики дозволяють відповісти на питання, як такий експеримент поставити, щоб максимально зменшити роботу дослідника як із постановки експерименту, так і з подальшої обробки експериментальних даних.

Математична статистика виникла (XVII ст.) та почала розвиватись паралельно з теорією ймовірностей. Подальшим розвитком (кінець XIX – початок XX ст.) математична статистика зобов'язана П. Л. Чебишову, А. А. Маркову, О. М. Ляпунову, а також К. Гауссу, Ф. Гальтону, К. Пірсону й іншим вченим.

У XX ст. найбільший вклад у математичну статистику зробили В. І. Романовський, Е. Е. Слуцький, А. Н. Колмогоров, Стьюдент (псевдонім В. Госсета), Е. Пірсон, Ю. Нейман, А. Вальд, А. В. Скороход.

3.1.2. Генеральна та вибіркова сукупності

Кожен об'єкт, який спостерігають, має декілька ознак. Розглядаючи лише одну ознаку кожного об'єкта, припускають, що інші ознаки несуттєві, тобто множина об'єктів однорідна.

Означення. Множину однорідних об’єктів називають **статистичною сукупністю. Вибірковою сукупністю (вибіркою)** називають сукупність випадково взятих об’єктів із статистичної сукупності.

Генеральною називають сукупність об’єктів, з яких зроблено вибірку. ***Обсягом сукупності (вибіркової або генеральної)*** називають кількість об’єктів цієї сукупності.

Вибірки бувають ***повторні*** або ***безповторні***. Альтернатива вибірки – *перепис* – обстеження, що мають на меті дослідження кожного елемента сукупності (генеральної сукупності), яку вивчають.

Означення. **Повторною** називають вибірку, за якої відібраний об’єкт повертають до генеральної сукупності перед відбором іншого об’єкта.

Означення. Вибірку називають **безповторною**, якщо взятий об’єкт до генеральної сукупності не повертається перед відбором наступного об’єкта.

Вибірка має бути ***репрезентативною***, тобто *представницькою* – її дані мають правильно відображати відповідну ознаку генеральної сукупності, яку вивчають.

Проста випадкова вибірка здійснюється за допомогою *простого випадкового відбору* (кожний елемент із однаковою ймовірністю може потрапити до вибірки).

Джерела даних у статистиці: вибіркові обстеження, спеціально поставлені експерименти і дані, що є результатом повсякденної роботи у бізнесі. Джерела даних бувають *первинними і вторинними.*

Первинні дані збираються спеціально для статистичного дослідження. Для цих даних є відомості про методи збирання, точність даних тощо.

Важливими є питання збору статистичної інформації про процес експлуатації об'єктів. Ці відомості можна одержати в результаті експериментального дослідження шляхом спостереження за роботою об’єктів за умовами реальної експлуатації або при випробуваннях на безвідмовну роботу. Дані випробувань зазвичай не можуть повністю замінити експлуатаційні дані. Реальна ж експлуатація являє собою експеримент, який в лабораторних умовах

за своїми масштабами не може бути досягнутим. Однак і при реальній експлуатації далеко не завжди вдається одержати потрібну інформацію.

По-перше, дані реальної експлуатації завжди стосуються морально старіючих об’єктів. Конструкція та технологія виготовлення сучасних технічних об’єктів змінюється так швидко, що дані про експлуатацію, які збираються тривалий термін, стають непотрібними завдяки надходженню нових зразків техніки і необхідності налагодження повторного збирання інформації про надійність. В той же час основною метою збирання та аналізу статистичних даних про відмови є підвищення надійності об’єктів шляхом їх модернізації, доробок чи повної заміни на нові з кращими характеристиками надійності.

По-друге, дані реальної експлуатації звичайно є неповними або недостатньо точними.

Це пояснюється рядом причин:

- організаційними труднощами збирання та обробки відомостей;

- трудомісткістю досліджень;

- недостатньою чутливістю та точністю контролюючої апаратури, яка викликає хибні відмови;

- не завжди високою кваліфікацією виконавців.

В багатьох випадках ці причини не дозволяють одержати ймовірні характеристики.

По-третє, іноді важко здійснювати спостереження за роботою деяких об'єктів при їх реальній експлуатації у зв’язку з неможливістю своєчасного виявлення прихованих відмов.

Перелічені причини визначають необхідність широкого застосування випробувань виробів на безвідмовну роботу і моделювання процесу експлуатації (рис. 3.1).

При проведенні цих випробувань зазвичай вдається перебороти більшість перелічених вище труднощів одержання відомостей про роботу об’єктів.

Рис. 3.1. Планування збору даних

***Основні недоліки експериментальних досліджень на надійність*:**

- тривалий період проведення (наприклад, при випробуванні технічних деталей на зношення);

- відносна дорожнеча випробувань;

- великі витрати самих досліджуваних об’єктів;

- неможливість одержання інформації на етапі проектування.

Шлях лабораторних досліджень дає можливість проводити експеримент протягом достатньо короткого часу, багаторазово повторювати та видозмінювати його. Крім цього можна в якійсь мірі досліджувати поведінку майбутніх об’єктів. В процесі експерименту не потрібно витрачати значну кількість досліджуваних об’єктів. Однак при моделюванні завжди необхідна

вхідна інформація, яка одержується з реальної експлуатації. Разом з цим при експлуатації деяких об'єктів існує багато ситуацій, коли майже неможливо проводити натурні експерименти і моделювання є єдиним шляхом експериментального дослідження.

Процес експериментального дослідження надійності включає такі основні етапи:

1) встановлення вимог до якості результатів експерименту;

2) планування експерименту (випробувань);

3) формування вибірки, включаючи ідентифікацію кожного з призначених до її складу об'єктів;

4) проведення власне експерименту, тобто одержання його безпосередніх результатів;

5) обробку безпосередніх результатів експериментів і прийняття рішення.

Експеримент планують виходячи зі встановлених вимог до якості його результатів. Планування полягає у виборі плану, визначенні його параметрів і складанні програми дослідження. Зміст плану залежить від цілі експериментального дослідження (кількісна оцінка показників надійності або якісний контроль рівня надійності за альтернативною ознакою).

Вторинні дані – це дані, використовувані у статистиці, але зібрані з іншою метою. Наприклад, робочі записи про діяльність фірм, офіційні статистичні звіти, журнали обліку успішності та відвідування занять студентами тощо.

Способи відбору

1. Вибір, який не потребує розділення генеральної сукупності на частини: простий випадковий безповторний відбір та простий випадковий повторний відбір.

2. Вибір, за якого генеральна сукупність розділяється на частини (розшарований випадковий відбір): типовий, механічний та серійний відбори.

В економічних дослідженнях іноді використовують *комбінований відбір*. Наприклад, спочатку поділяють генеральну сукупність на серії однакового обсягу, навмання відбирають декілька серій та, нарешті, з кожної серії навмання беруть окремі об'єкти.

Найкращий спосіб здійснення простої вибірки – *використання випадкових вибіркових чисел*. Ці числа складаються з цифр від 0 до 9, генеруються випадково (імовірність розміщення від будь-якої цифри у будь-якому місці таблиці не більше й не менше імовірності розміщення будь-якої іншої цифри з десяти наведених цифр) і записуються до спеціальної таблиці. Використання таблиць випадкових чисел гарантує, що не буде скоєно систематичної помилки, тобто помилки, яка робить дані не репрезентативними.

Означення*. Вибіркою обсягу *n називають n незалежних випадкових величин $X_1,\ X_2,\ \ldots,\ X_n,$ кожна з яких – копія випадкової величини X з функцією розподілу $F(x)$.

Числа $x_1,\ x_2,\ \ldots,\ x_n$ – реалізація вибірки обсягу n.

3.1.3. Статистичний розподіл вибірки. Емпірична функція розподілу. Полігон і гістограма

Дані у статистиці, отримані за допомогою спеціальних досліджень або із звичайних робочих записів у бізнесі, надходять до дослідника у вигляді неорганізованої маси незалежно від того, чи є вони даними із вибіркової сукупності, чи даними з генеральної сукупності. Тому постає питання щодо оброблення і впорядкування даних.

Значення x_i вибірки називають ***варіантами*.** Послідовність варіант, розміщених у порядку зростання, називають *варіаційним рядом*. Якщо при цьому x_i повторюється n_i разів $(i = 1,\ 2,\ \ldots,\ k;\ n_1 + n_2 + \ldots + n_k = n),$ то число n_i називають ***абсолютною частотою*** варіанти x_i, а $\frac{n_i}{n}$ – ***відносною частотою*** варіанти x_i.

***Означення.* Статистичним розподілом** вибірки називають перелік варіант та відповідних частот або відносних частот. Статистичний закон розподілу зручно задавати таблицею, що встановлює зв'язок між значеннями випадкової величини та їх частотами:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Закон розподілу такий: *в даній серії випробувань випадкова величина* x_i *набуває значення* n_i *разів (i = 1, 2, …, k).*

Часто статистичний закон розподілу задають аналогічною таблицею, що встановлює зв'язок між значеннями випадкової величини та її відносними частотами. Цей закон називають статистичним рядом відносних частот.

x_i	x_1	x_2	…	x_k
$\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_k}{n}$

Зауваження. У теорії ймовірностей під розподілом дискретної випадкової величини розуміють відповідність можливих значень випадкової величини їх імовірностям, а в математичній статистиці під розподілом розуміють відповідність спостережених значень (варіантів) їх частотам (або відносними частотам).

***Означення*. Емпірична функція розподілу** має такий вигляд:

$$F_n^*(x) = \begin{cases} 0, & x \le x_1; \\ \sum\limits_{i=1}^{k} \frac{n_i}{n}, & x_i \le x < x_{i+1}; \\ 1, & x > x_k. \end{cases} \qquad (3.1)$$

Функція $F_n^*(x)$ має всі властивості функції розподілу $F(x)$ випадкової величини X.

1. $0 \le F_n^*(x) \le 1$.

2. $F_n^*(x)$ – неспадна функція.

3. $F_n^*(x)$ – неперервна зліва.

Функцію називають ***емпіричною функцією розподілу,*** оскільки її знаходять емпіричним (дослідним) способом.

Функцію розподілу $F(x)$ генеральної сукупності називають ***теоретичною функцією розподілу.***

Теоретична функція розподілу визначає ймовірність події $(X < x)$, *а емпірична функція* $F^*(x)$ *– відносну частоту цієї події.*

Із закону великих чисел, зокрема з теореми Бернуллі, випливає, що відносна частота події $(X < x)$ збігається з імовірністю цієї події, тобто

$$\lim_{n\to\infty} P\left(\left|F^*(x) - F(x)\right| < \varepsilon\right) = 1.$$

Іншими словами, для великих значень n емпірична функція розподілу наближено представляє теоретичну функцію розподілу генеральної сукупності.

***Означення*. Полігоном частот** вибірки називають ламану з вершинами в точках $(x_i,\ n_i)$, $(i = 1,\ 2,\ ...,\ k)$. **Полігоном відносних частот** вибірки називають ламану з вершинами в точках $\left(x_i,\ \frac{n_i}{n}\right)$, $(i = 1,\ 2,\ ...,\ k)$.

Якщо X – неперервна випадкова величина, то її статистичний розподіл також подають у вигляді таблиць. Для оцінювання $f(x)$ за вибіркою $x_1,\ x_2, ...,\ x_n$ розбивають інтервал на частинні інтервали довжини h, у таблицю записують середину i-го інтервалу, n_i – кількість елементів вибірки i-го інтервалу. Прямокутники з основами h і висотами $\frac{n_i}{nh}$ у прямокутній системі координат утворюють фігуру, яку називають ***гістограмою*** вибірки.

Приклад 3.1. Побудувати емпіричну функцію розподілу за розподілом вибірки

x_i	0,2	0,4	0,8	1,0	1,2	1,6
n_i	5	8	12	13	10	2

Розв'язання. Обсяг вибірки $n = 50$.

Шуканий розподіл відносних частот має вигляд

x_i	0,2	0,4	0,8	1,0	1,2	1,6
$\frac{n_i}{n}$	$\frac{5}{50}$	$\frac{8}{50}$	$\frac{12}{50}$	$\frac{13}{50}$	$\frac{10}{50}$	$\frac{2}{50}$

або остаточно

x_i	0,2	0,4	0,8	1,0	1,2	1,6
$\frac{n_i}{n}$	0,1	0,16	0,24	0,26	0,2	0,04

Тоді записуємо емпіричну функцію розподілу за формулою (3.1):

$$F^*(x)=\begin{cases} 0, & x\le 0,2; \\ \frac{5}{50}=0,1, & 0,2<x\le 0,4; \\ \frac{5+8}{50}=0,26, & 0,4<x\le 0,8; \\ \frac{5+8+12}{50}=0,5, & 0,8<x\le 1,0; \\ \frac{5+8+12+13}{50}=0,76, & 1,0<x\le 1,24; \\ \frac{5+8+12+13+10}{50}=0,96, & 1,2<x\le 1,6; \\ 1, & x>1,6. \end{cases}$$

3.1.4. Точкові оцінки невідомих параметрів розподілів

У багатьох випадках потрібно дослідити кількісну ознаку X генеральної сукупності, використовуючи результати вибірки. Дослідник має дані вибірки, одержані в результаті спостережень. Іноді з певних міркувань вдається встановити закон розподілу випадкової величини X. Тоді потрібно вміти оцінювати параметри цього розподілу.

***Означення*. Статистичною оцінкою** невідомого параметра випадкової величини X генеральної сукупності називають функцію від випадкових величин (результатів вибірки), що спостерігаються.

Нехай $X_1,\ X_2,\ ...,\ X_n$ – вибірка з генеральної сукупності X із функцією розподілу $F(x,\ \theta)$, яка має відомий функціональний вигляд, але залежить від невідомого параметра θ.

Означення. Оцінка $\hat{\theta} = h(X_1,\ X_2,\ ...,\ X_n)$ параметра θ називається ***незміщеною***, якщо $M\hat{\theta} = \theta$ (математичне сподівання дорівнює параметру, який оцінюється).

Якщо $M\hat{\theta} = \theta + b(\theta)$, то $b(\theta)$ називають зсувом оцінки $\hat{\theta}$.

Означення. Послідовність оцінок $\hat{\theta}_n = h(X_1,\ X_2,\ \cdots,\ X_n),\quad n = 1,\ 2, ...$ параметра θ називається ***слушною (конзистентною)***, якщо $\hat{\theta}_n \xrightarrow{p} \theta,\ n \to \infty.$

Означення. Послідовність оцінок $\hat{\theta}_n = h(X_1,\ X_2,\ ...,\ X_n),\quad n = 1,\ 2,\ ...$ параметра θ називається ***сильно слушною (сильно конзистентною)***, якщо $\hat{\theta}_n \xrightarrow{p} \theta,\ n \to \infty$ з імовірністю 1.

Означення. Оцінка $\hat{\theta}$ параметра θ називається ***ефективною***, якщо вона за даного обсягу вибірки має найменшу можливу дисперсію.

Незміщеною сильно слушною *оцінкою математичного сподівання* випадкової величини X є ***вибіркове середнє*** $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ і реалізацію цієї оцінки позначають

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\,. \tag{3.2}$$

Вибіркове середнє для ***дискретного статистичного ряду*** обчислюють за формулою

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

Вибіркове середнє для ***інтервального статистичного ряду*** обчислюють за формулою

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i z_i,$$

де z_i – середина i-го інтервалу, $\sum_{i=1}^{k} n_i = n.$

Оцінкою дисперсії σ^2 випадкової величини X є ***вибіркова дисперсія***

$$\overline{s}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2, \tag{3.3}$$

яка є ***зміщеною*** оцінкою для σ^2.

Величина

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2,\ s^2 = \frac{n}{n-1}\overline{s}^2 \tag{3.4}$$

є ***незміщеною*** оцінкою дисперсії σ^2 випадкової величини X.

Якщо математичне сподівання a відоме, то *незміщеною сильно слушною оцінкою дисперсії* σ^2 випадкової величини X є оцінка

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - a\right)^2.$$

Вибіркову дисперсію для *дискретного статистичного ряду* обчислюють за формулою

$$\overline{s}^2 = \frac{1}{n}\sum_{i=1}^{k}\left(n_i x_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i {x_i}^2 - \overline{x}^2,$$

відповідно,

$$s^2 = \frac{n}{n-1}\overline{s}^2;\ s^2 = \frac{1}{n-1}\sum_{i=1}^{k}\left(n_i x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i {x_i}^2 - \overline{x}^2.$$

Вибіркову дисперсію для *інтервального статистичного ряду* обчислюють за формулою

$$\overline{s}^2 = \frac{1}{n}\sum_{i=1}^{k}\left(n_i z_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i {z_i}^2 - \overline{x}^2,$$

відповідно,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}\left(n_i z_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i {z_i}^2 - \overline{x}^2.$$

Значення $\overline{s} = \sqrt{\overline{s}^2}$ називається ***вибірковим середнім квадратичним відхиленням,*** $s = \sqrt{s^2}$ ***– вибірковим виправленим середнім квадратичним відхиленням*** (s^2 – вибіркова виправлена дисперсія).

При достатньо великих значеннях обсягу вибірки n вибіркова і виправлена дисперсія мало відрізняються. Тому на практиці використовують виправлену дисперсію зазвичай за умови $n < 30.$

Коефіцієнт варіації. Для порівняння величин розсіювання стосовно вибіркової середньої двох варіаційних рядів вводять безрозмірну характеристику, яку називають ***коефіцієнтом варіації***

$$c = \frac{\overline{s}}{\overline{x}} \cdot 100 \quad (\overline{x} \neq 0).$$

Із двох рядів той має більше розсіювання відносно вибіркової середньої, в якого V більший.

Зауваження. Властивістю ефективності часто нехтують, вибираючи в якості пріоритету раціональність обчислення оцінки:

1) якщо $(X_1,..., X_n)$ – незалежна вибірка з генеральної сукупності з розподілом $N(a, \sigma^2)$, то вибіркове середнє $\overline{x}$ є ефективною оцінкою для математичного сподівання і при відомому математичному сподіванні вибіркова дисперсія s^2 є також ефективною оцінкою дисперсії σ^2;

2) відносна частота $\hat{\theta} = \frac{\mu_n}{n}$ є ефективною оцінкою ймовірності появи події в схемі Бернуллі;

3) вибіркова середня закону Пуассона є ефективною оцінкою параметра λ.

***Означення*.** **Медіаною** M_e називають варіанту, яка ділить варіаційний ряд на дві частини, рівні за кількістю варіант.

Для дискретних статистичних рядів

$$M_e = \begin{cases} x_m, \ n = 2m-1; \\ \dfrac{x_m + x_{m+1}}{2}, \quad \text{çà} \ \ n = 2m. \end{cases}$$

Для інтервальних статистичних рядів

$$M_e = x_i + h_i \frac{\frac{n}{2} - \sum_{j=1}^{i} n_j}{n_i},$$

де x_i – початок медіанного інтервалу (йому відповідає перша з нагромаджених частот, що більша від половини всіх спостережень), h_i – довжина *i*-го інтервалу, n_i – частота медіанного інтервалу.

Означення. **Модою** називають варіанту, яка найчастіше трапляється у вибірці.

Для дискретних статистичних рядів

$$M_0 = x_j, \text{ якщо } n_j = \max_i n_i.$$

Для інтервальних статистичних рядів

$$M_0 = x_i + h\frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}},$$

де x_i – початок інтервалу із найбільшою частотою, n_i – частота *i*-го інтервалу.

Зауваження. Якщо кількість варіант досить велика, то для спрощення обчислень застосовують групування емпіричних даних. Інтервал, в якому лежать вибіркові дані, розбивають на k рівних частин довжини h. Якщо z_i – середина *i*-ї частини, n_i – кількість варіант, що потрапили в *i*-ту частину, то $\overline{x}$, $\overline{s}^2$ обчислюють за формулами

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i z_i; \ \overline{s}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i z_i^{\,2} - \overline{x}^2.$$

Для подальшого спрощення обчислень застосовують метод умовних варіант, які визначаються рівністю $u_i = \frac{z_i - C}{h}$, C – умовний нуль (новий початок відліку). Найбільшої простоти обчислень досягають тоді, коли за умовний нуль взяти варіанту, розміщену приблизно посередині варіаційного ряду і яка має найбільшу частоту. Легко знайти, що

$$\overline{x} = h\overline{u} + C, \ \overline{s}_x^{\,2} = h^2\overline{s}_u^{\,2}.$$

Якщо варіанти x_i вибірки рівновіддалені, тобто утворюють арифметичну прогресію з різницею h, то можна зразу вводити умовні варіанти, не застосовуючи методу групування.

Приклад 3.2. Обчислити незміщені оцінки математичного сподівання та дисперсії, вибіркову медіану та моду вибірки.

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

Розв'язання. Обсяг вибірки $n=5+15+50+16+10+4=100.$ Вибіркове середнє

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{k}n_i x_i=\frac{1}{100}\left(5\cdot 12+15\cdot 14+50\cdot 16+16\cdot 18+10\cdot 20+4\cdot 22\right)=16,46.$$

Щоб знайти значення $\overline{s}^2=\frac{1}{n}\sum_{i=1}^{k}n_i x_i^2-\overline{x}^2$ обчислюємо

$$\frac{1}{n}\sum_{i=1}^{k}n_i x_i^2=\frac{1}{100}\left(5\cdot 12^2+15\cdot 14^2+50\cdot 16^2+16\cdot 18^2+10\cdot 20^2+4\cdot 22^2\right)=275,8,$$

і тоді $\overline{s}^2=275,8-16,46^2=4,8684;$

$$s^2=\frac{n}{n-1}\overline{s}^2=\frac{100}{99}\cdot 4,8684=4,92.$$

Медіана та мода вибірки: $M_e=\frac{16+18}{2}=17;\ M_0=16.$

Відповідь. $\overline{x}=16,46;\ s^2=4,92;\ M_e=17;\ M_0=16.$

Приклад 3.3. Обчислити вибіркове середнє та дисперсію, медіану та моду для вибірки

інтервал	$[2,4)$	$[4,6)$	$[6,8)$	$[8,10)$	$[10,12)$
n_i	2	8	35	40	15

Розв'язання. Обсяг вибірки $n=2+8+35+40+15=100$.

Тоді

z_i	3	5	7	9	11
n_i	2	8	35	40	15

Знаходимо числові характеристики:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i z_i = \frac{1}{100}\left(2\cdot 3 + 8\cdot 5 + 35\cdot 7 + 40\cdot 9 + 15\cdot 11\right) = 8{,}16;$$

$$\overline{s}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i z_i^2 - \overline{x}^2 = \frac{1}{100}\left(2\cdot 9 + 8\cdot 25 + 35\cdot 49 + 40\cdot 81 + 15\cdot 121\right) - 8{,}16^2 =$$

$$= 69{,}88 - 66{,}5856 = 3{,}2944;$$

$$s^2 = \frac{n}{n-1}\overline{s}^2 = \frac{100}{99}\cdot 3{,}29 = 3{,}33;$$

$$M_e = 8 + 2\frac{50-45}{40} = 8{,}25;\ M_0 = 8 + 2\frac{40-35}{80-35-15} = 8{,}33.$$

Відповідь. $\overline{x} = 8{,}16;\ s^2 = 3{,}33;\ M_e = 8{,}25;\ M_0 = 8{,}33.$

3.1.5. Метод моментів та метод максимальної правдоподібності

Метод моментів точкової оцінки невідомих параметрів розподілу полягає у прирівнюванні теоретичних моментів розподілу та відповідних емпіричних моментів того самого порядку. Цей метод запропонований К. Пірсоном.

3.1.5.1. Метод моментів

Нехай $X_1,\ X_2,\ ...,X_n$ – вибірка з генеральної сукупності, $x_1,\ x_2,\ ...,\ x_n$ – реалізація вибірки.

***Означення*. Початковим емпіричним моментом порядку** k називається вираз

$$\grave{I}_{\ k} = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \tag{3.5}$$

зокрема $M_1 = \overline{x}$.

***Означення*. Центральним емпіричним моментом порядку** k називається вираз

$$m_k = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^k, \tag{3.6}$$

зокрема $m_2 = \overline{s}^2$.

***Означення*. Коефіцієнтом асиметрії A_s** називається відношення центрального емпіричного моменту третього порядку до куба середнього квадратичного відхилення:

$$A_s = \frac{m_3}{\sigma^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{s^3}. \qquad (3.7)$$

Означення*. Ексцесом *E називається зменшене на три одиниці відношення центрального моменту четвертого порядку до четвертого степеня середнього квадратичного відхилення:

$$E = \frac{m_4}{\sigma^4} - 3. \qquad (3.8)$$

Оцінка одного параметра. Припустимо, що нам відомий вигляд щільності розподілу $f(x, \theta)$ ознаки X, який визначається одним параметром θ. Розглянемо, як знайти точкову оцінку цього параметра. Для отримання оцінки одного параметра достатньо мати одне рівняння відносно цього параметра, тому прирівнюють початковий момент першого порядку $\nu_1 = M(X)$ і початковий емпіричний момент першого порядку $M_1 = \overline{x}$:

$$M(\theta) = \overline{x}. \qquad (3.9)$$

Розв'язавши рівняння, одержують оцінку параметра θ, яка є функцією від $\overline{x}$, отже, і від $x_1, x_2, ..., x_n$.

Приклад 3.4. Випадкова величина X – час роботи елемента, вона має показниковий розподіл з параметром λ. Отримано статистичний розподіл середнього часу роботи 200 елементів

x_i	2,5	7,5	12,5	17,5	22,5	27,5
n_i	133	45	15	4	2	1

де x_i – середній час роботи елемента в годинах, частота n_i – кількість елементів, які пропрацювали в середньому x_i годин. Знайти методом моментів точкову оцінку параметра λ.

Розв'язання. Прирівнявши теоретичний і емпіричний моменти першого порядку і враховуючи, що для показникового закону $M(X)=\frac{1}{\lambda}$, отримаємо $\lambda=\frac{1}{\bar{x}}$. Отже, точковою оцінкою параметра λ є $\lambda^*=\frac{1}{\bar{x}}$. Обчисливши

$$\bar{x}=\frac{1}{200}\sum_{i=1}^{6}n_i x_i=5,$$

одержимо $\lambda^*=\frac{1}{5}=0{,}2$.

Оцінка двох параметрів. Нехай щільність розподілу $f(x,\theta_1,\theta_2)$ генеральної сукупності X визначається двома невідомими параметрами θ_1,θ_2. Для їх визначення потрібно мати два рівняння. Прирівняємо теоретичний та емпіричний початкові моменти першого порядку $\nu_1=\nu_1^*$ і теоретичний та емпіричний центральні моменти другого порядку $m_2=m_2^*$. Розв'язавши систему

$$\begin{cases} M(\theta_1)=\bar{x}; \\ D(\theta_2)=\bar{s}^2, \end{cases} \tag{3.10}$$

одержимо оцінку параметрів θ_1, θ_2, які є функціями від x_1, x_2, ..., x_n.

Приклад 3.5. Випадкова величина X – відхилення контрольованого параметра виробу від переглянутого і зміненого нормативу, що підлягає нормальному закону розподілу з параметрами a та σ. Отримано статистичний розподіл відхилення 200 виробів

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,2	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

Знайти методом моментів точкові оцінки параметрів a та σ.

Розв'язання. Враховуючи, що для нормального розподілу $\nu_1=M(X)=a$, $m_2=D(X)=\sigma^2$, і прирівнюючи відповідні теоретичні й емпіричні моменти $\nu_1=\nu_1^*$, $m_2=m_2^*$, отримаємо вирази для точкових оцінок:

$$a^* = \bar{x} = \frac{1}{200}\sum_{i=1}^{11} n_i x_i = 1{,}266;$$

$$(\sigma^2)^* = \frac{1}{200}\sum_{i=1}^{11} n_i x_i^2 - \left(\frac{1}{200}\sum_{i=1}^{11} n_i x_i\right)^2 = 0{,}25,$$

звідки $\sigma^* = 0{,}5.$

Отже, $a^* = 1{,}266$ та $\sigma^* = 0{,}5.$

3.1.5.2. Метод максимальної правдоподібності

Ідея методу максимуму правдоподібності полягає в тому, що за оцінку $\hat{\theta}$ беруть таке значення параметра $\theta,$ для якого ймовірність отримання вже наявної вибірки максимальна.

Припустимо, що вигляд закону розподілу випадкової величини заданий, але параметр θ, яким визначається цей закон, невідомий. Потрібно знайти його точкову оцінку. Позначимо ймовірність того, що в результаті випробування випадкова величина X набуде значення x_i $(i = 1, 2, ..., n)$ через $p(x_i, \theta)$.

***Означення*. Функцією правдоподібності** дискретної випадкової величини X називають функцію аргумента θ:

$$L(x_1,\ x_2,\ ...,\ x_n,\ \theta) = p(x_1, \theta) \cdot p(x_2, \theta) \cdot \ldots \cdot p(x_n, \theta),$$

де $x_1,\ x_2,\ ...,\ x_n$ – вибіркові значення.

***Означення*. Оцінкою максимальної правдоподібності** параметра θ називається таке його значення $\hat{\theta} = \hat{\theta}(x_1,\ x_2,\ \ldots x_n),$ за якого функція правдоподібності досягає максимального значення.

Функцію $\ln L$ називають ***логарифмічною функцією правдоподібності***, а рівняння $\frac{d \ln L}{d\theta} = 0$ – ***рівнянням правдоподібності***. Якщо друга похідна

$$\frac{d^2 \ln L}{d\theta^2} < 0$$

за $\theta = \hat{\theta}$, то $\hat{\theta}$ – точка максимуму.

***Означення*. Функцією правдоподібності** неперервної випадкової величини X називають функцію аргумента θ:

$$L(x_1,\ x_2,\ ...,\ x_n,\ \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta),$$

де $x_1,\ x_2,\ ...,\ x_n$ – фіксовані числа.

Якщо щільність розподілу $f(x, \theta_1, \theta_2)$ неперервної випадкової величини X визначається двома параметрами θ_1 і θ_2, то функція правдоподібності – це функція двох аргументів θ_1 і θ_2:

$$L(x_1,\ x_2,\ ...,\ x_n, \theta_1, \theta_2) = f(x_1, \theta_1, \theta_2) \cdot f(x_2, \theta_1, \theta_2) \cdot \ldots \cdot f(x_n, \theta_1, \theta_2),$$

де $x_1,\ x_2,\ ...,\ x_n$ – значення X, які спостерігались.

Приклад 3.6. Знайти оцінку максимальної правдоподібності параметрів $a,\ \sigma$ нормального розподілу $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ за вибіркою $x_1,\ x_2,\ ...,\ x_n$.

Розв'язання. Складемо функцію правдоподібності

$$L(x_1,\ x_2,\ ...,\ x_n,\ a, \sigma) = \frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} e^{-\frac{\sum\limits_{i=1}^{n}(x_i-a)^2}{2\sigma^2}},$$

звідки $\ln L = -n\ln\sigma - \ln\left(\sqrt{2\pi}\right)^n - \dfrac{\sum\limits_{i=1}^{n}(x_i - a)^2}{2\sigma^2}.$

Розв'яжемо систему

$$\begin{cases} \dfrac{\partial \ln L}{\partial a} = 0; \\ \dfrac{\partial \ln L}{\partial \sigma} = 0; \end{cases} \quad \begin{cases} \sum\limits_{i=1}^{n} x_i - na = 0; \\ -n\sigma^2 + \sum\limits_{i=1}^{n}\left(x_i - a\right)^2 = 0; \end{cases} \quad \begin{cases} a = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i = \overline{x}; \\ \sigma^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - a)^2 = \overline{s}^2. \end{cases}$$

Знаходимо частинні похідні другого порядку для визначення точки екстремуму:

$$\frac{\partial^2 \ln L}{\partial a^2} = -\frac{n}{\sigma^2};\ \frac{\partial^2 \ln L}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3\sum\limits_{i=1}^{n}(x_i - a)^2}{\sigma^4};\ \frac{\partial^2 \ln L}{\partial a \partial \sigma} = -\frac{2\sum\limits_{i=1}^{n}(x_i - a)}{\sigma^3}.$$

Тоді $A = \left.\dfrac{\partial^2 \ln L}{\partial a^2}\right|_{\substack{a=\overline{x} \\ \sigma=\overline{s}}} = -\dfrac{n}{\overline{s}^2} < 0;\ \tilde{N} = \left.\dfrac{\partial^2 \ln L}{\partial \sigma^2}\right|_{\substack{a=\overline{x} \\ \sigma=\overline{s}}} = \dfrac{n}{\overline{s}^2} - \dfrac{3\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}{\overline{s}^4} = -\dfrac{2n}{\overline{s}^2};$

$$B = \frac{\partial^2 \ln L}{\partial a \partial \sigma}\Bigg|_{\substack{a=\overline{x} \\ \sigma=\overline{s}}} = -\frac{2\sum_{i=1}^{n}(x_i - \overline{x})}{\overline{s}^3} = 0, \text{ оскільки } \sum_{i=1}^{n}(x_i - \overline{x}) = 0.$$

Отже, $\Delta = AC - B^2 = \left(-\frac{n}{\overline{s}^2}\right)\cdot\left(-\frac{2n}{\overline{s}^2}\right) = \frac{2n^2}{\overline{s}^4} > 0$ і $A < 0$, тому функція $\ln L$ в точці $(\overline{x},\ \overline{s}^2)$ має максимум. Для параметрів (a, σ^2) оцінкою максимальної правдоподібності є $(\overline{x},\ \overline{s}^2)$.

Зауваження. Метод максимуму правдоподібності оцінювання параметрів розподілу має низку важливих переваг: вони слушні, асимптотично нормально розподілені (за великих n їх розподіл близький до нормального) і мають найменшу дисперсію порівняно з іншими асимптотично нормальними оцінками. Цей метод найбільш повно використовує дані вибірки для оцінювання параметрів, тому він особливо корисний для малих обсягів вибірки. Однак його недолік полягає у тому, що він вимагає складних обчислень.

3.1.6. Визначення числових характеристик із використанням табличного процесору Microsoft Excel

3.1.6.1. Визначення основних числових характеристик

Більшість числових характеристик у випадку незгрупованих даних можна обчислити з використанням табличного процесору Microsoft Excel. Основні вбудовані функції Excel, що застосовуються для таких розрахунків, надано у табл. 3.1. Щоб викликати потрібну функцію, слід обрати категорію *Статистические* та ім’я функції.

Часто використовуваними та корисними є такі функції:

− **НАИБОЛЬШИЙ** (массив, *k*) – надає *k*-е найбільше значення в ряді даних;

− **НАИМЕНЬШИЙ** (массив, *k*) – надає *k*-е найменше значення в ряді даних.

Приклад 3.7. За даними вибіркового дослідження відома заробітна платня (у грн.) 20-и службовців певної компанії (табл. 3.2). Знайти за допомогою вбудованих статистичних функцій Excel всі можливі числові характеристики за

даними таблиці.

Табл.3.2

3560	2190	2390	3400
2180	2400	3350	2340
2900	2570	3300	3150
3680	3250	2250	3240
2180	2600	2870	3050

Розв'язання. Запишемо в лист Excel вхідні дані і числові характеристики, які можна знайти (рис. 3.2). Для знаходження характеристик введемо:

1) середнє значення (математичне сподівання) − формула **СРЗНАЧ** (A2:D6);

2) медіана − формула **МЕДИАНА** (A2:D6);

3) дисперсія (незміщена оцінка) − формула **ДИСП** (A2:D6);

4) середнє квадратичне відхилення (незміщена оцінка) − формула **СТАНДОТКЛОН** (A2:D6); також можна ввести **КОРЕНЬ** (H4), тобто обчислення за означенням (за коміркою дисперсії);

5) максимальне значення − формула **МАКС** (A2:D6);

6) мінімальне значення − формула **МИН** (A2:D6).

Зауваження. Функція **ДИСП** оцінює дисперсію за вибіркою, тобто вважається,що аргументи є вибіркою із генеральної сукупності. Якщо дані представляють всю генеральну сукупність, то слід використовувати функцію **ДИСПР**.

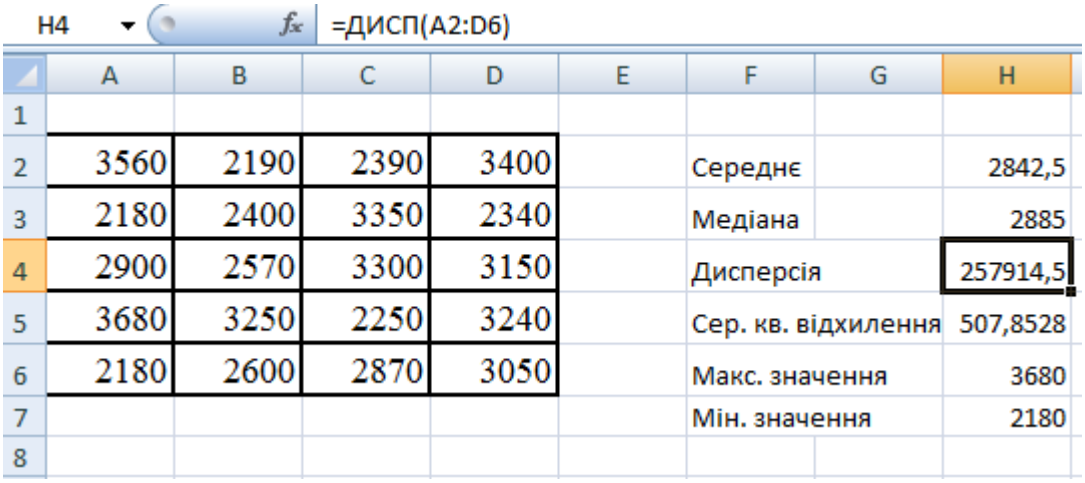


Рис. 3.2. Розрахунок числових характеристик

Табл. 3.1.

Статистичні функції Excel

Числові характеристики	Назва функції
Середнє	**СРЗНАЧ** (число1, число 2, …)
Середнє геометричне	**СРГЕОМ** (число1, число 2, …)
Мода	**МОДА** (число1, число 2, …)
Медіана	**МЕДИАНА** (число1, число 2, …)
Дисперсія	**ДИСП** (число1, число 2, …) **ДИСПР** (число1, число 2, …)
Середнє квадратичне відхилення	**СТАНДОТКЛОН** (число1, число 2, …)
Мінімальне значення	**МИН** (число1, число 2, …))
Максимальне значення	**МАКС** (число1, число 2, …)
Частота	**ЧАСТОТА** (массив_данных; массив_ интервалов)

3.1.6.2. Побудова гістограми

Табличний процесор Excel надає два способи побудови гістограми.

Для побудови гістограми *першим способом* необхідно:

1. Внести в лист Excel вхідні дані та інтервали, за якими ці дані будуть групуватися.

2. Знайти частоти попадання даних в інтервали за допомогою функції **ЧАСТОТА**, для чого:

− виділити діапазон комірок (на одну більше, ніж інтервалів), в яких будуть записані частоти;

− викликати f_x – *Статистические* **– ЧАСТОТА**;

− ввести посилання на комірки, що містять вхідні дані і інтервали;

− натиснути Ctrl+Shift+Enter.

3. Викликати *Вставка — Гистограмма*, появиться діалогове вікно (див. рис. 3.3).

4. Надати необхідні для побудови гістограми параметри:

− діапазон вхідних даних, спосіб їх групування (за рядками або стовпчиками) та імена рядів даних, якщо це потрібно;

− якщо імена рядів надано, відмітити *Добавить легенду* і вказати її розміщення;

− якщо потрібно, добавити *Имена рядов*, або (та) *Имена категорій*, або (та) *Значения*;

− якщо потрібно, добавити *Заголовок*, *Линии сетки*, *Оси*, *Таблицу данных*.

Рис. 3.3. Діалогове вікно майстра гістограм

Для побудови гістограми *другим способом* необхідно:

1. Внести в лист Excel вихідні дані.

2. Обрати в меню *Сервис – Анализ данных – Гистограмма*, появиться діалогове вікно (рис. 3.4).

3. Задати необхідні для побудови гістограми параметри:

входной диапазон – задати посилання на комірки, в яких знаходяться вхідні дані;

интервал карманов (параметр не є обов'язковим) – задати діапазон комірок і набір граничних значень в порядку зростання; якщо параметр не введений, то буде автоматично створений набір відрізків, рівномірно розподілених між мінімальним і максимальним значеннями даних;

выходной диапазон – ввести посилання на верхню ліву комірку діапазону, в який буде надано гістограму, або відмітити параметр *Новый рабочий лист* або *Новая рабочая книга*;

интегральный процент – якщо цей параметр відмічено, то будуть розраховані накопичені частоти і побудований їх графік;

вывод графика – якщо цей параметр відмічено, то буде створено автоматично діаграму, при цьому обов'язково задається значення *Новая книга*.

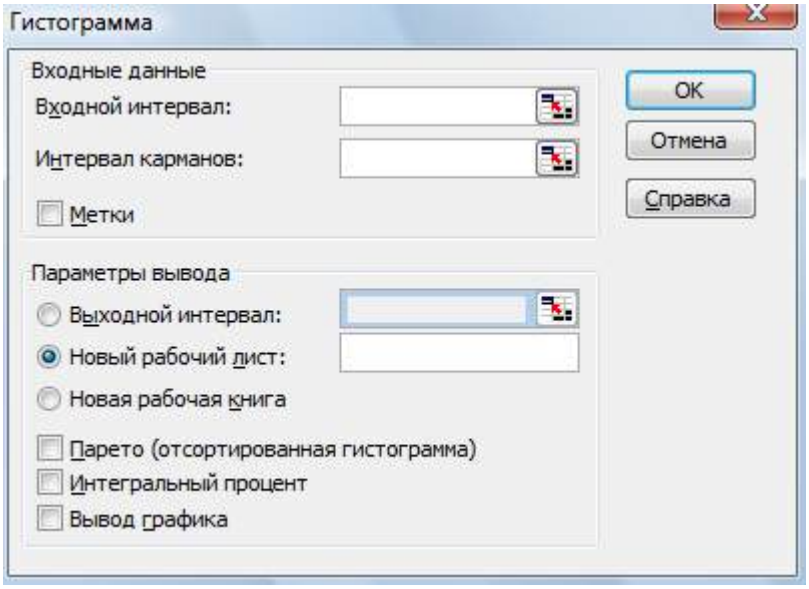


Рис. 3.4. Діалогове вікно для побудови гістограми

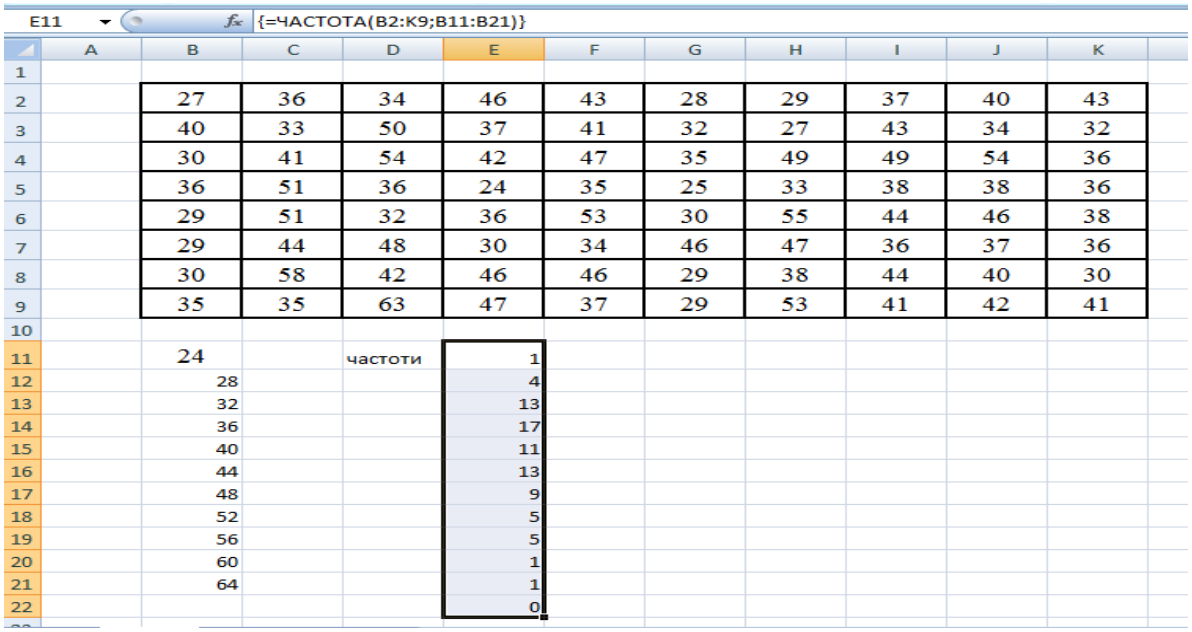
Приклад 3.8. За даними вибіркового дослідження відома кількість родин з дітьми дошкільного віку в селах деякої області (табл. 3.3). Побудувати за допомогою Excel гістограму за даними таблиці.

Табл. 3.3

27	36	34	46	43	28	29	37	40	43
40	33	50	37	41	32	27	43	34	32
30	41	54	42	47	35	49	49	54	36
36	51	36	24	35	25	33	38	38	36
29	51	32	36	53	30	55	44	46	38
29	44	48	30	34	46	47	36	37	36
30	58	42	46	46	29	38	44	40	30
35	35	63	47	37	29	53	41	42	41

Розв'язання. Запишемо в лист Excel вхідні дані завдання (рис. 3.5).

Розрахуємо частоти попадання в інтервали (див. зміст командного рядка на рис. 3.5).



	A	B	C	D	E	F	G	H	I	J	K
1											
2		27	36	34	46	43	28	29	37	40	43
3		40	33	50	37	41	32	27	43	34	32
4		30	41	54	42	47	35	49	49	54	36
5		36	51	36	24	35	25	33	38	38	36
6		29	51	32	36	53	30	55	44	46	38
7		29	44	48	30	34	46	47	36	37	36
8		30	58	42	46	46	29	38	44	40	30
9		35	35	63	47	37	29	53	41	42	41
10											
11		24		частоти	1						
12		28			4						
13		32			13						
14		36			17						
15		40			11						
16		44			13						
17		48			9						
18		52			5						
19		56			5						
20		60			1						
21		64			1						
22					0						

Рис. 3.5. Вхідні дані для побудови гістограми

Викличемо *Вставка – Гистограмма*, задамо діапазон даних. Для зручності читання діаграми добавимо *Заголовок* та *Значения*. Приберемо відмітку *Легенда*, оскільки імена рядів не було надано – розглядається тільки один тип даних.

Після побудови діаграми можна у разі необхідності змінити шрифти, ширину стовпчиків гістограми, їх колір та фон. Для внесення змін потрібно двічі натиснути правою кнопкою миші на відповідне поле гістограми.

Зауважимо, що на горизонтальній осі надаються не границі інтервалів (рис. 3.6), а їх порядковий номер.

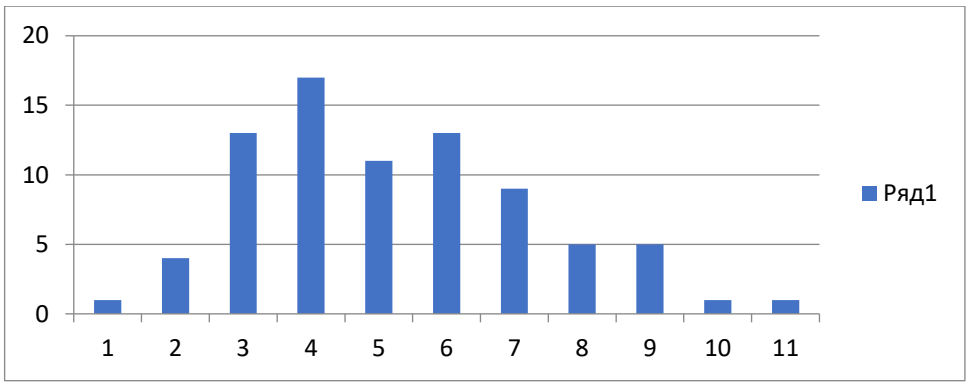


Рис. 3.6. Лист Excel із гістограмою

Приклад 3.9. Для вибірки

інтервал	$[2,4)$	$[4,6)$	$[6,8)$	$[8,10)$	$[10,12)$
n_i	2	8	35	40	15

побудувати емпіричну функцію розподілу.

Розв'язання. Обсяг вибірки $n = 2+8+35+40+15 = 100$. Тоді

z_i	3	5	7	9	11
n_i	2	8	35	40	15

При побудові емпіричної функції розподілу (кумуляти) для інтервального статистичного розподілу вибірки за основу береться припущення, що ознака на кожному частинному інтервалі має рівномірну щільність імовірностей. Тому кумулята матиме вигляд ламаної лінії, яка зростає на кожному частковому інтервалі та наближається до 1. Отже, емпірична функція розподілу

$$F^*(x) = \begin{cases} 0, & x \le 2, \\ 0{,}02, & 2 < x \le 4, \\ 0{,}1, & 4 < x \le 6, \\ 0{,}45, & 6 < x \le 8, \\ 0{,}85, & 8 < x \le 10, \\ 1, & 10 < x \le 12, \end{cases}.$$

Графік кумуляти показано на рис. 3.7.

Зауваження. З рисунка також визначається модальний інтервал (інтервал, що має найбільшу частоту появи), який дорівнює 8−10. Цей інтервал також є і медіанним, оскільки $F^*(8) < 0{,}5$, але $F^*(10) > 0{,}5$, то, беручи до уваги неперервність досліджуваної ознаки та властивість функції $F^*(x)$, яка є неспадною функцією, всередині інтервалу існує таке значення $X = M_e$, для якого $F^*(M_e) = 0{,}5$.

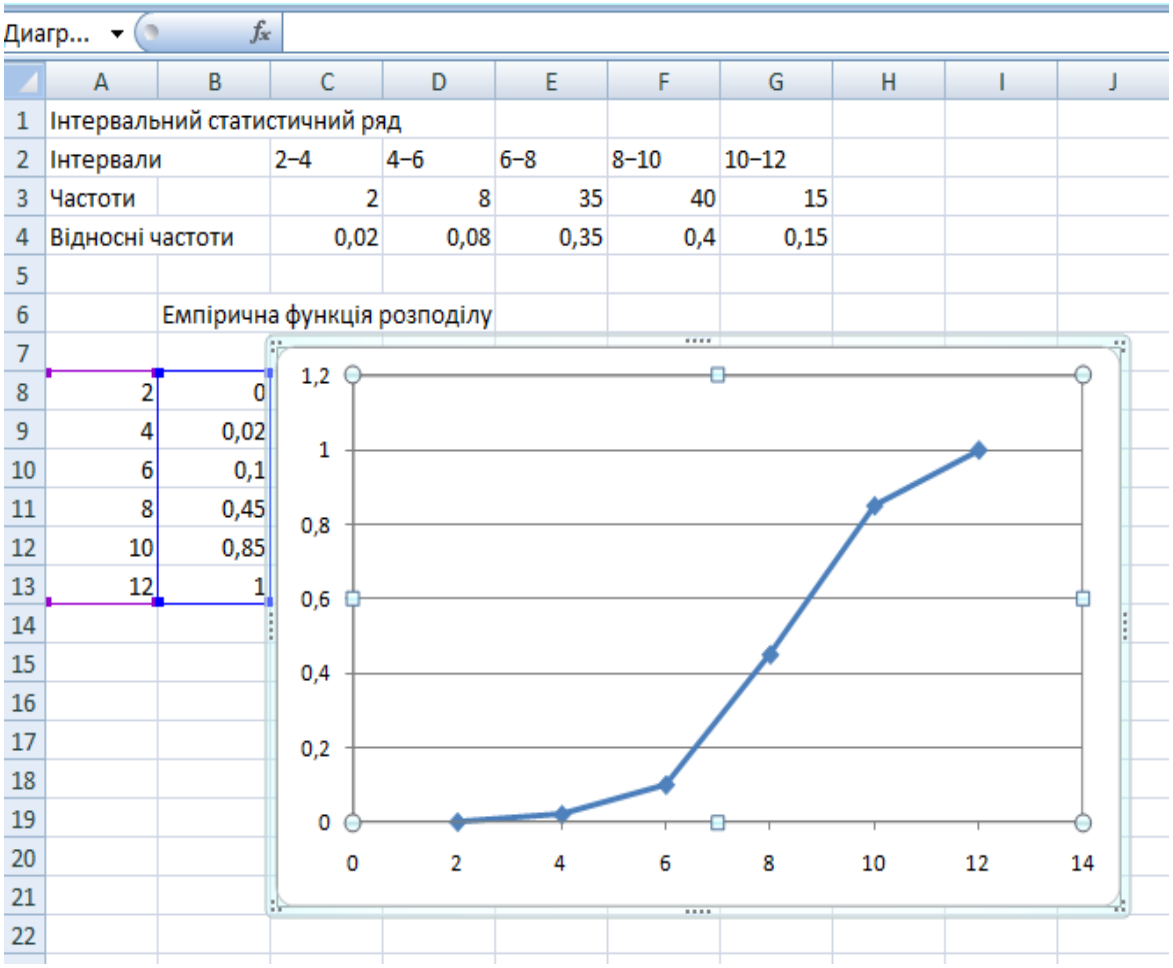


Рис. 3.7. Вигляд емпіричної функції розподілу

Завдання для самоконтролю

1. Основні задачі математичної статистики, їх прикладне застосування.

2. Генеральна та вибіркова сукупність, збір інформації.

3. Пояснити відмінність між статистичним рядом вибірки та інтервальним статистичним рядом.

4. Навести приклад побудови статистичного ряду відносних частот, побудувати полігон частот.

5. Емпірична функція розподілу, її властивості. Навести приклад її побудови.

6. Записати основні числові характеристики вибірки, пояснити їх обчислення за допомогою вбудованих статистичних функцій Excel.

7. Побудова гістограми.

8. В чому зміст методу моментів? Оцінка одного та двох параметрів.

9. Метод максимальної правдоподібності, його зміст.

10. Відомі дані про можливості збою проявлення пластини в автоматичному проявному процесорі (в секундах) (табл. 3.4). Побудувати статистичний ряд за даними вибірки, визначити середній час безаварійної роботи, дисперсію і середнє квадратичне відхилення часу. Побудувати полігон частот і гістограму.

Табл. 3.4

0,00	0,00	0,02	0,06	0,23	0,08	0,32	0,80	0,00	0,84
1,04	0,91	0,03	0,34	0,94	0,52	2,36	0,57	0,65	0,50
0,67	0,76	0,03	0,57	0,83	1,30	0,72	0,98	1,62	0,85
0,85	0,74	0,61	0,56	0,45	0,67	1,09	0,00	0,90	0,91
0,25	0,48	0,32	0,82	0,58	0,93	1,24	0,76	0,30	0,22

11. Відомі дані про місячний обсяг виробництва станків підприємств зварювального виробництва (табл. 3.5). Побудувати інтервальний статистичний ряд, полігон частот і гістограму. Знайти всі можливі числові характеристики. Побудувати графіки функції розподілу.

Табл. 3.5

11,240	18,545	17,750	22,560	18,355	20,424	20,650	10,780	15,590
13,720	28,505	23,170	20,360	22,450	21,590	14,565	24,295	25,140
27,655	17,786	27,045	28,650	18,670	31,445	18,540	15,598	19,720
15,230	21,240	19,535	12,934	18,195	19,074	17,037	19,610	20,970
22,075	15,090	20,754	10,195	13,580	21,490	13,987	22,645	21,218

11. В табл. 3.6 наведено значення прибутку 50 фірм, що належать одній корпорації (в 1000 у. од.). Знайти середнє значення прибутку, дисперсію і середнє квадратичне відхилення. Побудувати полігон частот і гістограму.

Табл. 3.6

4,744	9,127	7,201	8,650	11,534	9,013	10,390	9,268	7,354	10,255
6,232	6,739	6,088	8,671	15,103	9,124	11,902	10,216	10,954	11,470
7,351	9,832	7,126	9,715	10,744	10,687	10,582	12,271	11,047	13,190
5,536	8,917	9,823	8,383	14,212	15,031	13,001	11,089	12,091	10,321
9,766	5,854	2,917	6,379	6,748	7,024	11,587	11,101	10,954	10,387

12. Відомі дані про інтервал часу між телефонними повідомленнями в одного із операторів мобільного зв'язку. (табл. 3.7). Побудувати інтервальний статистичний ряд, знайти всі можливі числові характеристики. Побудувати графік функції розподілу випадкової величини X – інтервалу часу.

Табл. 3.7

0,0025	1,0042	0,0072	0,6121	0,0912	1,6922	1,5274	0,9081	2,5991	1,2925
4,1342	3,6471	0,3742	2,1504	0,7789	5,2239	3,3447	2,0016	3,4927	4,0116
3,5078	0,8389	0,1486	0,6184	0,7041	1,8535	3,2271	0,6531	3,6564	1,6536
1,7382	1,0694	0,4536	1,4479	1,6146	2,7421	1,3141	1,2112	1,9492	1,0014

13. Наведено результати дослідження річного обсягу споживання риби та рибної продукції в кг на душу населення (табл. 3.8). Побудувати інтервальний статистичний ряд, знайти всі можливі числові характеристики. Побудувати графік функції розподілу.

Табл. 3.8

14	10,5	9,8	12,5	11,5	13,5	12,4	10,4	11	12
10,2	11,5	9,8	13,6	12,5	9,5	10,2	11,9	11,4	12,8
11,6	12,3	10,8	9,9	11,1	10,7	11,4	10,9	12,9	11,1
13,8	14	13,1	10,7	12,9	13,9	10,1	12,2	11,2	13,1

Варіанти завдань для індивідуальної роботи

Завдання 1. Скласти статистичний ряд розподілу за зібраними статистичними даними. Знайти всі можливі числові характеристики.

Завдання 2. Обчислити вибіркове середнє та дисперсію, медіану та моду для вибірки та побудувати емпіричну функцію розподілу. Всі обчислення бажано виконати за допомогою Excel.

2.1.

інтервал	$[-2, 0)$	$[0, 2)$	$[2, 4)$	$[4, 6)$	$[6, 8)$
n_i	4	6	25	40	25

2.2.

інтервал	$[2, 3)$	$[3, 4)$	$[4, 5)$	$[5, 6)$	$[6, 7)$
n_i	12	8	35	40	15

2.3.

інтервал	$[-12,-10)$	$[-10,-8)$	$[-8,-6)$	$[-6,-4)$	$[-4,-2)$
n_i	7	13	32	38	10

2.4.

інтервал	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$
n_i	20	8	32	22	18

2.5.

інтервал	$[1,2;1,4)$	$[1,4;1,6)$	$[1,6;1,8)$	$[1,8;1,0)$	$[1,0;1,2)$
n_i	23	7	15	40	15

2.6.

інтервал	$[0,1)$	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$
n_i	11	45	5	4	35

2.7.

інтервал	$[-1,2)$	$[2,5)$	$[5,8)$	$[8,11)$	$[11,13)$
n_i	7	8	15	40	30

2.8.

інтервал	$[0,2;0,4)$	$[0,4;0,6)$	$[0,6;0,8)$	$[0,8;1)$	$[1;1,2)$
n_i	21	2	8	49	20

2.9.

інтервал	$[3,4)$	$[4,5)$	$[5,6)$	$[6,7)$	$[7,8)$
n_i	25	15	25	20	15

2.10.

інтервал	$[4,5)$	$[5,6)$	$[6,7)$	$[7,8)$	$[8,9)$
n_i	2	18	25	30	25

2.11.

інтервал	$[7,8)$	$[8,9)$	$[9,10)$	$[10,11)$	$[11,12)$
n_i	9	11	5	50	25

2.12.

інтервал	$[0;0,4)$	$[0,4;0,8)$	$[0,8;1,2)$	$[1,2;1,6)$	$[1,6;2)$
n_i	30	36	5	14	15

2.13.

інтервал	$[-1,2;-0,8)$	$[-0,8;-0,4)$	$[-0,4;0)$	$[0;0,4)$	$[0,4;0,8)$
n_i	23	17	15	40	5

2.14.

інтервал	$[-5,-4)$	$[-4,-3)$	$[-3,-2)$	$[-2,-1)$	$[-1,0)$
n_i	30	18	12	20	20

2.15.

інтервал	$[-3,1)$	$[1,5)$	$[5,9)$	$[9,13)$	$[13,17)$
n_i	7	18	25	10	40

Вказівка. 1. Обчислити зміщену та незміщену оцінку дисперсії, надати пояснення.

2. Побудова кумуляти наведено у прикладі (рис. 3.7).

3.2. Статистичні оцінки параметрів розподілу

Одна з основних задач математичної статистики – оцінювання параметрів розподілів випадкових величин. Точкові оцінки залежать від обсягу вибірки. Зокрема, якщо обсяг вибірки малий, то точкова оцінка може суттєво відрізнятися від оцінюваного параметра. З огляду на це доцільно користуватися інтервальними оцінками, тобто такими оцінками, які визначаються двома числами − кінцями інтервалу.

3.2.1. Надійні інтервали

Якщо обсяг вибірки досить великий, то точкові оцінки задовольняють практичні потреби точності.

Нагадаємо, що ***точковими оцінками*** параметрів розподілу генеральної сукупності називають такі оцінки, які визначаються одним числом.

***Означення*. Інтервальною оцінкою** параметрів розподілу генеральної сукупності називають оцінку, яка визначається двома числами – кінцями інтервалу.

Нехай $X_1,\ X_2,\ ...,X_n$ – вибірка з генеральної сукупності, θ – невідомий параметр генеральної сукупності X. У разі інтервального оцінювання невідомого параметра θ шукають такі дві функції $h_1 = h_1(X_1,\ X_2,\ \ldots,\ X_n)$ і $h_2 = h_2(X_1,\ X_2,\ \ldots,\ X_n)$, для яких завжди $h_1 < h_2$ і за заданого $\gamma \in (0,\ 1)$ виконується умова

$$P\left(h_1 < \theta < h_2\right) \geq \gamma\,. \tag{3.11}$$

Тоді інтервал (h_1, h_2) називають ***надійним інтервалом***, γ ***– рівнем надійності***, а числа h_1 і h_2 – ***нижньою та верхньою межами надійності***.

Рівень надійності задають наперед, найчастіше його беруть 0,95; 0,99.

Надійний інтервал для математичного сподівання будь-якого розподілу легко знайти за допомогою центральної граничної теореми.

Метод надійних інтервалів у статистиці започаткований Р. Фішером і Ю. Нейманом.

Розглянемо деякі задачі на побудову надійних інтервалів.

3.2.1.1. Надійні інтервали для оцінювання математичного сподівання нормального розподілу за відомого σ

Нехай відомо, що випадкова величина X розподілена нормально і σ − її середнє квадратичне відхилення. Потрібно побудувати інтервальну оцінку для невідомого математичного сподівання $M(X)=a.$ Точковою оцінкою для математичного сподівання є вибіркове середнє

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

Середнє вибіркове $\overline{x}$ різне для окремо взятих вибірок з генеральної сукупності, отже, його можна розглядати як випадкову величину $\overline{X}$, а значення x_i як однаково розподілені незалежні випадкові величини X_i $(i=\overline{1,n})$. Оскільки значення x_i незалежні, то

$$M\left(\overline{x}\right)=M\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n}\sum_{i=1}^{n}M\left(X_i\right)=M\left(X\right);$$

$$D\left(\overline{x}\right)=D\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D\left(X_i\right)=\frac{D\left(X\right)}{n}.$$

Вважаємо, що $\sqrt{D\left(\overline{x}\right)}=\frac{\sqrt{D\left(X\right)}}{\sqrt{n}}=\frac{\sigma}{\sqrt{n}}$ − відома величина. Нерівність $\left|\overline{x}-a\right|<\varepsilon$ має виконуватись із заданою ймовірністю γ:

$$P\left(\left|\overline{x}-a\right|<\varepsilon\right)=\gamma$$

або, замінивши її еквівалентною нерівністю, отримаємо:

$$P\left(a-\varepsilon<\overline{x}<a+\varepsilon\right)=\gamma, \tag{3.12}$$

Пригадаємо, що для нормально розподіленої випадкової величини X з параметрами a і σ ймовірність попадання в інтервал (α,β) визначається формулою

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right),$$

де $\Phi(z)$ − функція Лапласа (табульована).

Тоді співвідношення (3.12) можна переписати так:

$$\Phi\left(\frac{a+\varepsilon-a}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{a-\varepsilon-a}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-\varepsilon}{\sigma/\sqrt{n}}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right) = \gamma\,.$$

Позначивши $t = \dfrac{\varepsilon\sqrt{n}}{\sigma}$, маємо рівняння $2\Phi(t) = \gamma$ і остаточно отримаємо

$$P\left(\overline{x} - \frac{t\sigma}{\sqrt{n}} < a < \overline{x} + \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi(t) = \gamma. \qquad (3.13)$$

Отже, побудований надійний інтервал

$$\left(\overline{x} - \frac{t\sigma}{\sqrt{n}},\ \overline{x} + \frac{t\sigma}{\sqrt{n}}\right) \qquad (3.14)$$

містить невідомий параметр *a* (математичне сподівання) з імовірністю γ. Число *t* за заданого значення γ знаходимо із таблиці значень функції Лапласа.

Висновки:

1) зі збільшенням обсягу *n* вибірки число $\varepsilon = \dfrac{\sigma t}{\sqrt{n}}$ зменшується, тобто точність оцінки підвищується;

2) зростання надійності $\gamma = 2\Phi(t)$ веде до збільшення *t*, отже до зростання ε або до зниження точності.

Приклад 3.10. Нехай $\sigma = 2,\ n = 25,\ \gamma = 0{,}95$. Знайти надійний інтервал для *a*, якщо $\overline{x} = 5$.

Розв'язання. Для обчислення *t* використаємо рівняння

$$2\Phi(t) = \gamma \ \Rightarrow\ 2\Phi(t) = 0{,}95 \ \Rightarrow \Phi(t) = 0{,}475.$$

Із таблиці значень функції Лапласа знаходимо $t = 1{,}96$.

Отже, $\varepsilon = \dfrac{2 \cdot 1{,}96}{\sqrt{25}} = 0{,}784$ й отримаємо інтервал $(\overline{x} - 0{,}784;\ \overline{x} + 0{,}784)$ або $(4{,}216;\ 5{,}784)$.

Більшість числових характеристик у випадку незгрупованих даних можна обчислити із використанням табличного процесору Microsoft Excel.

Ширину надійного інтервалу для генерального середнього можна знайти за допомогою вбудованої статистичної функції Excel **ДОВЕРИТ** (*альфа, станд_откл, размер*). Параметр *альфа* – рівень значущості, $\alpha = 1 - \gamma$; параметр *станд_откл* – вибіркове середнє квадратичне відхилення *S*; параметр *размер* – обсяг вибірки.

Аналогічний результат маємо при використанні обчислень, як на рис. 3.8.

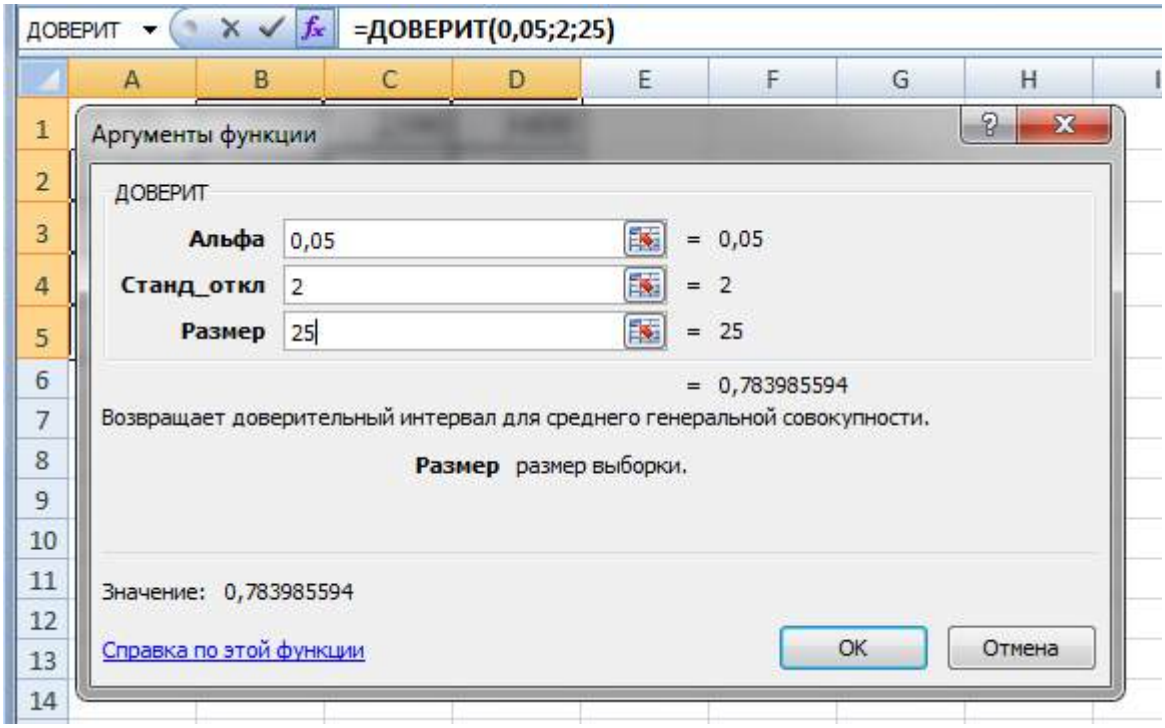


Рис. 3.8. Обчислення надійного інтервала

За відомою формулою отримаємо такий самий інтервал $(4{,}216;\ 5{,}784)$.

Цей результат можна трактувати так: якщо зроблена достатньо велика кількість вибірок, то в 95% випадків значення *a* належить знайденому інтервалі, а в 5% це значення *a* може вийти за межі інтервалу.

3.2.1.2. Надійні інтервали для оцінювання математичного сподівання нормального розподілу за невідомого σ

У цьому випадку для малих обсягів вибірки $(n < 30)$ використовують ***розподіл Стьюдента***. Середнє вибіркове $\bar{x}$ та виправлене середнє квадратичне

відхилення s є різними для окремо взятих вибірок з генеральної сукупності, отже, їх можна розглядати як випадкові величини $\overline{X}$ і S. За даними вибірки обсягу n можна побудувати випадкову величину

$$T = \frac{\overline{X} - a}{S/\sqrt{n}} \tag{3.15}$$

(її можливі значення позначимо через t), яка має ***розподіл Стьюдента*** з $(n-1)$ степенями вільності. Перевагою цього розподілу є те, що він визначається одним параметром − обсягом вибірки n і не залежить від невідомих параметрів a , σ. Щільність ***розподілу Стьюдента*** має вигляд

$$f(t,n) = \frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\Gamma((n-1)/2)}\left(1 + t^2/(n-1)\right)^{-n/2}, \tag{3.16}$$

де $\Gamma(z) = \int_0^{+\infty} u^{z-1}e^{-u}du$ − гамма-функція.

Оскільки $f(t,n)$ парна функція від t, то ймовірність виконання нерівності

$$\left|\frac{\overline{X} - a}{S/\sqrt{n}}\right| < t_\gamma$$

можна замінити рівносильною їй подвійною нерівністю вигляду

$$P\left\{\overline{X} - t_\gamma\frac{S}{\sqrt{n}} < a < \overline{X} + t_\gamma\frac{S}{\sqrt{n}}\right\} = \gamma, \tag{3.17}$$

Для конкретної вибірки обсягу n випадкові величини $\overline{X}$ і S замінимо невипадковими $\overline{x}$ і s. Отже, використовуючи ***розподіл Стьюдента***, знайдемо надійний інтервал для оцінювання математичного сподівання нормального розподілу за невідомого σ:

$$\left(\overline{x} - t_\gamma\frac{s}{\sqrt{n}};\ \overline{x} + t_\gamma\frac{s}{\sqrt{n}}\right), \tag{3.18}$$

який покриває невідомий параметр a з надійністю γ, $t_\gamma = t(\gamma, k)$ − табульоване значення випадкової величини, розподіленої за ***законом Стьюдента***, $k = n-1$ – кількість степенів вільності.

Зауваження. У статистиці кількістю степенів вільності певної величини називають різницю між кількістю випробувань і кількістю величин, обчислених завдяки цим випробуванням.

Приклад 3.11. Нехай X − нормально розподілена випадкова величина, яка має такі параметри: $n=25,\ \overline{x}=20,\ s=0,4.$ Знайти надійний інтервал для оцінювання математичного сподівання, якщо $\gamma=0,95.$

Розв'язання. Знаходимо t_γ з таблиці: $t(0,95;\ 24)=2,064$. Тоді

$$t_\gamma \frac{s}{\sqrt{n}} = 2,064 \cdot \frac{0,4}{5} \approx 0,165.$$

Отже, $(\overline{x}-0,165;\ \overline{x}+0,165)$ − надійний інтервал або $(19,835;\ 20,165)$.

Значення t_γ обчислюють також засобами програмного забезпечення, як на рис. 3.9, вказавши ймовірність $P=1-p=1-0,95=0,05.$

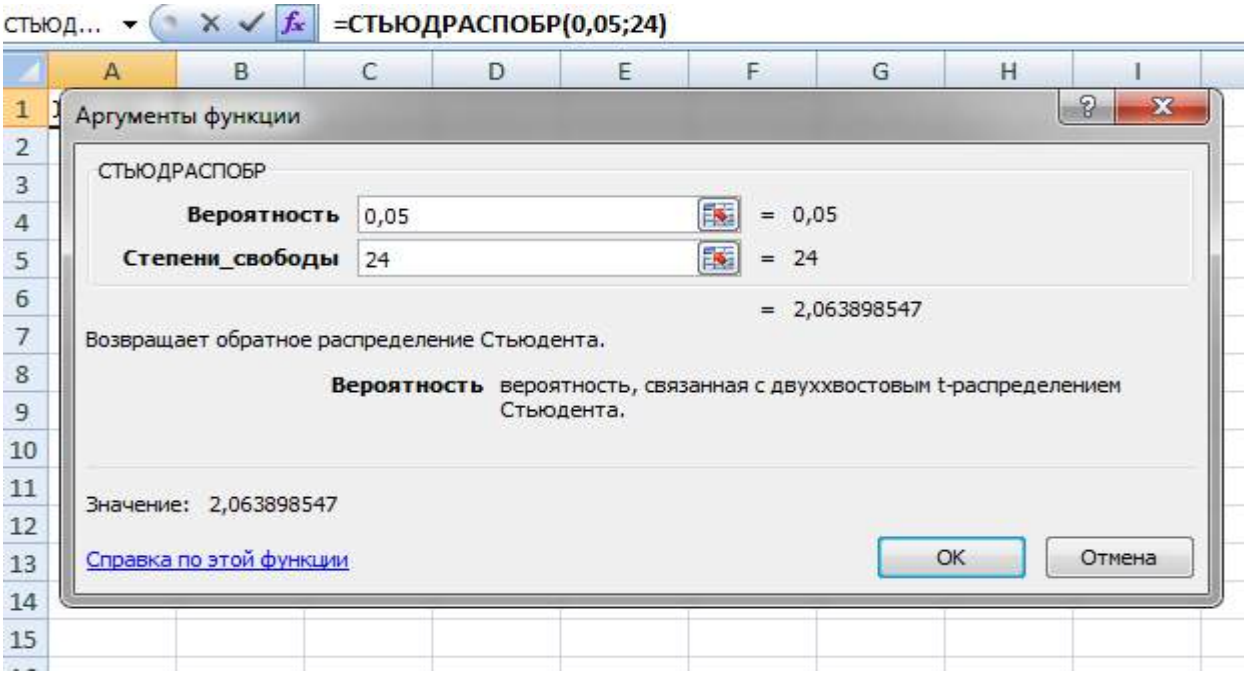


Рис. 3.9. Обчислення t_γ-характеристики

Також далі проводять обчислення надійного інтервала, рис. 3.10, задавши відповідно до формули (3.12) потрібні математичні дії.

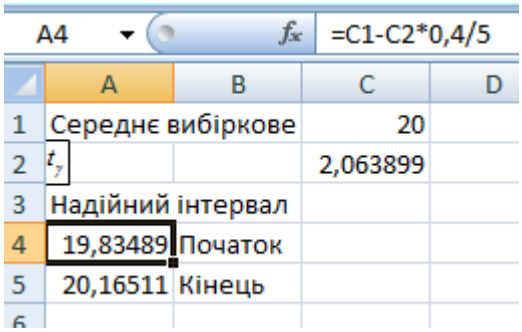


Рис. 3.10. Обчислення надійного інтервала

3.2.1.3. Надійні інтервали для оцінки середнього квадратичного відхилення σ нормального розподілу

Нехай випадкова величина X розподілена нормально. За даними вибірки обсягу n обчислено «виправлене» середнє квадратичне відхилення

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x}_b)^2}\,. \tag{3.19}$$

Знайдемо надійний інтервал, який буде містити в собі параметр σ з надійністю γ.

Очевидно, нерівність $|\sigma-s|<\varepsilon$ повинна виконуватися з ймовірністю $P\{|\sigma-s|<\varepsilon\}=\gamma$ або $P\{s-\varepsilon<\sigma<s+\varepsilon\}=\gamma$. З останньої рівності маємо: $P\left(s\left(1-\frac{\varepsilon}{s}\right)<\sigma<s\left(1+\frac{\varepsilon}{s}\right)\right)=\gamma$ або $P\big(s(1-q)<\sigma<s(1+q)\big)=\gamma$,

Отже, отримали інтервал надійний інтервал

$$s(1-q)<\sigma<s(1+q) \tag{3.20}$$

з невідомою величиною $q=\frac{\varepsilon}{s}$, яку і далі визначимо.

Нехай X_i $(i=\overline{1,n})$ незалежні нормально розподілені випадкові величини, для яких $M(X_i)=0$, $\sigma(X_i)=1$. Тоді сума квадратів цих величин $\sum_{i=1}^{n}X_i^2$, де $\sum_{i=1}^{n}X_i=n\overline{X}$, розподілена за законом χ^2 (законом Пірсона) з $(n-1)$ степенями вільності.

Щільність цього розподілу

$$f(x)=\begin{cases}0, & x\le 0,\\ \dfrac{1}{2^{(n-1)/2}\Gamma((n-1)/2)}e^{-x/2}x^{(n-3)/2}, & x>0\end{cases} \tag{3.21}$$

тобто цей розподіл визначається одним параметром n − обсягом вибірки і зі зростанням n наближається до нормального. Доведено, що випадкова величина

$$V=\frac{S^2}{\sigma^2}(n-1) \tag{3.22}$$

розподілена за законом χ^2 з $(n-1)$ степенем вільності. Позначимо

$$\chi=\sqrt{V}=\frac{S}{\sigma}\sqrt{n-1.} \tag{3.23}$$

Знайдемо щільність розподілу $g(\chi)$ функції $\chi=\sqrt{V}=\varphi(X)=\sqrt{X}$, $(\chi>0)$ за формулою $g(\chi)=f[\varphi^{-1}(\chi)]\,|[\varphi^{-1}(\chi)]'|$. Оскільки обернена функція $x=\varphi^{-1}(\chi)=\chi^2$ і $[\varphi^{-1}(\chi)]'=2\chi$, то

$$g(\chi)=\begin{cases}0, & \chi\le 0,\\ \dfrac{1}{2^{(n-1)/2}\Gamma((n-1)/2)}e^{-\chi^2/2}(\chi^2)^{(n-3)/2}\cdot 2\chi, & \chi>0\end{cases}$$

або після спрощення

$$g(\chi)=\frac{1}{2^{(n-3)/2}\Gamma((n-1)/2)}e^{-\chi^2/2}\chi^{n-2} \text{ для } \chi>0 \tag{3.24}$$

Для того, щоб знайти $q=\dfrac{\varepsilon}{s}$, введемо випадкову величину $\chi=\dfrac{S}{\sigma}\sqrt{n-1.}$

Перетворимо нерівність (3.20), припустивши, що $q<1$: $\dfrac{1}{S(1+q)}<\dfrac{1}{\sigma}<\dfrac{1}{S(1-q)}$

або, помноживши почленно на $S\sqrt{n-1}$, отримаємо $\dfrac{\sqrt{n-1}}{1+q}<\dfrac{S\sqrt{n-1}}{\sigma}<\dfrac{\sqrt{n-1}}{1-q}$

або

$$\frac{\sqrt{n-1}}{1+q}<\chi<\frac{\sqrt{n-1}}{1-q}. \tag{2.25}$$

Імовірність виконання нерівності (3.25) а, отже, і рівносильної їй нерівності (3.20), дорівнює

$$\int_{\sqrt{n-1}/(1+q)}^{\sqrt{n-1}/(1-q)} g(\chi)d\chi = \gamma \qquad (3.26)$$

З рівняння (3.26) за даними значеннями n і γ знаходять q. Значення $q(\gamma, n)$ протабульовані.

Таким чином, побудований інтервал (3.20).

Приклад 3.12. Нехай $n=16$, обчислено $s=0,7$ і задано $\gamma=0,95$. Знайти надійний інтервал для σ.

Розв'язання. З таблиці маємо $q(\gamma, n)=q(0,95;\ 16)=0,44$. Отже,

$$0,7(1\text{-}0,44)<\sigma<0,7(1+0,44) \text{ або } 0,392<\sigma<1,008.$$

Зауваження. Побудований інтервал у випадку $q<1$. Якщо $q>1$, то (3.20) матиме вигляд

$$0<\sigma<s(1+q),$$

або $\frac{\sqrt{n-1}}{1+q}<\chi<\infty$, практично для відшукання значень $q>1$ теж використовують ту ж таблицю.

Приклад 3.13. Відомо, що кількісна ознака *X* генеральної сукупності розподілена нормально. За вибіркою обсягом $n=10$ обчислено «виправлене» середнє квадратичне відхилення $s=0,16$, задано $\gamma=0,999$. Знайти надійний інтервал, який покриє генеральне середнє квадратичне відхилення σ.

Розв'язання. За таблицею маємо $q(\gamma, n)=q(0,999;\ 10)=1,8 \quad (q>1)$. Отже,

$$0<\sigma<0,16\left(1+1,8\right), \quad 0<\sigma<0,448.$$

3.2.1.4. Оцінка ймовірності (біномного розподілу) за відносною частотою

Нехай проводяться незалежні випробування з невідомою ймовірністю p настання події *A* в кожній спробі. Оцінимо невідому ймовірність p за відносною частотою $k_n=\frac{m}{n}$.

Зауваження. За точкову оцінку ймовірності *p* беруть відносну частоту

$k_n = \dfrac{m}{n}$ (n – кількість проведених випробувань, у яких подія A відбулася m разів).

Ця оцінка незміщена. Дійсно, оскільки $M(m) = np,$ то

$$M(k_n) = \frac{M(m)}{n} = \frac{np}{n} = p,$$

тобто математичне сподівання оцінки дорівнює значенню оцінюваного параметра.

Дисперсія оцінки (оскільки $D(m) = npq$)

$$D(k_n) = \frac{D(m)}{n^2} = \frac{pq}{n},$$

звідки $\sigma_{k_n} = \sqrt{\dfrac{pq}{n}}.$

Розглянемо *інтервальну оцінку*.

Якщо n досить велике і $0 < p < 1,$ тоді для відносної частоти k_n можна користуватися формулою

$$P(|X - a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

В нашому випадку

$$P\{|k_n - p| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sigma_{k_n}}\right) = \gamma \quad \text{або} \quad P\{|k_n - p| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) = \gamma.$$

Позначимо $t = \dfrac{\varepsilon\sqrt{n}}{\sqrt{pq}},$ де t визначається з рівняння $2\Phi(t) = \gamma.$

Отже, $|k_n - p| < \dfrac{t\sqrt{pq}}{\sqrt{n}}$ або

$$|k_n - p| < \frac{t\sqrt{p(1-p)}}{\sqrt{n}}. \qquad (3.27)$$

Розв’яжемо останню нерівність (3.27) відносно $p.$

Якщо $k_n > p,$ то $k_n - p < \dfrac{t\sqrt{p(1-p)}}{\sqrt{n}}.$ Піднісши до квадрата, отримаємо:

$$k_n^2 - 2k_n p + p^2 < \frac{t^2}{n}(p - p^2) \quad \text{або} \quad \left(\frac{t^2}{n} + 1\right)p^2 - 2\left(k_n + \frac{t^2}{2n}\right)p + k_n^2 < 0.$$

Дискримінант тричлена нерівності, точніше

$$\frac{1}{4}\Delta=\left(k_n+\frac{t^2}{2n}\right)^2-k_n^2\left(\frac{t^2}{n}+1\right)=\frac{t^2}{n}\left(\frac{t^2}{4n}+k_n(1-k_n)\right)>0,\quad (0<k_n<1)$$

додатна, тому корені дійсні і різні.

Знаходимо менший корінь p_1:

$$p_1=\frac{n}{t^2+n}\left[k_n+\frac{t^2}{2n}-t\sqrt{\frac{k_n(1-k_n)}{n}+\left(\frac{t}{2n}\right)^2}\right] \tag{3.28}$$

і більший корінь p_2:

$$p_2=\frac{n}{t^2+n}\left[k_n+\frac{t^2}{2n}+t\sqrt{\frac{k_n(1-k_n)}{n}+\left(\frac{t}{2n}\right)^2}\right] \tag{3.29}$$

нерівності, тоді шуканий надійний інтервал

$$p_1<p<p_2. \tag{3.30}$$

Приклад 3.14. Знайти надійний інтервал для оцінки ймовірності p настання деякої події з надійністю $\gamma=0,95,$ якщо у 80 спробах ця подія настала 20 разів.

Розв'язання. За умовою $\gamma=0,95,\ n=80,\ m=20.$ Корінь рівняння $2\Phi(t)=\gamma$ знаходимо з таблиці додатку:

$$\Phi(t)=0,475,\quad t=1,96.$$

Обчислимо відносну частоту настання події $k_n=\frac{m}{n}=\frac{1}{4}=0,25.$

За формулами (3.20) та (3.21) знаходимо $p_1=0,167;\ p_2=0,353.$

Отже, шуканий надійний інтервал $0,167<p<0,353.$

Зауваження. Із формули (3.14) можна знаходити мінімальний обсяг вибірки, за якого із вказаними надійністю і точністю виконується рівність $\bar{x}=a,$ якщо випадкова величина розподілена за нормальним законом із параметром $\sigma.$

Приклад 3.15. Знайти мінімальний обсяг вибірки, за якого із надійністю 0,95 точність оцінки математичного сподівання нормально розподіленої генеральної сукупності за вибірковою середньою дорівнює 0,2, якщо відоме середнє квадратичне відхилення генеральної сукупності $\sigma=1,5.$

Розв'язання. Із формули (3.14) видно, що точність оцінки дорівнює $\frac{t\sigma}{\sqrt{n}}$, яка за умовою дорівнює 0,2, тобто $\frac{t\sigma}{\sqrt{n}} = 0,2,$ тому

$$n = \frac{t^2\sigma^2}{0,2^2}; \quad \gamma = 0,95; \quad \Phi(t) = 0,475.$$

Згідно з таблицею $t = 1,96,$ тоді $n = \frac{1,96^2 \cdot 1,5^2}{0,2^2} = 217$ – мінімальний обсяг вибірки.

3.2.2. Вибіркові характеристики зв'язку

3.2.2.1. Обробка вибірки методом найменших квадратів

Нехай відома функціональна залежність між випадковими величинами X та Y вигляду

$$Y = f(X, \alpha_1, \alpha_2, \ldots, \alpha_m)$$

з невідомими параметрами $\alpha_1, \alpha_2, \ldots, \alpha_m$.

Наприклад, можна розглядати залежність між собівартістю продукції (ознака *Y*) та обсягом продукції (ознака *X*) деякої кількості однотипних підприємств.

Нехай унаслідок *n* незалежних випробувань отримано варіанти ознак Y та *X*, які оформлено у статистичній таблиці вигляду

Номери випробувань	1	2	…	k	…	n
X	x_2	x_2	…	x_k	…	x_n
Y	y_1	y_2	…	y_k	…	y_n

Для визначення оцінок параметрів функціональної залежності $\alpha_1, \alpha_2, \ldots, \alpha_m$ згідно з методом найменших квадратів найімовірніші значення параметрів $\alpha_1, \alpha_2, \ldots, \alpha_m$ мають давати мінімум функції

$$\Phi = \sum_{k=1}^{n} \left(y_k - f(x_k, \alpha_1, \alpha_2, \ldots, \alpha_m) \right)^2.$$

Якщо функція $f(x_k, \alpha_1, \alpha_2, \ldots, \alpha_m)$ має неперервні частинні похідні першого порядку відносно невідомих параметрів $\alpha_1, \alpha_2, \ldots, \alpha_m$, то необхідною умовою існування мінімуму функції Φ буде система m рівнянь з m невідомими

$$\frac{\partial \Phi}{\partial \alpha_k} = 0,\ k = 1,\ 2,\ \ldots,\ m.$$

Визначення функціональної залежності між випадковими величинами X та Y з використанням даних випробувань називають ***вирівнюванням емпіричних даних*** уздовж кривої $y = f(x, \alpha_1, \alpha_2, \ldots, \alpha_m)$.

Задача. Отримано експериментальні дані вимірювань динаміки зміни крайових кутів змочування на поверхні зразків паперу з різним покриттям. Оцінити їх здатність сприймати водні чорнила струменевого принтера (змочування, розтікання), дані вимірювань наведено в табл. 3.9.

Табл. 3.9. Середні значення крайового кута

Тип ЗМ	Маса, г/м^2	$t_0 = 0c$	$t_1 = 7{,}5c$	$t_2 = 15c$	$t_3 = 22{,}5c$	$t_1 = 30c$
1. Папір крейдований глянцевий						
№ 1	130	51,13	50,17	49,60	49,27	48,43
№ 2	200	52,03	51,83	50,83	50,40	49,63
№ 3	250	65,60	61,22	60,37	60,00	59,33
№ 4	300	65,50	59,53	57,40	56,83	55,97
2. Папір крейдований, матовий						
№ 1	130	58,97	47,40	45,73	44,07	42,40
№ 2	170	94,33	84,93	84,13	83,43	81,90
№ 3	250	55,87	51,07	50,87	49,50	49,73
№ 4	350	79,63	74,57	73,90	73,43	72,93
№ 5	400	74,20	58,90	58,50	59,87	57,13
3. Папір офсетний						
№ 1	150	75,50	69,10	68,00	67,40	66,30
№ 2	170	73,37	66,83	60,83	51,40	46,33
№ 3	250	71,63	65,40	53,43	33,50	24,23
4. Папір дизайнерський						
№ 1	280	105,33	106,93	105,80	105,33	105,20
№ 2	280	80,13	80,87	81,07	80,27	79,97
№ 3	280	70,27	68,53	68,53	68,33	68
5. Плівка						
№ 1		69,97	71,07	70,60	70,30	70,37
№ 2		65,63	67,00	67,03	66,57	66,30

Розв'язання. В статті [12] показано використання теорії стосовно практичного оцінювання за методом найменших квадратів.

Будемо шукати для паперу крейдованого, глянцевого та матового залежність крайового кута змочування як функцію, лінійну залежність у вигляді $y = at + b.$ Згідно з методом найменших квадратів складаємо вираз

$$S(a,b) = \sum_{i=1}^{5} \left(y_i - \left(at_i + b \right) \right)^2.$$

Для розв'язання системи рівнянь (із необхідних умов мінімуму функції двох змінних)

$$\begin{cases} \sum_{i=1}^{5} y_i t_i - a \sum_{i=1}^{5} t_i^2 - b \sum_{i=1}^{5} t_i = 0, \\ \sum_{i=1}^{5} y_i - a \sum_{i=1}^{5} t_i - 5b = 0, \end{cases}$$

слід обчислювати всі потрібні коефіцієнти. Так, для паперу №1 типу 3 (вагою 250 г/м^2) за поданою таблицею значень

T	0	7,5	15	22,5	30
y	65,6	61,22	60,37	60,00	59,33

знаходимо $\sum_{i=1}^{5} y_i t_i = 4494,6;$ $\sum_{i=1}^{5} t_i^2 = 1687,5;$ $\sum_{i=1}^{5} y_i = 306,52;$ $\sum_{i=1}^{5} t_i = 75,$ тоді, відповідна система рівнянь

$$\begin{cases} 1687,5a + 75b = 4494,6; \\ 75a + 5b = 306,52; \end{cases}$$

дає значення $a \approx -0,183$ та $b \approx 64,06.$

Отже, шукана пряма має вигляд

$$y = -0,183t + 64,06.$$

Аналогічно отримано лінійні залежності для решти видів цих паперів, але з використанням вбудованої функції Microsoft Excel **ЛИНЕЙН** (Известные_значения_y; Известные_значения_x; Конст; Статистика), яка повертає параметри лінійного наближення за методом найменших квадратів.

Отже, лінійні залежності крайового кута змочування від часу для паперу крейдованого глянцевого для всіх №№1-4 мають відповідний запис: 1) $y = -0,084t + 50,98$; 2) $y = -0,08307t + 52,19$; 3) $y = -0,18347t + 64,056$; 4) $y = -0,29013t + 63,398$; також для паперів крейдованих матових за №№ 1-5: 1) $y = -0,48627t + 55,008$; 2) $y = -0,35147t + 91,016$; 3) $y = -0,18467t + 54,168$; 4) $y = -0,19387t + 77,8$; 5) $y = -0,44227t + 68,354$.

На рис. 3.11 графічно відображено кінетику зміни крайового кута змочування дистильованою водою при дослідженні офсетного паперу різної маси.

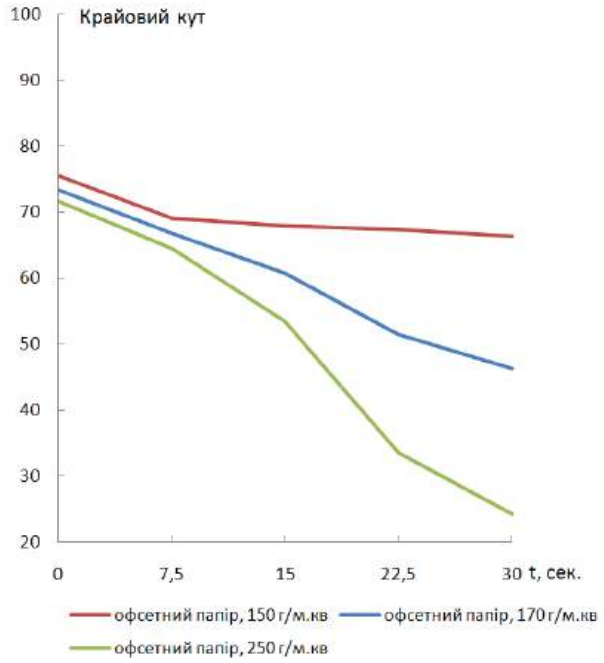


Рис. 3.11. Кінетика зміни крайового кута змочування дистильованою водою при дослідженні офсетного паперу

Також виконане припущення стосовно квадратичної залежності, тобто наближення многочленом другого степеня $F(t) = a_0 + a_1 t + a_2 t^2$. Тоді для функції

$$S(a_0, a_1, a_2) = \sum_{i=1}^{5} \left(y_i - \left(a_2 t_i^2 + a_1 t_i + a_0 \right) \right)^2$$

система лінійних алгебраїчних рівнянь, в загальному випадку, для визначення чисел a_0, a_1, a_2 має вигляд

$$\begin{cases} 5a_0 + a_1\sum\limits_{i=1}^{5} t_i + a_2\sum\limits_{i=1}^{5} t_i^2 = b\sum\limits_{i=1}^{5} y_i, \\ a_0\sum\limits_{i=1}^{5} t_i + a_1\sum\limits_{i=1}^{5} t_i^2 + a_2\sum\limits_{i=1}^{5} t_i^3 = \sum\limits_{i=1}^{5} t_i y_i, \\ a_0\sum\limits_{i=1}^{5} t_i^2 + a_1\sum\limits_{i=1}^{5} t_i^3 + a_2\sum\limits_{i=1}^{5} t_i^4 = \sum\limits_{i=1}^{5} t_i^2 y_i. \end{cases}$$

Для офсетного паперу №2 типу 3 (вагою170 г/м^2) за поданою таблицею значень система рівнянь набуває вигляду

$$\begin{cases} 5a_0 + 75a_1 + 1687,5a_2 = 298,76; \\ 75a_0 + 1687,5a_1 + 42187,5a_3 = 3960,075; \\ 1687,5a_0 + 42187,5a_1 + 1120078a_3 = 85164,19; \end{cases}$$

з якої отримано $F(t) = -0,00062t^2 - 0,90813t + 73,584$, причому близьким до нуля є коефіцієнт $a_0 = -0,00062$. Визначимо середню похибку апроксимації за відомою формулою

$$E = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{F(t_i) - y_i}{F(t_i)}\right|.$$

Провівши обчислення, маємо $E = 6,062\%$, тобто помилка складає приблизно 6%, причому коефіцієнт кореляції (детермінації) в даному випадку є близьким до одиниці: $R^2 = 0,993$.

Дослідження того ж типу паперу за лінійною залежністю приводить до рівняння прямої вигляду $y = -0,9268t + 73,654$ з помилкою, меншою, ніж $E = 1,216\%,\ R^2 = 0,994.$

3.2.2.2. Основні характеристики зв'язку

Нехай X і Y – випадкові величини, зв'язок між якими потрібно вивчити. В результаті n випробувань отримали n точок $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Вибіркові середні та дисперсії позначимо $\overline{x}$, $\overline{y}$, $\overline{s}_x^{\,2}$, $\overline{s}_y^{\,2}$. За оцінку коефіцієнта кореляції $\rho = \rho(X, Y)$ можна взяти

$$r = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\overline{s}_x\overline{s}_y} = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\overline{xy}}{n\overline{s}_x\overline{s}_y} \qquad (3.31)$$

у разі невеликої кількості випробувань.

Аналогічно визначаються рівняння вибіркових прямих регресії:

–прямої регресії Y на X: $y-\overline{y}=r\dfrac{\overline{s}_y}{\overline{s}_x}(x-\overline{x})$;

–прямої регресії X на Y: $x-\overline{x}=r\dfrac{\overline{s}_x}{\overline{s}_y}(y-\overline{y})$.

Прямі регресії доцільно шукати у тому разі, коли точки $(x_i,\ y_i)$ $(i=1,\ 2,\ \ldots,\ n)$ групуються навколо деякої прямої.

Якщо кількість випробувань велика, то для спрощення обчислень дані групують і застосовують метод умовних варіант. Вважатимемо, що дані уже згруповані і що пара чисел $(x_i,\ y_j)$ спостерігалась n_{ij} разів $(i=1,\ 2,\ \ldots,\ l;$ $j=1,\ 2,\ \ldots,\ k)$. Ці дані записують у вигляді кореляційної таблиці:

Y \ X	x_1	x_2	...	x_l	$\sum$
y_1	n_{11}	n_{21}	…	n_{l1}	n_1
y_2	n_{12}	n_{22}	…	n_{l2}	n_2
…	…	…	…	…	…
y_k	n_{1k}	n_{2k}	…	n_{lk}	n_k
$\sum$	m_1	m_2	…	m_l	n

де $m_i=\sum\limits_{j=1}^{k}n_{ij}$; $n_j=\sum\limits_{i=1}^{l}n_{ij}$; $\sum\limits_{i=1}^{l}m_i=\sum\limits_{j=1}^{k}n_j=n$ (n – кількість усіх спостережень). Тоді формула обчислення r набуває вигляду $r=\dfrac{\sum\limits_{i,j}n_{ij}x_iy_j-n\overline{x}\overline{y}}{n\overline{s}_x\overline{s}_y}$.

Нехай $x_{i+1}-x_i=h_1(i=1,\ 2,\ \ldots,\ l-1)$; $y_{j+1}-y_j=h_2(j=1,\ 2,\ \ldots,\ k-1)$ – кроки таблиці (варіанти рівновіддалені). Введемо умовні варіанти

$$u_i=\frac{x_i-C_1}{h_1};\ \ \nu_j=\frac{y_j-C_2}{h_2},$$

де C_1, C_2 – умовні нулі.

Тоді $x_i = h_1 u_i + C_1;\ y_j = h_2 \nu_j + C_2$ і формула набуває вигляду

$$r = \frac{\sum\limits_{i,j} n_{ij} u_i \nu_j - n\overline{u}\,\overline{\nu}}{n\overline{s}_u \overline{s}_\nu}. \qquad (3.32)$$

З огляду на можливості сучасної обчислювальної техніки формулою (3.32) користуються набагато рідше, ніж формулою (3.31).

3.2.2.3. Метод найменших квадратів для прямих регресії

Якщо $y = \alpha_1 x + \alpha_0$ – рівняння вибіркової прямої регресії Y на X, то згідно з означенням

$$\Phi(\alpha_1,\ \alpha_0) = \sum_{i=1}^{n} \left(y_i - \alpha_1 x_i - \alpha_0 \right)^2 \to \min.$$

Диференціюючи за α_1 і α_0 функцію $\Phi(\alpha_1,\ \alpha_0)$, отримуємо систему рівнянь для визначення α_1 і α_0, з якої

$$\begin{cases} \alpha_0 = \overline{y} - \alpha_1 \overline{x}; \\ \alpha_1 = \dfrac{\sum\limits_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum\limits_{i=}^{n} x_i^2 - n\overline{x}^2} = \dfrac{\sum\limits_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}. \end{cases}$$

Якщо значення x_i відомі без похибок, а значення y_i незалежні й точні, оцінку дисперсії (похибку вимірювань) величин y_i визначають за формулою

$$\sigma^2 = \frac{\Phi_{\min}}{n-2}, \text{ де } \Phi_{\min} = \sum_{i=1}^{n} \left(y_i - \alpha_1 x_i - \alpha_0 \right)^2.$$

Оцінки дисперсій коефіцієнтів α_1 і α_0 визначають за формулами

$$\sigma_{\alpha_0}^2 = \frac{\sum\limits_{i=1}^{n} x_i^2}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i \right)^2} \cdot \frac{\Phi_{\min}}{n-2},\ \sigma_{\alpha_1}^2 = \frac{n}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i \right)^2} \cdot \frac{\Phi_{\min}}{n-2}.$$

Якщо величини y_i мають нормальний розподіл, то для коефіцієнтів α_1 і α_0 справджуються такі надійні інтервали:

$$\left(\alpha_k - t(n-2, \gamma)\sigma_{\alpha_k},\ \alpha_k + t(n-2, \gamma)\sigma_{\alpha_k} \right),\ k = 0,\ 1,$$

де α_k – оцінки, отримані методом найменших квадратів, а число $t(n-2,\gamma)$ знаходять відповідно до таблиці Стьюдента за кількості степенів вільності $k=n-2$ і $\gamma = P\left(|t| < t(n-2,\gamma)\right)$.

Приклад 3.16. Для торгових агентів компанії зареєстровано дві ознаки: X – витрати на представництво, Y – обсяг продажів товарів за певний час (в тис. грн). Дані наведено в таблиці:

x_i	6,5	6,5	6,2	6,7	6,9	6,5	6,1	6,7
y_i	105	125	110	120	140	135	95	130

Знайти вибірковий коефіцієнт кореляції, рівняння прямої регресії Y на X, похибку вимірювань, надійні інтервали для коефіцієнтів прямої регресії за $\gamma = 0,95.$ Обсяг вибірки n = 8.

Розв'язання. Проміжні результати обчислень заносимо у таку таблицю:

№	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	$\hat{y}_i$	$\left(y_i-\hat{y}_i\right)^2$
1	6,5	105	682,5	42,25	11025	119,4120	207,7057
2	6,5	125	812,5	42,25	15625	119,4120	31,2257
3	6,2	110	682,0	38,44	12100	105,2981	22,1079
4	6,7	120	804,0	44,89	14400	128,8212	77,8136
5	6,9	140	966,0	47,61	19600	138,2305	3,1311
6	6,5	135	877,5	42,25	18225	119,4110	242,9857
7	6,1	95	579,5	37,21	9025	100,5934	31,2861
8	6,7	130	871,0	44,89	16900	128,8212	1,3896
$\sum$	52,1	960	6275	339,79	116900	–	617,6454

$$\overline{x} = \frac{1}{8}\cdot 52,1 = 6,5125;\quad \overline{y} = \frac{1}{8}\cdot 960 = 120;$$

$$\overline{s}_x^2 = \frac{1}{8}\cdot 339,79 - 6,5125^2 = 0,0611;\quad \overline{s}_x = 0,2472;$$

$$\overline{s}_y^2 = \frac{1}{8}\cdot 116900 - 120^2 = 212,5;\quad \overline{s}_y = 14,5774;$$

$$r = \frac{6275 - 8 \cdot 6,5125 \cdot 120}{8 \cdot 0,2472 \cdot 14,5774} = 0,7978.$$

Записуємо пряму регресії Y на X за формулою $y - \overline{y} = r\frac{\overline{s}_y}{\overline{s}_x}(x - \overline{x})$, тому

$$\alpha_1 = r\frac{\overline{s}_y}{\overline{s}_x} = 0,7978 \cdot \frac{14,5774}{0,2472} = 47,0463;$$

$$\alpha_0 = \overline{y} - \alpha_1\overline{x} = 120 - 47,0463 \cdot 6,5125 = -186,3890;$$

$$\widehat{y} = 47,0463x - 186,3890.$$

Обчислимо:

$$\widehat{y}_1 = 47,0463 \cdot 6,5 - 186,389 = 119,4120;$$

$$\widehat{y}_2 = 47,0463 \cdot 6,5 - 186,389 = 119,4120;$$

$$\widehat{y}_3 = 47,0463 \cdot 6,2 - 186,389 = 105,2981;$$

$$\widehat{y}_4 = 47,0463 \cdot 6,7 - 186,389 = 128,8212;$$

$$\widehat{y}_5 = 47,0463 \cdot 6,9 - 186,389 = 138,2305;$$

$$\widehat{y}_6 = 47,0463 \cdot 6,5 - 186,389 = 119,4120;$$

$$\widehat{y}_7 = 47,0463 \cdot 6,1 - 186,389 = 100,5934;$$

$$\widehat{y}_8 = 47,0463 \cdot 6,7 - 186,389 = 128,8212;$$

$$\Phi_{\min} = 617,6454.$$

Знаходимо

$$\Delta = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 = 8 \cdot 339,79 - 52,1^2 = 3,91.$$

Тоді $\sigma^2 = \frac{617,6454}{8-2} = 102,94044;$

$$\sigma_{\alpha_0}^2 = \frac{339,79}{3,91} \cdot 102,9409 = 8945,8537;\ \sigma_{\alpha_0} = 94,5825;$$

$$\sigma_{\alpha_1}^2 = \frac{8}{3,91} \cdot 102,9409 = 210,6208;\ \sigma_{\alpha_1} = 14,1278.$$

За таблицею розподілу Стьюдента знаходимо $t(6;0,95) = 2,45$. Тоді з імовірністю 0,95

$$-186,389 - 2,45 \cdot 94,5825 < \alpha_0 < -186,389 + 2,45 \cdot 94,5825,$$

тобто $-418{,}1161 < \alpha_0 < 45{,}3381;$

$$47{,}0463 - 2{,}45 \cdot 14{,}1278 < \alpha_1 < 47{,}0463 + 2{,}45 \cdot 14{,}1278,$$

тобто $12{,}4332 < \alpha_1 < 81{,}6594.$

3.2.3. Поняття кореляційного зв’язку між досліджуваними величинами

Означення. **Кореляцією** ***(кореляційним зв’язком***) між випадковими величинами (ознаками) називають наявність статистичного або ймовірнісного зв’язку між ними.

При цьому закономірна зміна певних ознак призводить до закономірної зміни середніх значень інших, пов’язаних з ними ознак.

Означення. Кореляційним аналізом називають сукупність методів виявлення кореляційного зв’язку.

Кореляційний аналіз можна застосовувати для формалізованого подання моделей зв’язків між окремими компонентами системи або між окремими процесами, що відбуваються в ній. Його наявність не означає існування причинно-наслідкового зв’язку між досліджуваними ознаками. Він може бути зумовлений тим, що обидві ознаки мають причинно-наслідковий зв’язок з певним іншим фактором. Кореляція також може бути випадковою.

В багатьох прикладних задачах необхідно виявити залежність між двома властивостями (ознаками) X та Y одного і того ж об'єкту, або між певними ознаками різних об’єктів. Якщо вказані ознаки допускають кількісне вимірювання, і, з погляду теорії, виходячи з характеристики об'єкту, ознака Y залежить від ознаки X, тоді X можна назвати незалежною змінною, або факторною ознакою (просто фактором), а Y – залежною змінною або результативною ознакою.

Якщо кожному значенню факторної ознаки X відповідає одне і тільки одне значення результативної ознаки Y, то говорять, що між цими ознаками існує функціональний зв’язок: $Y = f(X)$. Якщо кожному значенню факторної ознаки

X відповідає безліч значень результативної ознаки *Y*, то між цими ознаками існує статистичний зв’язок.

Вивчення статистичного зв'язку дуже складний і трудомісткий процес, у якому потрібно аналізувати багатомірні таблиці даних. Тому зазвичай вивчається не статистичний, а кореляційний зв'язок між *X* та *Y*. Якщо кожному значенню факторної ознаки *X* відповідає певне середнє значення результативної ознаки *Y*, то між цими ознаками існує кореляційний зв’язок, тобто кореляційною є функціональна залежність між значеннями *X* і середніми значеннями *Y*: $\overline{Y} = f(X)$.

Наприклад, відомо, що з однакових за площею ділянок землі за однакових кількостей внесеного добрива отримують різний урожай. Тому, якщо *Y* – урожайність зерна, а *X* – кількість внесеного добрива, то функціонального зв’язку між *X* та *Y* немає. Це пояснюється впливом таких випадкових факторів, як температура повітря, кількість опадів і т. ін. Однак досвід показує, що середній урожай є функцією від кількості добрива, тобто між *X* та *Y* існує кореляційний зв’язок.

Основними задачами кореляційного аналізу є:

− вивчення сили зв’язку між двома і більше ознаками досліджуваного об’єкту;

− встановлення факторів, що найбільш суттєво впливають на результативну ознаку;

− виявлення невідомих причинно-наслідкових зв’язків між ознаками об’єкту.

Групування даних для кореляційного аналізу

Вибіркові дані для вивчення кореляційного зв’язку між ознаками *X* та *Y* мають вигляд пар їх значень: $(x_1; y_1)$, $(x_2; y_2)$, …, $(x_n; y_n)$, x_i – значення величини *X*, y_i – значення *Y*, *n* – кількість пар значень, $i = \overline{1,n}$.

Якщо кількість пар значень достатньо велика (принаймні, $n > 20$), то для зручності розрахунків дані групуються.

Для групування даних необхідно:

1. Розбити множини значень X та Y на інтервали, їх кількість для X та Y може бути різною (позначення: k – кількість інтервалів для X; m – кількість інтервалів для Y).

2. Зобразити дані графічно: побудувати на площині точки з координатами $(x_i; y_j)$. В результаті отримується площина, розбита на прямокутники, в кожному з яких може бути множина точок (рис. 3.12).

Вказане графічне зображення вибіркових даних називається *полем кореляції.*

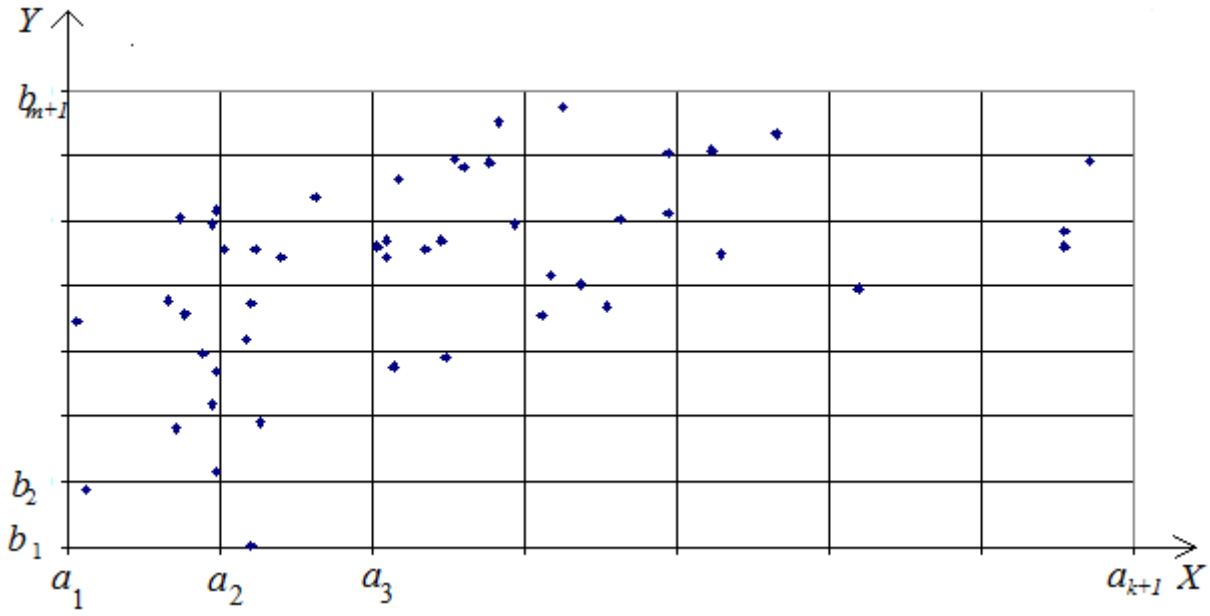


Рис. 3.12. Поле кореляції

3. Побудувати кореляційну таблицю (табл. 3.2). В першому рядку, розбитому на дві частини, записуються інтервали $[a_i; a_{i+1})$ для X та їх середини x_i. У першому стовпчику, розбитому на дві частини, записуються інтервали $[b_j; b_{j+1})$ для Y та їх середини y_j. В центральній частині таблиці записуються частоти n_{ij} – кількість точок, що потратили в прямокутник, обмежений по осі X інтервалом $[a_i; a_{i+1})$ і по осі Y інтервалом $[b_j; b_{j+1})$. В останньому рядку таблиці записуються частоти n_i для X – кількість точок, що потратили в прямокутники,

які відповідають інтервалу $[a_i; a_{i+1})$, тобто $n_i = \sum_{j=1}^{m} n_{ij}$ – сума частот n_{ij} в стовпчики з номером *i*. В останньому стовпчику таблиці записуються частоти n_j для *Y* – кількість точок, що потратили в прямокутники, які відповідають інтервалу $[b_j; b_{j+1})$, тобто $n_j = \sum_{i=1}^{k} n_{ij}$ – сума частот n_{ij} в рядку з номером *j*.

Кореляційну таблицю можна розглядати як своєрідний подвійний статистичний ряд (табл. 3.10).

Табл. 3.10

X (інтервали і їх середини) / *Y* (інтервали і їх середини)		$[a_1; a_2)$	$[a_2; a_3)$	…	$[a_k; a_{k+1})$	$n_j = \sum_{i=1}^{k} n_{ij}$
		x_1	x_2	…	x_k	
$[b_1; b_2)$	y_1	n_{11}	n_{21}	…	n_{k1}	n_1
$[b_2; b_3)$	y_2	n_{12}	n_{22}	…	n_{k2}	n_2
…	…	…	…	…	…	…
$[b_m; b_{m+1})$	y_m	n_{1m}	n_{2m}	…	n_{km}	n_m
$n_i = \sum_{j=1}^{m} n_{ij}$		n_1	n_2	…	n_k	

4. За даними кореляційної таблиці будується ряд, що відображає залежність середнього значення *Y* від *X* (табл. 3.11). В першому рядку таблиці записуються середини інтервалів x_i, в другому – відповідні середні значення $\overline{y}_{x_i}$, що знаходяться за формулами:

$$\overline{y}_{x_1} = \frac{y_1 n_{11} + y_2 n_{12} + \ldots + y_m n_{1m}}{n_1}; \overline{y}_{x_2} = \frac{y_1 n_{21} + y_2 n_{22} + \ldots + y_m n_{2m}}{n_2}; \ldots$$

$$\overline{y}_{x_k} = \frac{y_1 n_{k1} + y_2 n_{k2} + \ldots + y_m n_{km}}{n_k}.$$

Табл. 3.11

x_i	x_1	x_2	…	x_k
$\overline{y}_{x_i}$	$\overline{y}_{x_1}$	$\overline{y}_{x_2}$	…	$\overline{y}_{x_k}$
n_i	n_1	n_2	…	n_k

В результаті отримується статистичний ряд, що містить значення *X*, відповідні середні значення *Y* та частоти. За даними такого ряду проводиться кореляційний аналіз.

3.2.4. Коефіцієнт кореляції Пірсона

Методику кількісного оцінювання кореляції між ознаками вперше було запропоновано британським географом, антропологом та психологом Френсисом Гальтоном в 1888 р. Універсальною характеристикою ступеня тісноти зв'язку між кількісними ознаками є коефіцієнт детермінації.

Вибірковий коефіцієнт детермінації певної ознаки y за вектором незалежних ознак $X = (X_1,\ X_2,\ \ldots,\ X_p)$ можна розрахувати так:

$$K_d\left(y, X\right) = 1 - \frac{s_\varepsilon^2}{s_y^2}, \quad s_y^2 = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - \bar{y}\right)^2,$$

n – кількість спостережень, а вибіркове значення дисперсії нев'язок є обчислюють за однією з таких формул:

$$s_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f^2\left(X_i\right)\right)^2,$$

де $f^2\left(X_i\right)$ – є статистичною оцінкою невідомого значення функції регресії $f\left(X_i\right)$ у точці X_i або:

$$s_\varepsilon^2 = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{\nu_j}\sum_{i=1}^{\nu_j}\left(y_{ji} - \bar{y}_{j_*}\right)^2,$$

де ν_j – кількість даних, що потрапили до *j*-го інтервалу групування; y_{ji} – значення *i*-го спостереження досліджуваної ознаки, що потрапило до *j*-го інтервалу; $\bar{y}_{j_*} = \dfrac{\sum_{i=1}^{\nu_j} y_{ji}}{\nu_j}$ – її середнє значення за спостереженнями, які потрапили до *j*-го інтервалу; m – кількість інтервалів.

Останню формулу обчислення s_ε^2 використовують, якщо умова $D\left(\varepsilon \mid X\right) = \sigma_\varepsilon^2 = \text{const}$ не виконується (умовна дисперсія є залежна від X), а

також у всіх випадках, коли обчислення здійснюють за згрупованими даними. У цьому випадку необхідно попередньо здійснити групування даних. Для цього їх впорядковують за зростанням значень однієї з ознак (X). Потім задають кількість та межі інтервалів для цієї ознаки. Підраховують кількість точок, що потрапили до кожного інтервалу, тобто ν_j, для змінної *Y* обчислюють загальне середнє $\bar{y}$ та середні за інтервалами $\bar{y}_{j_*}$, тоді й розраховують значення коефіцієнта детермінації за наведеними формулами.

Величина коефіцієнта детермінації може змінюватися в межах від нуля до одиниці. Вона відображає частку загальної дисперсії досліджуваної ознаки, яка зумовлена зміною функції регресії $f(X)$. При цьому нульове значення коефіцієнта детермінації відповідає відсутності будь-якого зв'язку, а його рівність одиниці – наявності строго функціонального зв'язку.

Інші поширені характеристики ступеня тісноти зв'язку між ознаками можна розглядати як окремі випадки коефіцієнта детермінації, отримані для конкретних математичних моделей зв'язку.

Розрізняють парні та частинні кореляційні характеристики. Парні характеристики розраховують за результатами вимірювань тільки досліджуваної пари ознак.

Для кількісних ознак найчастіше застосовують **коефіцієнти кореляції Пірсона і Фехнера.** ***Коефіцієнт кореляції Пірсона*** *(коефіцієнт кореляційного відношення Пірсона, парний коефіцієнт кореляції, вибірковий коефіцієнт кореляції, коефіцієнт Бравайса – Пірсона)* вимірює ступінь лінійного кореляційного зв'язку між кількісними скалярними ознаками (запропонований К. Пірсоном у 1896 р).

Отже, для оцінки тісноти (або сили) зв'язку між *X* та *Y* слугує коефіцієнт кореляції Пірсона. Його використовують у випадку, коли між *X* та *Y* існує *лінійний зв'язок та вибіркові дані розподілені за нормальним законом* (також називають *параметричним коефіцієнтом кореляції*).

Як відомо, за формулою (3.31) коефіцієнт кореляції Пірсона також можна записати у вигляді:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{s_x \cdot s_y}, \tag{3.33}$$

де $\overline{x}$ – вибіркове середнє величини X; $\overline{y}$ – вибіркове середнє величини Y; $\overline{xy}$ – вибіркове середнє величини XY; s_x – вибіркове середнє квадратичне відхилення величини X; s_y – вибіркове середнє квадратичне відхилення величини Y.

Враховуючи формули для знаходження вибіркових середніх і середніх квадратичних відхилень, а саме:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i ; \quad \overline{y} = \frac{1}{n}\sum_{j=1}^{m} y_j n_j ; \quad \overline{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m} x_i y_j n_{ij} ;$$

$$s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \left(\frac{1}{n}\sum_{i=1}^{k} x_i n_i\right)^2} ; \quad s_y = \sqrt{\frac{1}{n}\sum_{j=1}^{m} y_j^2 n_j - \left(\frac{1}{n}\sum_{j=1}^{m} y_j n_j\right)^2} ;$$

отримують більш зручну для розрахунків формулу:

$$r = \frac{n\sum_{i=1}^{k}\sum_{j=1}^{m} x_i y_j n_{ij} - \left(\sum_{i=1}^{k} x_i n_i\right)\left(\sum_{j=1}^{m} y_j n_j\right)}{\sqrt{n\sum_{i=1}^{k} x_i^2 n_i - \left(\sum_{i=1}^{k} x_i n_i\right)^2}\sqrt{n\sum_{j=1}^{m} y_j^2 n_j - \left(\sum_{j=1}^{m} y_j n_j\right)^2}}. \tag{3.34}$$

У випадку незгрупованих даних розрахункова формула суттєво спрощується:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\cdot\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}. \tag{3.35}$$

Властивості коефіцієнта кореляції Пірсона

1. Коефіцієнт кореляції Пірсона приймає значення на проміжку $[-1;1]$, тобто $-1 \le r \le 1$.

2. Якщо $|r| \le 0,5$, то зв'язок вважається слабким; якщо $0,5 < |r| \le 0,7$, то зв'язок вважається середнім; $|r| > 0,7$, то зв'язок вважається сильним.

3. Якщо $r > 0$, то зв'язок називається додатнім, тобто зі збільшенням значень *X* значення *Y* також збільшуються. Якщо $r < 0$, то зв'язок називається від'ємним, тобто зі збільшенням значень *X* значення *Y* зменшуються.

Зауваження. Слід пам'ятати, що коефіцієнт кореляції Пірсона показує силу лінійного зв'язку. Якщо між *X* та *Y* існує сильний нелінійний зв'язок, коефіцієнт кореляції Пірсона може дорівнювати нулю.

Оскільки сила зв'язку між *X* та *Y* оцінюється за вибірковими даними, то потрібна перевірка її статистичної значущості, тобто оцінка можливості розповсюдити отримані результати на всю генеральну сукупність.

Зауваження. Застосування коефіцієнта Пірсона як міри зв'язку є обґрунтованим лише за умови, що спільний розподіл пари ознак є нормальним. Тому перед його розрахунком слід перевірити виконання цієї гіпотези. Якщо вона справедлива, то квадрат коефіцієнта кореляції Пірсона дорівнює коефіцієнту детермінації.

Перевірка статистичної значущості коефіцієнта кореляції Пірсона здійснюється за допомогою так званої t-статистики, яка розраховується за формулою

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}. \qquad (3.36)$$

Розраховане значення t-статистики порівнюється з критичним значенням $t_{крит}$ – табличне значення розподілу Стьюдента (додаток), яке також можна знайти за допомогою вбудованої статистичної функції Excel **СТЬЮДРАСПОБР** (α; *l*), де α – обраний дослідником рівень значущості, *l* – степені вільності, $l = n - 2$.

Якщо розраховане значення t-статистики більше критичного $|t| > t_{крит}$, то коефіцієнт кореляції вважається значимим на обраному рівні α.

Приклад 3.17. За наявними даними про рівень механізації праці *X* (%) і продуктивності праці *Y* (од. продукції/год.) для 14 однотипних підприємств

(табл. 3.12) оцінити тісноту зв'язку між X та Y. Визначити можливість розповсюдження результатів розрахунків на всі підприємства такого типу.

Табл. 3.12

X	32	30	36	40	41	47	56	54	60	55	61	67	69	76
Y	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Розв'язання. Дані табл. 3.12 є вибіркою значень X та відповідних значень Y. Оскільки кількість даних невелика ($n=14$), то їх можна не групувати. Для оцінки тісноти зв'язку між X та Y розрахуємо коефіцієнт кореляції Пірсона за формулою (3.35) для незгрупованих даних. Розрахунки для зручності оформимо у вигляді таблиці (табл. 3.13).

Табл. 3.13

x_i	y_i	x_i^2	y_i^2	$x_i\, y_i$
32	20	1024	400	640
30	24	900	576	720
36	28	1296	784	1008
40	30	1600	900	1200
41	31	1681	961	1271
47	33	2209	1089	1551
56	34	3136	1156	1904
54	37	2916	1369	1998
60	38	3600	1444	2280
55	40	3025	1600	2200
61	41	3721	1681	2501
67	43	4489	1849	2881
69	45	4791	2025	3105
76	48	5779	2304	3848
Суми				
724	492	40134	18138	26907

Отже, за відомими формулами обчислення коефіцієнта кореляції отримуємо його числове значення.

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2}} = \frac{14 \cdot 26907 - 724 \cdot 492}{\sqrt{14 \cdot 40134 - 724^2}\sqrt{14 \cdot 18138 - 492^2}} =$$

$$= \frac{20490}{\sqrt{37700}\sqrt{11868}} \approx 0{,}969.$$

Таке ж значення отримується за допомогою вбудованої функції **КОРРЕЛ**(массив1, массив2) (рис. 3.13, 3.14).

За значенням коефіцієнта кореляції можна зробити висновок, що між *X* та *Y* існує сильний додатній зв'язок.

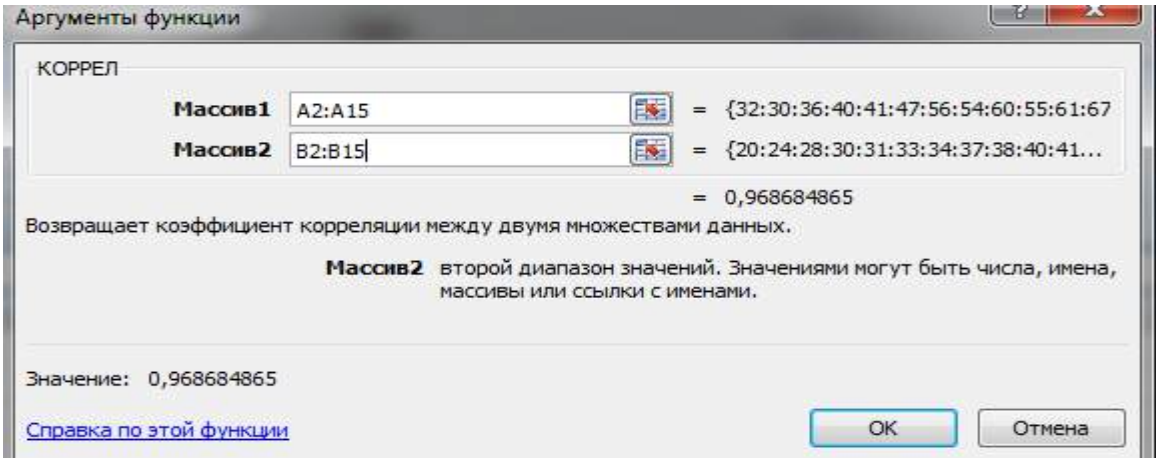
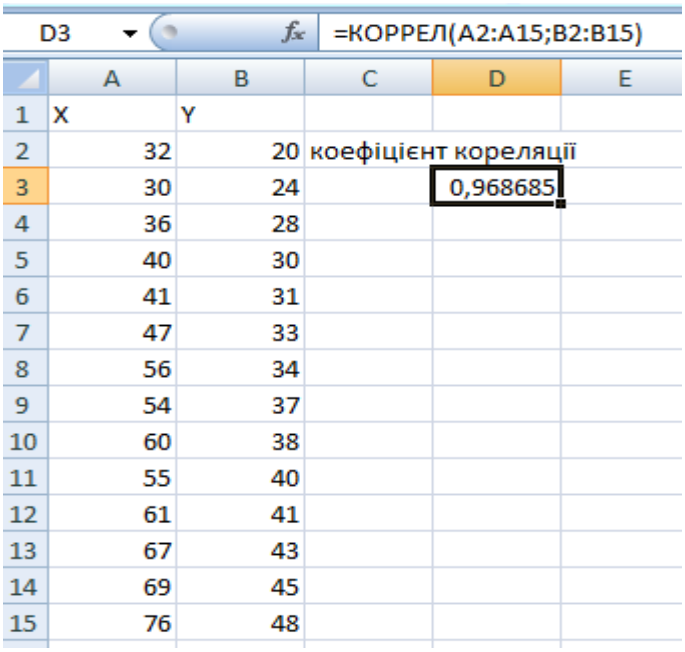


Рис. 3.13. Обчислення за вбудованою функцією **КОРРЕЛ**



	A	B	C	D	E
1	X	Y			
2	32	20	коефіцієнт кореляції		
3	30	24		0,968685	
4	36	28			
5	40	30			
6	41	31			
7	47	33			
8	56	34			
9	54	37			
10	60	38			
11	55	40			
12	61	41			
13	67	43			
14	69	45			
15	76	48			

Рис. 3.14. Обчислення коефіцієнт кореляції

Перевіримо статистичну значущість знайденого коефіцієнта кореляції Пірсона. Розрахуємо t-статистику за формулою (3.36):

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}969\sqrt{14-2}}{\sqrt{1-0{,}969^2}} \approx 13{,}59.$$

Знайдемо $t_{крит}$, враховуючи, що $l = n-2 = 14-2 = 12$. Оберемо рівень значущості α=0,01. Тоді $t_{крит}$=**СТЬЮДРАСПОБР** (0,01; 12) =3,055.

Оскільки розраховане значення t-статистики більше критичного $13{,}59 > 3{,}055$, то коефіцієнт кореляції можна вважати значимим на обраному рівні α=0,01.

Висновок. Між рівнем механізації праці та її продуктивністю на підприємствах, що досліджувалися, існує сильний додатній зв'язок: чим більше рівень механізації праці, тим вище її продуктивність. Висновок дійсний для всіх підприємств такого типу.

3.2.5. Коефіцієнти кореляції рангів

Нехай елементи генеральної сукупності мають якісні ознаки *A* та *B*, тобто ознаки, які не піддаються кількісним оцінкам, але дають змогу порівнювати елементи між собою і, отже, розміщувати їх у порядку погіршення якості.

Нехай вибірка обсягу *n* містить незалежні елементи, що мають якісні ознаки *A* та *B*. Розмістимо елементи в порядку погіршення якості за ознакою *A*. Поставимо у відповідність елементу, що стоїть на *i*-му місці, число – ранг x_i, рівний порядковому номеру елемента $x_i = i$. Потім розмістимо елементи в порядку спадання якості за ознакою *B* і припишемо кожному з них ранг (порядковий номер) y_i, причому для зручності порівняння рангів індекс *i* дорівнює порядковому номеру елемента за ознакою *A*. Наприклад, $y_3 = 4$, означає, що за ознакою *A* елемент стоїть на третьому місці, а за ознакою *B* – на четвертому.

В результаті отримаємо дві послідовності рангів:

– за ознакою *A*: $x_1, x_2, \ldots, x_n$;

– за ознакою *B*: $y_1, y_2, \ldots, y_n$.

Позначають $d_i = x_i - y_i$ – різниця рангів.

Означення. Коефіцієнт щільності зв'язку між *A* і *B,* який визначається формулою

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} \tag{3.37}$$

називають **коефіцієнтом кореляції рангів Спірмена.**

Запропонований британським психологом Чарльзом Едвардом Спірменом у 1904 р. Його використовують, якщо досліджується зв'язок між рядами даних, виміряними за порядковою шкалою. Його можна застосовувати також і для кількісних даних, але, як правило, це буває недоцільним. У найпростішому випадку досліджувані об'єкти класифікують за двома ознаками. Наприклад, ми можемо спочатку впорядкувати групу студентів за їх здібностями до вищої математики, а потім – до іноземних мов.

Значення коефіцієнта можуть змінюватися в межах від −1 до +1, при цьому значення −1 відповідає повній протилежності послідовностей рангів, а +1 – їх повному збігу. Коефіцієнт рангової кореляції Спірмена можна застосовувати як показник некорельованості вибірок. Для великих за обсягом вибірок (n > 50) статистика цього критерію наближається до розподілу Стьюдента з (n-2) степенями вільності.

Якщо не можна визначити рангову відмінність деяких елементів, то беруть середній ранг. У цьому разі використовують ***коефіцієнт кореляції рангів Кендала*** **(**запропонований британським статистиком Маурісом Кендалом у 1938 р):

$$\tau = \frac{\frac{n^3 - n}{6} - \left(T_x + T_y\right) - \sum_{i=1}^{n} d_i^2}{\sqrt{\left(\frac{n^3 - n}{6} - 2T_x\right)\left(\frac{n^3 - n}{6} - 2T_y\right)}}, \tag{3.38}$$

де $T_x = T_y = \frac{\sum_{i=1}^{l} (t_i^3 - t_i)}{12}$, t_i – число об'єднаних рангів для X та Y.

Коефіцієнт рангової кореляції Кендала призначений для визначення сили кореляційного зв'язку між двома рядами даних за тих самих умов, що і

коефіцієнт рангової кореляції Спірмена. Як і для коефіцієнта Спірмена, його значення можуть змінюватися в межах від −1 до +1, при цьому: −1 відповідає повній протилежності послідовностей рангів, а +1 – їх повному збігу. Слід зазначити, що обчислення коефіцієнта Кендала є більш трудомістким, але з іншого боку, він має ряд переваг порівняно із коефіцієнтом Спірмена.

Основними з них є такі:

– кращий рівень вивченості його статистичних властивостей, зокрема його вибіркового розподілу;

– можливість його застосування для визначення частинної кореляції;

– більша зручність перерахунку при додаванні нових даних.

Абсолютна величина коефіцієнта кореляції рангів не більше одиниці:

$$|\rho| \le 1.$$

Що ближче до нуля $|\rho|$, то залежність між якісними ознаками *A* та *B* менша.

Приклад 3.18. Два інспектори *A* і *B* перевірили 12 водіїв на швидкість реакції і розмістили їх у порядку погіршення реакції (в дужках вказано порядкові номери водіїв з однаковою реакцією).

A	1	(2, 3, 4)	5	(6, 7, 8)	9	10	11	12
B	3	1 2 6	4	5 7 8	11	10	9	12

Знайти коефіцієнти кореляції рангів Спірмена і Кендела.

Розв'язання. Ранги водіїв з номерами 2, 3, 4 дорівнюють $\frac{2+3+4}{3} = 3$, водіїв з номерами 6, 7, 8 відповідно $\frac{6+7+8}{3} = 7.$ Запишемо послідовність рангів

A	1	3	3	3	5	7	7	7	9	10	11	12
B	3	1	2	6	4	5	7	8	11	10	9	12

$$\sum d_i^2 = 4+4+1+9+1+4+1+4+4 = 32.$$

Коефіцієнт кореляції рангів Спірмена

$$\rho = 1 - \frac{6 \cdot 32}{12^3 - 12} = 0{,}8881.$$

У першій послідовності маємо трьох водіїв з однаковими рангами 3 і трьох водіїв з однаковими рангами 7. Тоді

$$T_x = \frac{(3^3 - 3) + (3^3 - 3)}{12} = 4,\ T_y = 0.$$

Коефіцієнт кореляції рангів Кендалла

$$\tau = \frac{\frac{12^3 - 12}{6} - (4 + 0) - 32}{\sqrt{\left(\frac{12^3 - 12}{6} - 8\right)\left(\frac{12^3 - 12}{6} - 0\right)}} = \frac{250}{281{,}9716} = 0{,}8866.$$

3.2.6. Множинний та частинний коефіцієнти кореляції

Про множинну кореляцію мова йде в тому випадку, коли певна ознака може бути пов'язана не з однією, а із сукупністю декількох інших ознак.

У реальних дослідженнях можлива ситуація, коли на певну ознаку може впливати не одна, а декілька інших. В таких випадках парні показники кореляції будуть давати неправильну інформацію щодо наявності зв'язку між відповідними показниками, оскільки їх значення будуть викривлятися невраховуваними ознаками.

Для уникнення помилок використовують частинні показники кореляції, що усувають такий вплив. Ідея введення таких показників вперше була висунута Г.У. Юлом у 1896 р., а пізніше розвинена ним та К. Пірсоном.

Якщо досліджувані ознаки задовольняють багатовимірний нормальний розподіл, при фіксованих значеннях інших ознак розраховують то ***частинний коефіцієнт кореляції*** між двома ознаками *i* та *j*.

Частинні коефіцієнти кореляції мають всі властивості парних коефіцієнтів кореляції. Вони є показниками наявності лінійного зв'язку між двома незалежними ознаками, який не залежить від впливу інших ознак. Тісноту зв'язку між декількома змінними у випадку множинної регресії можна оцінити за допомогою ***коефіцієнта множинної кореляції.***

Нехай досліджуваний об'єкт або явище характеризується більш ніж двома ознаками X_1, X_2, …, X_k, та необхідно вивчати множинні залежності. Для оцінки сили зв'язку між певною ознакою X_i та усіма іншими ознаками слугує ***множинний коефіцієнт кореляції***, який позначається R_i.

Для розрахунку ***множинного коефіцієнта кореляції*** потрібно:

1. Побудувати матрицю парних коефіцієнтів кореляції $r_{ij}, i=\overline{1,k}$ між ознаками X_i та X_j:

$$A=\begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & r_{kk} \end{pmatrix}. \tag{3.39}$$

2. Знайти визначник $|A|$ матриці A та алгебраїчне доповнення A_{ii} елемента r_{ii} цієї матриці.

3. Розрахувати множинний коефіцієнт кореляції за формулою:

$$R_i = \sqrt{1-\frac{|A|}{A_{ii}}}\,. \tag{3.40}$$

Перевірка статистичної значущості множинного коефіцієнта кореляції здійснюється за допомогою t-статистики, яка розраховується за формулою:

$$t=\frac{R^2(n-k)}{(1-R^2)(k-1)}, \tag{3.41}$$

де n – кількість взаємопов'язаних значень ознак $X_i, i=\overline{1,k}$.

Розраховане значення t-статистики порівнюється з критичним значенням $F_{крит}$. $F_{крит}$ – табличне значення розподілу Фішера (додаток), яке також можна знайти за допомогою вбудованої статистичної функції Excel **FРАСПОБР** (α; l_1; l_2), де $\alpha-$ обраний дослідником рівень значущості, l_1; l_2- степені вільності, $l_1=k-1$; $l_2=n-k$.

Якщо розраховане значення t-статистики більше критичного $|t| > F_{крит}$, то множинний коефіцієнт кореляції вважається значимим на обраному рівні значущості α.

У випадку, коли необхідно дослідити кореляційний зв'язок між ознаками X_i та X_j, $i = \overline{1,k}$, $j = \overline{1,k}$, із множини ознак $X_1, X_2, ..., X_k$ досліджуваного об'єкту або явища, вільний від впливу всіх інших ознак, розраховується *частинний коефіцієнт* кореляції, який позначається R_{ij}.

Для розрахунку ***частинного коефіцієнта кореляції*** потрібно:

1. Побудувати матрицю парних коефіцієнтів кореляції A.

2. Знайти алгебраїчні доповнення A_{ii}, A_{jj}, A_{ij} елементів r_{ii}, r_{jj}, r_{ij} відповідно.

3. Розрахувати частинний коефіцієнт кореляції за формулою:

$$R_{ij} = \frac{-A_{ij}}{\sqrt{A_{ii}A_{jj}}}. \qquad (3.42)$$

Перевірка статистичної значущості частинного коефіцієнта кореляції здійснюється за допомогою t-статистики, яка розраховується за формулою:

$$t = \frac{R_{ij}\sqrt{n-k+2}}{\sqrt{1-R_{ij}^{\,2}}}, \qquad (3.43)$$

де n – кількість взаємопов'язаних значень ознак X_i, $i = \overline{1,k}$.

Розраховане значення t-статистики порівнюється з критичним значенням $t_{крит}$. $t_{крит}$ – табличне значення розподілу Стьюдента, яке також можна знайти за допомогою вбудованої статистичної функції Excel **СТЬЮДРАСПОБР** (α; l), де α – обраний дослідником рівень значущості, l – степені вільності, $l = n - k + 2$.

Якщо розраховане значення t-статистики більше критичного $|t| > t_{крит}$, то частинний коефіцієнт кореляції вважається значимим на обраному рівні значущості α.

Зауваження.

1. Вважається, що для коректного використання множинного і частинного коефіцієнтів кореляції необхідно, щоб вибіркові дані мали сумісний нормальний

розподіл, однак перевірка цієї умови на практиці зазвичай не виконується, оскільки пов'язана зі значними труднощами у розрахунках.

2. Замість парного коефіцієнта кореляції Пірсона можна використовувати також парний коефіцієнт кореляції Спірмена.

3. Кореляційна матриця завжди симетрична відносно головної діагоналі, оскільки $r_{ij} = r_{ji}, i = \overline{1,k},\ j = \overline{1,k}$. Елементи головної діагоналі завжди дорівнюють 1, оскільки вони є коефіцієнтами кореляції X_i та X_i.

Приклад 3.19. Для вивчення залежності урожайності зернових культур *Z* (ц/га) від якості пашні *X* (бали) і кількості внесеного добрива *Y* (кг/га) було проведено дослідження 6 фермерських господарств, результати якого надано у табл. 3.14. Визначити тісноту зв'язку між *Z* та *X* та *Y*, використовуючи множинний коефіцієнт кореляції. Порівняти силу зв'язку між *Z* та *X* та між *Z* та *Y* за частинними коефіцієнтами кореляції.

Табл. 3.14

X	26	35	36	40	41	45
Y	2,1	2,3	2,4	2,6	2,9	3
Z	18	21	22,1	25,3	28	28,5

Розв'язання. За умов задачі необхідно для об'єкту, що характеризується трьома ознаками *X, Y* та *Z* ($k = 3$), розрахувати множинний коефіцієнт кореляції R_Z і частинні коефіцієнти кореляції R_{XZ} та R_{YZ} на основі 6 взаємопов'язаних трійок вибіркових даних $(x_i, y_i, z_i),\ i = \overline{1,n},\quad n = 6$.

Побудуємо матрицю парних коефіцієнтів кореляції, які обчислимо за формулою (3. 35) (випадок незгрупованих даних).

Розрахунки для зручності оформимо у вигляді таблиці (табл. 3.15).

Табл. 3.15

Розрахункова таблиця							Суми
x_i	26	35	36	40	41	45	223
y_i	2,1	2,3	2,4	2,6	2,9	3	15,3
z_i	18	21	22,1	25,3	28	28,5	142,9

x_i^2	676	1225	1296	1600	1681	2025	8503
y_i^2	4,41	5,29	5,76	6,76	8,41	9	39,63
z_i^2	324	441	488,41	640,09	784	812,25	3489,75
$x_i\, y_i$	54,6	80,5	86,4	104	118,9	135	579,4
$x_i\, z_i$	468	735	795,6	1012	1148	1282,5	5441,1
$y_i\, z_i$	37,8	48,3	53,04	65,78	81,2	85,5	371

Отже, за формулами маємо:

$$r_{XY} = r_{YX} = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right)}{\sqrt{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2}\sqrt{n\sum\limits_{i=1}^{n} y_i^2 - \left(\sum\limits_{i=1}^{n} y_i\right)^2}} = \frac{6\cdot 579,4 - 223\cdot 15,3}{\sqrt{6\cdot 8503 - 223^2}\sqrt{6\cdot 39,63 - 15,3^2}} \approx$$

$$\approx 0,935;$$

$$r_{XZ} = r_{ZX} = \frac{n\sum\limits_{i=1}^{n} x_i z_i - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} z_i\right)}{\sqrt{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2}\sqrt{n\sum\limits_{i=1}^{n} z_i^2 - \left(\sum\limits_{i=1}^{n} z_i\right)^2}} = \frac{6\cdot 5441,1 - 223\cdot 142,9}{\sqrt{6\cdot 8503 - 223^2}\sqrt{6\cdot 3489,75 - 142,9^2}} \approx$$

$$\approx 0,954;$$

$$r_{YZ} = r_{ZY} = \frac{n\sum\limits_{i=1}^{n} y_i z_i - \left(\sum\limits_{i=1}^{n} y_i\right)\left(\sum\limits_{i=1}^{n} z_i\right)}{\sqrt{n\sum\limits_{i=1}^{n} y_i^2 - \left(\sum\limits_{i=1}^{n} y_i\right)^2}\sqrt{n\sum\limits_{i=1}^{n} z_i^2 - \left(\sum\limits_{i=1}^{n} z_i\right)^2}} = \frac{6\cdot 371,62 - 15,3\cdot 142,9}{\sqrt{6\cdot 39,63 - 15,3^2}\sqrt{6\cdot 3489,75 - 142,9^2}} \approx$$

$$\approx 0,991;$$

Таким чином, побудовано кореляційну матрицю вигляду:

$$A = \begin{pmatrix} 1 & 0,935 & 0,954 \\ 0,935 & 1 & 0,991 \\ 0,954 & 0,991 & 1 \end{pmatrix}.$$

Знайдемо визначник $|A|$ матриці A та алгебраїчне доповнення $A_{ZZ} = A_{33}$:

$$|A| = \begin{vmatrix} 1 & 0,935 & 0,954 \\ 0,935 & 1 & 0,991 \\ 0,954 & 0,991 & 1 \end{vmatrix} = 1 + 2\cdot 0,935\cdot 0,991\cdot 0,954 - 0,954^2 - 0,991^2 -$$

$$- 0,935^2 \approx 0,0015;$$

$$A_{ZZ} = A_{33} = (-1)^{3+3}\begin{vmatrix} 1 & 0{,}935 \\ 0{,}935 & 1 \end{vmatrix} = 1 - 0{,}935^2 \approx 0{,}1258;$$

тоді $R_Z = R_3 = \sqrt{1 - \frac{|A|}{A_{33}}} = \sqrt{1 - \frac{0{,}0015}{0{,}1258}} \approx 0{,}994.$ Значення множинного коефіцієнта кореляції R_z показує, що величина *Z* сильно пов'язана з *X* та *Y*.

Зауваження. Можна значно спростити розрахунки, використовуючи вбудовану математичну функцію **МОПРЕД**, яка дозволяє знайти визначник заданої матриці (рис. 3.15), а також задавати формули для обчислення множинного коефіцієнта кореляції

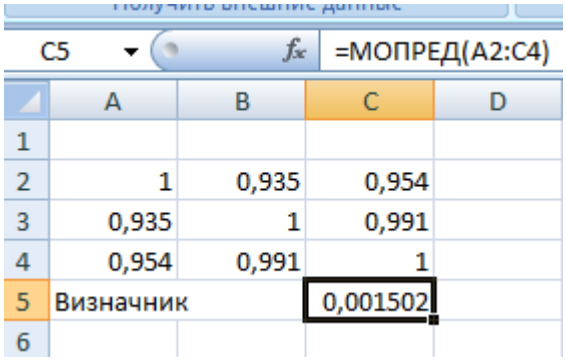


Рис. 3.15. Обчислення визначника матриці

Перевіримо статистичну значущість множинного коефіцієнта кореляції R_Z. Знайдемо t-статистику за формулою (3.41):

$$t = \frac{R^2(n-k)}{(1-R^2)(k-1)} = \frac{0{,}994^2(6-3)}{(1-0{,}994^2)(3-1)} \approx 124{,}09.$$

Знайдемо $F_{крит}$, враховуючи, що $l_1 = k - 1 = 3 - 1 = 2, l_2 = n - k = 6 - 3 = 3.$ Оберемо рівень значущості α=0,01. Тоді $F_{крит}$=**FРАСПОБР** (0,01; 2; 3)=30,82. Оскільки t> $F_{крит}$, то множинний коефіцієнт кореляції R_Z є статистично значимим на рівні значущості α=0,01.

Для обчислення частинних коефіцієнтів кореляції $R_{XZ} = R_{13}$ та $R_{YZ} = R_{23}$ знайдемо алгебраїчні доповнення:

$$A_{13} = (-1)^{1+3}\begin{vmatrix} 0{,}935 & 1 \\ 0{,}954 & 0{,}991 \end{vmatrix} = 0{,}935 \cdot 0{,}991 - 0{,}954 \approx -0{,}027;$$

$$A_{23}=(-1)^{2+3}\begin{vmatrix}1 & 0,935\\ 0,954 & 0,991\end{vmatrix}=(-1)(0,991-0,935\cdot 0,954)\approx -0,099;$$

$$A_{11}=(-1)^{1+1}\begin{vmatrix}1 & 0,991\\ 0,991 & 1\end{vmatrix}=(1-0,991^2)\approx 0,018;$$

$$A_{22}=(-1)^{2+2}\begin{vmatrix}1 & 0,954\\ 0,954 & 1\end{vmatrix}=(1-0,954^2)\approx 0,09.$$

Тоді за формулою (3.42) маємо:

$$R_{13}=\frac{-A_{13}}{\sqrt{A_{11}A_{33}}}=\frac{-(-0,027)}{\sqrt{0,018\cdot 0,126}}\approx 0,577;\qquad R_{23}=\frac{-A_{23}}{\sqrt{A_{22}A_{33}}}=\frac{-(-0,099)}{\sqrt{0,09\cdot 0,126}}\approx 0,929.$$

Значення частинних коефіцієнтів кореляції показують, що величина *Z* пов'язана з величиною *Y* сильніше, ніж з величиною *X*.

Перевіримо статистичну значущість частинного коефіцієнта кореляції R_{13}. Знайдемо t-статистику за формулою (3.43):

$$t=\frac{R_{ij}\sqrt{n-k+2}}{\sqrt{1-R_{ij}^{\ 2}}}=\frac{0,577\sqrt{6-3+2}}{\sqrt{1-0,577^2}}\approx 1,581.$$

Знайдемо критичне значення $t_{крит}$, враховуючи, що $l=n-k+2=5$. Оберемо рівень значущості α=0,01.

Тоді $t_{крит}$= **СТЬЮДРАСПОБР** (0,01;5) =4,032 (рис. 3.16). Оскільки розраховане значення t-статистики менше критичного $|t|<t_{крит}$, то частинний коефіцієнт кореляції R_{13} не є значимим на рівні значущості α=0,01.

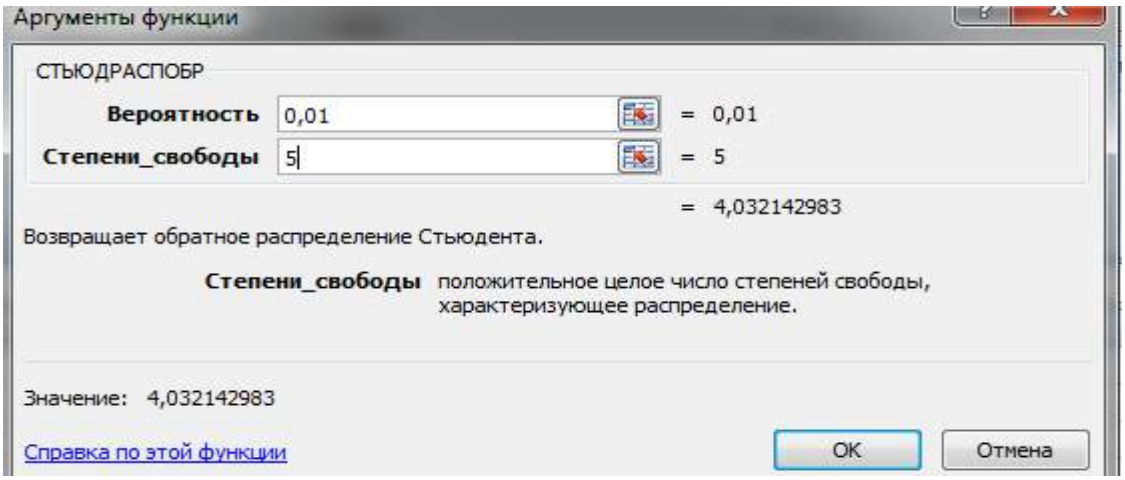


Рис. 3.16. Обчислення за допомогою функції Excel **СТЬЮДРАСПОБР**

Якщо $|t| > t_{кр}$, то вважають коефіцієнт кореляції значущим, якщо ж $|t| < t_{кр}$ − то незначущим.

Відповідно до рівня значущості розглядається *часткова класифікація кореляційних зв'язків:*

1) висока значима кореляція – для *r*, який відповідає рівню статистичної значущості $\alpha \le 0{,}01$;

2) значима кореляція – для *r*, який відповідає рівню статистичної значущості $\alpha \le 0{,}05$;

3) тенденція достовірного зв'язку – для *r*, який відповідає рівню статистичної значущості $\alpha \le 0{,}1$;

4) незначима кореляція − при *r*, який не досягає рівня статистичної значущості.

Перевіримо статистичну значущість частинного коефіцієнта кореляції R_{23}. Знайдемо t-статистику:

$$t = \frac{R_{ij}\sqrt{n-k+2}}{\sqrt{1-R_{ij}^{\ 2}}} = \frac{0{,}929\sqrt{6-3+2}}{\sqrt{1-0{,}929^2}} \approx 5{,}614\,.$$

Оскільки розраховане значення t-статистики більше критичного $|t| > t_{крит}$, то частинний коефіцієнт кореляції R_{23} є значимим на рівні значущості α=0,01.

Висновок. Урожайність зернових культур сильно пов'язана з якістю пашні і кількістю внесеного добриву. При цьому урожайність значно сильніше залежіть від кількості добрива, ніж від якості пашні. Сила зв'язку між урожайністю та якістю пашні середня і не є статистично значимою.

Зауваження. Вбудовані сервісні функції Microsoft Excel дозволяють розраховувати коефіцієнти кореляції Пірсона.

Приклад і результати розрахунків парних коефіцієнтів кореляції надано на рис. 3.17. Клітинки матриці, що розташовані вище головної діагоналі звичайно надаються незаповненими, оскільки матриця симетрична відносно головної діагоналі.

	A	B	C	D	E	F	G	H	I
1									
2		Кореляційний аналіз							
3	Вхідні дані								
4		Значення					Стовп. 1	Стовп. 2	Стовп. 3
5	1	1	328	0,054		Рядок 1	1		
6	2	2	329	0,101		Рядок 2	0,8914	1	
7	3	3	329	0,099		Рядок 3	0,563423	0,692214	1
8	4	4	345	0,019					
9	5	5	352	0,065					
10	6	6	370	0,053					
11	7	7	377	0,178					
12	8	8	385	0,174					
13	9	9	396	0,289					
14	10	10	399	0,195					
15	11	11	390	0,102					
16	12	12	373	0,138					
17									

Рис. 3.17. Результати розрахунку коефіцієнтів кореляції

Зауваження. Використання описаної теорії для опрацювання, дослідження експериментальних даних − проведено розрахунок величин кореляції та встановлення кореляційних зв'язків [11], а саме: досліджено залежності найбільшого опору проникнення води $T_{\max}$ від показника пористості по точці (T_{95}) і поверхневим об'ємом води W. Відповідно, позначено ці величини, як Z, X та Y, їх результати вимірювання та всі потрібні розрахунки для зручності оформлено у вигляді таблиці (табл. 3.16).

Табл. 3.16

Розрахункова таблиця							$\sum$
x_i	0,99	0,58	0,66	1,06	4,69	11,37	19,35
y_i	0,16	0,03	0,04	0,32	6,13	21,4	28,08
z_i	2,28	0,19	0,19	2,36	3,01	4,88	12,91
x_i^2	0,9801	0,3364	0,4356	1,1236	21,9961	129,2769	154,1487
y_i^2	0,0256	0,0009	0,0016	0,1024	37,5769	457,96	495,6674
z_i^2	5,1984	0,0361	0,0361	5,5696	9,0601	23,8144	43,7174
$x_i\, y_i$	0,1584	0,0174	0,0264	0,3392	28,7497	243,318	272,6091
$x_i\, z_i$	2,2572	0,1102	0,1254	2,5016	14,1169	55,4856	74,5969
$y_i\, z_i$	0,3648	0,0057	0,0076	0,7552	18,4513	104,432	124,0166

Кореляційна матриця має вигляд:

$$A=\begin{pmatrix} 1 & 0,9959 & 0,862 \\ 0,9959 & 1 & 0,8347 \\ 0,862 & 0,8347 & 1 \end{pmatrix}.$$

Обчислено визначник $|A|$ матриці A (вбудована математична функція Microsoft Excel **МОПРЕД)** та алгебраїчне доповнення $A_{ZZ}=A_{33}$.

$$|A|=\begin{vmatrix} 1 & 0,9959 & 0,862 \\ 0,9959 & 1 & 0,8347 \\ 0,862 & 0,8347 & 1 \end{vmatrix}=0,0156;$$

$$A_{ZZ}=A_{33}=(-1)^{3+3}\begin{vmatrix} 1 & 0,9959 \\ 0,9959 & 1 \end{vmatrix}=1-0,9959^2\approx 0,0083.$$

Тоді за відомою формулою

$$R_Z=R_3=\sqrt{1-\frac{|A|}{A_{33}}}=\sqrt{1-\frac{0,0156}{0,0083}}\approx 0,9008.$$

Значення множинного коефіцієнта кореляції R_z показує, що величина Z сильно пов'язана з X та Y.

Перевірено статистичну значущість множинного коефіцієнта кореляції R_Z. Знайдено t-статистику за формулою:

$$t=\frac{R^2(n-k)}{(1-R^2)(k-1)}=\frac{0,9008^2(6-3)}{(1-0,9008^2)(3-1)}\approx 6,45,$$

також $F_{крит}$=**ГРАСПОБР**(0,085;2; 3)=6,259. Оскільки $t>F_{крит}$, то множинний коефіцієнт кореляції R_Z є статистично значимим на обраному рівні значущості (тенденція достовірного зв'язку).

Для частинних коефіцієнтів кореляції $R_{XZ}=R_{13}$ та $R_{YZ}=R_{23}$ обчислено алгебраїчні доповнення: $A_{13}\approx -0,0308$; $A_{23}\approx 0,0238$; $A_{11}\approx 0,3032$; $A_{22}\approx 0,2569$.

Тоді, відповідно, за формулою (3.42):

$$R_{13} = \frac{-A_{13}}{\sqrt{A_{11}A_{33}}} = \frac{-(-0,0308)}{\sqrt{0,3032 \cdot 0,008}} \approx 0,615;$$

$$R_{23} = \frac{-A_{23}}{\sqrt{A_{22}A_{33}}} = \frac{-0,0238}{\sqrt{0,2569 \cdot 0,008}} \approx -0,5158.$$

Значення частинних коефіцієнтів кореляції показують, що величина *Z* пов'язана з величиною *X* сильніше, ніж з величиною *Y*.

Перевірка статистичної значущості, t-статистики за формулою (3.43) дає

$$t_1 = \frac{R_{13}\sqrt{n-k+2}}{\sqrt{1-R_{13}^{\ 2}}} = \frac{0,615 \cdot \sqrt{6-3+2}}{\sqrt{1-0,615^2}} \approx 1,744;\ t_2 \approx -1,346.$$

Критичне значення $t_{крит}$, обрано за рівнем значущості α=0,05 та $\alpha = 0,01$. Тоді $t_{крит}$=**СТЬЮДРАСПОБР**(0,05;5)=2,57 і, відповідно,

$t_{крит}$=**СТЬЮДРАСПОБР**(0,01; 5) =4,03.

Оскільки розраховані значення t-статистики для R_{13} та R_{23} менші критичного $|t| < t_{крит}$, то частинні коефіцієнти кореляції не є значимими на обраних рівнях значущості. Однак, для $\alpha = 0,25$ $t_{крит}$=**СТЬЮДРАСПОБР(**0,25; 5) =1,3, $|t| > t_{крит}$, то частинні коефіцієнти кореляції R_{13} та R_{23} є значимими на рівні значущості α=0,25, причому наявна незначима кореляція.

Завдання для самоконтролю

1. Інтервальна оцінка параметрів розподілу генеральної сукупності, надійний інтервал, рівень надійності.

2. Побудова надійного інтервалу для оцінювання математичного сподівання нормального розподілу за відомого середнього квадратичного відхилення, навести приклад.

3. Побудова надійного інтервалу для оцінювання математичного сподівання нормального розподілу за невідомого середнього квадратичного відхилення, навести приклад.

4. Показати на прикладі побудову надійного інтервалу для оцінки середнього квадратичного відхилення σ нормального розподілу.

5. Оцінити ймовірність біномного розподілу за відносною частотою.

6. Які вбудовані статистичні функції Excel використовують для побудови надійних інтервалів, пояснити їх доцільність використання на прикладах.

7. Метод найменших квадратів ,його практичне застосування.

8. Навести приклад побудови ліній регресії.

9. На якій властивості корельованих ознак ґрунтується коефіцієнт рангової кореляції Спірмена?

10. На якій властивості корельованих ознак ґрунтується коефіцієнт рангової кореляції Кендалла?

11. Яких значень можуть набувати показники рангової кореляції і про що свідчать їх значення?

12. Для заданого набору (вказати самостійно) даних розрахувати значення коефіцієнтів рангової кореляції Спірмена й Кендалла і зробити висновок про наявність кореляційного зв'язку.

13. Пояснити побудову множинного коефіцієнта кореляції.

14. Побудова частинного коефіцієнтів кореляції.

15. Пояснити доцільність використання вбудованих статистичних функцій Excel **FРАСПОБР**, **СТЬЮДРАСПОБР**.

16. Пояснити застосування функцій програмного забезпечення Excel для розв'язування практичних задач кореляційного аналізу.

17. За даною вибіркою $X=\left(-2,1,2,0,3,4,2,5,0,1\right), \sigma=1$ знайти точкову оцінку математичного сподівання, а також зміщену і незміщену оцінки дисперсії та середньоквадратичного відхилення. Побудувати надійний інтервал із ймовірністю 95%, тобто $\gamma=0,95$ для вказаного σ і надійний інтервал у випадку невідомої дисперсії із значущістю 5%. Порівняти побудовані надійні інтервали.

18. Відомо, що $\sigma=2, \quad \bar{x}=5,4, \quad n=10, \quad \gamma=0,95$. Знайти надійний інтервал для оцінки математичного сподівання із заданою надійністю.

Відповідь. $4,16<a<6,64$.

19. Побудувати надійний інтервал із ймовірністю 95% при $\sigma=3$ і надійний інтервал у випадку невідомої дисперсії із ймовірністю 95%. Порівняти

побудовані надійні інтервали.

20. Відомо, що $s=1,5,\ \bar{x}=16,8,\ n=12,\ \gamma=0,95.$ Користуючись розподілом Стьюдента, знайти надійний інтервал для оцінки невідомого математичного сподівання із заданою надійністю.

Відповідь. $15,85<a<17,75.$

21. За даними 16-ти незалежних вимірювань фізичної величини обчислено $s=0,4,\ \bar{x}=23,161.$ Оцінити істинне значення математичного сподівання вимірюваної величини і точність вимірювань середньоквадратичного відхилення з надійністю 0,95.

Відповідь. $22,948<a<23,374,\quad 0,224<\sigma<0,576.$

22. Знайти мінімальний обсяг вибірки, за якого з надійністю 0, 95 точність оцінки математичного сподівання нормального розподілу за вибірковим середнім значенням буде дорівнювати 0,2, якщо середнє квадратичне відхилення є 2.

Відповідь. $n=385.$

23. За даними вибіркового дослідження відома заробітна платня (у грн.) 20-и службовців певної компанії (табл. 3.17). Знайти за допомогою вбудованих статистичних функцій Excel всі можливі числові характеристики за даними таблиці та побудувати надійний інтервал для генерального середнього

Табл. 3.17

3560	2190	2390	3400
2180	2400	3350	2340
2900	2570	3300	3150
3680	3250	2250	3240
2180	2600	2870	3050

Вказівка. Знаходження надійного інтервалу для генерального середнього виконати за допомогою функції **ДОВЕРИТ.**

24. В табл. 3.18 наведено дані про роздрібний товарообіг *Z* (млрд. грн.), середню кількість населення *X* (млн. осіб) та середній дохід *Y* (млн. грн.). Проаналізувати зв'язок між *Z* та *X* і *Y* за частинними і множинним коефіцієнтами кореляції.

Табл. 3.18

Z	1,2	1,3	2,5	1,4	1,2	0,2	2,4	4,1	1,1
X	1,4	1,4	2,5	1,5	1,3	0,3	2,6	4,2	1,1
Y	1,3	1,3	1,4	1,8	1,5	1,6	1,8	1,9	1,6

25. Для дослідження впливу капіталовкладень *X* (млн. грн.) на отриманий річний прибуток *Y* (млн. грн.) було зібрано статистичні дані за 20 великими підприємствами (табл. 3.19). Визначити силу зв'язку між означеними факторами.

Табл. 3.19

X / *Y*	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
1,5 – 2,5	1	-	-	-	-
2,5 – 3,5	2	5	2	-	-
3,5 – 4,5	-	3	3	2	-
4,5 – 5,5	-	-	-	2	-

Варіанти завдань для індивідуальної роботи

Завдання 1.

За даною вибіркою X знайти точкову оцінку математичного сподівання, а також зміщену і незміщену оцінки дисперсії та середньоквадратичного відхилення. Побудувати надійний інтервал із ймовірністю 95%, тобто $\gamma = 0{,}95$ для вказаного σ і надійний інтервал у випадку невідомої дисперсії для рівня із значущістю 5%. Порівняти побудовані надійні інтервали.

1.1. $X = (-2,-1,-2,0,1,0,2,2,4,1),\ \sigma = 4.$

1.2. $X = (-2,1,2,0,3,4,2,5,0,1),\ \sigma = 1.$

1.3. $X = (2,-1,0,-3,0,2,1,4,0,1),\ \sigma = 1.$

1.4. $X = (4,2,0,3,0,1,2,3,0,1),\ \sigma = 2.$

1.5. $X = (2,-1,2,0,3,0,-2,-2,0,-1),\ \sigma = 1.$

1.6. $X = (2,-1,2,0,3,0,-2,-2,4,-1),\ \sigma = 1.$

1.7. $X = (-2,1,-2,0,-3,2,-2,4,0,1),\ \sigma = 2.$

1.8. $X = (3,1,2,0,3,1,2,3,0,1),\ \sigma = 2.$

1.9. $X=(2,-1,-1,0,0,-2,2,-4,0,1),\ \sigma=1.$

1.10. $X=(2,1,4,0,4,0,2,4,0,1),\ \sigma=2.$

1.11. $X=(3,1,3,3,0,2,2,3,0,1),\ \sigma=2.$

1.12. $X=(4,1,4,0,3,4,-2,4,-2,1),\ \sigma=2.$

1.13. $X=(-1,1,-1,0,-1,0,2,1,0,1),\ \sigma=1.$

1.14. $X=(1,1,0,1,0,2,2,1,0,1),\ \sigma=1.$

1.15. $X=(2,1,2,3,3,3,2,4,4,1),\ \sigma=2.$

Завдання 2.

Користуючись теоретичними знаннями та засобами програмного забезпечення Excel, за даними варіанту обчислити середні значення, середні квадратичні відхилення та коефіцієнт кореляції.

Варіант 1		
n	X_i	Y_i
1	1,8	36,1
2	2,4	38,3
3	2,5	30,6
4	2,3	32,1
5	2,3	37,6
6	2,5	34,8
7	2,4	34,2
8	2,5	34,2
9	2,1	32,5

Варіант 2		
n	X_i	Y_i
1	2	38,2
2	2,1	36,9
3	2,3	39,7
4	2,6	37,2
5	2,8	31,7
6	2,8	30,1
7	2,2	39,9
8	2,3	38,8
9	2,6	38,4

Варіант 3		
n	X_i	Y_i
1	3	29,4
2	2,6	35,4
3	2,3	39,7
4	2,5	37,1
5	2,2	35,7
6	2,4	40,2
7	2,2	39,4
8	2,6	43,7
9	2,6	38,4

Варіант 4		
n	X_i	Y_i
1	2,8	14
2	2,4	17,1
3	2,3	18,2
4	2,5	17,4
5	2,7	16,1
6	2,4	18,8
7	2,3	32,2
8	1,9	31
9	2,3	32,4

Варіант 5		
n	X_i	Y_i
1	2	35
2	2,3	43,7
3	2,7	31,9
4	2,2	37,3
5	2,4	40,9
6	2,3	38,8
7	2,3	35,7
8	2,6	43,2
9	2,7	30,5

Варіант 6		
n	X_i	Y_i
1	2,4	40,2
2	2,2	39,4
3	2,6	43,7
4	2,6	38,4
5	2,3	38,8
6	2,2	39,9
7	2,8	30,1
8	2,8	31,7
9	2,8	37,2

Варіант 7		
n	X_i	Y_i
1	2,3	32,1
2	1,9	31
3	2,3	32,4
4	2,5	33,2
5	2,6	31,2
6	2	34,8
7	1,9	35,4
8	2,4	33
9	2,2	34,8

Варіант 8		
n	X_i	Y_i
1	2,8	13,8
2	2,7	14,8
3	2,4	16,9
4	2,3	16,8
5	2,5	14,8
6	2,5	17,9
7	2,5	17,6
8	2,4	15,7
9	2,3	15,2

Варіант 9		
n	X_i	Y_i
1	2,7	14,9
2	2,5	16,1
3	2,1	19,7
4	2,8	14
5	2,4	17,1
6	2,3	18,2
7	2,5	17,4
8	2,7	16,1
9	2,4	18

Варіант 10		
n	X_i	Y_i
1	30,2	5000
2	32	5200
3	32	5350
4	37	5880
5	30	5430
6	30	5430
7	30	5350
8	29	5740
9	33	5570

Варіант 11		
n	X_i	Y_i
1	29	5350
2	33	2740
3	31	5570
4	30	5530
5	34	6020
6	38	7010
7	31	6420
8	39	7150
9	39,5	7190

Варіант 12		
n	X_i	Y_i
1	5,1	15,7
2	7,2	17
3	2	5,2
4	5,1	17,5
5	10	22,1
6	12	25,8
7	9,4	23
8	6,9	17
9	3,4	9,1

Варіант 13		
n	X_i	Y_i
1	5,4	3,7
2	7,6	8
3	2,3	3,2
4	5,9	2,5
5	11	4,9
6	12,6	4,8
7	10,4	5,1
8	4,9	2,2
9	2,4	1,1

Варіант 14		
n	X_i	Y_i
1	1,4	12,7
2	2,6	18
3	2,3	16,2
4	5,1	25,5
5	6	24,1
6	5,6	24,8
7	10,4	55
8	4,9	33
9	2,4	18,1

Варіант 15		
n	X_i	Y_i
1	0,4	13,7
2	0,6	18
3	0,3	6,2
4	0,9	15,5
5	1	24,1
6	1,6	24,8
7	1,4	25
8	0,6	13
9	0,3	8,1

Завдання 3.

Зібрати статистичні дані (опитування однієї з груп), проаналізувати зв’язок між вибраними означеними факторами за частинними і множинним коефіцієнтами кореляції., побудувати лінії регресії.

Вказівка. Як приклад, даними можуть слугувати бали з вищої математики, фізики, іноземної мови чи будь-якої іншої навчальної дисципліни за семестри; зріст студента, вага студента; середній бал атестата, бал из предметів ЗНО тощо.

3.3. Статистична перевірка гіпотез

На практиці часто необхідно знати закон розподілу генеральної сукупності. Якщо закон розподілу невідомий, але є припущення його певного вигляду, тоді висувають гіпотезу, що випадкова величина має цей розподіл і виникає потреба у перевірці гіпотези. Розглянемо одну з основних задач математичної статистики – перевірку правдоподібності гіпотез.

3.3.1. Перевірка статистичних гіпотез

На основі статистичних даних при розв'язанні практичних задач потрібно зробити припущення про вигляд закону розподілу випадкової величини X. При цьому для остаточного вирішення питання про вигляд закону розподілу доцільно перевірити, наскільки зроблене припущення узгоджується з дослідними даними. Через обмежену кількість спостережень емпіричний закон розподілу, зазвичай, в деякій мірі, відрізняється від передбачуваного, навіть, якщо припущення про вигляд закону розподілу виявилося правильним. В зв'язку з цим виникає наступна задача [1]: чи розбіжність між емпіричним і передбачуваним (теоретичним) законом розподілу є наслідком обмеженості кількості спостережень, а чи вона є істотною і пов'язана з тим, що істинний закон розподілу випадкової величини суттєво відрізняється від передбачуваного. Для розв'язування цієї задачі служать так звані «критерії згоди».

При застосуванні певних статистичних методів обробки даних вибірки часто ставляться вимоги до розподілу даних або до числових характеристик.

***Означення*. Статистичною гіпотезою** називають будь-яке твердження про вигляд або властивості розподілу випадкових величин, що спостерігаються в експерименті.

За змістом статистичні гіпотези можна віднести до таких типів:

1. Гіпотези про вид закону розподілу досліджуваної величини.
2. Гіпотези про числові характеристики досліджуваної величини.
3. Гіпотези про рівність числових характеристик досліджуваних величин.

4. Гіпотези про належність досліджуваних величин до однієї генеральної сукупності.

5. Гіпотези про вид моделі, що описує взаємозв'язок між досліджуваними величинами.

6. Гіпотези про належність досліджуваних величин до одного класу.

Означення. **Нульовою** (основною) називають висунуту гіпотезу H_0. **Конкуруючою** (альтернативною) називають гіпотезу H_1, яка суперечить нульовій.

Означення. **Простою** називають гіпотезу, що містить лише одне твердження. **Складною** називають гіпотезу, яка складається зі скінченної або нескінченної кількості простих.

Прийняття основної або однієї з альтернативних гіпотез здійснюється на основі дослідження статистичних даних. Дослідження проводиться за певним ***критерієм***, який обирається відповідно до змісту гіпотези і виду наявних статистичних даних.

Для перевірки нульової гіпотези вибирається деяка випадкова величина K, розподіл якої відомий, і вона називається **статистичним критерієм** перевірки нульової гіпотези.

Означення. **Критичною областю** називають множину значень критерію, за яких нульову гіпотезу відхиляють. ***Областю прийняття гіпотези*** називають множину значень критерію, за яких нульову гіпотезу приймають.

Означення. **Критичними точками** $k_{кр}$ називають точки, які відокремлюють критичну область від області прийняття гіпотези.

Означення. **Правосторонньою** називають критичну область, що визначається нерівністю $K > k_{кр}$.

Для визначення правосторонньої критичної області достатньо знайти критичну точку $k_{кр}$. Для цього задають достатньо малу ймовірність α; число α називається ***рівнем значущості критерію.***

Тоді знаходять критичну точку, виходячи з того, що за умови справедливості нульової гіпотези виконується рівність:

$$P\left(K > k_{кр}\right) = \alpha\,. \qquad (3.44)$$

Порядок дій у разі перевірки статистичних гіпотез

Для перевірки правильності основної статистичної гіпотези H_0 потрібно:

1) визначити гіпотезу H_1, альтернативну до гіпотези H_0;

2) обрати статистичну характеристику перевірки;

3) визначити допустиму ймовірність похибки першого роду, тобто рівень значущості α;

4) знайти за відповідною таблицею критичну область (критичну точку) для обраної статистичної характеристики.

До критичної області належать такі значення статистичної характеристики, за яких гіпотеза H_0 відхиляється на користь альтернативної гіпотези H_1.

Зауваження. Якщо гіпотеза H_0 правильна, то з імовірністю α значення вибіркової функції будуть належати критичній області.

Якщо сформульовані гіпотези H_0 – основна та H_1, як альтернативна (конкуруюча), і обраний критерій перевірки справедливості основної гіпотези, то прийняття H_0 означає відкидання H_1, а відкидання H_0 означає справедливість гіпотези H_1.

Оскільки прийняття гіпотези здійснюється на основі статистичних даних, то завжди існує ймовірність помилки.

Означення. Імовірність відкидання гіпотези H_0, якщо вона справедлива, називається ймовірністю ***помилки першого роду*** або ***рівнем значущості***, який позначається α. Величина $\left(1-\alpha\right)$ є ймовірністю прийняття справедливої гіпотези і називається ***рівнем надійності***.

Імовірність прийняття гіпотези H_0, якщо вона не вірна, називається ймовірністю ***помилки другого роду*** і позначається β. Величина $\left(1-\beta\right)$ є ймовірністю відкидання невірної гіпотези і називається ***потужністю критерію***.

Помилки прийняття статистичних рішень подано нижче таблицею.

Прийняття рішення на основі критерію	Реальний стан дійсності (невідомий)	
	H_0 істинна	H_0 хибна
H_0 прийнято	Правильне рішення	Помилка 2-го роду
H_0 відхилено	Помилка 1-го роду	Правильне рішення

Потужність критерію – його здатність виявляти відмінності, тобто відхиляти нульову гіпотезу про відсутність відмінностей, якщо вона помилкова. Потужність визначають емпіричним шляхом. Відомо, що деякі критерії дозволяють виявити відмінності там, де інші неспроможні, тому пропонують використовувати більш потужні критерії. Проте підставою для вибору критерію може бути не лише потужність, але й інші характеристики, наприклад, ширший діапазон застосування до даних, обмеженість обсягів вибірки або їхня неоднаковість за обсягом, завелика інформативність результатів. Тоді і використовують менш чи більш потужні критерії.

Чим менше рівень значущості, тим менше ймовірність відкинути вірну гіпотезу. Зазвичай рівень значущості обирається дослідником, покладають його рівним 0,1; 0,05; 0,01 або 0,001. Якщо, наприклад, обраний рівень значущості $\alpha = 0,01$, то ризик відкинути вірну гіпотезу виникає в одному випадку із ста.

Зауваження. Перевірка статистичної гіпотези не надає точного висновку щодо її вірності або невірності. *Прийняття гіпотези означає, що на прийнятому рівні значущості вона не протирічить статистичним даним.*

Отже, перевірка статистичних гіпотез зазвичай здійснюється за такими ***етапами:***

1. Висунення припущень про вид розподілу досліджуваної величини (величин) або про її числові характеристики.

2. Формулювання статистичних гіпотез.

3. Вибір критерію перевірки відповідно до змісту гіпотез і статистичних даних.

4. Вибір рівня значущості залежно від вимог до точності результатів дослідження.

5. Розрахунок значення обраного критерію за статистичними даними.

6. Порівняння розрахованого значення критерію з його критичним значенням і прийняття або відкидання основної гіпотези.

3.3.2. Перевірка гіпотези про вид закону розподілу досліджуваної величини. Критерії згоди

3.3.2.1. Статистичні рішення на основі p-значень

Стандартні процедури прийняття (відхилення) нульової гіпотези H_0 ґрунтуються на фіксації факту потрапляння значень емпіричного критерію $\psi_{емп}$ у критичну область $\psi_{кр}$, яка визначена фіксованим рівнем значущості α. Проте можна виконувати і обернену дію: визначити ймовірність $p_{емп}$, яка відповідає емпіричному критерієві $\psi_{емп}$. Нульова гіпотеза H_0 відхиляється, якщо $p_{емп} \le \alpha$ (для однобічних гіпотез), $p_{емп} \le \frac{\alpha}{2}$ (для двобічних гіпотез).

Як приклад, розглянемо прийняття статистичного рішення щодо нульової гіпотези за статистичним Z-критерієм.

Зауваження. Z-критерій − це будь-який статистичний критерій, для якого розподіл значень статистики тесту при нульовій гіпотезі приблизно відповідає нормальному розподілу.

Припустимо, що $z_{емп} = 2{,}19$ і розглянемо випадок однобічних гіпотез.

За допомогою функції MS Excel **НОРМСТРАСП** отримано значення (рис. 3.18): $1 - p_{емп} = 0{,}9857$. Відповідно, $p_{емп} \approx 0{,}0143$.

За проведеними обчисленнями можна зробити висновок: оскільки $p_{емп} < 0{,}05 \quad (0{,}0143 < 0{,}05)$ H_0 відхиляється на рівні $\alpha = 0{,}05$, проте $p_{емп} > 0{,}01 \quad (0{,}0143 > 0{,}01)$, H_0 приймається на рівні $\alpha = 0{,}01$.

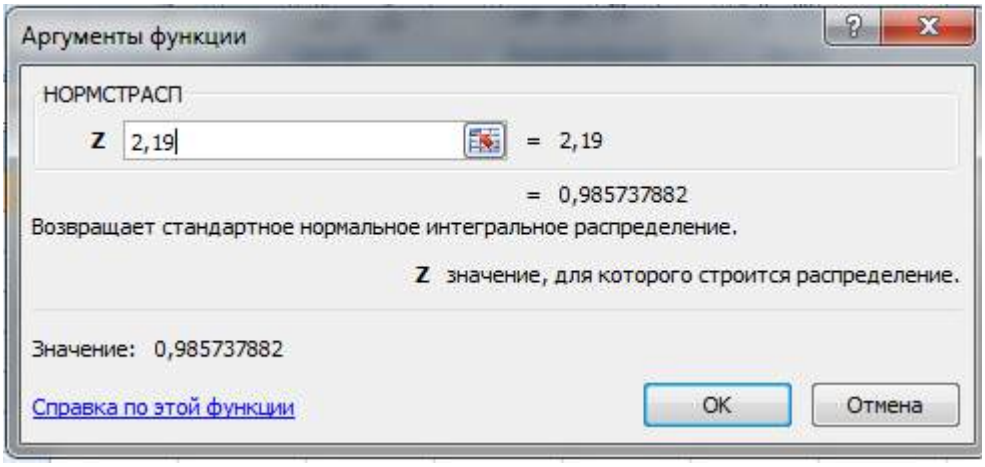


Рис. 3.18. Вигляд функції

За проведеними обчисленнями можна зробити висновок: оскільки $p_{емп} < 0,05 \quad (0,0143 < 0,05)$ H_0 відхиляється на рівні $\alpha = 0,05$, проте $p_{емп} > 0,01 \quad (0,0143 > 0,01)$, H_0 приймається на рівні $\alpha = 0,01$.

3.3.2.2. Перевірка гіпотези про закони розподілу

Перевірка гіпотези про вид закону розподілу досліджуваної величини має велике значення для прикладних досліджень. Необхідність такої перевірки виникає при виборі критерію, оскільки для багатьох з них висувається вимога нормального розподілу статистичних даних. Означені гіпотези перевіряються при проектуванні систем масового обслуговування, перевірки якості продукції або праці та т. ін.

Припустимо, що з деякої генеральної сукупності *X*, яка розглядається як випадкова величина, обрана вибірка $\{x_1, x_2, \ldots x_n\}$. За даними вибірки побудовано статистичний ряд (табл. 3.20), що містить варіанти x_i та відповідні частоти n_i, $i = \overline{1,k}$, k – кількість варіант у випадку дискретного ряду. У випадку інтервального ряду x_i – середини інтервалів, k – кількість інтервалів.

Табл. 3.20

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Отриманий на основі вибіркових даних статистичний ряд називається ***емпіричним законом розподілу*** величини *X*.

За даними статистичного ряду можна знайти числові характеристики, які є вибірковими параметрами закону розподілу *X*. Вид закону розподілу визначається відповідно до умов здобуття вибірки або залежно від виду графіка емпіричної щільності розподілу (гістограми) у випадку неперервної випадкової величини *X* і полігону частот, якщо величина *X* дискретна. Параметри обраного закону розподілу заміняються відповідними вибірковими параметрами.

Означення. Закон розподілу випадкової величини *X*, параметрами якого є відповідні вибіркові числові характеристики, називається ***теоретичним законом розподілу***.

При здійсненні такої заміни немає впевненості, що закон розподілу обраний правильно. Тому розроблено процедуру, яка дозволяє оцінити ступінь відповідності обраного закону даним вибірки. Критерії, в яких формулюється лише одна гіпотеза H_0, і необхідно перевірити, чи узгоджуються статистичні дані з цією гіпотезою, чи ні, називають ***критеріями згоди***. Найбільш відомим з яких є **критерій Пірсона** χ^2 (хі-квадрат).

Критерій Пірсона χ^2 обчислюється за формулою:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}}, \qquad (3.45)$$

де n_i' – частоти, отримані за теоретичним законом розподілу (теоретичні частоти).

З формули (3.45) видно, що у випадку, коли відповідні теоретичні та емпіричні частоти співпадають, $\chi^2 = 0$. Отже, чим ближче χ^2 до нуля, тим краще узгоджуються вибіркові дані та обраний теоретичний закон розподілу.

Розраховане значення критерію χ^2 порівнюється з його критичним значенням $\chi^2_{\alpha,l}$, яке знаходиться за статистичними таблицями (додаток) або за допомогою вбудованої статистичної функції Excel **ХИ2ОБР**(α,*l*). Параметри функції **ХИ2ОБР**: α – рівень значущості; *l* – степені вільності, $l = k - r - 1$, де $k -$ кількість груп емпіричного розподілу, *r* є кількістю параметрів теоретичного розподілу (наприклад, для нормального розподілу $r = 2$, оскільки параметрів

два: a та σ). Якщо $\chi^2 < \chi^2_{\alpha,l}$, то гіпотеза про закон розподілу приймається. У противному випадку гіпотеза відкидається.

За критерієм Романовського розглядають величину

$$a = \frac{\left|\chi^2 - l\right|}{\sqrt{2l}},$$

де l – степені вільності. Якщо $a \geq 3,$ то розбіжність між теоретичними та дослідними даними слід вважати не випадковою. Якщо $a < 3$, то таку розбіжність можна вважати випадковою, тобто емпіричні дані узгоджуються з обраним теоретичним розподілом.

Обґрунтування критерія Романовського базується на тому, що математичне сподівання χ дорівнює числові l, а дисперсія – подвоєному числові степенів вільності $2l$. В цьому випадку ймовірність відхилення величини хі-квадрат на $3\sigma^2 = 3\sqrt{2l}$ близька до одиниці.

Зауваження. В деяких статистичних таблицях критичне значення χ^2 надається залежно від рівня надійності γ, $\gamma = 1 - \alpha.$

Перевірка гіпотези про закон розподілу величини X здійснюється за такими пунктами:

1. З генеральної сукупності X беруть вибірку і будується статистичний ряд.

2. Висувається гіпотеза про закон розподілу випадкової величини *X.*

3. Знаходяться вибіркові параметри обраного закону розподілу.

4. Розраховуються теоретичні частоти.

5. Розраховується критерій χ^2 за формулою (3.45).

6. Обирається рівень значущості α (або рівень надійності γ) і знаходиться критичне значення $\chi^2_{\alpha,l}$ (або $\chi^2_{\gamma,l}$).

7. Порівнюються розраховане і критичне значення критерію χ^2 і робиться висновок про справедливість висунутої гіпотези.

Зауваження. Критерій згоди χ^2 можна використовувати для будь-яких розподілів. Розглянемо його для перевірки гіпотези H_0: генеральна сукупність має нормальний закон розподілу $N\left(a,\ \sigma^2\right)$, де $a=\overline{x},\ \sigma^2=s^2$.

Розбиваємо числову вісь на r інтервалів, що не перетинаються: $h_1,\ h_2,\ \ldots,\ h_r$; n_i – кількість елементів вибірки, що потрапили в інтервал h_i $(i=1,\ 2,\ \ldots,\ r)$. Позначимо через $p_i=P\left(X\in h_i\right)$, тоді np_i – теоретичні частоти. За формулою обчислення ймовірності $P\left(\alpha<X<\beta\right)=\Phi\left(\dfrac{\beta-a}{\sigma}\right)-\Phi\left(\dfrac{\alpha-a}{\sigma}\right)$ для нормального розподілу, якщо $a=\overline{x},\ \sigma=s,\ h_i=(x_i,\ x_{i+1})$, отримаємо:

$$p_i=P\left(x_i<X<x_{i+1}\right)=\Phi\left(\frac{x_{i+1}-\overline{x}}{s}\right)-\Phi\left(\frac{x_i-\overline{x}}{s}\right).$$

За критерій перевірки гіпотези H_0 беруть величину

$$\chi^2=\sum_{i=1}^{r}\frac{\left(n_i-np_i\right)^2}{np_i}. \tag{3.46}$$

Кількість k степенів вільності дорівнює $k=r-l-1$, де l – кількість параметрів розподілу. Для нормального розподілу $k=r-3$.

Якщо нульова гіпотеза справедлива за даним рівнем значущості α і кількістю степенів вільності k, відповідно до таблиці розподілу χ^2 знаходимо критичну точку $\chi^2_{k;\alpha}$ з рівності $P\left(\chi^2>\chi^2_{k;\alpha}\right)=\alpha$. Якщо $\chi^2<\chi^2_{k;\alpha}$, то гіпотезу H_0 приймають, якщо $\chi^2\geq\chi^2_{k;\alpha}$, то гіпотезу H_0 відхиляють.

3.3.2.3. Критерій незалежності χ^2

Нехай дані, одержані в результаті спостережень над парою випадкових величин X і Y, подано у вигляді кореляційної таблиці:

Y \ X	x_1	x_2	…	x_l	$\sum$
y_1	n_{11}	n_{21}	…	n_{l1}	n_1
y_2	n_{12}	n_{22}	…	n_{l2}	n_2
…	…	…	…	…	…
y_k	n_{1k}	n_{2k}	…	n_{lk}	n_k
$\sum$	m_1	m_2	…	m_l	n

Потрібно перевірити гіпотезу H_0 про незалежність випадкових величин X та Y.

За критерій перевірки гіпотези H_0 беруть величину

$$\chi_n^2 = \sum_{i=1}^{l}\sum_{j=1}^{k} \frac{\left(n_{ij} - \dfrac{m_i n_j}{n}\right)^2}{\dfrac{m_i n_j}{n}}. \qquad (3.47)$$

Відповідно до таблиці розподілу χ^2 за даним числом α і кількістю степенів вільності $m=(l-1)(k-1)$ знаходимо число $\chi^2_{m;\alpha}: P\left\{\chi^2(m) \geq \chi^2_{m;\alpha}\right\} = \alpha.$ Якщо $\chi_n^2 \geq \chi^2_{m;\alpha}$, то гіпотезу H_0 відхиляють, якщо $\chi_n^2 < \chi^2_{m;\alpha}$, то гіпотезу H_0 приймають.

Приклад 3.20. Проведено 200 спостережень над випадковими величинами X та Y:

Y \ X	1	2	$\sum$
1	25	52	77
2	50	41	91
3	25	7	32
$\sum$	100	100	200

Перевірити за допомогою критерію χ^2 незалежність випадкових величин X та Y за $\alpha = 0,05.$

Розв'язання. Обчислюємо за формулою (3.47):

$$\chi_n^2 = \frac{(25-38,5)^2}{38,5} + \frac{(52-38,5)^2}{38,5} + \frac{(50-45,5)^2}{45,5} + \frac{(41-45,5)^2}{45,5} + \frac{(25-16)^2}{16} +$$

$$+ \frac{(7-16)^2}{16} = 20,48.$$

Кількість степенів вільності $m=(2-1)(3-1)=2$, тоді маємо за таблицею значень

$$\chi^2_{2;\,0,05} = 5,99.$$

Оскільки 20,4 > 5,99, то гіпотезу незалежності відхиляють.

3.3.3. Перевірка гіпотез про рівність математичних сподівань та дисперсій для нормальних сукупностей

3.3.3.1. Гіпотеза про рівність математичних сподівань за відомих дисперсій

Нехай X та Y – незалежні випадкові величини, кожна з яких має нормальний розподіл $N(a_1, \sigma_x^2)$ і $N(a_2, \sigma_y^2)$. В результаті спостережень цих величин отримано дві вибірки $x_1, x_2, \ldots, x_n$ та $y_1, y_2, \ldots, y_m$ обсягів n і m. Потрібно перевірити гіпотезу $H_0: a_1 = a_2$ проти альтернативної гіпотези H_1: $a_1 \neq a_2$; σ_x, σ_y – відомі.

За критерій перевірки нульової гіпотези беруть випадкову величину

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}. \qquad (3.48)$$

Критерій Z – нормована випадкова величина, яка має нормальний розподіл, тому двостороння критична область симетрична відносно нуля. Критичну точку z_α знаходиться з рівності $\Phi(z_\alpha) = \frac{1-\alpha}{2}$. За реалізації вибірок обчислюють

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}.$$

Якщо $|Z| < z_\alpha$, то приймають гіпотезу H_0, якщо $|Z| \geq z_\alpha$, то приймають гіпотезу H_1.

3.3.3.2. Гіпотеза про рівність математичних сподівань за невідомих дисперсій

Нехай потрібно перевірити гіпотезу H_0: $a_1 = a_2$ проти альтернативної гіпотези H_1: $a_1 \neq a_2$, але $\sigma_x^2 = \sigma_y^2 = \sigma^2$ і величина σ^2 – невідома.

За критерій перевірки нульової гіпотези беруть випадкову величину

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}}. \qquad (3.49)$$

Величина T за умови правильності H_0 має розподіл Стьюдента з $k = n + m - 2$ степенями вільності.

Двостороння критична область симетрична відносно нуля, її знаходять з умови, що ймовірність прийняття гіпотези H_1 дорівнює взятому рівню значущості α, коли правильна H_0. Критичну точку $t_{n+m-2,\,\alpha}$ знаходять відповідно до таблиці розподілу Стьюдента за кількості степенів вільності $k = n + m - 2$ та $P\{|t| < t_{k,\alpha}\} = 1 - \alpha$. За реалізації вибірок обчислюють

$$T = \frac{\overline{x} - \overline{y}}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}},$$

де $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$, $s_y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \overline{y})^2$.

Якщо $|T| < t_{k,\alpha}$, то приймають гіпотезу H_0; якщо $|T| \geq t_{k,\alpha}$, то приймають гіпотезу H_1.

Приклад 3.21. Випадкові величини X та Y мають нормальний розподіл з рівними дисперсіями. В результаті спостережень над X та Y отримано вибірки:

$$X: 45,\ 48,\ 53,\ 44,\ 59,\ 60,\ 41,\ 43,\ 57;$$

$$Y: 51,\ 50,\ 42,\ 44,\ 39,\ 40,\ 48,\ 38,\ 59,\ 55,\ 55,\ 51.$$

Чи можна вважати, що випадкові величини мають однакові математичні сподівання? Вважати похибку першого роду $\alpha = 0,05$.

Розв'язання. За умовою $n = 9,\ m = 11$.

$$\overline{x} = \frac{1}{9}(45 + 48 + 53 + 44 + 59 + 60 + 41 + 43 + 57) = 50,\ s_x^2 = 54,25;$$

$$\overline{y} = 47,\ s_y^2 = 47,8.$$

$$|T| = \frac{|50 - 47|}{\sqrt{\left(\frac{1}{9} + \frac{1}{11}\right)\frac{8 \cdot 54,25 + 10 \cdot 47,8}{2 + 11 - 2}}} = 0,938.$$

Кількість степенів вільності $k = 9 + 11 - 2 = 18;\ \gamma = 1 - \alpha = 0{,}95,$ тоді $t_{18;\,0{,}05} = t(18;\ 0{,}95) = 2{,}10.$ Оскільки $0{,}938 < 2{,}10$, то можна вважати, що випадкові величини X та Y мають однакові математичні сподівання.

3.3.3.3. Гіпотеза про рівність дисперсій за невідомих математичних сподівань

Нехай потрібно перевірити гіпотезу H_0: $\sigma_x^2 = \sigma_y^2$ проти альтернативної гіпотези $H_1 : \sigma_x^2 > \sigma_y^2$.

За критерій перевірки гіпотези беруть відношення більшої виправленої дисперсії s_v^2 до меншої виправленої дисперсії $s_u^2,$ тобто величину

$$F = \frac{s_v^2}{s_u^2}. \tag{3.50}$$

У разі справедливості гіпотези H_0 випадкова величина має розподіл Фішера-Снедекора з $(n_1 - 1, n_2 - 1)$ степенями вільності, де n_1 – обсяг вибірки, за якою обчислюють $s_v^2,$ n_2 – обсяг вибірки, за якою обчислюють s_u^2.

Зауваження. *Я*кщо U та V – незалежні випадкові величини, що мають розподіл χ^2 з k_1 та k_2 степенями вільності, то величина

$$F = \frac{U / k_1}{V / k_2}$$

має ***розподіл Фішера-Снедекора*** з $(k_1,\ k_2)$ степенями вільності.

Правосторонню критичну область знаходять з умови, що ймовірність прийняття гіпотези H_1 дорівнює взятому рівню значущості α, коли правильна H_0:

$$P\left(F > F_{(\alpha, k_1, k_2)}\right) = \alpha,$$

де $k_1 = n_1 - 1,\ k_2 = n_2 - 1.$ Критичну точку $F_{(\alpha, k_1, k_2)}$ знаходять за таблицею розподілу Фішера-Снедекора.

За вибірками обчислюємо $F = \frac{s_v^2}{s_u^2}$, тоді:

– якщо $F < F_{(\alpha, k_1, k_2)}$, то гіпотезу H_0 приймають;

– якщо $F \geq F_{(\alpha, k_1, k_2)}$, то гіпотезу H_0 відхиляють.

3.3.3.4. Гіпотеза про рівність дисперсій за відомих математичних сподівань

Цю гіпотезу перевіряють аналогічно до попередньої, але в цьому разі

$$F = \frac{s_v^2}{s_u^2}, \tag{3.51}$$

де $s_v^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i - a_1)^2$, $s_u^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(y_i - a_2)^2$.

Якщо правильна гіпотеза H_0: $\sigma_x^2 = \sigma_y^2$, то випадкова величина має розподіл Фішера-Снедекора з (k_1, k_2) степенями вільності. Правосторонню критичну область знаходять з умови

$$P\left(F > F_{(\alpha, k_1, k_2)}\right) = \alpha,$$

коли правильна гіпотеза H_0. Критичну точку $F_{(\alpha, k_1, k_2)}$ знаходять за таблицею розподілу Фішера-Снедекора.

За вибірками обчислюємо значення $F = \frac{s_v^2}{s_u^2}$, тоді:

– якщо $F < F_{(\alpha, k_1, k_2)}$, то гіпотезу H_0 приймають;

– якщо $F \geq F_{(\alpha, k_1, k_2)}$, то гіпотезу H_0 відхиляють.

Приклад 3.22. За двома вибірками з нормальних сукупностей обсягу $n_1 = 11$, $n_2 = 15$ підраховано $s_1^2 = 0,76$ і $s_2^2 = 0,36$. За $\alpha = 0,05$ перевірити гіпотезу про рівність дисперсій цих двох нормальних сукупностей.

Розв'язання. Враховуючи $s^2 = \frac{n}{n-1}\overline{s}^2$, обчислюємо:

$$s_1^2 = \frac{11 \cdot 0,76}{10} = 0,836;\ s_2^2 = \frac{15 \cdot 0,36}{14} = 0,386.$$

Тоді

$$F = \frac{s_1^2}{s_2^2} = \frac{0,836}{0,386} = 2,17;\ F_{(0,05;10;14)} = 2,6.$$

Оскільки $2,17 < 2,6$, гіпотезу про рівність дисперсій приймають.

3.3.3.5. Перевірка статистичних гіпотез із використанням Microsoft Excel

Двохвибірковий F-тест для дисперсій

Перевірку гіпотези про рівність генеральних дисперсій за F-критерієм (Фішера) можна виконати за допомогою пакету аналізу Microsoft Excel. Для використання пакету необхідно:

1. Вибрати в меню послідовно пункти ***Сервис – Анализ данных***, після чого з'явиться вікно для вибору інструменту аналізу (рис. 3.19).

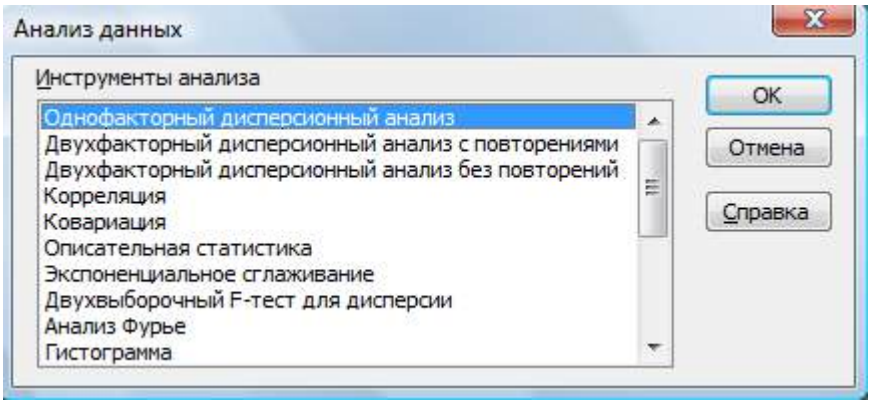


Рис. 3.19. Діалогове вікно аналізу даних

2. Вибрати у діалоговому вікні інструмент ***Двухвыборочный F-тест для дисперсии***, після чого з'явиться вікно для надання параметрів (рис. 3.20).

3. Задати всі необхідні параметри і натиснути ОК.

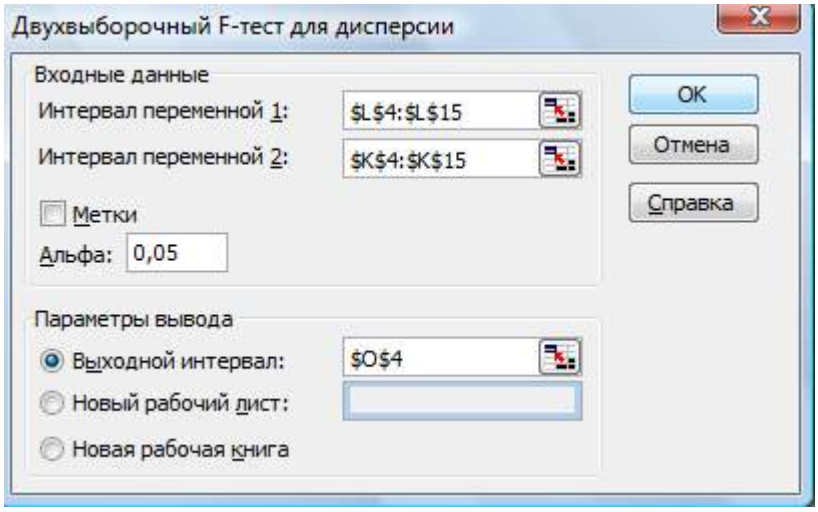
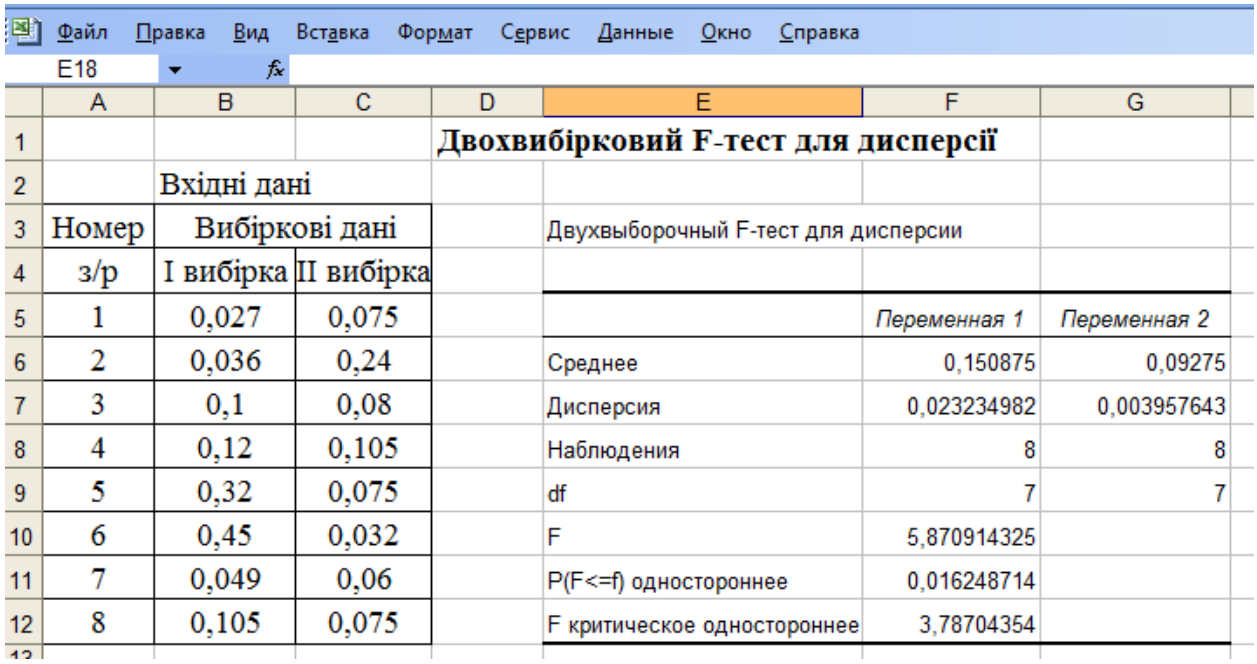


Рис. 3.20. Вікно надання параметрів

Приклад і результат роботи тесту наведено на рис. 3.21.



Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

E18

	A	B	C	D	E	F	G
1				**Двохвибірковий F-тест для дисперсії**			
2		Вхідні дані					
3	Номер	Вибіркові дані			Двухвыборочный F-тест для дисперсии		
4	з/р	I вибірка	II вибірка				
5	1	0,027	0,075			*Переменная 1*	*Переменная 2*
6	2	0,036	0,24		Среднее	0,150875	0,09275
7	3	0,1	0,08		Дисперсия	0,023234982	0,003957643
8	4	0,12	0,105		Наблюдения	8	8
9	5	0,32	0,075		df	7	7
10	6	0,45	0,032		F	5,870914325	
11	7	0,049	0,06		P(F<=f) одностороннее	0,016248714	
12	8	0,105	0,075		F критическое одностороннее	3,78704354	

Рис. 3.21. Перевірка рівності генеральних дисперсій

Двохвибірковий t-тест для середніх

Перевірку гіпотез про рівність генеральних середніх за критерієм Стьюдента можна також виконати за допомогою пакету аналізу даних Microsoft Excel. Для здійснення перевірки необхідно у вікні для вибору інструменту аналізу (рис. 3.19) вибрати:

1) у випадку рівних генеральних дисперсій – інструмент ***Двухвыборочный t-тест с одинаковыми дисперсиями***;

2) у випадку різних генеральних дисперсій – інструмент ***Двухвыборочный t-тест с разными дисперсиями***;

3) у випадку залежних вибірок – ***Парный двухвыборочный t-тест для средних***.

Після вибору інструменту аналізу з’явиться вікно для надання параметрів аналізу, аналогічне зображеному на рис. 3.20, в якому необхідно задати масиви комірок із вхідними вибірковими даними і рівень значущості (стандартний рівень – 0,05). Приклад і результат роботи тесту наведено на рис. 3.22.

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

C22

	A	B	C	D	E	F	G
1					**Двохвибірковий t-тест для середніх**		
2		Вхідні дані					
3	Номер	Вибіркові дані			Двухвыборочный t-тест с различными дисперсиями		
4	з/р	I вибірка	II вибірка				
5	1	0,027	0,075			*Переменная 1*	*Переменная 2*
6	2	0,036	0,24		Среднее	0,150875	0,09275
7	3	0,1	0,08		Дисперсия	0,023234982	0,003957643
8	4	0,12	0,105		Наблюдения	8	8
9	5	0,32	0,075		Гипотетическая разность средних	0	
10	6	0,45	0,032		df	9	
11	7	0,049	0,06		t-статистика	0,996970694	
12	8	0,105	0,075		P(T<=t) одностороннее	0,172413357	
13					t критическое одностороннее	1,833112923	
14					P(T<=t) двухстороннее	0,344826713	
15					t критическое двухстороннее	2,262157158	
16							

Рис. 3.22. Перевірка рівності генеральних середніх

3.3.4. Перевірка законів розподілу випадкової величини на прикладах

Наближенням (оцінкою) невідомої щільності ймовірностей можуть служити гістограма або полігон відносних частот. Гістограма і полігон відносних частот призначені для геометричного зображення групованого варіаційного ряду.

Гістограма відносних частот представляється у вигляді прямокутників, які примикають один до одного з основою $\Delta = \frac{R}{N}$, рівними ширині інтервалів угруповань, і висотами

$$h_i = \frac{\overline{n}_i}{\Delta} = \frac{n_i}{n \cdot \Delta}.$$

Для гістограми відносних частот площа ступінчастої фігури відповідає сумі ймовірностей і дорівнює 1. Площа будь-якого прямокутника гістограми дорівнює ймовірності потрапляння значень аналізованої випадкової величини в інтервал, відповідний основі прямокутника.

При збільшенні обсягу вибірки та зменшенні довжин інтервалів гістограма і полігон відносних частот наближаються до графіка невідомої функції *f(x)* – щільності ймовірності генеральної сукупності.

Гістограма − графік статистичної щільності розподілу випадкової величини у вигляді ступінчатого багатокутника. Будується вона наступним чином: на осі абсцис відкладаються варіантні інтервали Δx_i. На кожному з них будується прямокутник з ординатою, яка рівна значенню досліджуваної величини. По виду гістограми або полігону частот можна висунути гіпотезу про вид розподілу генеральної сукупності (рис. 3.23).

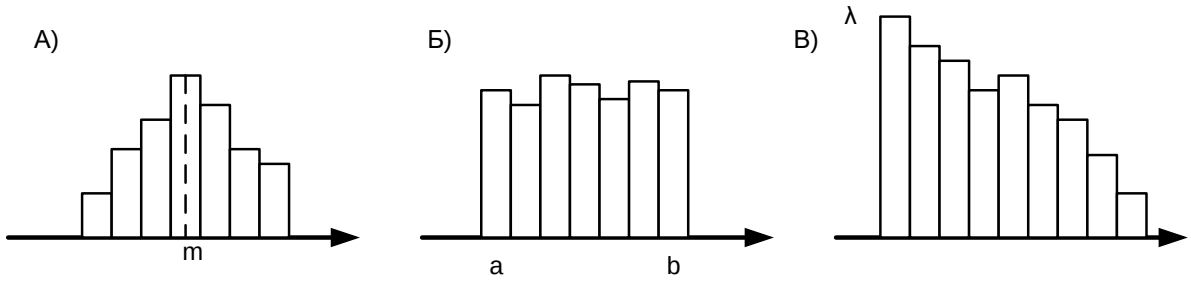


Рис. 3.23. Види гістограм

Наприклад, якщо гістограма має вид, представлений на рис. 3.23 а), то можна припустити, що генеральна сукупність має нормальний закон розподілу з щільністю ймовірностей

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

Якщо гістограма має вид рис. 3.23 б), тоді вона характеризує рівномірний розподіл з щільністю ймовірностей

$$f(x)=\begin{cases}\dfrac{1}{b-a}, x\in[a;b]\\ 0, x\notin[a,b]\end{cases}.$$

Якщо гістограма має вид рис. 3.23 в), можна робити припущення стосовно показового розподілу з щільністю ймовірностей

$$f(x)=\begin{cases}\lambda\cdot e^{-\lambda x}, x>0\\ 0, x\le 0\end{cases}.$$

Отже, порівнюючи за зовнішнім виглядом ці криві з відповідними теоретичними кривими, приймають ***гіпотезу*** про даний закон розподілу ймовірності.

Перевірку допустимості вибраного теоретичного закону (узгодженість експериментальної і теоретичної кривих) проводять за критеріями згоди, з яких найбільш розповсюджені критерії Колмогорова і χ^2 (хі-квадрат) Пірсона.

Критерії Колмогорова використовується у випадку, коли параметри розподілу вже відомі перед дослідом і потрібно після досліду перевірити узгодженість теоретичного і експериментального розподілів.

Критерій χ^2 (хі-квадрат) Пірсона використовується для невідомих параметрів розподілу.

3.3.4.1. Закон розподілу Пуассона

*З*як відомо, випадкова величина розподілена за ***законом Пуассона,*** якщо ймовірність обчислюється за формулою:

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0,1,\ldots;\lambda > 0, \tag{3.52}$$

де $\lambda -$ параметр розподілу. Крім того, для статистичних даних відомо, що $\overline{x} = \lambda; \quad s^2 = \lambda$. Отже, для встановлення параметра λ достатньо знайти $\overline{x}$, обчислити s^2 та порівняти їх. Далі розраховуються теоретичні частоти за формулою закону, також критерій χ^2. За обраним рівнем значущості α (або рівнем надійності γ) знаходиться критичне значення $\chi^2{}_{\alpha,l}$ (або $\chi^2_{\gamma,l}$), відповідно, висновок про справедливість висунутої гіпотези.

Приклад 3.23. На одній з міських АТС фіксувалася кількість телефонних дзвінків в годину. Спостереження велися протягом 100 годин, їх результати представлені в табл. 3.21. Чи можна вважати навантаження на АТС стандартним?

Табл. 3.21

Кількість викликів в годину	0	1	2	3	4	5	6	7
Кількість спостережень	6	27	26	20	10	5	5	1

Розв'язання. Навантаження на АТС можна вважати стандартним, якщо випадкова величина X – кількість телефонних дзвінків, що поступили, відповідає закону розподілу Пуассона. Тому необхідно перевірити гіпотезу про закон розподілу випадкової величини. Сформулюємо гіпотези:

H_0 – випадкова величина X розподілена за законом Пуассона;

H_1 – випадкова величина X не розподілена за законом Пуассона.

Випадкова величина X – кількість викликів в годину; тоді кількість спостережень – це відповідні значенням X частоти n_i, а таблиця 3.21 є статистичним рядом і емпіричним законом розподілу величини X. Знайдемо $\overline{x}$ та s^2. Для зручності обчислення подано у вигляді таблиці (табл. 3.22).

Табл. 3.22

x_i	0	1	2	3	4	5	6	7	Суми
n_i	6	27	26	20	10	5	5	1	100
$x_i\, n_i$	0	27	52	60	40	25	30	7	241
$\left(x_i - \overline{x}\right)^2 n_i$	34,85	53,68	4,37	6,96	25,28	33,54	64,44	21,07	244,19

Отже,

$$\overline{x} = \frac{1}{n}\sum_{i=1} x_i n_i = \frac{1}{100}\cdot 241 = 2,41; \quad s^2 = \frac{1}{n}\sum_{i=1}(x_i - \overline{x})^2 n_i = \frac{1}{100}\cdot 244,19 \approx 2,44.$$

Оскільки повинна виконуватися рівність $\overline{x} = \lambda; s^2 = \lambda,$ то як параметр можна вибрати або $\overline{x}$, або s^2, або їх середнє арифметичне. Виберемо $\lambda = \dfrac{\overline{x} + s^2}{2} \approx 2,426$. Таким чином, гіпотеза H_0 – це припущення, що величина X розподілена згідно із законом Пуассона (3.52).

Перевіримо правильність гіпотези за допомогою критерію Пірсона. Знайдемо теоретичні частоти, використовуючи формулу (3.52):

$$p_k = \frac{e^{-2,426}\, 2,426^k}{k!}, \quad k = 0,1,\ldots,$$

де за k візьмемо значення X, тобто x_i.

Відмітимо, що p_i – ймовірність того, що X прийме значення x_i, тобто статистично вони є відносними частотами, теоретичні частоти знаходитимемо за формулою: $n_i' = np_i$.

Для зручності при обчисленні теоретичних частот продовжимо таблицю 3.22, складену на основі статистичного ряду (табл. 3.21).

Отже, за результатами розрахунків

$$\chi^2 = \sum_{i=0}^{7} \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}} = 5,739 .$$

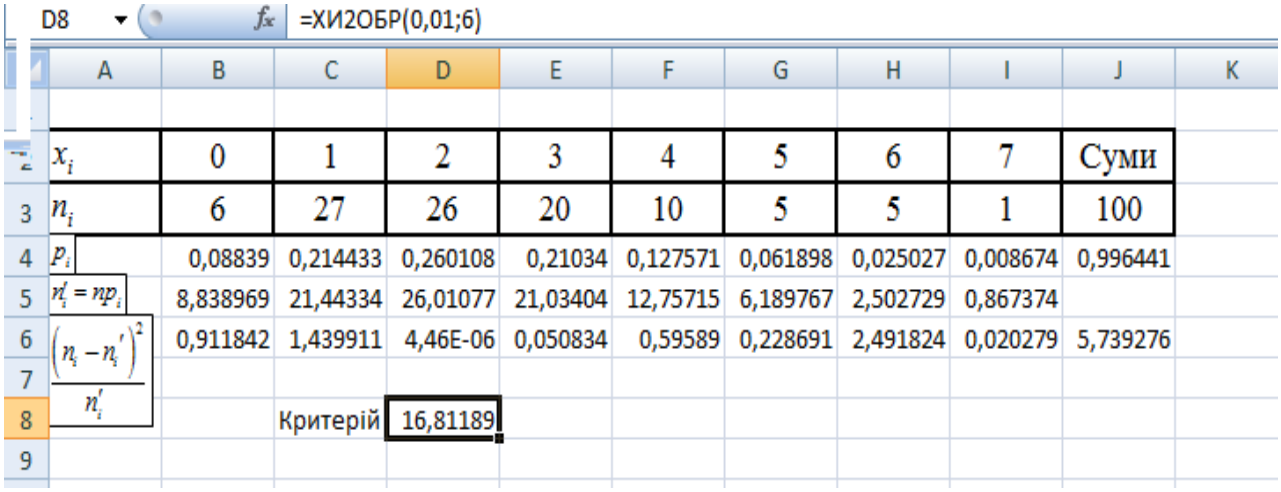
Для даного завдання $l = k - r - 1 = 8 - 1 - 1 = 6$. Виберемо рівень значущості α=0,01 і знайдемо за допомогою таблиць або функції **ХИ2ОБР** табличного процесора Excel значення $\chi^2_{\alpha,l}$: $\chi^2_{0,01;6}$=16,812. Оскільки для такого рівня надійності, тобто (табл. 3.23) $\gamma = 0,99$: $\chi^2 < \chi^2_{\alpha,l}$, то гіпотезу H_0 про розподіл Пуассона можна прийняти.

Табл. 3.23

x_i	0	1	2	3	4	5	6	7	Суми
n_i	6	27	26	20	10	5	5	1	100
$p_i = \frac{e^{-2,426} 2,426^{x_i}}{x_i!}$	0,09	0,21	0,26	0,21	0,13	0,06	0,03	0,01	
$n_i' = np_i$	8,84	21,44	26,01	21,03	12,76	6,19	2,50	0,87	
$\frac{\left(n_i - n_i'\right)^2}{n_i'}$	0,91	1,44	4,65E-06	0,05	0,59	0,23	2,49	0,02	5,739

Зауваження. В таблицю записані для ймовірності p_i наближені значення, а на рис. 3.24 показано точніші значення, обчислені за допомогою статистичної функції **ПУАССОН** табличного процесора Excel. Також показано використання вбудованої функції **ХИ2ОБР.**

Висновок. Навантаження на АТС можна вважати стандартним.



	A	B	C	D	E	F	G	H	I	J	K
2	x_i	0	1	2	3	4	5	6	7	Суми	
3	n_i	6	27	26	20	10	5	5	1	100	
4	p_i	0,08839	0,214433	0,260108	0,21034	0,127571	0,061898	0,025027	0,008674	0,996441	
5	$n_i' = np_i$	8,838969	21,44334	26,01077	21,03404	12,75715	6,189767	2,502729	0,867374		
6	$\frac{(n_i - n_i')^2}{n_i'}$	0,911842	1,439911	4,46E-06	0,050834	0,59589	0,228691	2,491824	0,020279	5,739276	
7											
8			Критерій	16,81189							
9											

Рис. 3.24. Використання функції **ХИ2ОБР**

3.3.4.2. Рівномірний закон розподілу

Щільність розподілу випадкової величини, яка є ***рівномірно розподіленою***, має вигляд

$$f(x) = \begin{cases} \dfrac{1}{b-a}, a \le x \le b, \\ 0, \qquad x < a, x > b, \end{cases}, \tag{3.53}$$

де a; b – параметри розподілу.

Відомо, що $\overline{x} = \dfrac{a+b}{2}$; $s^2 = \dfrac{(b-a)^2}{12}$, тобто для встановлення параметрів a та b потрібно знайти $\overline{x}$ і s^2, після чого розв'язати систему

$$\begin{cases} \overline{x} = \dfrac{a+b}{2}, \\ s^2 = \dfrac{(b-a)^2}{12}. \end{cases} \tag{3.54}$$

Приклад 3.24. З метою впорядкування роботи міського суспільного транспорту фіксувався час очікування в хвилинах пасажирами тролейбусів на декількох маршрутах. Було проведено 200 вимірювань, їх результати представлені в табл. 3.24. Чи можна вважати, що перевезення за перевіреними маршрутами забезпечені раціонально?

Табл. 3.24

Час очікування	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13
Кількість спостережень	25	30	48	35	42	20

Розв'язання. Можна вважати, що перевезення за перевіреними маршрутами забезпечені раціонально, якщо випадкова величина X – час очікування пасажирами транспорту відповідає рівномірному закону розподілу. Отже, задача зводиться до перевірки гіпотези про рівномірний закон розподілу випадкової величини. Сформулюємо гіпотези:

H_0 – випадкова величина X розподілена рівномірно;

H_1 – випадкова величина X не розподілена рівномірно.

Оскільки за умови задачі випадкова X – час очікування транспорту, то кількість спостережень – це відповідні значенням X частоти n_i, а табл. 3.24 – це інтервальний статистичний ряд і емпіричний закон розподілу X. Знайдемо $\overline{x}$ і s^2. За x_i візьмемо середини відповідних інтервалів. Для зручності обчислення оформимо у вигляді таблиці (табл. 3.25).

Табл. 3.25

$[a_i; a_{i+1})$	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13	Суми
x_i	2	4	6	8	10	12	
n_i	25	30	48	35	42	20	200
$x_i\, n_i$	50	120	288	280	420	240	1398
$\left(x_i - \overline{x}\right)^2 n_i$	622,5	268,2	47,045	35,704	380,52	502	1856

Отже, $\overline{x} = \frac{1}{n}\sum_{i=1}^{6} x_i n_i = \frac{1}{200}\cdot 1398 = 6{,}99$;

$$s^2 = \frac{1}{n}\sum_{i=1}^{6}(x_i - \overline{x})^2 n_i = \frac{1}{200}\cdot 1856 \approx 9{,}2799.$$

Складемо систему (3.54) для визначення параметрів рівномірного розподілу:

$$\begin{cases} 6{,}99 = \dfrac{a+b}{2} \\ 9{,}2799 = \dfrac{(b-a)^2}{12} \end{cases} \Rightarrow \begin{cases} a+b = 13{,}98 \\ (b-a)^2 = 111{,}3588 \end{cases} \Rightarrow \begin{cases} a+b = 13{,}98 \\ b-a \approx 10{,}55 \end{cases} \Rightarrow \begin{cases} a \approx 1{,}715 \\ b \approx 12{,}265 \end{cases}.$$

Таким чином, гіпотеза H_0 – це припущення, що X розподілена за рівномірним законом із щільністю розподілу:

$$f(x)=\begin{cases}\dfrac{1}{12,265-1,715}=\dfrac{1}{10,55}, & 1,715\le x\le 12,265\\ 0, & x<1,715,\quad x>12,265\end{cases}.$$

Перевіримо справедливість гіпотези H_0 за допомогою критерію Пірсона. Для знаходження теоретичних частот використаємо формулу $n_i'=np_i$, а ймовірності потрапляння в інтервали p_i знайдемо за формулою

$$P(\alpha<X<\beta)=\frac{\beta-\alpha}{b-a}.$$

Для зручності обчислень теоретичних частот складено таблицю (табл. 3.26). Врахуємо, що $\beta-\alpha=2;\quad b-a=12,265-1,715=10,55$ для всіх інтервалів, окрім першого і останнього.

Для першого інтервалу $\beta-\alpha=\beta-a=3-1,715=1,285$; для останнього інтервалу $\beta-\alpha=b-\alpha=12,265-11=1,265$.

Табл. 3.26

$[a_i;a_{i+1})$	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13	Суми
n_i	25	30	48	35	42	20	200
$\beta-\alpha$	1,285	2	2	2	2	1,265	
$p_i=\dfrac{\beta-\alpha}{10,55}$	0,122	0,19	0,19	0,19	0,19	0,12	
$n_i'=np_i$	24,360	37,915	37,915	37,915	37,915	23,981	
$\dfrac{\left(n_i-n_i'\right)^2}{n_i'}$	0,017	1,6522	2,6827	0,2241	0,4402	0,6609	5,677

Отже, $\chi^2=\sum_{i=0}^{7}\dfrac{\left(n_i-n_i^{'}\right)^2}{n_i^{'}}=5,677$. Таку ж таблицю формується в Excel набагато швидше, задаючи вихідні дані та описуючи відповідні функції для розрахунків, одну з яких видно на рис. 3.25, а перевірка критерію на рис. 3.26.

Для даного завдання $l=k-r-1=6-2-1=3$. Виберемо рівень значущості α=0,01 і знайдемо за допомогою таблиць або функції **ХИ2ОБР** табличного

процесора Excel значення $\chi^2_{\alpha,l}$: $\chi^2_{0,01;3}$=11,34 (рис. 3.26). Оскільки для такого рівня надійності (0,99) $\chi^2 < \chi^2_{\alpha,l}$, то гіпотезу про рівномірний розподіл приймаємо.

C7 =((C1-C5)^2)/C5

	A	B	C	D	E	F	G	H	I
1	n_i	25	30	48	35	42	20		
2	$\beta-\alpha$	1,285	2	2	2	2	1,265		
3	$p_i=\frac{\beta-\alpha}{10,55}$	0,121801	0,189573	0,189573	0,189573	0,189573	0,119905		
4									
5	$n_i'=np_i$	24,36019	37,91469	37,91469	37,91469	37,91469	23,98104		
6	$\frac{(n_i-n_i')^2}{n_i'}$								
7		0,016804	1,652192	2,682692	0,224067	0,440192	0,660885	5,676832	сума
8									
9									

Рис. 3.25. Задання функції для обчислення

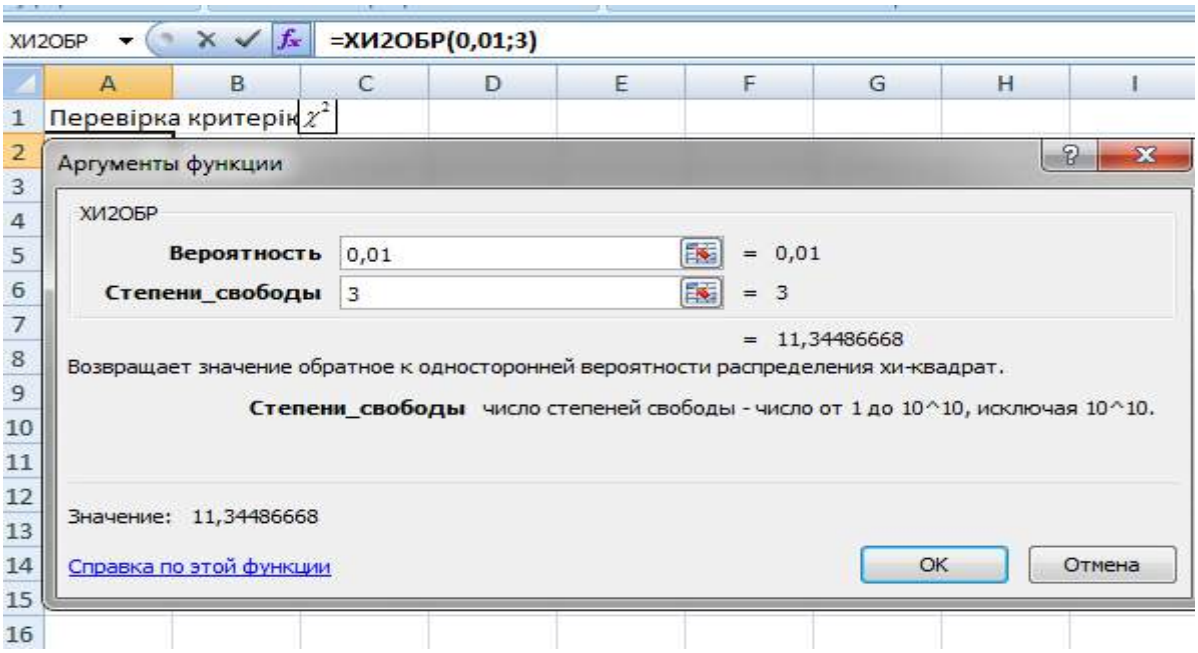


Рис. 3.26. Перевірка критерію

Висновок. Перевезення за перевіреними маршрутами організовані раціонально.

3.3.4.3. Приклади перевірки статистичних гіпотез про нормальний закон розподілу

Критерій згоди χ^2 можна використовувати для будь-яких розподілів. Розглянемо його для перевірки гіпотези H_0: генеральна сукупність має нормальний закон розподілу $N(a, \sigma^2)$, де $a = \bar{x}, \sigma^2 = s^2$.

Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, треба:

1. Обчислити вибіркову середню $\bar{x}_в$ і вибіркове середнє квадратичне відхилення σ.

2. Обчислити теоретичні частоти.

Відомо, що за умови неперервного розподілу ймовірностей ознаки X генеральної сукупності, теоретичні частоти обчислюють за формулою:

$$n_i' = np_i, \tag{3.55}$$

де n − обсяг вибірки, а p_i − імовірність того, що випадкова величина X потрапить в i-й частковий інтервал. Вона обчислюється в загальному випадку за формулами того закону розподілу, який припускаємо на основі обробки статистичного розподілу вибірки.

Якщо є підстави для припущення, що ознака генеральної сукупності X має нормальний закон розподілу, то теоретичні частоти можна обчислювати за формулами:

$$n_i' = \frac{nh}{\sigma} \cdot \varphi(u_i) = \frac{nh}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \bar{x}_B)^2}{2\sigma^2}}, \tag{3.56}$$

де n − обсяг вибірки; h − довжина *часткового* інтервалу; $\bar{x}_в$ − вибіркова середня величина; σ − вибіркове середнє квадратичне відхилення; $\varphi(u_i)$ – щільність імовірностей для загального нормального закону розподілу

або

$$n_i' = n \cdot \left(\Phi\left(\frac{x_{i+1} - \bar{x}_в}{\sigma} \right) - \Phi\left(\frac{x_i - \bar{x}_в}{\sigma} \right) \right), \tag{3.57}$$

де $\Phi\left(\frac{x_{i+1}-\overline{x}_в}{\sigma}\right)$, $\Phi\left(\frac{x_i-\overline{x}_в}{\sigma}\right)$ – функції Лапласа.

3. Порівняти емпіричні і теоретичні частоти за допомогою критерію Пірсона. Для цього обчислити значення критерію за формулою (3.46) або її аналогом.

4. За таблицею критичних точок розподілу χ^2 за заданим рівнем значущості α та кількості степенів вільності знаходять критичну точку $\chi^2_{k;\alpha}$, відповідно, висновки стосовно зробленої гіпотези.

Приклад 3.25. За заданим інтервальним статистичним розподілом випадкової величини X – маса новонароджених дітей (табл. 3.27),

Таблиця 3.27

H=0,5	1–1,5	1,5–2	2–2,5	2,5–3	3–3,5	3,5–4	4–4,5
n_i	10	20	50	35	28	15	12

за рівнем значущості $\alpha = 0,01$ перевірити правильність H_0 про нормальний закон розподілу ознаки X – маси новонароджених дітей.

Розв'язання. Для визначення теоретичних частот $n'_i = np_i$ потрібно обчислити $\overline{x}_в$, σ.

Дискретний статистичний розподіл буде таким:

x_i	1,25	1,75	2,25	2,75	3,25	3,75	4,25
n_i	10	20	50	35	28	15	12

$$n = \sum n_i = 170.$$

$$\overline{x}_в = \frac{\sum x_i n_i}{n} = = \frac{12,5+35+112,5+96,25+91+56,25+51,0}{170} = \frac{454,5}{170} = 2,67;$$

$$\frac{\sum x_i^2 n_i}{n} = \frac{1,25^2 \cdot 10 + 1,75^2 \cdot 20 + 2,25^2 \cdot 50 + 2,75^2 \cdot 35 + 3,25^2 \cdot 28 +}{170}$$

$$\frac{+3,75^2 \cdot 15 + 4,25^2 \cdot 12}{170} = \frac{1318,125}{170} = 7,75;$$

$$s^2 = \frac{\sum x_i^2 n_i}{n} - (\overline{x}_{\text{В}})^2 = 7,75 - (2,67)^2 = 7,75 - 7,1289 = 0,6211;$$

$$\sigma = \sqrt{s^2} = \sqrt{0,6211} \approx 0,79.$$

Обчислення теоретичних частот за формулою (3.57) подано в табл. (3.28).

Табл. 3.28

x_i	x_{i+1}	n_i	$z_i = \frac{x_i - \overline{x}_{\text{В}}}{\sigma_{\text{В}}}$	$z_{i+1} = \frac{x_{i+1} - \overline{x}_{\text{В}}}{\sigma_{\text{В}}}$	$\Phi(z_i)$	$\Phi(z_{i+1})$	$n_i' = n(\Phi(z_{i+1}) - \Phi(z_i))$
1	1,5	10	– 2,11	– 1,48	– 0,4821	– 0,4306	9
1,5	2	20	– 1,48	– 0,85	– 0,4306	– 0,3023	22
2	2,5	50	– 0,85	– 0,22	– 0,3023	0,0871	37
2,5	3	35	– 0,22	0,42	– 0,0871	0,1628	43
3	3,5	28	0,42	1,05	0,1628	0,3531	32
3,5	4	15	1,05	1,68	0,3531	0,4535	17
4	4,5	12	1,68	2,32	0,4535	0,4898	6

Обчислення за допомогою Excel (рис. 3.27):

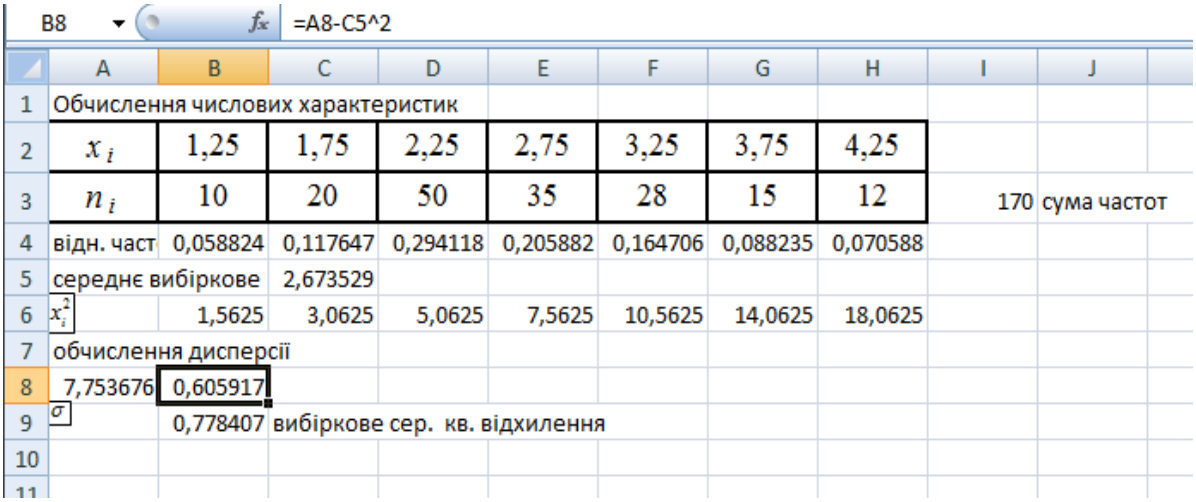


Рис. 3.27. Числові характеристики

Обчислення спостережуваного значення статистичного критерію χ^2 дається на рис. 3.28, де теоретичні частоти обчислені із використанням локальної формули Лапласа та, відповідно, використанням для функції **НОРМРАСП** логічного значення 0. Також критерій задано спрощеною формулою через відносні частоти.

Отже, маємо: $\chi^2 = \sum \frac{(n_i - np_i)^2}{np_i} = 14{,}51$.

За таблицею знаходимо значення

$$\chi^2_{\text{кр}}(\alpha = 0{,}01;\ k = 7 - 2 - 1 = 4) = \chi^2_{\text{кр}}(0{,}01;\ 4) = 13{,}3.$$

C10 =J9*170

	A	B	C	D	E	F	G	H	I	J
1		Дискретний статистичний ряд								
2	Середина інтервала		1,25	1,75	2,25	2,75	3,25	3,75	4,25	Сума
3	Частота		10	20	50	35	28	15	12	170
4	Відносна частота		0,058824	0,117647	0,294118	0,205882	0,164706	0,088235	0,070588	1
5	Середнє вибіркове		2,6735	2,6735	2,6735	2,6735	2,6735	2,6735	2,6735	
6	Виб. кв. відхилення		0,7784	0,7784	0,7784	0,7784	0,7784	0,7784	0,7784	
7	Лок. ф. Лапласа		0,096271	0,253547	0,442008	0,510047	0,389581	0,196967	0,065917	
8	Теорет. частоти		0,048135	0,126774	0,221004	0,255023	0,19479	0,098483	0,032958	0,977169
9	(Pi-Wi)^2/Pi		0,002373	0,000657	0,024188	0,009469	0,004646	0,001066	0,042963	0,085363
10	X^2		14,51173							
11	Хи2обр		13,2767							
12	Гіпотеза не приймається									

Рис. 3.28. Обчислення для статистичного критерію

Правобічна критична область показана на рис. 3.29.

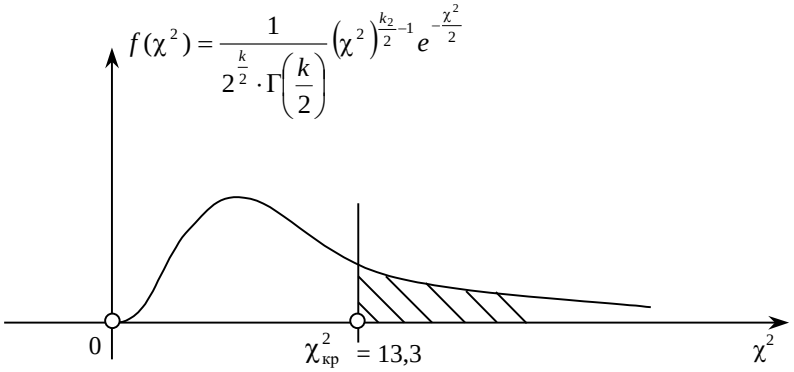


Рис. 3.29. Зображення критичної області та точки

Висновок. Оскільки $\chi^2 \in [0;\ 13{,}3]$, то не маємо підстав для прийняття H_0 про нормальний закон розподілу ознаки генеральної сукупності X.

Приклад 3.26. За наданим інтервальним статистичним рядом (табл. 3.29) знайти закон розподілу випадкової величини X.

Табл. 3.29

$[a_i; a_{i+1})$	$[-2; -1,2)$	$[-1,2; -0,4)$	$[-0,4; 0,4)$	$[0,4; 1,2)$	$[1,2; 2)$
n_i	6	11	21	7	5

Розв'язання. Для визначення виду закону розподілу побудуємо гістограму за даними табл. 3.29 (рис. 3.30). За видом гістограми висуваємо гіпотезу про нормальний закон розподілу даної випадкової величини:

H_0 – випадкова величина X розподілена за нормальним законом;

H_1 – випадкова величина X не розподілена за нормальним законом.

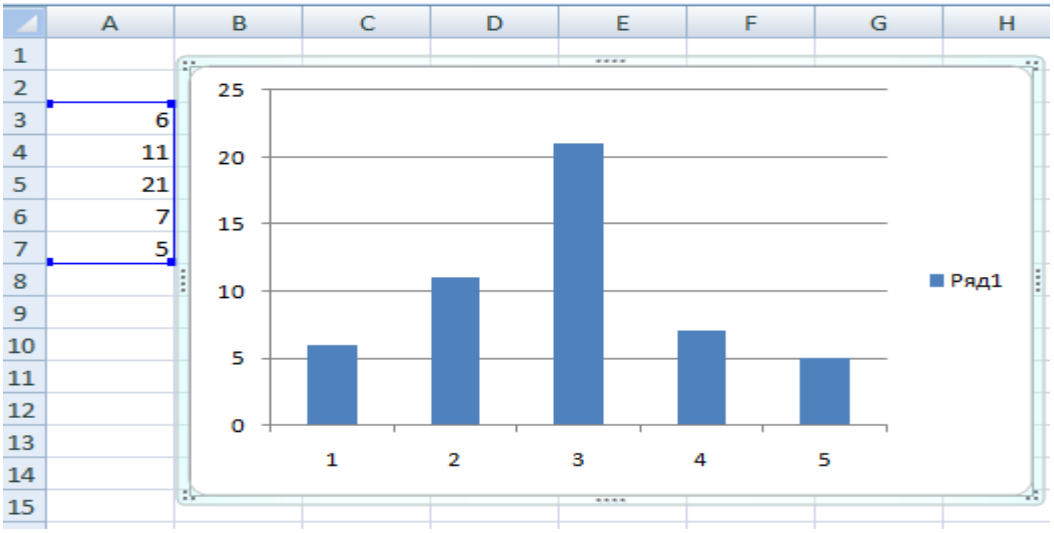


Рис. 3.30. Гістограма за даними таблиці

Щільність розподілу випадкової величини, розподіленої за нормальним законом має вид $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$, де a і σ – параметри розподілу. Знайдемо означені параметри, враховуючи, що $\bar{x} = a; \quad S^2 = \sigma^2$. Розрахунки оформимо у вигляді таблиці (табл. 3.30).

Табл. 3.30

$[a_i; a_{i+1})$	$[-2;-1,2)$	$[-1,2;-0,4)$	$[-0,4;0,4)$	$[0,4;1,2)$	$[1,2;2)$
n_i	6	11	21	7	5
x_i	-1,6	-0,8	0	0,8	1,6
$x_i n_i$	-9,6	-8,8	0	5,6	8

$\left(x_i-\overline{x}\right)^2 n_i$	13,572	5,452	0,194	5,620	14,382

Знайдемо вибіркове середнє, вибіркову дисперсію і вибіркове середнє квадратичне відхилення за відповідними формулами:

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{k}x_i n_i=\frac{1}{50}(-1,6\cdot 6-0,8\cdot 11+0\cdot 21+0,8\cdot 7+1,6\cdot 5)=-0,096;$$

$$s^2=\frac{1}{n}\sum_{i=1}(x_i-\overline{x})^2 n_i=\frac{1}{50}(13,572+5,452+0,194+5,620+14,382)=0,7844$$

$$\sigma=\sqrt{s^2}\approx 0,886.$$

Отже, параметрами теоретичного закону розподілу є величини: $\overline{x}=a=-0,096;\quad s=\sigma=0,886.$

Для знаходження значення критерію χ^2 розрахуємо теоретичні частоти n_i'. Для зручності обчислень побудуємо табл. 3.31.

Табл. 3.31

$[a_i;a_{i+1})$	[–2;–1,2)	[–1,2;–0,4)	[–0,4;0,4)	[0,4;1,2)	[1,2;2)
n_i	6	11	21	7	5
x_i	-1,6	0,8	0	0,8	1,6
$x_i n_i$	-9,6	8,8	0	5,6	8
$\left(x_i-\overline{x}\right)^2 n_i$	13,572	5,452	0,194	5,620	14,382
$\frac{a_i-\overline{x}}{S}$	-2,1498	-1,2465	-0,3433	0,56	1,4633
$\Phi(\frac{a_i-\overline{x}}{S})$	-0,958	-0,785	-0,266	0,425	0,856
p_i	0,0856	0,2595	0,3455	0,2155	0,063
$n_i'=np_i$	4,325	12,975	17,275	10,775	3,15
$\frac{\left(n_i-n_i'\right)^2}{n_i'}$	0,649	0,301	0,803	1,323	1,087

За формулою критерію маємо

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}} \approx 4{,}16.$$

Знайдемо критичне значення $\chi^2_{\alpha,l}$, враховуючи, що $l = k - r - 1$=5-2-1=2. Рівень значущості α оберемо рівним 0,1. За допомогою Excel знаходимо **ХИ2ОБР**(0,1; 2)=4,6.

Отже, оскільки $\chi^2 < \chi^2_{\alpha,l}$, гіпотеза H_0 про нормальний розподіл приймається, гіпотеза H_1 відкидається.

Завдання для самоконтролю

1. Дати означення нульової та альтернативної гіпотез.

2. Помилки першого і другого роду, їх означення.

3. Що називають простою та складною статистичнимигіпотезами?

4. Область прийняття нульової гіпотези, критична область, критична точка.

5. Що таке рівень значущості α?

6. Пояснити етапи перевірки статистичної гіпотези.

7. Описати критерії згоди, їх застосування та використання.

8. Які існують підстави для висунення гіпотези про закон розподілу ознаки генеральної сукупності.

9. Записати формулу для обчислення теоретичних частот, якщо припускається, що ознака X генеральної сукупності має експоненціальний закон розподілу.

10. Записати формулу для обчислення теоретичних частот, якщо припускається, що ознака X генеральної сукупності має рівномірний закон розподілу.

11. Перевірка гіпотези про рівномірний закон розподілу, навести приклад.

12. Пояснити вигляд гістограми для розподілу Пуассона, записавши вигляд його диференціальної функції розподілу та вказавши параметр λ.

13. Записати порядок дій перевірки гіпотези про нормальний розподіл генеральної сукупності.

14. Виписати формули обчислення теоретичних частот для нормального розподілу та пояснити використання відповідної вбудованої функції в Excel для їх знаходження.

15. Відомі дані про продуктивність праці (одиниць продукції за зміну) двох груп працівників: група 1 складається з працівників, що пройшли спеціальний навчальний курс; група 2 – із працівників, що не пройшли курсу (табл. 3.32). Враховуючи, що дані розподілені за нормальним законом, перевірити гіпотезу про рівність дисперсій.

Табл. 3.32

	Група 1					Група 2				
Продуктивність праці	34	85	96	102	103	63	69	83	89	106
Кількість працівників	5	2	11	8	4	2	6	8	3	1

Відповідь. Значення F-критерію: $F = \frac{S_1^2}{S_2^2} = \frac{612,46}{119,40} \approx 5,13 > F_{крит}$,

навчальний курс суттєво впливає на продуктивність праці робітників.

16. Для виробництва кожної з 10 деталей за першою технологією було витрачено, у середньому, 30с. Дисперсія часу складала $1с^2$. Для виробництва кожної з 16 деталей за другою технологією було витрачено, у середньому, 28с із дисперсією часу $2с^2$. Чи можна вважати, що у середньому для виробництва деталей за першою технологією потрібно більше часу?

Вказівка. Прийняти гіпотези: H_0 – середні двох генеральних сукупностей рівні, тобто $\overline{x_1} = \overline{x_2}$;H_1 − середні двох генеральних сукупностей не рівні, тобто $\overline{x_1} \neq \overline{x_2}$; за критерієм Фішера генеральні дисперсії можна вважати рівними.

Відповідь. Для виготовлення деталей за першою технологією потрібно, в середньому, більше часу.

17. Для виробництва кожної з 51 деталі за першою технологією було витрачено, у середньому, 30с. Дисперсія часу складала $6с^2$. Для виробництва кожної з 41 деталей за другою технологією було витрачено, у середньому, 25с із

дисперсією часу $3c^2$. Чи можна вважати, що у середньому для виробництва деталей за першою технологією потрібно більше часу?

Відповідь. Для виготовлення деталей за першою технологією потрібно, в середньому, більше часу.

18. Знайти закон розподілу величини X – кількості відвідувачів ресторану швидкого харчування за даними табл. 3.33.

Табл. 3.33

Час роботи	9.00 – 11.00	11.00 – 13.00	13.00 – 15.00	15.00 – 17.00	17.00 – 19.00	19.00 – 21.00
Кількість відвідувачів	25	53	68	73	52	39

Варіанти завдань для індивідуальної роботи

Завдання 1.

За даними побудувати гістограму розподілу, перевірити правильність гіпотези H_0 – випадкова величина X розподілена згідно із законом Пуассона за допомогою критерію Пірсона та Романовського.

№	X	0	1	2	3	4	5	6	7	8	9	10
1	n	37	37	18	6	2	0	0	0	0	0	0
2		29	36	22	9	3	1	0	0	0	0	0
3		22	33	25	13	5	2	0	0	0	0	0
4		17	30	27	16	7	2	1	0	0	0	0
5		14	27	27	18	9	4	1	0	0	0	0
6		11	24	27	20	11	5	2	0	0	0	0
7		8	21	26	21	13	7	3	1	0	0	0
8		6	18	24	22	15	8	4	2	1	0	0
9		5	15	22	22	17	10	5	2	1	0	0
10		4	13	20	22	18	12	6	3	1	1	0
11		0	4	10	20	28	34	32	26	18	12	8
12		0	4	12	22	32	34	32	26	16	10	6
13		2	6	14	26	34	34	30	22	16	8	4
14		2	6	16	28	36	36	30	20	14	8	4
15		2	8	20	30	36	34	28	20	12	6	4

Завдання 2.

Користуючись теоретичними знаннями та засобами програмного забезпечення Excel:

1. Побудувати таблицю статистичного розподілу.

2. Побудувати гістограму.

3. Вирівняти дослідні дані за допомогою нормального закону розподілу.

4. Перевірити узгодженість між емпіричними та теоретичними даними за допомогою критерія Пірсона та Романовського.

Варіант 1

2,37	-0,94	0,58	-0,38	-0,72	0,76	1,55	-0,53	1,41	1,03
-0,06	0,29	0,06	1,40	0,32	1,65	0,61	2,72	-1,03	0,48
0,61	2,05	1,12	-0,94	0,46	1,18	0,93	0,48	0,34	0,48
-0,50	1,58	1,39	2,30	0,83	1,19	-0,48	0,93	1,07	0,84
1,34	1,14	1,48	3,08	2,73	-1,14	-0,48	1,63	1,31	0,08
-0,10	1,52	2,41	0,16	0,31	0,60	2,75	-0,01	0,33	-0,13
0,51	0,81	-0,23	1,26	1,89	0,89	1,93	1,02	2,26	0,31
1,97	2,48	1,88	1,96	1,67	0,08	0,66	0,98	1,91	-0,11
0,67	1,18	2,30	3,15	1,24	0,81	0,73	0,65	0,79	0,63
1,45	1,31	1,42	1,23	1,84	1,99	2,05	2,50	2,55	2,90

Варіант 2

1,60	0,07	0,90	0,39	0,19	0,76	1,29	0,14	1,01	0,27
1,15	0,70	0,56	1,05	0,65	1,37	0,83	1,76	-0,04	0,72
0,96	1,70	1,13	0,19	0,70	1,06	0,96	0,71	0,68	0,66
0,20	1,29	1,01	1,55	0,80	1,19	0,35	0,77	1,15	0,85
1,19	1,11	1,16	2,02	1,88	-0,02	0,50	1,49	1,20	0,54
0,31	1,42	1,54	0,74	0,59	0,83	1,91	0,38	0,56	0,37
0,77	0,80	0,28	1,23	1,26	0,76	1,40	1,15	1,53	0,62
1,36	1,62	1,36	1,40	1,43	0,64	0,84	0,78	1,47	0,47
0,93	1,01	1,54	2,15	1,07	0,81	0,82	0,78	0,85	0,97
1,06	1,16	1,04	1,24	1,35	1,44	1,66	1,69	1,97	1,79

Варіант 3

3,33	0,18	1,52	0,99	0,38	1,74	2,78	0,23	2,20	0,67
2,19	1,46	1,03	2,09	1,38	2,83	1,81	3,53	-0,47	1,25
1,61	3,36	2,41	0,21	1,27	2,29	1,73	1,01	1,49	1,21
0,22	2,67	2,28	3,19	1,62	2,28	0,98	1,88	2,38	1,54
2,18	2,41	2,41	4,29	3,58	-0,27	0,83	2,88	2,21	1,16
0,86	2,85	3,07	1,36	1,33	1,95	3,96	0,67	1,02	0,85
1,96	1,61	0,64	2,46	2,56	1,90	2,88	2,45	3,42	1,23
2,62	3,05	2,64	2,59	2,60	1,28	1,90	1,97	2,85	0,84
1,94	2,28	3,09	4,17	2,02	1,82	1,84	1,97	1,66	1,86

2.40	2,01	2,41	2,47	2,99	2,67	3,47	3,38	3,94	3,54

Варіант 4

2,57	1,10	1,82	1,42	1,14	1,94	2,41	1,07	2,07	1,27
2,25	1,54	1,52	2,16	1,66	2,48	1,89	2,77	0,75	1,71
1,78	2,63	2,01	1,24	1,61	2,19	1,79	1,68	1,65	1,61
1,24	2,40	2,04	2,55	1,82	2,00	1,33	1,84	2,05	1,78
2,23	2,23	2,22	3,18	2,80	0,86	1,42	1,64	2,32	2,55
1,38	2,39	2,70	1,59	1,52	1,77	2,87	1,28	1,57	1,33
1,89	1,78	1,26	2,13	2,35	1,99	2,32	2,06	2,51	1,72
2,40	2,65	2,34	2,26	2,29	1,69	1,84	1,93	2,37	1,31
1,97	2,06	2,69	3,23	2,10	1,76	1,83	1,82	1,91	1,83
2,01	2,13	2,03	2,06	2,49	2,37	2,55	2,65	2,89	2,78

Варіант 5

4,24	1,32	2,80	1,95	1,47	2,55	3,90	1,18	3,36	1,79
3,04	2,11	2,30	3,08	2,32	3,87	2,61	4,90	0,93	2,29
2,64	4,26	3,01	1,06	2,13	3,39	2,60	2,07	2,46	2,22
1,08	3,79	3,17	4,18	2,51	3,39	1,90	2,63	3,37	2,82
3,28	3,04	3,37	5,32	4,73	0,62	1,99	3,97	3,27	2,02
1,65	3,72	4,26	2,00	2,25	2,55	4,78	1,63	2,07	1,92
2,84	2,56	1,75	3,31	3,56	2,67	3,54	3,16	4,06	2,45
3,99	4,43	3,83	3,69	3,56	2,05	2,86	2,61	3,69	1,65
2,96	3,04	4,37	5,23	3,40	2,54	2,54	2,59	2,60	2,77
3,20	3,06	3,14	3,32	3,94	3,94	4,43	4,23	4,98	4,64

Варіант 6

3,71	2,16	2,98	2,40	2,14	2,79	3,29	2,08	3,22	2,31
3,20	2,53	2,56	3,11	2,65	3,26	2,93	3,75	1,86	2,52
2,90	3,59	3,04	2,19	2,56	3,14	2,86	2,51	2,69	2,67
2,03	3,42	3,05	3,54	2,77	3,11	2,48	2,87	3,02	2,81
3,20	3,10	3,04	4,20	3,85	1,93	2,50	3,27	3,17	2,58
2,32	3,48	3,52	2,72	2,51	2,80	3,95	2,46	2,57	2,33
2,85	2,89	2,43	3,22	3,26	2,78	3,36	3,17	3,51	2,73
3,41	3,59	3,28	3,32	3,41	2,54	2,98	2,91	3,42	2,29
2,99	3,08	3,67	4,06	3,06	2,91	2,96	2,95	2,89	3,00
3,24	3,15	3,18	3,16	3,40	3,45	3,61	3,51	3,85	3,76

Варіант 7

5,24	2,29	3,84	2,61	2,01	3,66	4,71	2,36	4,11	2,92

4,44	3,38	3,18	4,08	3,43	4,56	3,83	5,57	1,61	3,14
3,61	5,48	4,34	2,27	3,24	4,37	3,84	3,35	3,16	3,01
2,33	4,97	4,05	5,07	3,55	4,31	2,95	3,53	4,08	3,65
4,43	4,29	4,11	6,07	5,88	1,57	2,82	4,98	4,01	3,34
2,58	4,64	5,19	3,06	3,40	3,66	5,82	2,90	3,30	2,61
3,64	3,97	2,90	4,46	4,89	3,62	4,51	4,45	5,45	3,08
4,65	5,25	4,82	4,85	4,98	3,08	3,80	3,62	4,59	2,86
3,96	4,37	5,41	6,40	4,03	3,92	3,77	3,53	3,54	3,66
4,40	4,50	4,02	4,34	4,99	4,95	5,24	5,37	5,46	5,75

Варіант 8

4,52	3,18	3,86	3,33	3,24	3,82	4,31	3,22	4,22	3,38
4,17	3,71	3,51	4,09	3,71	4,33	3,82	4,93	2,93	3,71
3,81	4,62	4,24	3,05	3,53	4,16	3,86	3,65	3,57	3,63
3,01	4,40	4,10	4,54	3,95	4,03	3,48	3,77	3,86	4,04
4,25	4,05	4,08	5,15	4,89	2,90	3,28	4,28	4,20	3,61
3,45	4,42	4,55	3,63	3,64	3,78	4,86	3,37	3,64	3,40
3,91	3,78	3,36	4,06	4,42	3,93	4,47	4,19	4,50	3,59
4,48	4,64	4,46	4,41	4,36	3,55	3,99	3,78	4,26	3,45
3,86	4,24	4,72	5,11	4,07	3,78	3,99	3,84	3,82	3,99
4,24	4,22	4,17	4,23	4,33	4,34	4,61	4,60	4,81	4,94

Варіант 9

6,25	3,44	4,61	3,67	3,14	4,52	5,88	3,18	5,07	3,94
5,09	4,20	4,38	5,07	4,32	5,55	4,53	6,73	2,92	4,13
4,55	6,42	5,03	3,47	4,01	5,11	4,66	4,14	4,02	4,01
3,37	5,84	5,24	6,05	4,57	5,03	3,66	4,85	5,48	4,72
5,26	5,33	5,45	7,38	6,74	2,60	3,83	5,79	5,21	4,28
3,51	6,01	6,29	4,41	4,15	4,73	6,81	3,68	4,47	4,02
4,66	4,81	3,92	5,15	5,87	3,98	4,98	5,109	6,48	4,07
5,74	6,26	5,56	5,54	5,98	4,22	4,94	4,92	5,97	3,67
4,81	4,45	6,31	7,19	5,25	4,56	4,79	4,86	4,88	4,56
5,02	5,29	5,02	5,01	5,65	5,66	6,11	6,49	6,72	6,86

Варіант 10

5,53	4,12	4,97	4,41	4,16	4,93	5,31	4,18	5,06	4,29
5,23	4,45	4,54	5,24	4,65	5,31	4,77	5,92	3,88	4,61
4,83	5,74	5,13	4,17	4,45	5,14	4,77	4,68	4,66	4,55
4,15	5,42	5,19	5,65	4,83	5,05	4,48	4,95	5.24	4,84
5,15	5,16	5,12	6,07	5,99	3,85	4,43	5,42	5,03	4,56
4,38	5,35	5,61	4,60	4,64	4,80	5,98	4,27	4,63	4,30
4,88	4,78	4,29	5,14	5,35	4,85	5,28	5,03	5,72	4,69

5,35	5,51	5,38	5,34	5,27	4,61	4,96	4,79	5,42	4,50
4,85	5,02	5,70	6,22	5,24	4,80	4,76	4,96	4,84	4,97
5,20	5,01	5,45	5,03	5,49	5,36	5,57	5,56	5,94	5,86

Варіант 11

7,17	4,34	5,87	4,65	4,12	5,81	6,85	4,29	6,47	4,69
6,29	5,14	5,07	6,20	5,27	6,86	5,66	7,98	3,59	5,15
5,76	7.16	6,24	4,13	5,11	6,04	5,86	5,04	5,10	5,17
4,31	6,58	6,49	7,48	5,88	6,12	4,56	5,81	6,19	5,93
6,09	6,23	6,38	8,37	7,56	3,90	4,69	6,75	6,05	5,20
4,65	6,96	7,19	5,35	5,23	5,94	7,55	4,72	5,41	4,89
5,86	5,94	6,67	6,07	6,58	5,62	6,74	6,25	7,14	5,11
6,80	7,19	4,61	6,67	6,90	5,12	5,74	5,77	6,57	4,93
5,88	6,41	4,34	8,05	6,18	5,68	5,59	5,69	5,79	4,67
6,07	6,13	6,34	6,05	6,84	6,65	7,05	7,06	7,97	7,88

Варіант 12

6,62	5,14	5,78	5,40	5,13	5,98	6,49	5,19	6,03	5,38
6,02	5,74	5,63	6,20	5,71	6,33	5,91	6,91	4,78	5,59
5,94	6,55	6,22	5,11	5,74	6,10	5,82	5,54	5,64	5,57
5,20	6,48	6,03	6,66	5,93	6,03	5,32	5,82	6,21	5,81
6,02	6,10	6,18	7,23	6,76	4,83	5,49	6,39	6,17	5,68
5,35	6,31	6,51	5,68	5,74	5,82	6,83	5,31	5,66	5,46
5,81	5,97	5,27	6,01	6,29	5,91	6,47	5,68	6,37	5,60
6,33	6,62	6,35	6,33	6,47	5,61	5,84	5,87	6,37	5,31
5,84	6,12	6,65	7,17	6,04	5,75	5,95	5,87	5,84	5,87
6,09	6,13	5,75	6,21	6,42	6,29	6,75	6,59	6,98	6,80

Варіант 13

9,58	8,09	9,00	8,31	8,01	8,81	9,25	8,05	9,24	8,45
9,07	8,53	8,64	9,22	8,71	9,40	8,86	9,88	7,79	8,65
8,93	9,60	9,07	8,01	8,70	9,11	8,79	8,63	8,68	8,74
8,18	9,37	9,16	9,68	8,84	9,12	8,33	8,76	9,00	8,88
9,07	9,19	9,18	10,03	9,93	7,96	8,45	9,30	9,22	8,54
8,30	9,34	9,74	8,68	8,74	8,99	9,91	8,31	8,66	8,35
8,87	8,78	8,33	9,17	9,29	9,00	9,47	9,11	9,37	8,60
9,38	9,68	9,35	9,26	9,47	8,66	8,84	8,89	9,46	8,50
8,87	9,05	9,65	10,20	9,04	8,96	8,95	8,87	8,78	8,82
9,16	9,19	9,11	9,18	9,42	9,41	9,75	9,69	9,83	9,80

Варіант 14

11,36	8,41	9,94	8,78	8,37	9,56	10,58	8,37	10,45	8,75
10,29	9,29	9,38	10,31	9,05	10,62	9,64	11,83	7,87	9,28
9,79	11,04	10,43	8,09	9,07	10,40	9,65	9,14	9,39	9,07
8,14	10,59	10,02	11,47	9,69	10,19	8,78	9,86	10,50	9,59
10,39	10,37	10,42	12,10	11,65	7,99	8,90	10,60	10,34	9,08
8,52	10,91	11,11	9,34	9,06	9,79	11,56	8,79	9,02	8,46
9,90	9,71	8,67	10,05	10,89	9,87	10,59	10,20	11,41	9,46
10,81	11,28	11,20	10,56	10,77	9,46	9,98	10,01	10,92	8,97
9,93	10,05	10,79	12,24	10,21	9,93	9,56	9,74	9,56	9,81
10,41	10,05	10,36	10,19	10,95	10,59	8,75	11,47	11,64	11,79

Варіант 15

14,24	12,29	13,84	12,61	12,01	13,66	14,71	12,36	14,11	12,92
13,44	11,38	13,18	14,08	13,43	14,56	13,83	15,57	11,61	13,14
12,61	12,48	14,34	12,27	13,24	14,37	13,84	13,35	13,16	13,01
11,33	14,97	14,05	15,07	13,55	14,31	12,95	13,53	14,08	13,65
13,43	13,29	14,11	16,07	15,88	11,57	12,82	14,98	14,01	13,34
11,58	13,64	15,19	13,06	13,40	13,66	15,82	12,90	13,30	12,61
12,64	12,97	12,90	14,46	14,89	13,62	14,51	14,45	15,45	13,08
13,65	14,25	14,82	14,85	14,98	13,08	13,80	13,62	14,59	12,86
12,96	13,37	15,41	16,40	14,03	13,92	13,77	13,53	13,54	13,66
13,40	13,50	14,02	14,34	14,99	14,95	15,24	15,37	15,46	15,75

Розділ 4. Аналіз даних за допомогою пакету STATISTICA

Пакет STATISTICA – це універсальний пакет статистичного аналізу, в якому реалізовані основні математичні методи аналізу даних. Розробником пакету є фірма StatSoft, Inc (США). У 2014 р. цю фірму поглинула корпорація Dell, яка і включила пакет до власного програмного забезпечення проблематики великих даних. STATISTICA дозволяє проводити різні процедури (модулі) обробки статистичних даних (в термінології програми – аналізи):

1. Розрахунок описових статистик.
2. Аналіз динамічних рядів й прогнозування.
3. Множинна регресія.
4. Дискримінантний аналіз.
5. Аналіз відповідностей.
6. Кластерний аналіз.
7. Факторний аналіз.
8. Дисперсійний аналіз та ін.

Крім загальних статистичних і графічних засобів STATISTICA має спеціалізовані модулі: для проведення соціологічних або біомедичних досліджень, вирішення технічних і, що дуже важливо, промислових завдань: карти контролю якості, аналіз процесів і планування експерименту.

Практичний досвід роботи з пакетом свідчить про можливість доступу до нових, нетрадиційних методів аналізу даних (а STATISTICA надає такі можливості у повній мірі) допомагає знаходити нові шляхи перевірки робочих гіпотез та дослідження даних [9].

4.1. Початкові відомості для роботи з пакетом

У загальному випадку робота із системою має послідовність дій.

1. Визначити структуру даних.

2. Ввести первинні дані.

3. Провести дослідження даних на помилки.

4. За необхідності здійснити попереднє перетворення даних, наприклад, групування або ранжування.

5. Розрахувати описові статистики.

6. Здійснити візуалізацію даних.

7. Застосувати конкретний метод аналізу.

В програмному пакеті STATISTICA використовують віконну форму роботи, так само як і в програмах Windows. Вікна досить схожі на вікна інших програм Windows, таких як MS Word, MS Exel та інші.

Головне вікно програми

Початковий варіант роботи із системою визначається у налаштуваннях системи. Це може бути створення нового файлу даних, завантаження існуючого файла і т. ін.

Вікно системи має наступну структуру: верхній заголовок Statistica – Basic Statistics and Tables (основні статистики і таблиці), тобто включений модуль Basic Statistics and Tables (рис. 4.1, 4.2). Рядок меню, панель інструментів і робоча область займає більшу частину вікна. В робочу область виводять документи системи, які отримують в процесі аналізу.

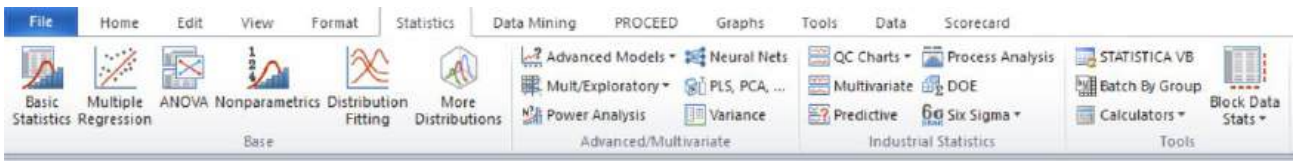


Рис. 4.1. Початок роботи з пакетом

Початок роботи з запуску пакету STATISTICA: Start → Programs → Statistica 12 (послідовність команд може бути трохи іншою, якщо при установці пакету були вибрані інші опції). Після завантаження програми з’являється вікно з контекстною панеллю для роботи з даними, а також відкривається таблиця нового або одного із стандартних, вже існуючих файлів.

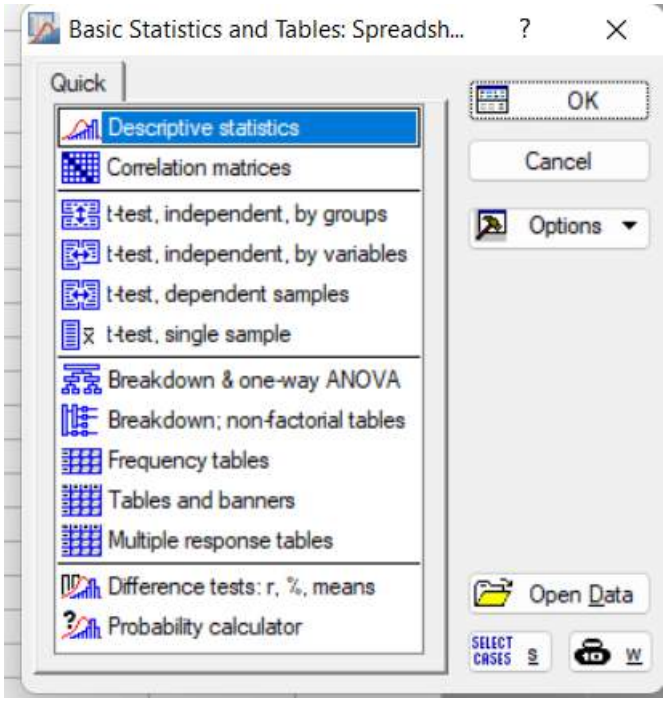


Рис. 4.2. Вигляд на екрані

Головне правило користувача: нічого не змінювати у файлах-прикладах пакету, або файлах інших користувачів. Тому, натиснення на «X» у правому верхньому куті відкритої таблиці даних, приведе до її закриття. Виберемо на верхній панелі Statistics. У меню (рис. 10.1) наявний ряд модулів для подальшої роботи. Модуль ‒набір статистичних засобів для роботи з певною специфічною інформацією, кожен з модулів полегшує вирішення певних статистичних задач.

Короткий опис модулів

Basic Statistics	Описова статистика (способи зображення (візуалізації) даних, оцінка параметрів, деякі парметричні тести)
Nonparametrics	Непараметричні тести
Distribution fitting	Можливість візуально підбирати криву розподілу до існуючої гістограми
Multiple Regression	Багатофакторна (множинна) регресія, використовують в разі залежності одного із показників від багатьох інших
Advanced Linear/ Nonlinear Models → Nonlinear estimation	Нестандартні типи регресій

Advanced Linear/ Nonlinear Models → Time Series	Аналіз часових рядів: визначення законів циклічності, періодичності, тренду, дослідження стохастичної компоненти
Multivariate Exploratory Techniques → Discriminant Analysis	Дискримінантний аналіз: знаходження визначальних показників для класифікації об'єктів у задані групи
Multivariate Exploratory Techniques → Cluster Analysis	Кластерний аналіз: поділ даних на групи за певними ознаками (наприклад, поділ країн на групи за показником ВВП)
Multivariate Exploratory Techniques → Factor Analysis	Факторний аналіз: проблеми класифікації і вибору показників, які є головними для опису даного явища

4.1.1. Введення первинних даних

Подання даних у програмі має табличний вигляд і зовні дуже схоже з електронною таблицею Excel. При цьому в рядках таблиці розташовуються спостереження (Cases), а у стовпцях – змінні (Variables). Заголовки рядків (спостереження) нумеруються 6 арабськими цифрами, а заголовки стовпців містять ім’я змінної.

Виберемо File → New → Speadsheet. (рис. 4.3). У діалоговому вікні, що відкрилося (рис. 4.4), можна вибрати кількість змінних та випадків, а також розміщення таблиці даних, яку потрібно створити.

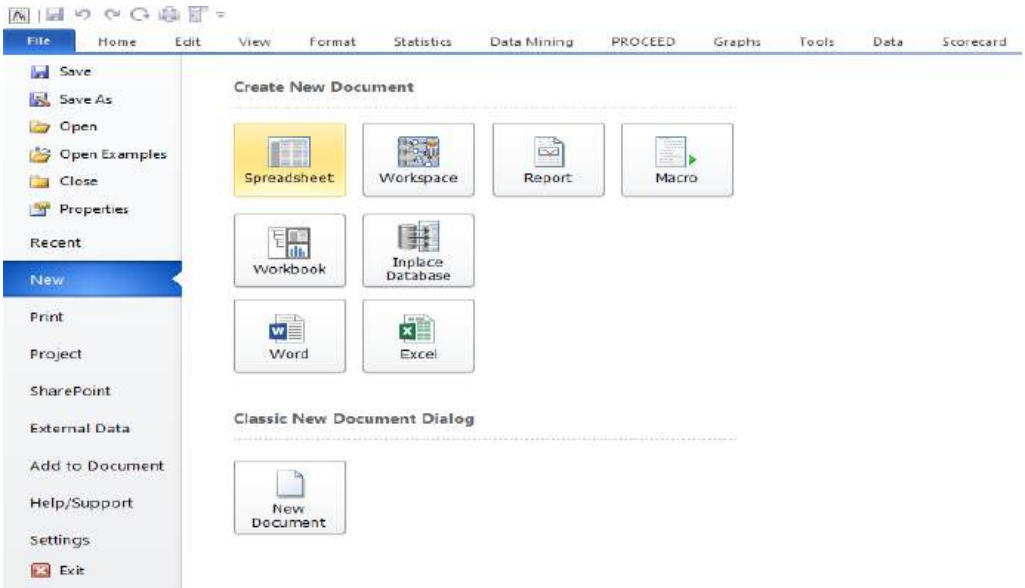


Рис. 4.3. Початок створення файлу

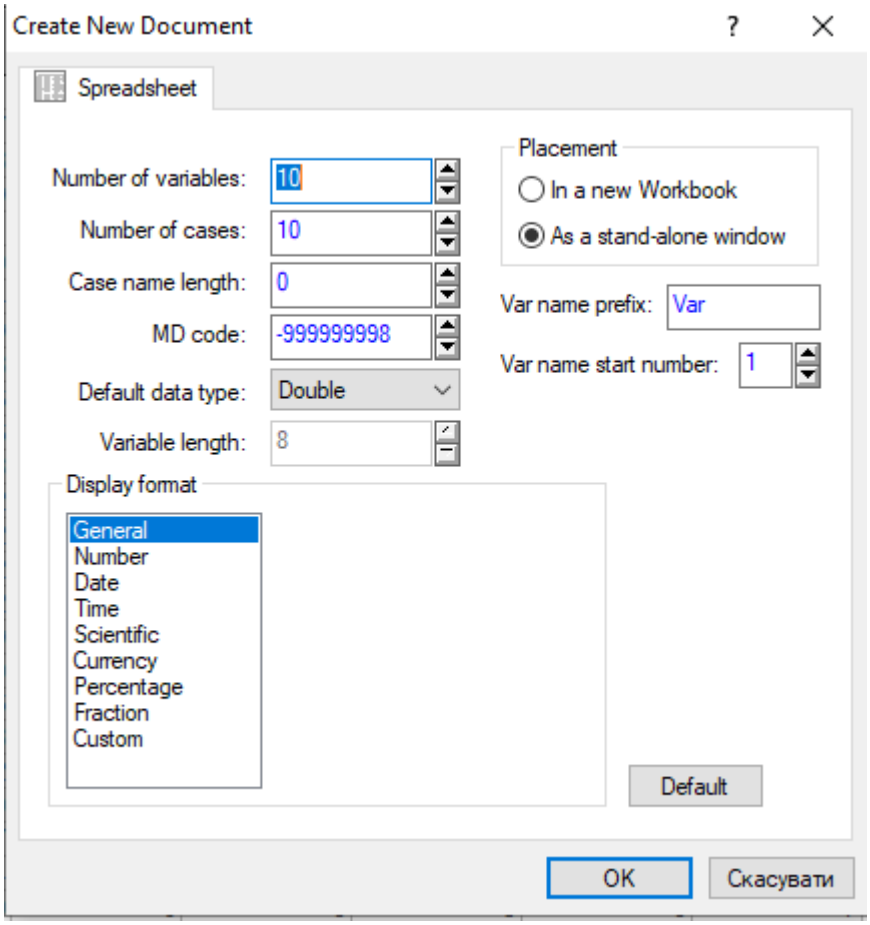


Рис. 4.4. Підготовка створення нового документу

Стандартна таблиця даних має розмір 10x10, де стовпці відповідають змінним (VAR1, VAR2.....VAR10), а рядки −випадкам зі значеннями, які змінні набувають.

Опція In a new Workbook розмістить новостворену таблицю у робочій книзі, в яку також будуть записані всі графіки, діаграми і таблиці, отримані у процесі роботи з даними. Опція As a standalone window створить таблицю в окремому вікні так, що дані можна буде зберегти окремо, виконавши File → Save при активізованій таблиці. Слід зазначити, що STATISTICA 12 оперує з «робочими книгами»−спеціальними файлами, в яких зберігається, залежно від обраних користувачами опцій, та чи інша інформація і результати роботи. Виконання певних дій в пакеті автоматично приводить до створення файлу звіту. На закладці Report можна вибрати розміщення даних звіту (опції аналогічні згаданим раніше).

Натискання ОК внизу вікна Create new document (рис. 4.4) дає нову порожню таблицю. За замовчуванням новий файл даних створюється для 10 спостережень і 10 змінних, але ці значення можна у вікні змінити. Файли даних у системі називаються «*Spreadsheet*».

Стовпці електронної таблиці з вихідними даними називаються Variables (змінні), а рядки Cases (випадки). Змінними звичайно виступають величини, що досліджуються, а випадки – це значення, які приймають змінні в деяких вимірах.

Змінні можна додавати, вилучати, переміщувати і копіювати за допомогою контекстного меню, яке потрібно викликати на *імені змінної* у рядку заголовка. (Add Variables, Delete Variables, Move Variables, Copy Variables).

Змінити кількість змінних − на панелі кнопка VARS →Add. У вікні, що з'явилося (рис. 4.5), показано скільки змінних ми хочемо додати і після якої змінної вставити нові змінні. Також можна обрати ім'я змінної за замовчуванням, тип даних і довге ім'я.

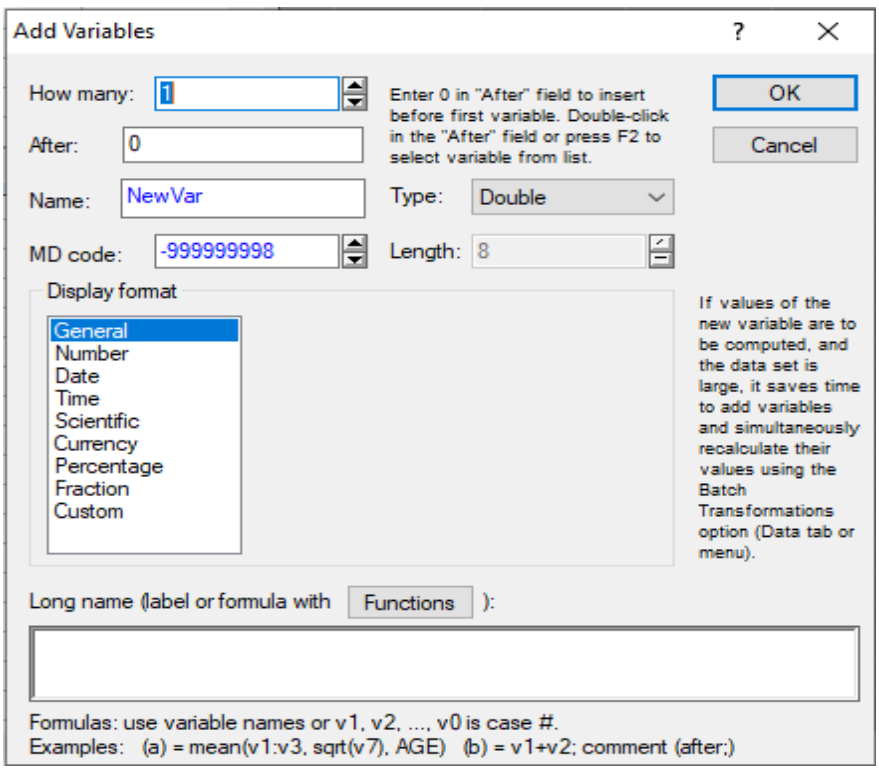


Рис. 4.5. Додавання змінних

Перемістити змінні можна таким чином: натиснути на верхній панелі кнопку VARS → Move (або виділити змінну і натиснути праву кнопку миші для

отримання контекстного меню, в якому вибираємо Move Variables). Вказуємо, з якої по яку змінну ми хочемо перемістити, також після якої змінної їх вставити (рис. 4.6).

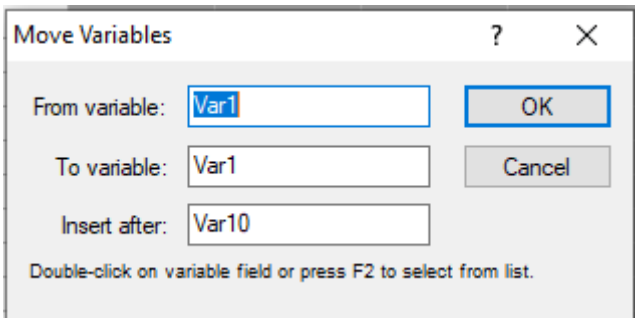


Рис. 4.6. Переміщення змінних

Аналогічно, операції VARS → Copy, VARS Delete (або Copy Variables і Delete Variables в контекстному меню) дають змогу скопіювати певні змінні, вставивши їх після вказаної змінної, та видалити вказані користувачем змінні.

Деякі зауваження стосовно заповнень значень змінних

1. Для заповнення значень змінної можна використовувати послідовність команд RC → Fill/Standartize Block → Fill Random Values. У результаті змінна буде заповнена випадковими значеннями.

2. Якщо потрібно заповнити весь стовпець (весь рядок) одним й тим самим значенням, то набирають це значення в першій клітинці. Далі, починаючи з цієї ж клітинки, виділяють вниз (вправо) стовпець (рядок) і натискають RC → Fill/Standartize Block → Fill/Copy Down (RC → Fill/Standartize Block → Fill/Copy Right).

3. Якщо потрібно заповнити стовпець, наприклад, арифметичною прогресією, то в перших двох клітинках вводять два перші члени арифметичної прогресії, виділяють ці клітинки, переміщують курсор у правий нижній кут виділеної області, доки він не змінить форму хрестика і протягують його вниз із натиснутою лівою кнопкою миші до тієї клітинки, до якої потрібно заповнити стовпець.

4. Для зміщення всіх даних у змінній, як одне ціле, на кілька позицій використовують на верхній панелі кнопку VARS → Shift (Lag). Для стандартизації змінних використовують на верхній панелі кнопку VARS → Standardize.

4.1.1.1. Копіювання даних

Якщо є дані в Excel і потрібно частину з цих даних скопіювати у файл в Statistica, то існує два способи:

1. За допомогою звичайних операцій Copy в Excel і Paste в Statistica.

2. За допомогою Copy в Excel та Edit → Paste special → Paste Link в Statistica (рис. 4.7).

Другий спосіб має перевагу над першим, оскільки при зміні даних в таблиці Excel, дані в файлі Statistica теж будуть відповідно змінюватися, а при першому способі цього не відбудеться. Якщо після внесення даних в файл з допомогою Paste Link подивитися на Edit → Link, то видно запис про те, з яким файлом встановлено динамічний зв'язок на файл Excel.

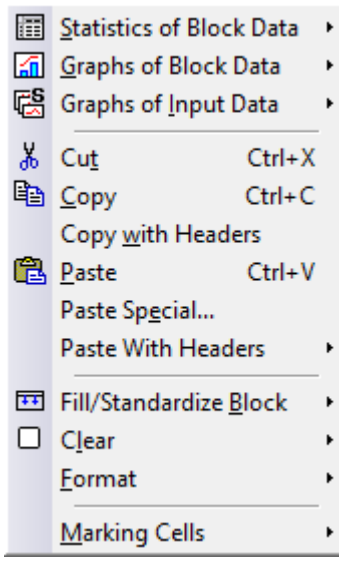


Рис. 4.7. Перенесення даних

Більшість згаданих операцій для змінних виконуються з випадками за допомогою меню Cases верхньої панелі. Наприклад, виділимо якусь змінну, натиснувши на її ім'я лівою кнопкою миші (LC), далі натиснемо праву кнопку

миші (RC) та виберемо Variable Specs. На екрані з'явиться вікно опису даної змінної (рис. 4.8).

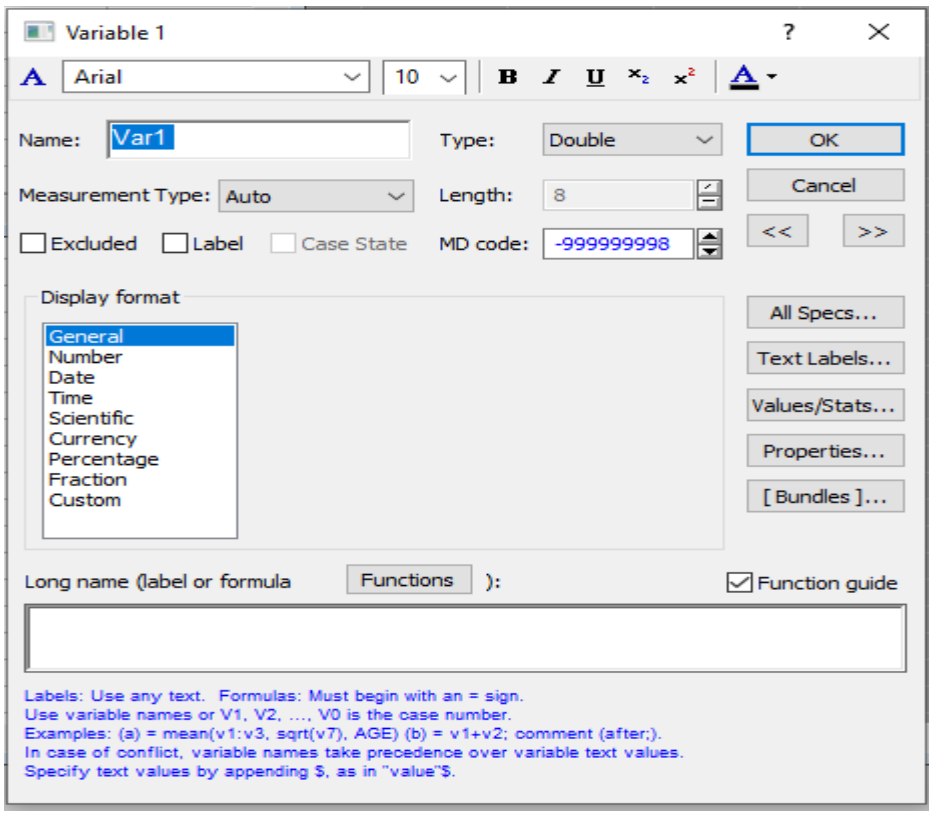


Рис. 4.8. Вікно опису вказаної змінної

Опис полів для заповнення подано далі в таблиці.

A	Опції для вибору різних характеристик шрифтів
Name	Ім'я змінної
Type	Вибрати тип даних – число подвійної точності, байти, ціле число, текст
Length	Ширина колонки даної змінної (для тексту)
Excluded	Виключити змінну з подальшого аналізу
Label	Використовувати значення змінної як тексові мітки, наприклад, для точок графіку
MD Code (missing data code)	Значення, яке за замовчуванням відсутнє з якихось причин
Display format	Вибір формату відображення числа (як дата, тощо) Дуже багато різних опцій–досліджуйте!
Long name	Поле, в якому можна задавати формулу для обчислення значення даної змінної

Кнопки (рис. 4.8) мають такі функції:

« , » −для переходу до попередньої та наступної змінної, яку відображає даний діалог;

All specs відкриває таблицю з усіма специфікаціями змінних;

Values/Stats дає змогу дізнатися «швидку статистику»−значення окремих випадків, середнє арифметичне, стандартне відхилення та інше (рис. 4.9).

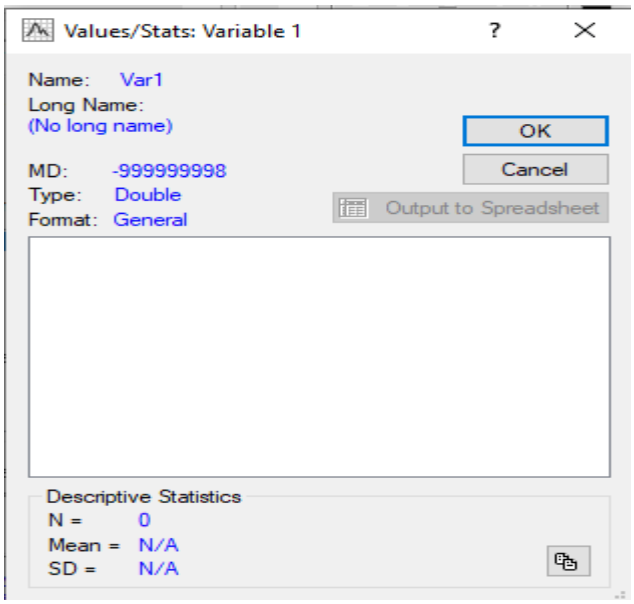


Рис. 4.9. Деякі статистичні дані

Кнопка Functions відкриває вікно вибору функцій для формули, яка визначатиме значення змінної. Деякі змінні можна обчислити на основі значень інших змінних. Для цього застосовують розрахункову формулу, яка повинна починатися зі знаку «=». Вона може містити імена змінних, математичні і логічні операції, функції тощо.

Наочним прикладом може слугувати проста арифметична дія додавання змінних. Для цього заповнюють колонки перших двох змінних потрібними чи довільними значеннями. Далі виділяють третю змінну натисканням RC−Variable Specs. У полі Long Name запис формули: =vl+v2, двічі натискання ОК дає бажаний результат: змінна VAR3 є сумою VAR1 та VAR2.

Зауваження. Імена v1, v2,… за замовчуванням присвоюються по порядку першій, другій, і т.д. змінним. Якщо ім’я першої змінної буде VARI0, то v1 буде

відповідати саме змінній VARI0. Якшо змінюються значення незалежних змінних, то перерахувати незалежну змінну можна натисканням на верхній панелі кнопки VARS →Recalculate або кнопки *x*=?

4.1.1.2. Сортування спостережень

Для роботи з усіма змінними таблиці даних або кількома файлами даних використовують меню *Data*. Наприклад, щоб транспонувати таблицю даних виконуємо Data →Transpose →File.

Спостереження у таблиці можна впорядковувати (сортувати) за значеннями однієї або відразу кількох змінних. Для цього слід зробити такі послідовні дії.

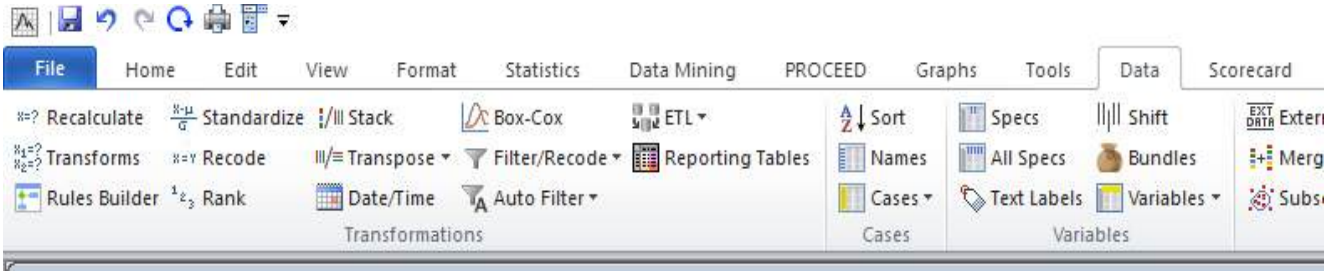
− Виконати команду *«Data Sort»* (Відкриється вікно «*Параметры сортировки*» («*Sort Options*»).

− Визначити змінну (або змінні), за якою слід впорядкувати дані, напрямок сортування (за зростанням або за спаданням). Для змінних текстового типу можна також вибрати варіант сортування: за текстом або за числовим кодом, що відповідає текстовому опису змінної.

− Натиснути «ОК».

Приклад 4.1. Порівняти значень змінних, а саме заміну значення змінної на її відносне місце у варіаційному ряді за зростанням чи спаданням.

Розв'язання. Натискаємо Data → Rank. У вікні, що з'явиться (рис. 4.10), вибираємо: найбільшому чи найменшому значенню присвоїти ранг 1 (за спаданням чи за зростанням будуть впорядковані значення), а також обираємо опцію Mean (якщо хочемо, щоб однакові значення мали однаковий усереднений ранг) або Sequential (якщо хочемо, що однакові значення мали послідовні значення рангу).



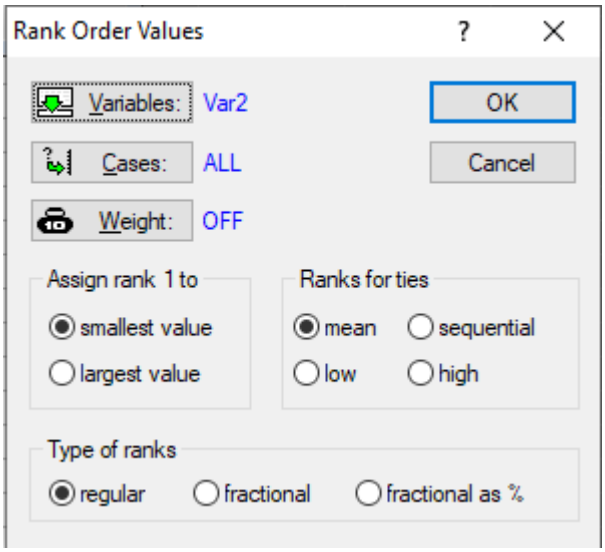


Рис. 4.10. Упорядкування змінних

Інколи буває потрібно розбити значення змінних на групи (наприклад, якщо певний показник більший за певну величину або менший за цю величину). Тоді такий порядок дій:

1. Виділяємо, наприклад, першу змінну, далі натискаємо Data → Recode.

2. У першому полі Include if пишемо умову для потрапляння значення змінної у групу: vl<5. У другому полі Include If пишемо: vl>=5.

3. У полях New Value 1 та New Value 2 вибираємо 1 та 2 відповідно (рис. 4.11).

4. Натискаємо OK і зберігаємо змінені значення. Бачимо, що всі значення, що були менші за 5, отримали нове значення 1, а ті, що були не менші за 5, отримали значення 2.

Оскільки візуально краще працювати з текстом, який би вказував назви груп, на які проведено розбиття значення змінної, то виконують: Data → Text Labels Editor, для значень 1 та 2 вводять назви груп, наприклад, male та female і далі Enter. Натискання кнопки Show/Hide Text Labels на верхній панелі показує у всіх клітинках заміну числових значень на назви груп.

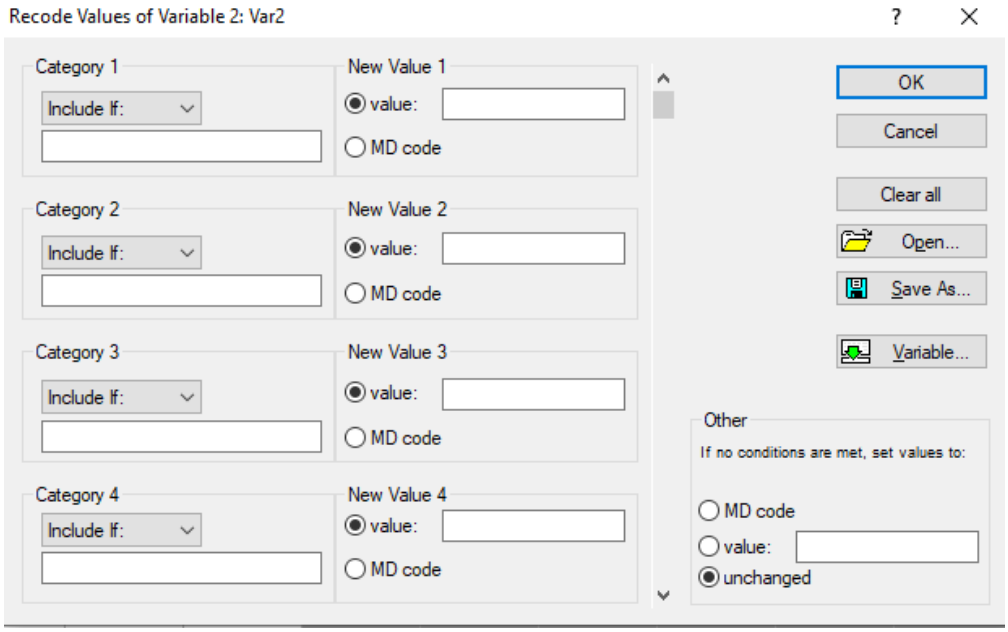


Рис. 4.11. Розбиття на групи

4.1.2. Сервісні функції

Система дає змогу швидко відобразити усі властивості однієї або всіх змінних. Для цього слід встановити курсор на стовпець, що містить потрібну змінну і виконати команду *«Данные Спецификации переменной» або «Данные Все спецификации переменных» («Data All Variables Spec»).*

Якщо дані були сформовані в інших статистичних пакетах, наприклад, SPSS або в електронній таблиці, то їх можна скопіювати до таблиці з даними за допомогою стандартної дії вставки. Наприклад, виділити потрібні дані, скопіювати їх до буферу обміну, а потім у STATISTICA виконати команду *«Правка, Вставить», «Edit Paste»* або натиснути на стандартній панелі.

Збереження даних

Збереження файлів відбувається стандартним чином за командою *«Файл Сохранить», «File Save»* або натисканням піктограми. Якщо потрібно зберегти існуючий файл з іншим ім’ям, то це здійснюється за стандартною командою «*Файл Сохранить как…», «File Save As…».* До імені файлу STATISTICA автоматично додає розширення *STA*. Такий файл даних є набором файлів, оскільки він містить і автоматично зберігає інформацію про всі додаткові файли (графіки, звіти, програми), що використовуються з поточним набором даних.

Крім цього зберегти таблицю даних можна також у форматі електронної таблиці Excel, у вигляді веб-сторінки, текстового файлу і т. ін.

Створені дані можна зберігати в різних форматах. Для цього при виділеному окремому вікні з даними натискають: File → Save as. Далі є можливість вибрати формат, в якому буде збережено інформацію. Якщо дані містяться в робочій книзі, тоді правою клавішею миші на назві таблиці в дереві документів (зліва) і вибирають Save item as. Потрібну таблицю, графік також можна виокремити з робочої книги за допомогою команди Extract as a standalone, якщо відмітити їх у робочій книзі і натиснути праву клавішу миші.

Для створення звіту −файлу, в якому будуть записані результати всіх дій (таблиці, графіки тощо) потрібно виконати: File → Save As (рис. 4.12).

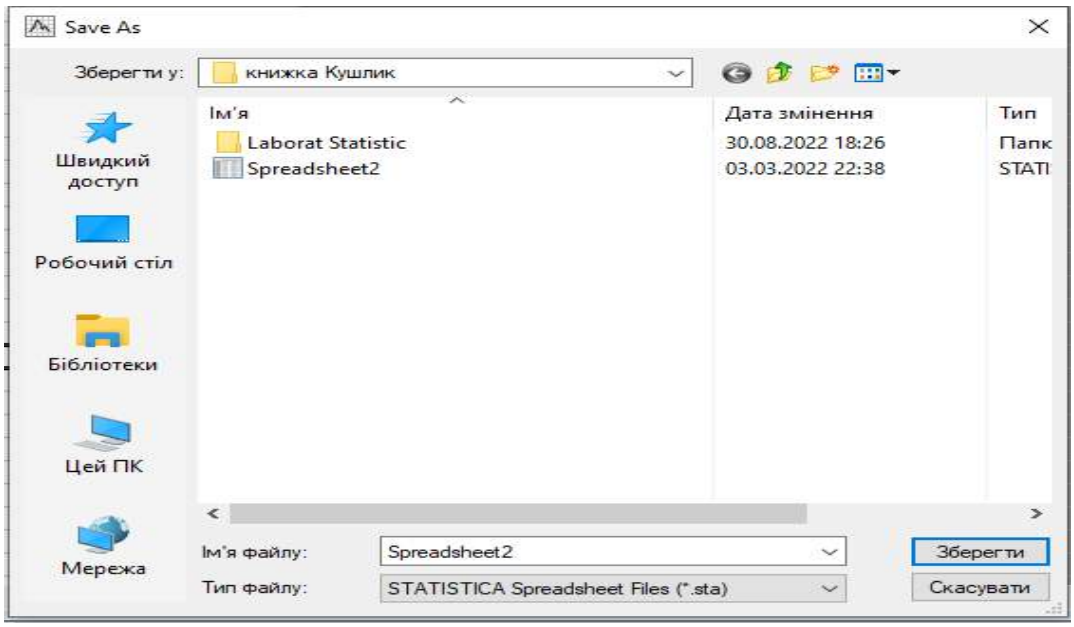


Рис. 4.12. Збереження файлу

Вибирають опцію, щоб інформація автоматично відсилалась до файлу звіту, потім визначають, чи створюватиметься окремий звіт для кожного графіку/аналізу, чи в одному вікні зливатиметься вся звітність, чи створиться файл звіту із вказаним іменем. У верхній частині вікна можна вибрати опції щодо використання робочих книг.

Про інформацію, що міститься у файлах, свідчить його розширення.

.sta	Файли з даними у вигляді таблиць

.stw	Файли робочих книг
.stg	Файли з графіками
.Str	Файли звіту
.svb, .svx	Файли STATISTICA Visual Basic чи макроси
.stm	Файли матриць
.snn	Файли нейронних мереж
.sdm	Файли проектів модулю Data Miner
.sti	Файли на віддалених серверах

Відкриття даних, збереження інформації

За певних налаштувань під час завантаження системи автоматично відкривається останній файл, з яким виконувалась робота. Відкрити файл даних можна й під час роботи з системою за командою *«File Open Data».* При цьому в робочій області може бути тільки один файл з даними.

Під час роботи з даними доцільно встановити режим автоматичного збереження інформації. Для цього слід виконати команду *«Сервис Параметры», «Tools Options»* і на вкладці «Загальні» вікна «*Параметри*» встановити числове значення для поля «Зберігати кожні ... хвилин».

4.2. Практичне застосування пакета

4.2.1. Описова статистика. Обчислення статистичних характеристик і побудова гістограм

4.2.1.1. Знаходження значень функції розподілу, побудова графіків

Якщо потрібно знайти значення теоретичної функції розподілу в певній точці, або, вказавши значення функції розподілу, знайти квантиль, то для цього можна скористатися імовірнісним калькулятором: Statistics → Probability Calculator.

Панель Probability Distibutioп Calculator (рис. 4.13) дає змогу подивитися, як виглядає один із даних нам розподілів. Змінюючи значення параметрів розподілу, видно автоматичну зміну щільності розподілу (Density Functioп) та функції розподілу (Distribution Function). Задавши значення в полі *X*, після натискання Compute в полі *p* з'явиться значення функції розподілу *F(x)*. Аналогічно в полі *p* можна задати ймовірність від 0 до 1, тоді після натискання Compute в полі *X* з'явиться значення квантилі рівня *p*.

Якщо відмітити Create Graph та Send to Report і натиснути Compute, то в окремих вікнах отримаємо, відповідно, графік та звіт (рис. 4.13).

4.2.1.2. Аналіз модуля Basic Statistics/Tables

Розглянемо модуль Basic Statistics/Tables. Натиснемо Statistics → Basic Statistics/Tables (рис. 4.14), зайдемо в розділ Descriptive Statistics.

Відкриємо файл Adstudy.sta (...\STAT1ST1CA 12\ExatpIes\ Datasets\), у якому зібрані дані про оцінки чоловіками та жінками реклами напоїв Pepsi та Coke. Кожен опитуваний оцінював рекламу за різними показниками, виставляючи оцінку від 0 до 9 (рис. 4.15).

Активізуємо вікно Descriptive Statistics з нижньої панелі. В полі Variables вкажемо 3-Measure01. В закладці Quick знаходяться найбільш вживані описові статистики, таблиця частот, гістограма, а також «коробка з вусами» (рис. 4.16).

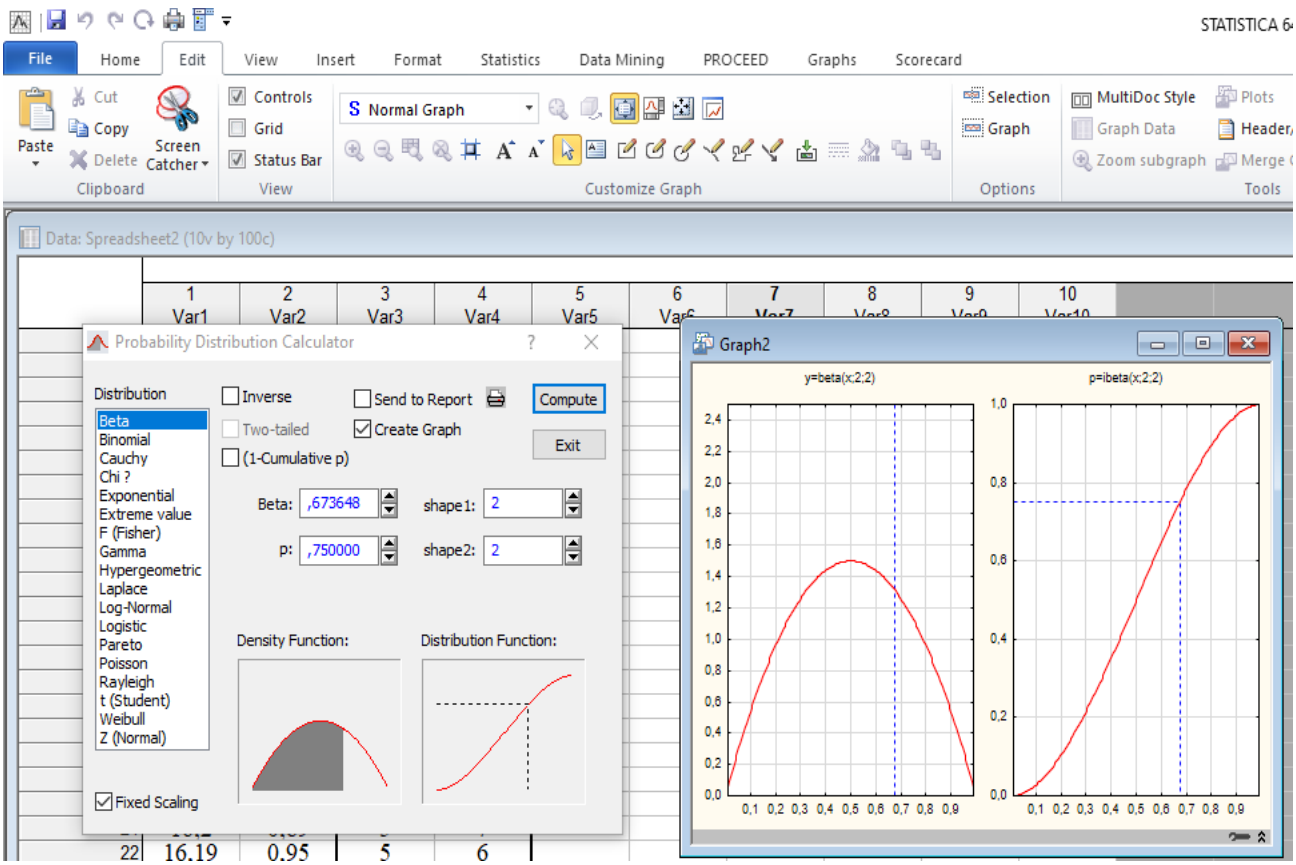


Рис. 4.13. Графіки щільності та функції розподілу

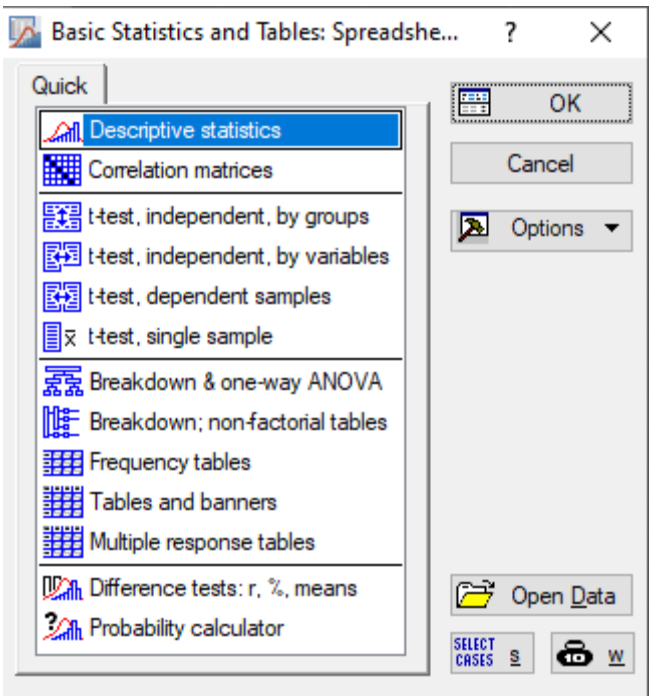
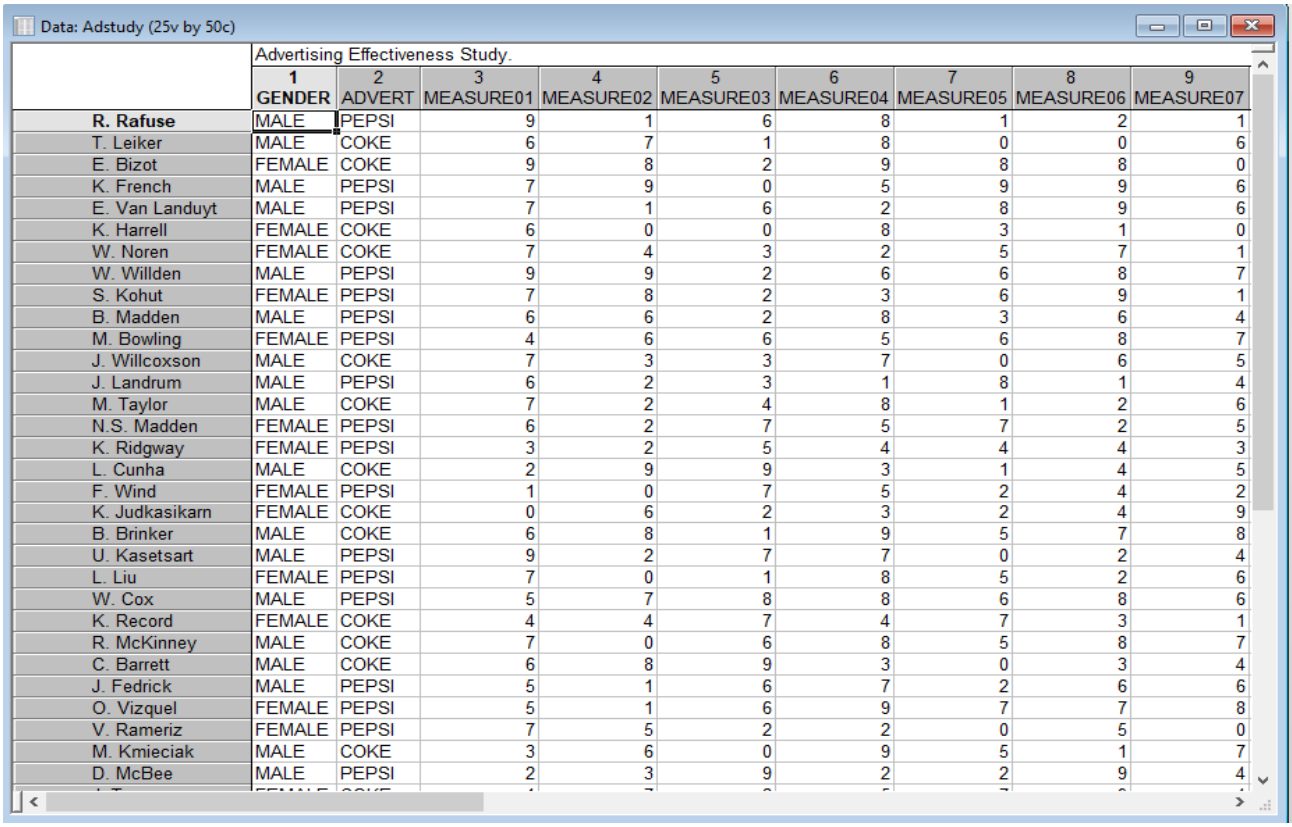


Рис. 4.14. Початок роботи з модулем



Data: Adstudy (25v by 50c)

Advertising Effectiveness Study.

	1 GENDER	2 ADVERT	3 MEASURE01	4 MEASURE02	5 MEASURE03	6 MEASURE04	7 MEASURE05	8 MEASURE06	9 MEASURE07
R. Rafuse	MALE	PEPSI	9	1	6	8	1	2	1
T. Leiker	MALE	COKE	6	7	1	8	0	0	6
E. Bizot	FEMALE	COKE	9	8	2	9	8	8	0
K. French	MALE	PEPSI	7	9	0	5	9	9	6
E. Van Landuyt	MALE	PEPSI	7	1	6	2	8	9	6
K. Harrell	FEMALE	COKE	6	0	0	8	3	1	0
W. Noren	FEMALE	COKE	7	4	3	2	5	7	1
W. Willden	MALE	PEPSI	9	9	2	6	6	8	7
S. Kohut	FEMALE	PEPSI	7	8	2	3	6	9	1
B. Madden	MALE	PEPSI	6	6	2	8	3	6	4
M. Bowling	FEMALE	PEPSI	4	6	6	5	6	8	7
J. Willcoxson	MALE	COKE	7	3	3	7	0	6	5
J. Landrum	MALE	PEPSI	6	2	3	1	8	1	4
M. Taylor	MALE	COKE	7	2	4	8	1	2	6
N.S. Madden	FEMALE	PEPSI	6	2	7	5	7	2	5
K. Ridgway	FEMALE	PEPSI	3	2	5	4	4	4	3
L. Cunha	MALE	COKE	2	9	9	3	1	4	5
F. Wind	FEMALE	PEPSI	1	0	7	5	2	4	2
K. Judkasikarn	FEMALE	COKE	0	6	2	3	2	4	9
B. Brinker	MALE	COKE	6	8	1	9	5	7	8
U. Kasetsart	MALE	PEPSI	9	2	7	7	0	2	4
L. Liu	FEMALE	PEPSI	7	0	1	8	5	2	6
W. Cox	MALE	PEPSI	5	7	8	8	6	8	6
K. Record	FEMALE	COKE	4	4	7	4	7	3	1
R. McKinney	MALE	COKE	7	0	6	8	5	8	7
C. Barrett	MALE	COKE	6	8	9	3	0	3	4
J. Fedrick	MALE	PEPSI	5	1	6	7	2	6	6
O. Vizquel	FEMALE	PEPSI	5	1	6	9	7	7	8
V. Rameriz	FEMALE	PEPSI	7	5	2	2	0	5	0
M. Kmieciak	MALE	COKE	3	6	0	9	5	1	7
D. McBee	MALE	PEPSI	2	3	9	2	2	9	4

Рис. 4.15. Внесення статистичних даних

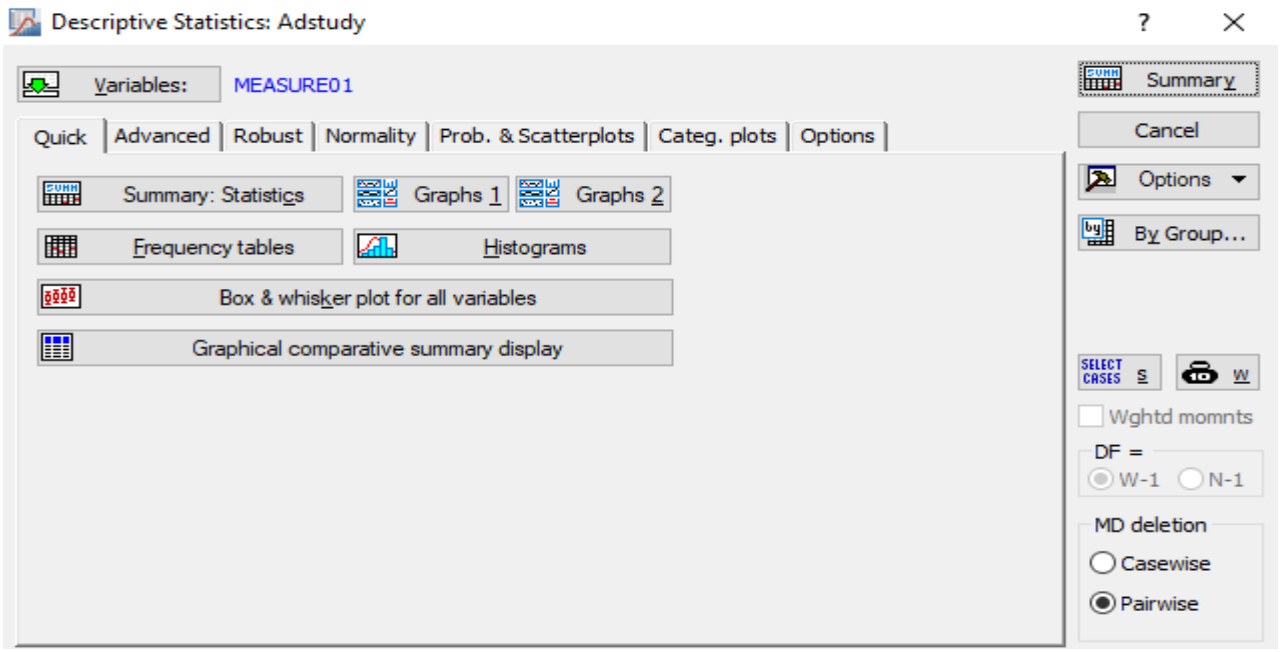


Рис. 4.16. Вигляд закладки Descriptive Statistics

Для того, щоб побачити описові статистики, слід натиснути Summary. В інших закладках можна налаштувати і подивитись детальніші характеристики.

Так, якщо вибрати закладку Advanced, а потім натиснути Select all stats та Summary, то отримаємо інші характеристики даних (рис. 4.17).

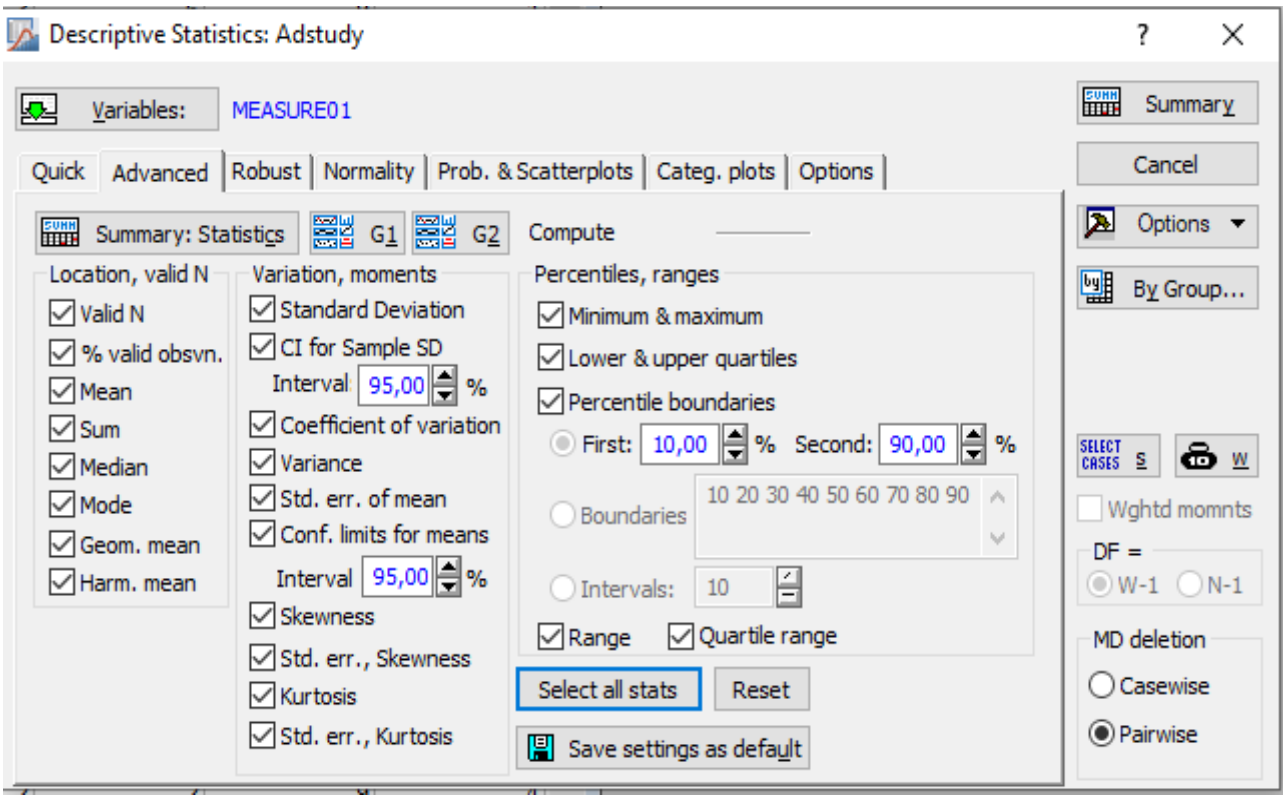


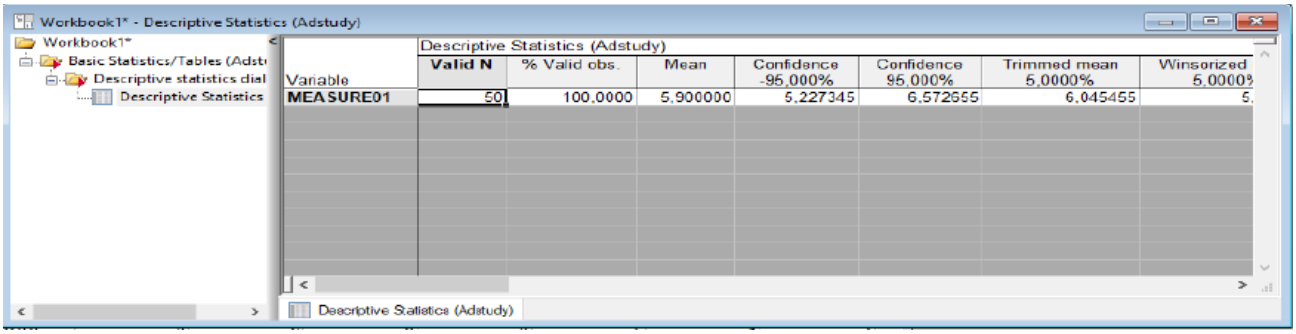
Рис. 4.17. Вигляд на екрані різних характеристик даних

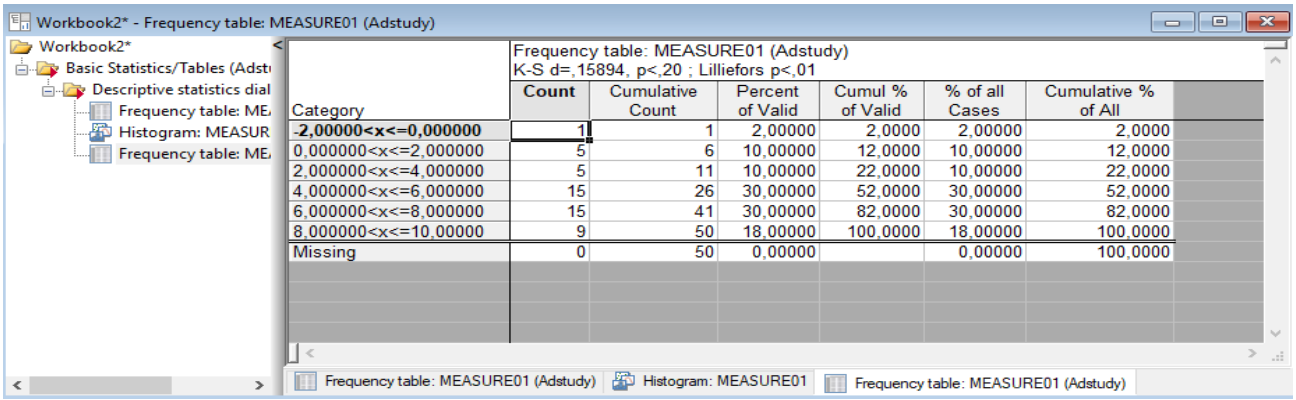
Зокрема, значення Skewness показує коефіцієнт асиметрії, тобто наскільки розподіл «скособочений», значення Kurtosis показує наскільки розподіл «пікоподібний». Для стандартного нормального розподілу Skewness та Kurtosis дорівнюють нулеві.

Повернемось до закладки Quick (рис. 4.16). Натиснувши кнопку Frequency Tables отримують таблицю частот для вибірки. Натиснувши кнопку Histograms отримують гістограму, на якій червоною лінією зображено підігнану криву нормального розподілу (рис. 4.18).

В закладці Options ми можемо обрати тип «коробки з вусами». Оберемо перший тип Median/Quart/Range. Повернемось у закладку Quick і натиснемо Box & whisker plot for all variables –з'явиться вікно з рисунком коробки з вусами, в якій маленький прямокутник відповідає значенню медіани, великий

прямокутник –нижній та верхній квартилі, а вуса –найменшому та найбільшому значенню вибірки (рис. 4.19).





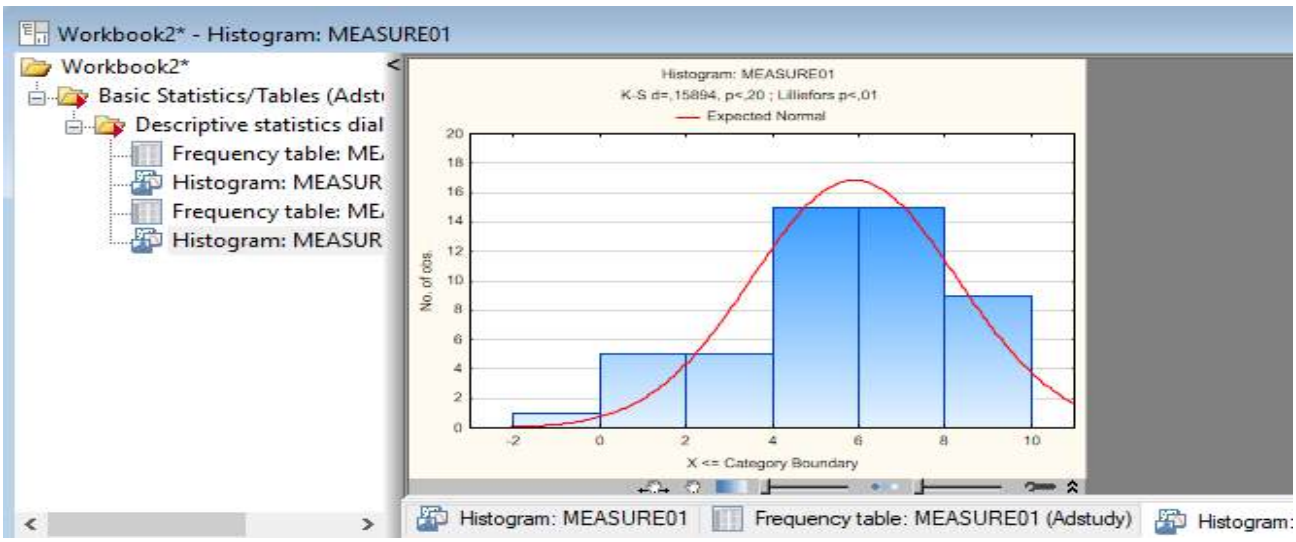


Рис. 4.18. Вигляд гістограми

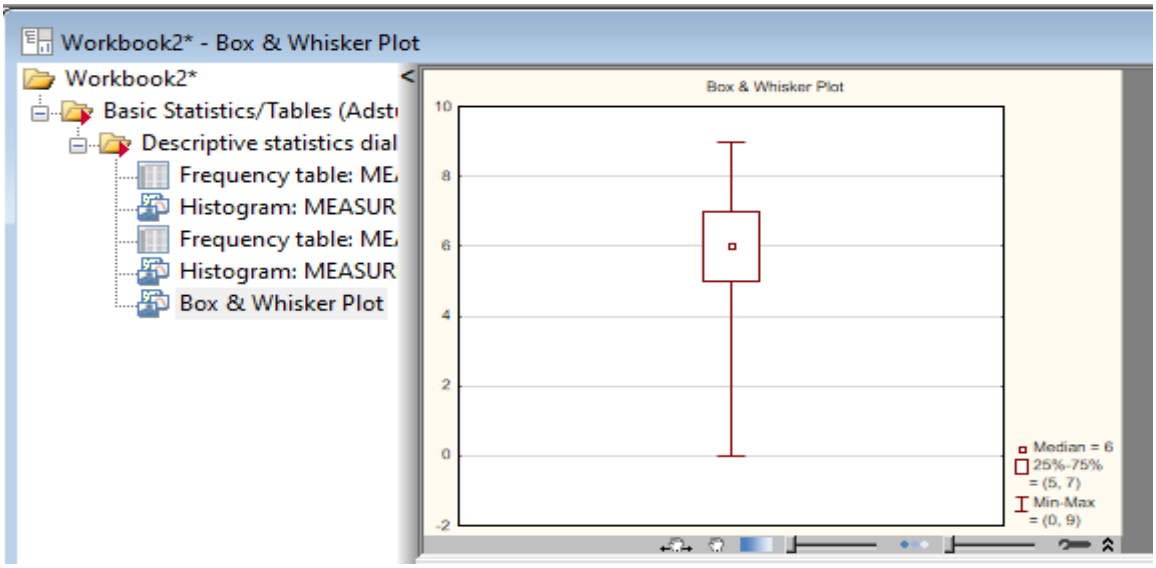


Рис. 4.19. Вигляд «коробки з вусами»

Проводять також порівняльну характеристику даних вибірок. В закладці Categorized Plots натискають Categorized box & whisker plots (рис. 4.20). Далі обирають I-GENDER як першу змінну, а 2-ADVERT як другу. Третю змінну для описаного вище прикладу не вказують, тому натискають ОК. З'явиться вікно Select Codes far the grouping variables. Для Gender і для Advert вибирають ALL (рис. 4.20), далі ОК.

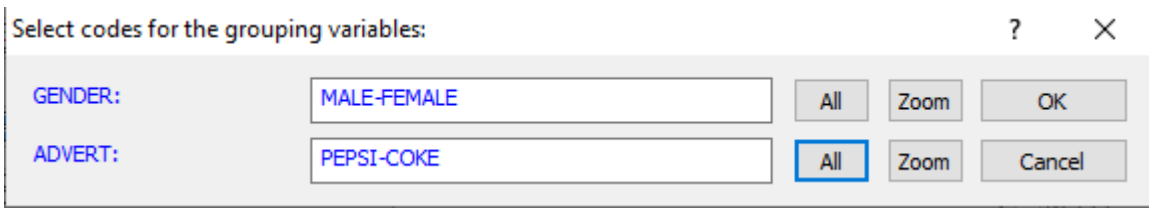


Рис. 4.20. Підготовка до порівняння

У вікні, що з’явиться (рис. 4.21), отримуємо «коробки з вусами» окремо для Pepsi і Coke, і окремо для чоловіків і жінок.

Бачимо, що реклама Pepsi подобається чоловікам більше, ніж жінкам, а реклама Coke навпаки. При цьому у жінок розкид уподобань більший. Зауважимо, що отримані дані стосуються першої вибірки.

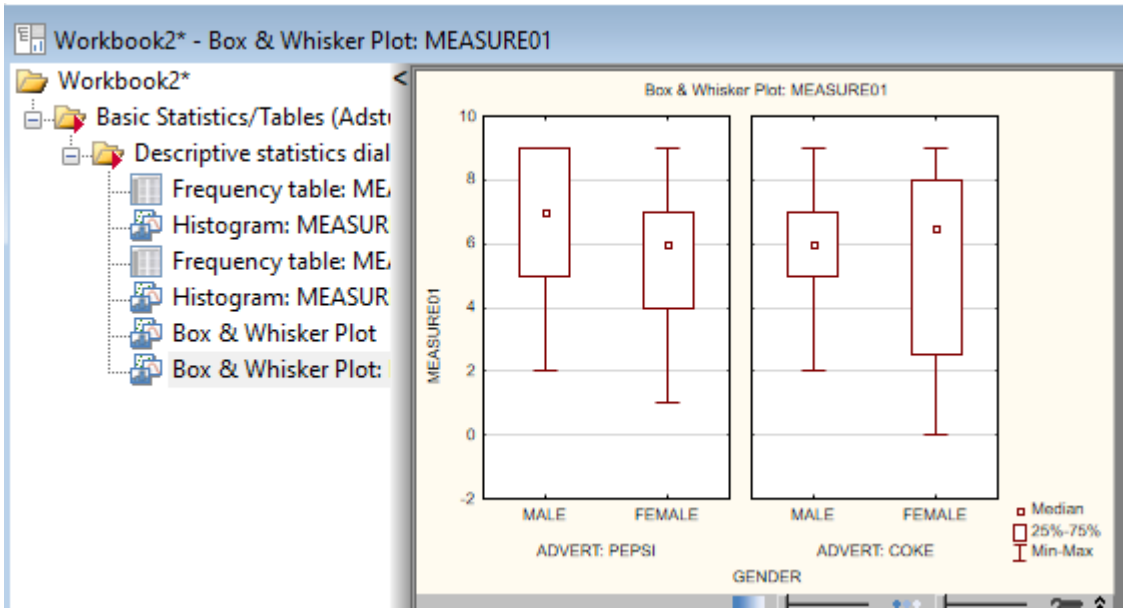


Рис. 4.21. Вигляд на екрані порівняння уподобань

Приклад 4.2. За наведеними результатами 50-ти вимірів (дані подано таблицею в Excel, рис. 4.22) значень деякої неперервної випадкової величини потрібно:

1) згрупувати результати спостережень (побудувати інтервальний статистичний ряд);

2) побудувати гістограму частот та емпіричну функцію розподілу;

3) обчислити основні числові характеристики.

	A	B	C	D	E	F	G	H	I	J	K
1											
2		24	23	20	28	32	28	12	19	18	40
3		17	39	38	22	24	23	31	22	30	23
4		15	14	24	30	23	12	34	27	28	22
5		13	11	24	8	24	26	19	7	24	27
6		22	29	24	18	26	25	24	15	20	25
7											

Рис. 4.22. Дані спостережень

Розв'язання. На даному прикладі показано опрацювання даних за допомогою теоретичних базових знань та із використанням програмного продукту Microsoft Excel і пакету STATISTICA.

1) Побудова інтервального статистичного ряду (угрупування). Знайдемо розмах вибірки: $R = x_{\max} - x_{\min} = 40 - 7 = 33,$ попередньо обчисливши найменше та найбільше числове значення вибірки (рис. 4.23).

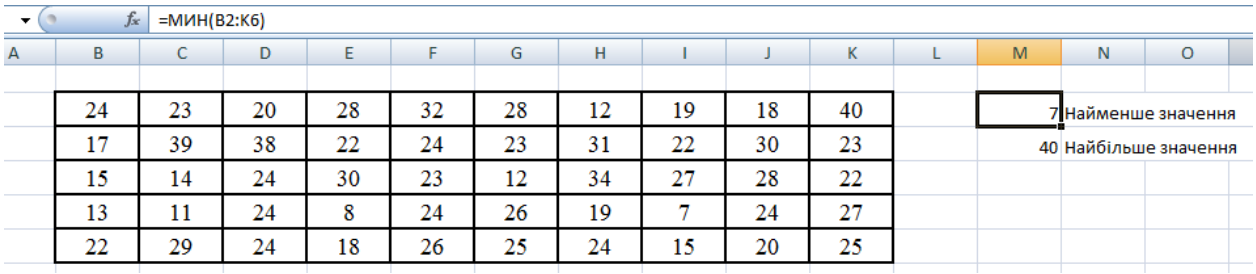


Рис. 4.23. Обчислення найменшого та найбільшого значень вибірки

Кількість інтервалів для інтервального ряду знаходять за формулою Стерджесса: $k = 1 + 3{,}32 \cdot \lg n \approx 7$, де число n є обсягом вибірки. Якщо використовувати Excel, то потрібно задати формулу в такому вигляді:

=**ОКРВВЕРХ**(1+3,32*LOC10(**СЧЕТ**(B2:K6);1).

Отже, розіб'ємо проміжок (7 ; 40) на $k = 7$ інтервалів, причому крок (ширину) інтервалу знайдемо за формулою:

$$h = \frac{R}{k} = \frac{33}{7} = 4{,}7 \approx 5.$$

Подальші дослідження передбачають обчислення кількості значень, які потрапляють у вказані проміжки розбиття. Їх можна прорахувати вручну, можна із використанням вбудованої функції

ЧАСТОТА категорії «Статистические» згідно з її описанням (массив_данных;массив_интервалов).

Варто зауважити, що ця функція вводиться як формула масиву після виділення інтервалу суміжних комірок, в які потрібно повернути отриманий масив даних, причому кількість елементів для повернення в комірки повинна

бути заданою на одиницю більшою «масиву інтервалів». Для даного прикладу (рис. 4.24) прораховуємо частоти, дані заносимо в таблицю (рис. 4.25).

На рис. 4.25 також показано побудову значень емпіричної функції розподілу (кумуляти), яку і зображено графічно (рис. 4.26).

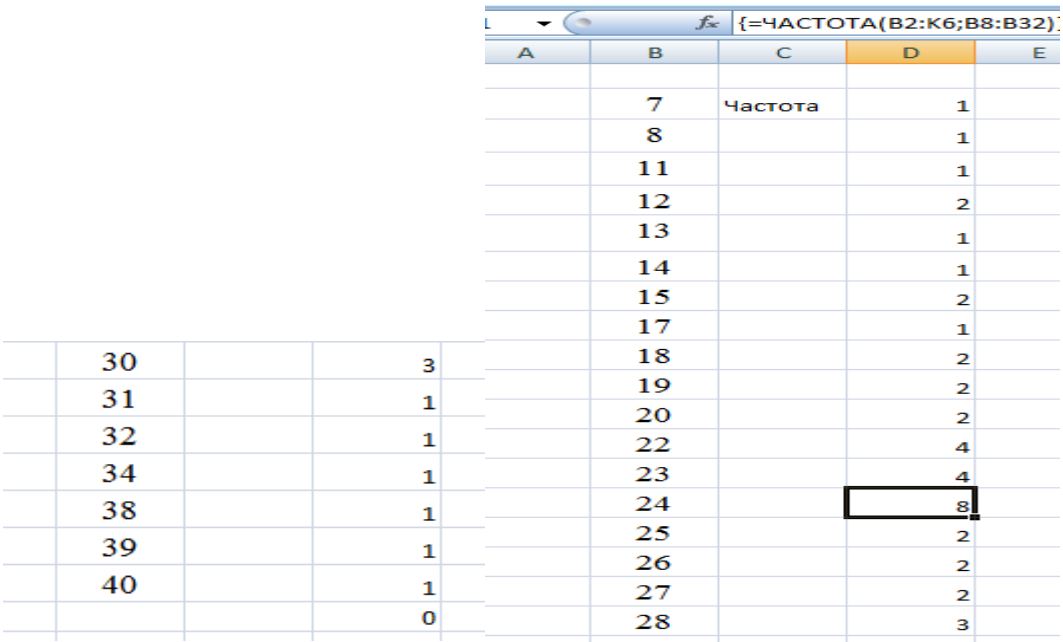


Рис. 4.24. Використання функції **ЧАСТОТА**

	1	2	3	4	5	6	7	Разом
I_i	[7-12)	[12-17)	[17-22)	[22-27)	[27-32)	[32-37)	[37-42)	
n_i	3	6	9	19	8	2	3	50
wi-відн. Ч	0,06	0,12	0,18	0,38	0,16	0,04	0,06	1
F_i^*	0,06	0,18	0,36	0,74	0,9	0,94	1	

Рис. 4.25. Статистичний інтервальний ряд частот(відносних частот)

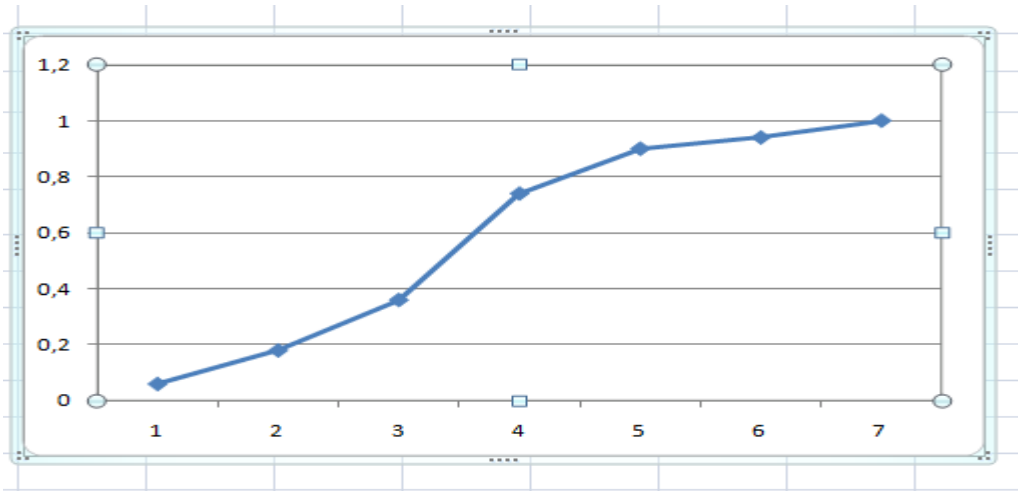


Рис. 4.26. Емпірична функція розподілу

За відомою теорією п. 3.1.6.2 розділу 3 для даного інтервального статистичного ряду побудовано гістограму відносних частот (рис. 4.27).

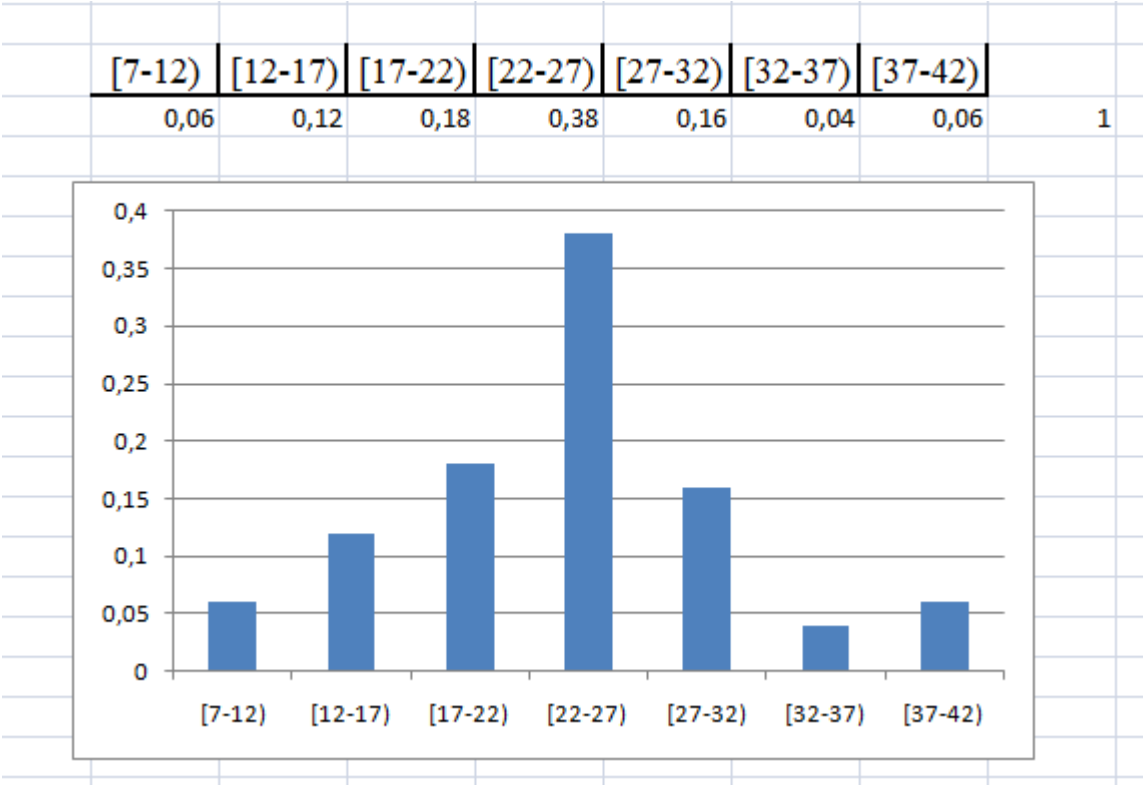


Рис. 4.27. Гістограма розподілу

Обчислення моди і медіани

Моду для інтервального статистичного ряду обчислюють за формулою

$$M_0 = x_0 + h\frac{n_0 - n_{0-1}}{(n_0 - n_{0-1}) + (n_0 - n_{0+1})},$$

де x_0 – початок модального інтервалу, h – довжина інтервалу, n_0 – частота модального інтервалу, n_{0-1} –частота перед модальним інтервалом, n_{0+1} –частота після модального інтервалу.

Модальним є інтервал [22, 27) (там, де найбільша частота), тому отримано:

$$M_0 = 22 + 5 \cdot \frac{19 - 9}{(19 - 9) + (19 - 8)} = 24{,}38.$$

Знайдемо *медіану.*

Медіанним буде також інтервал [22, 27) і обчислення проводяться за формулою

$$M_e = x_e + h \cdot \frac{\frac{\sum n}{2} - \sum_{i=1}^{l-1} n_i}{n_e},$$

де x_e–початок медіального інтервалу, h-довжина інтервалу, $\frac{\sum n}{2}$–півсума частот, $\sum_{i=1}^{l-1} n_i$ – сума частот до медіанного інтервалу, n_e– частота медіанного інтервалу. Отже,

$$M_e = 22 + 5 \cdot \frac{25-18}{19} = 22{,}36.$$

Зручніше використати для обчислень моди та медіани вбудовані функції (рис. 4.28) програмного пакету Microsoft Excel.

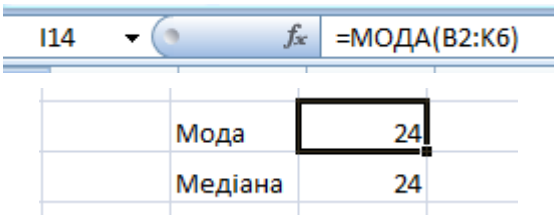


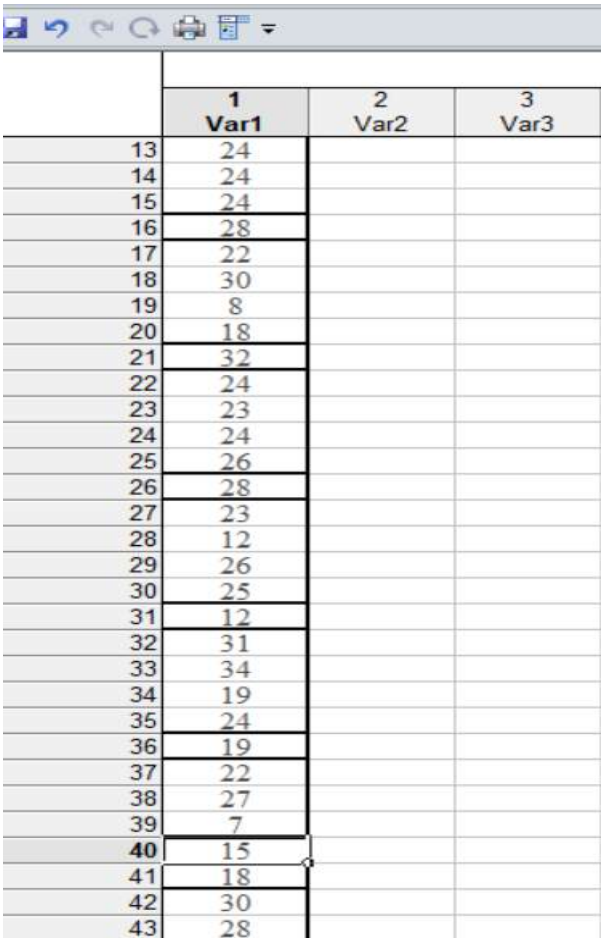
Рис. 4.28. Обчислення моди та медіани

Проведені дослідження є підтвердженням зручності використання готових програмних пакетів. З точки зору ще більшої економії часу є зручнішим пакет STATISTICA.

Для початку потрібно перенести дані з таблиці Excel. Зручно виконати за допомогою звичайних операцій Copy в Excel та Paste в Statistica (рис. 4.29), вибравши одну змінну VAR1.

Послідовне натискання Statistics → Basic Statistics → Descriptive Statistics (рис. 4.30) та кнопки ОК (рис. 4.31) приведе до початку виконання задачі та отримання результатів досліджень.

Обравши Graphs 1 та змінну Var 1 (рис. 4.32) отримуємо зразу всі обчислення поставленої задачі (рис. 4.33).



	1 Var1	2 Var2	3 Var3
13	24		
14	24		
15	24		
16	28		
17	22		
18	30		
19	8		
20	18		
21	32		
22	24		
23	23		
24	24		
25	26		
26	28		
27	23		
28	12		
29	26		
30	25		
31	12		
32	31		
33	34		
34	19		
35	24		
36	19		
37	22		
38	27		
39	7		
40	15		
41	18		
42	30		
43	28		

Рис. 4.29. Фрагмент таблиці даних

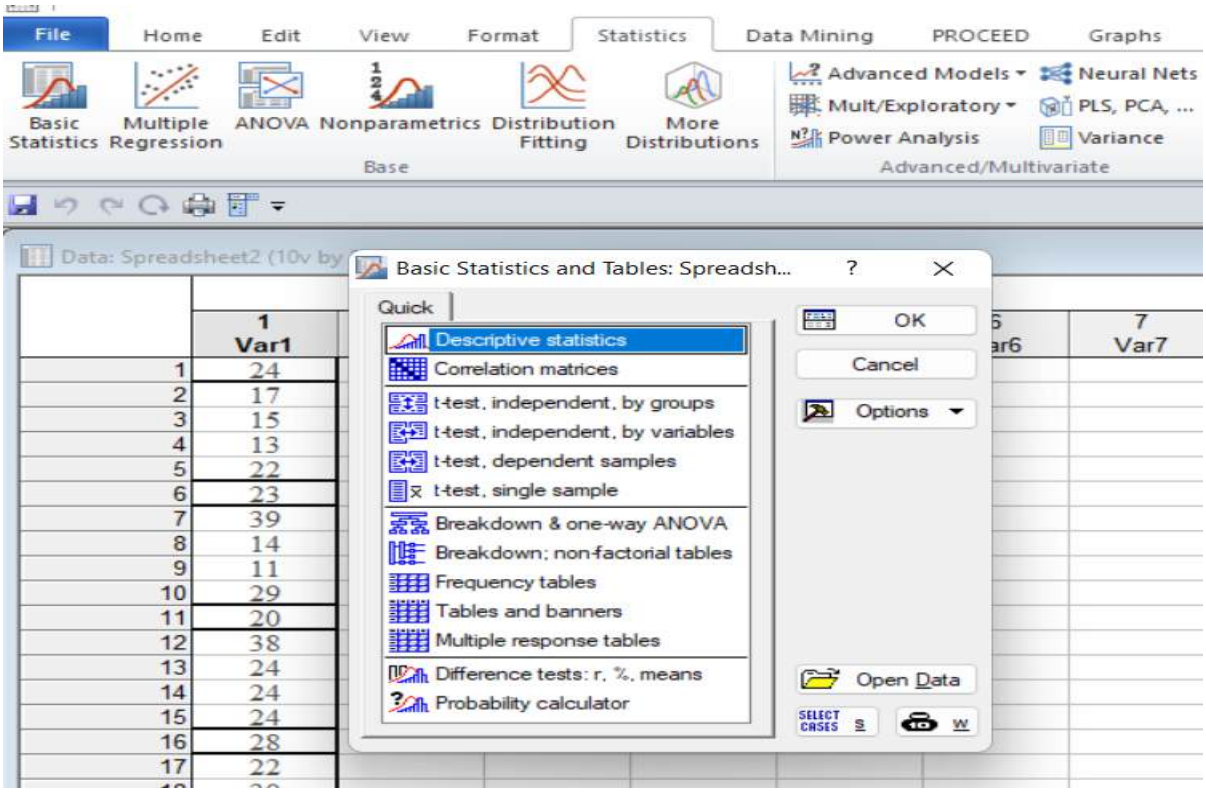


Рис. 4.30. Початок виконання прикладу.

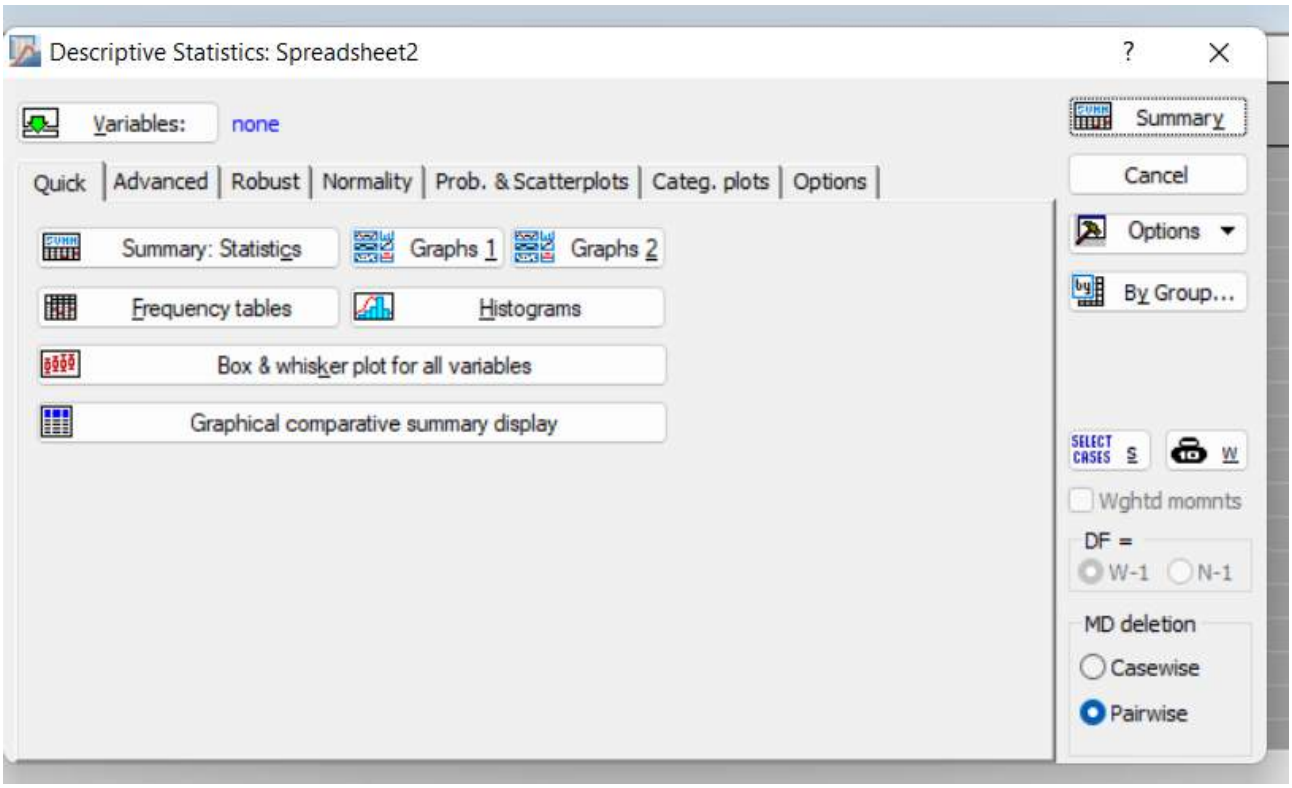


Рис. 4.31. Вигляд на екрані для вибору кнопки

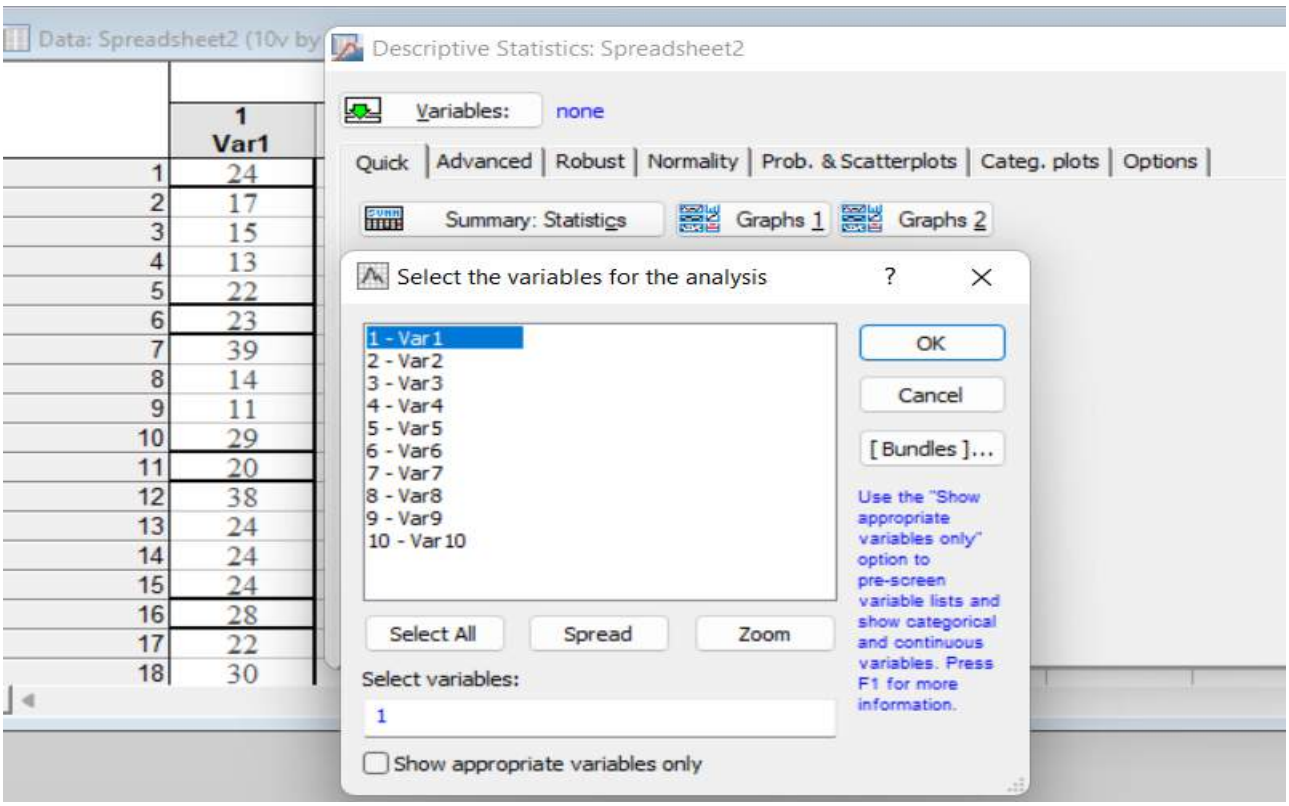


Рис. 4.32. Вибір Graphs 1 та даних

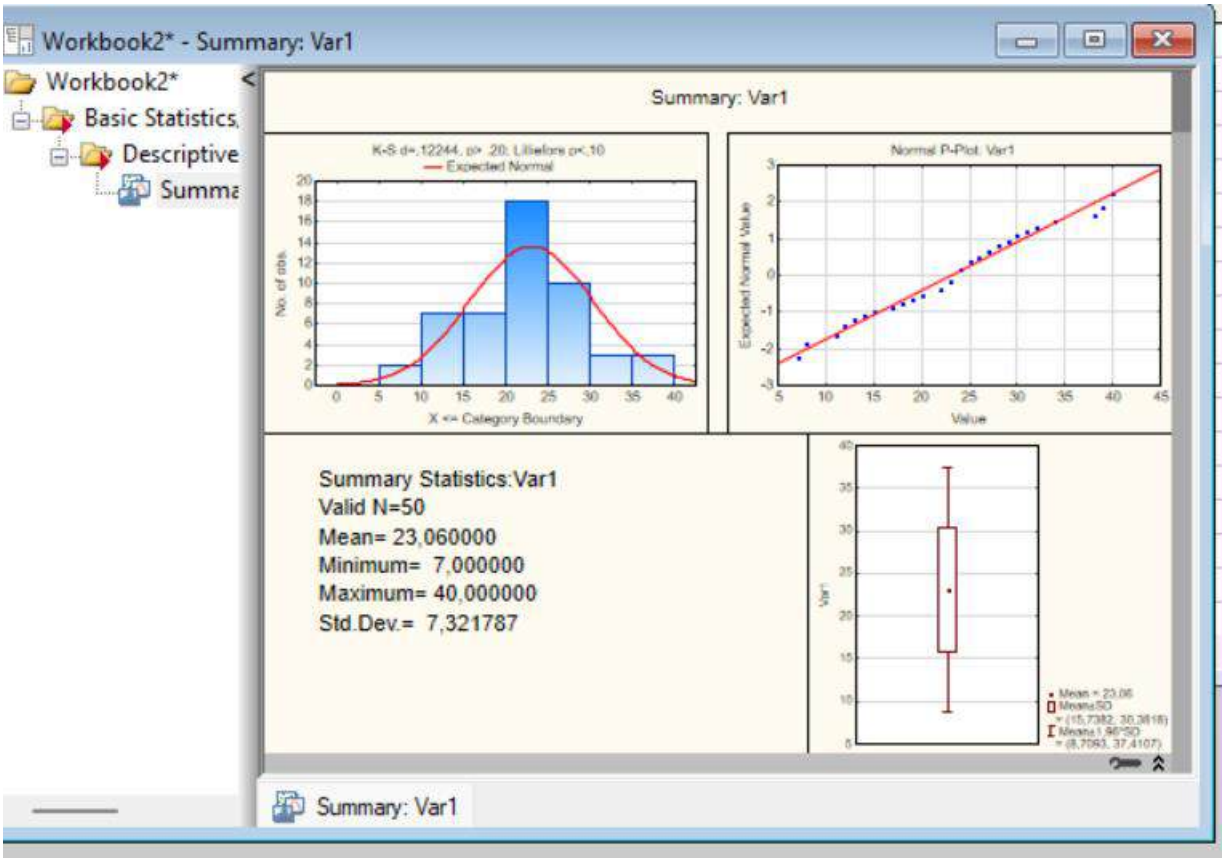


Рис. 4.33. Вигляд обчислень

На рис. 4.33 побудована гістограма частот, вказано обсяг вибірки, медіана, найменше та найбільше значення, середнє квадратичне відхилення; зображена «коробка з вусами», також є картинка з точками та прямою, яка є підказкою до визначення закону розподілу досліджуваних даних (буде розглянуто детальніше далі).

Відносно побудови інтервального статистичного ряду, то також розбиття (рис. 4.34, Frequency table) на інтервали з кроком 5, але кількість інтервалів є на одиницю більшою, ніж за формулою Стерджесса (перший інтервал не є інформативний, так як мінімальне значення вибірки дорівнює 7). За даними (рис. 4.34) можна побудувати варіаційний ряд та емпіричну функцію розподілу.

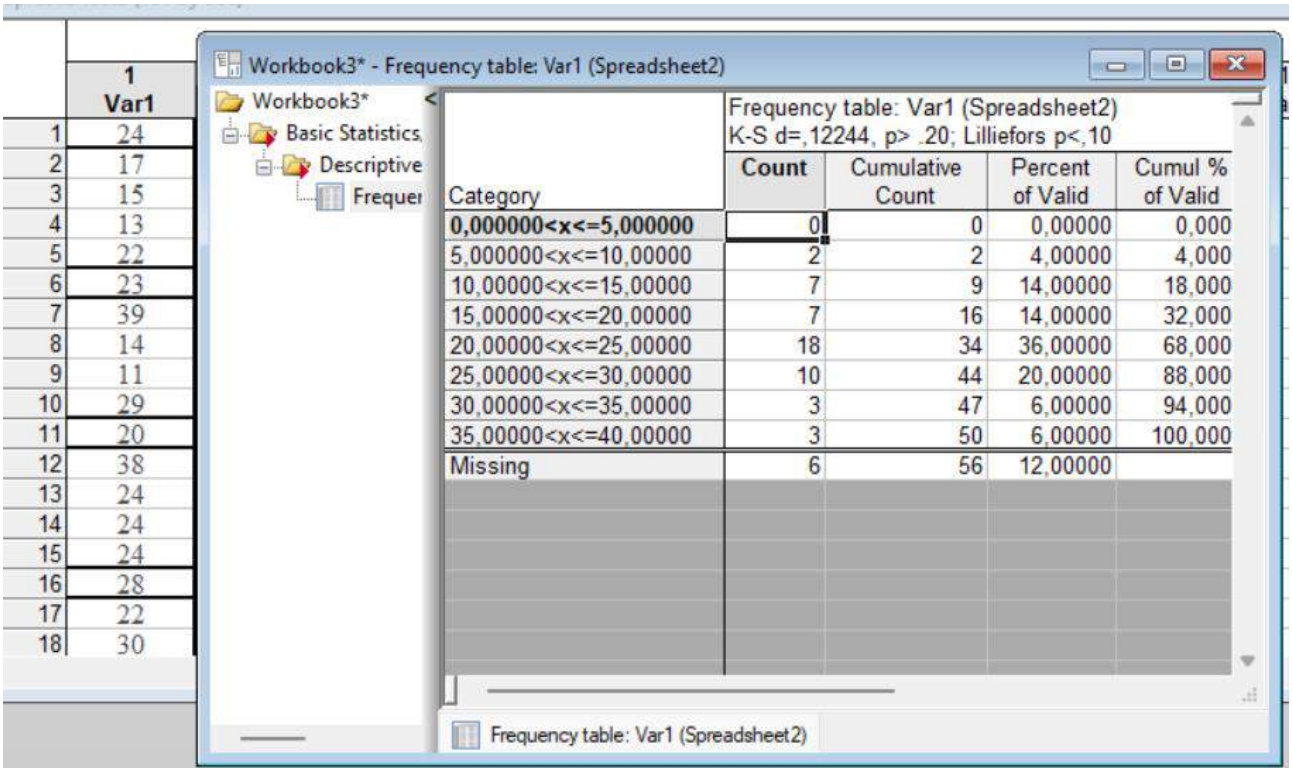


Рис. 4.34. Вигляд на екрані з роботою Frequency table

4.2.2. Перевірка гіпотез

Для обрання типу розподілу, з яким візуально найкраще узгоджується вибірка, послідовно виконують: Graphs → 2D Graphs → Quantile-Quantile Plots (рис. 4.35).

Далі потрібно перейти на закладку Advanced, за допомогою кнопки Variables задати потрібну змінну. У полі Distribution вибрати розподіл, на відповідність якому перевіряється вибірка (рис. 4.36) і, відповідно, ОК.

Як приклад, показано візуальне не підтвердження експоненціального закону розподілу.

У результаті отримуємо *Q-Q* графік, який аналізують таким чином: чим ближче точки графіка розміщені до прямої, тим краще заданий тип розподілу описує дані вибірки. Наприклад, з рис. 4.37 видно, що експоненціальний розподіл візуально не підходить до даних вибірки.

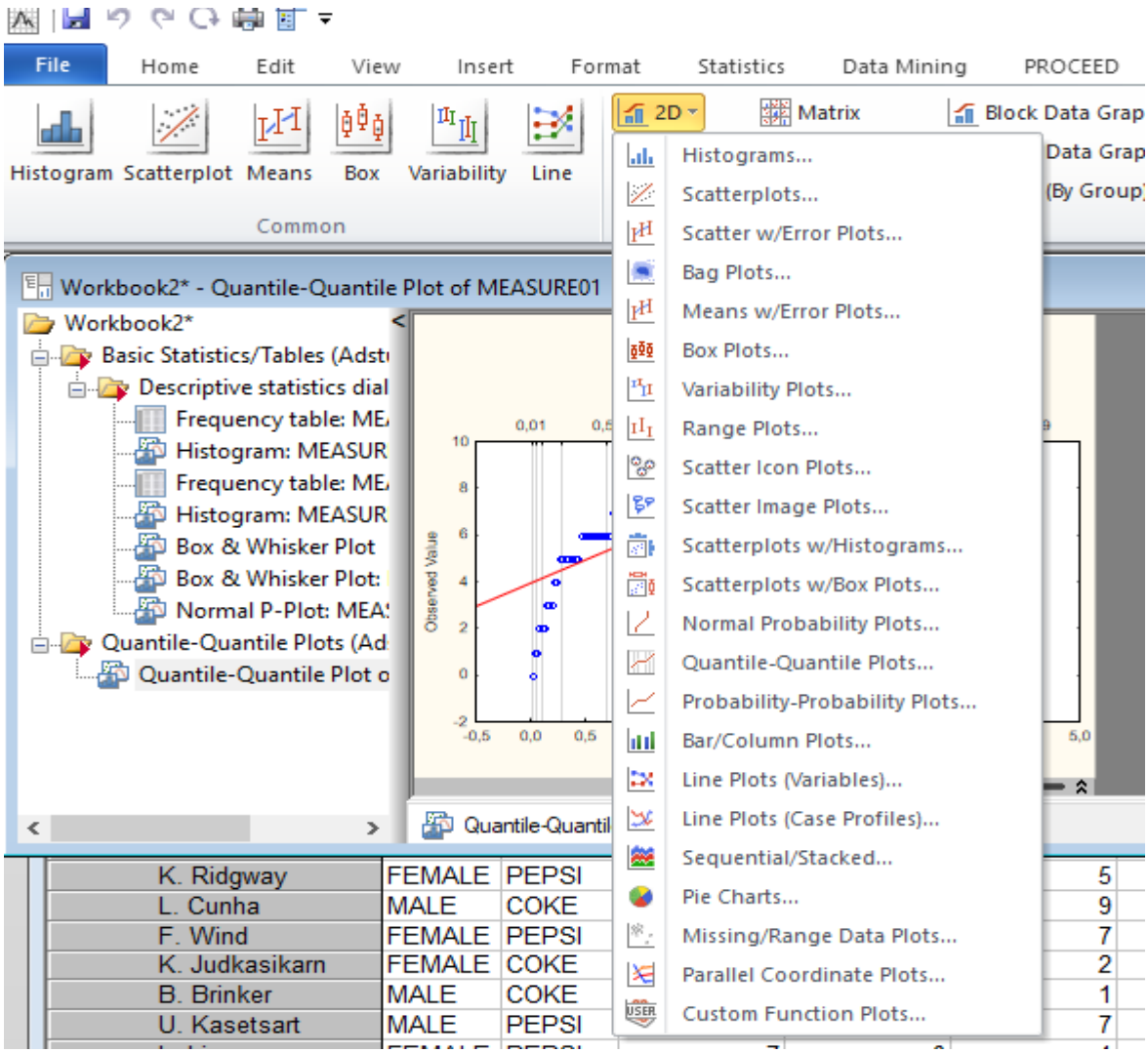


Рис. 4.35. Обрання типу розподілу

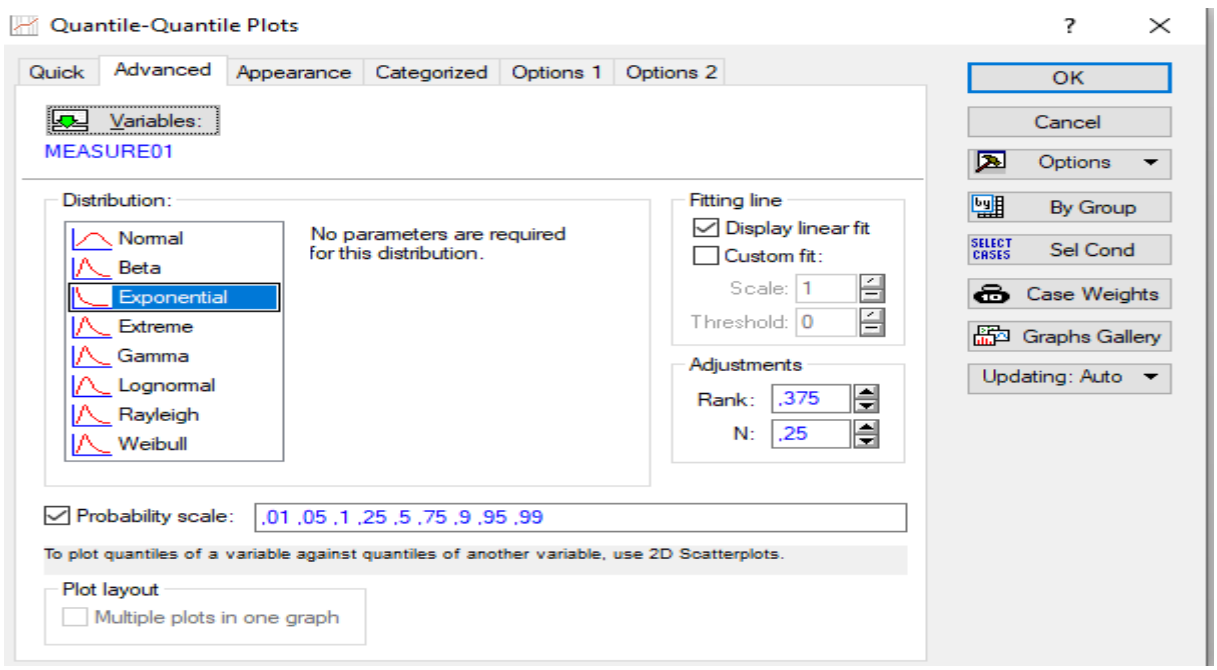


Рис. 4.36. Вибір експоненціального закону розподілу

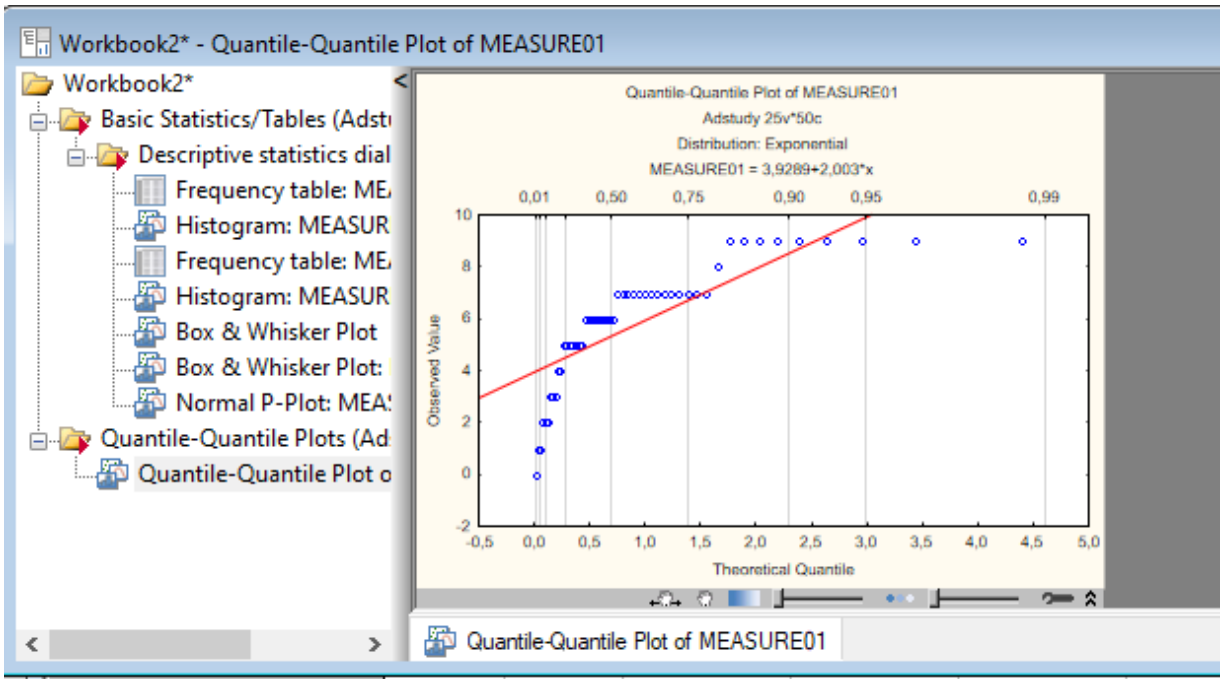


Рис. 4.37. Графічна апроксимація на екрані

4.2.2.1. Перевірка гіпотези на нормальний закон розподілу

В 3.3 «Статистична перевірка гіпотез» описано теоретично і практично використання критеріїв згоди для перевірки статистичних гіпотез про нормальний закон розподілу. Найчастіше перевірку відповідності вибраного теоретичного закону проводять за критеріями згоди Колмогорова і χ^2 (хі-квадрат) Пірсона (найбільш універсальний, його можна застосувати до розподілу будь-якого виду).

Дослідивши емпіричний розподіл статистичної вибірки значень, необхідно також встановити, яким теоретичним законом він описується. Часто в таких випадках у дослідника виникають певні гіпотези про відповідність емпіричного розподілу тому чи іншому закону, наприклад нормальному або логнормальному. В такому разі висувається основна і альтернативна статистичні гіпотези наступного виду:

- H_0: розподіл значень відповідає закону розподілу A;
- H_1: розподіл не відповідає закону розподілу A.

В якості A може виступати будь-який закон розподілу.

В 3.2.2.2 (перевірка гіпотези про закони розподілу) за пунктами описано послідовність дій (алгоритм) для перевірки за критерієм згоди χ^2, детально показано для перевірки гіпотези H_0: генеральна сукупність має нормальний закон розподілу $N(a,\ \sigma^2)$, де $a=\overline{x},\ \sigma^2=s^2$. Також наведено приклади перевірки із використанням програмного продукту Microsoft Excel.

Розглянемо послідовність виконання розрахунків за допомогою програми STATISTICA.

Перевірка гіпотези на нормальний розподіл за критеріями Колмогорова-Смірнова та χ^2 (хі-квадрат)

1. Вибираємо в меню програми розрахунковий модуль «Distribution Fitting», тобто Statistics → Distribution Fitting, далі нормальний розподіл (рис. 4.38) і OK.

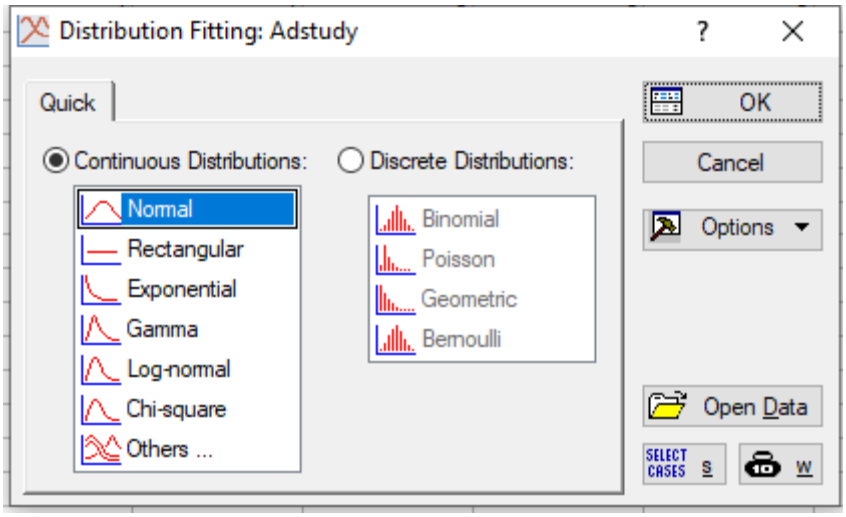


Рис. 4.38. Вибір закону розподілу

Зауваження. За допомогою меню з 2 груп запропонованих розподілів (рис. 4.38) «Continuous Distributions» − неперервні розподіли (6 основних типів і можливість додаткового вибору «Others») та «Discrete Distributions» −дискретні розподіли (4 типи) вибирають один з теоретичних законів розподілу, який, враховуючи висновки візуального оцінювання емпіричного розподілу досліджуваного ряду, найбільше підходить до емпіричного розподілу.

2. Після цього натискаємо кнопку ОК, вибираємо розрахунковий модуль «Fitting Continuous Distribution» (рис. 4.39), в якому вказуємо досліджуваний ряд, як Variable обираємо measure1.

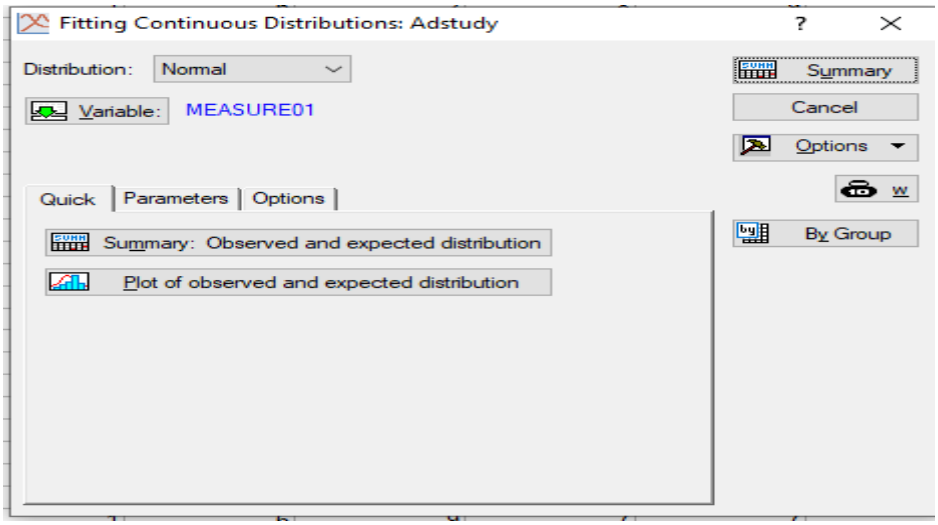


Рис. 10.39. Вибір ряду для дослідження

3. У закладці Options відмітимо Yes (continuous) у розділі Kolmogorov-Smirnov test та кнопку Summary (доступ до розрахунку емпіричних («observed») і теоретичних («expected») частот розподілу (рис. 4.39)).

Зауваження. В цьому режимі можна почергово підбирати усі доступні у меню програми теоретичні закони і вибрати оптимальний за найменшим значенням критерію χ^2.

4. Натискання кнопки Summary (рис. 4.40) і приводить зразу до статистики для тестів Колмогорова-Смірнова та χ^2.

5. Для наочності натискання кнопки Plot of observed and expected distribution на закладці Quick дає графічне порівняння гістограми з щільністю обраного нормального розподілу (рис. 4.41).

6. Аналіз отриманих результатів. Перевірка гіпотези про відповідність емпіричного розподілу теоретичному закону (нормальному) за результатами розрахунку критерію Пірсона, Колмогорова, які відображаються на екрані і в таблиці розподілу частот, і на гістограмі.

Variable: MEASURE01, Distribution: Normal (Adstudy)
Kolmogorov-Smirnov d = 0,15894, p < 0,20, Lilliefors p < 0,01
Chi-Square = 13,18127, df = 4 (adjusted) , p = 0,01042

Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumu Expe
<= 0,00000	1	1	2,00000	2,0000	0,316894	0,31689	0,63379	0
1,00000	2	3	4,00000	6,0000	0,643827	0,96072	1,28765	1
2,00000	3	6	6,00000	12,0000	1,524375	2,48510	3,04875	4
3,00000	3	9	6,00000	18,0000	3,026925	5,51202	6,05385	11
4,00000	2	11	4,00000	22,0000	5,040955	10,55298	10,08191	21
5,00000	7	18	14,00000	36,0000	7,041017	17,59399	14,08203	35
6,00000	8	26	16,00000	52,0000	8,248521	25,84252	16,49704	51
7,00000	14	40	28,00000	80,0000	8,104725	33,94724	16,20945	67
8,00000	1	41	2,00000	82,0000	6,679156	40,62640	13,35831	81
9,00000	9	50	18,00000	100,0000	4,616597	45,24299	9,23319	90
< Infinity	0	50	0,00000	100,0000	4,757007	50,00000	9,51401	100

Рис. 4.40. Статистика опрацьованих даних

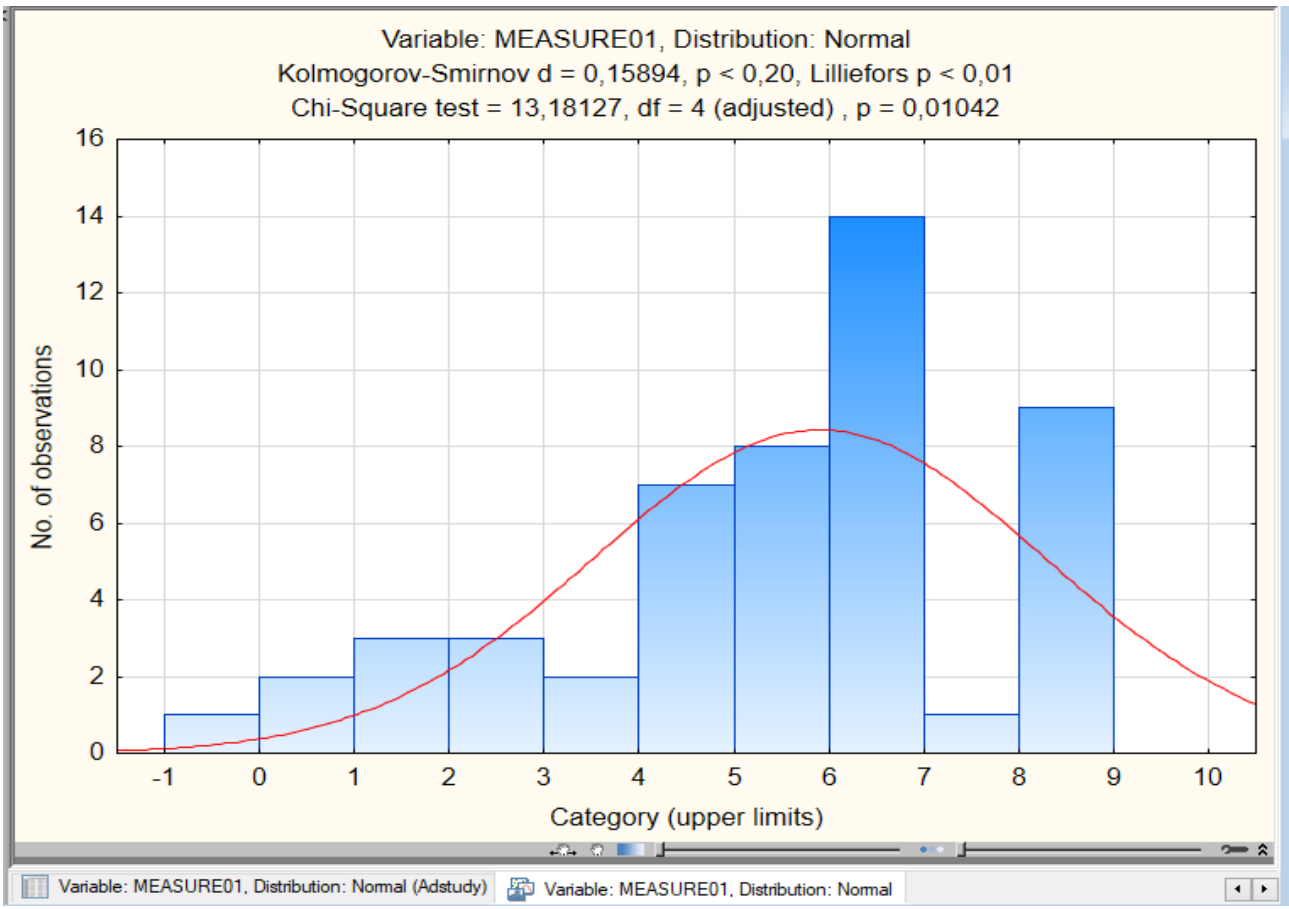


Рис. 4.41. Гістограма і графік щільності розподілу

Якщо тести Kolmogorov-Smirnov та χ^2 (Chi-Square) видають значення *p*, які відмінні від 0 або є запис «n.s.», то це означає, що ймовірність помилки при відхиленні гіпотези про даний тип розподілу велика, і гіпотезу потрібно прийняти.

Зауваження. Можна переконатись ще і перевіркою отриманого значення $\chi^2 = 13{,}18127$ із критичним значенням, використовуючи таблицю значень χ^2 в залежності кількості степенів вільності та рівня значущості (Додаток). Враховують тоді в даному прикладі $s = df = 4$, $\alpha = p = 0{,}01$, відпвідно, за таблицею $\chi^2_{\text{кр}} = 13{,}3$, $\chi^2 < \chi^2_{\text{кр}}$, гіпотеза приймається.

Зауваження. Для того, щоб візуально побачити, наскільки вибірка відповідає нормальному закону розподілу у вікні Descriptive Statistics в закладці Prob. & Scatterplots натискають Normal Probability Plot (рис. 4.42).

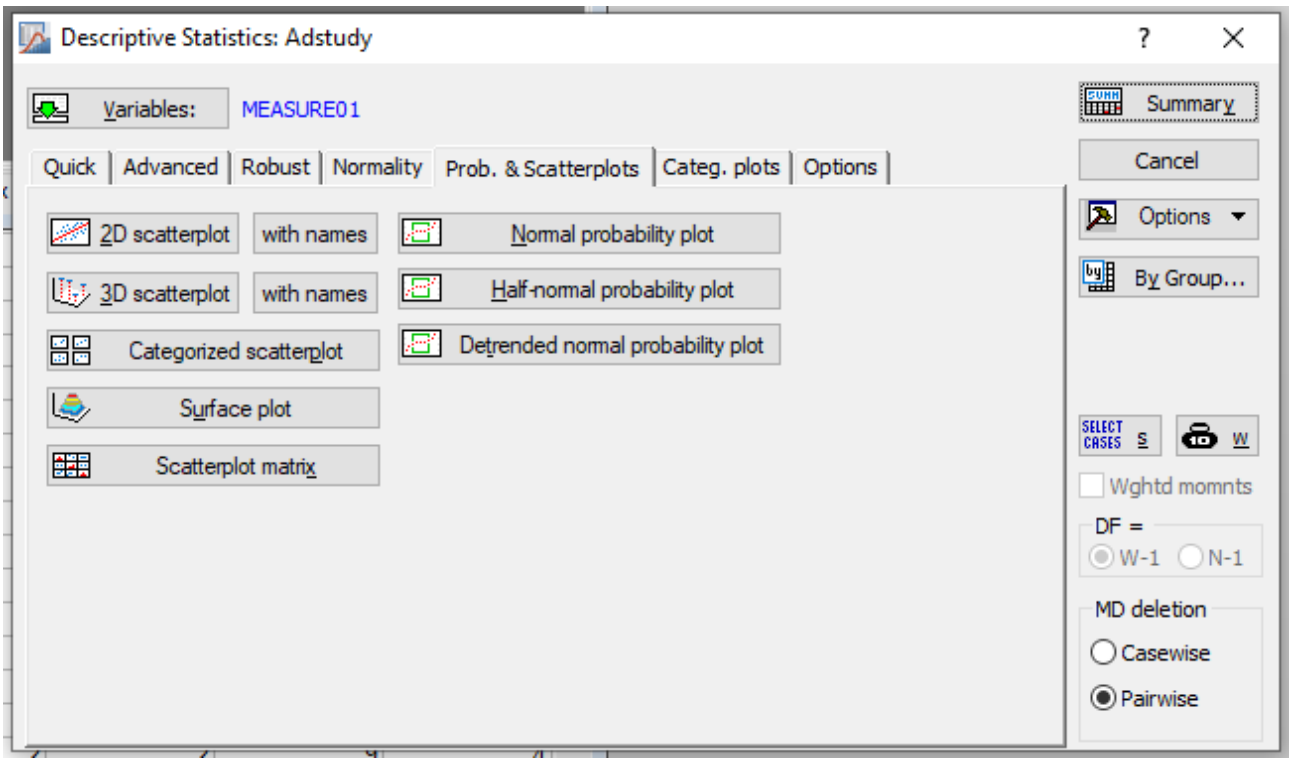


Рис. 4.42. Візуальна перевірка

Чим ближче точки розміщені до прямої, на графіку, який з'явився на екрані, тим краще закон Гаусса описує розподіл наших даних (рис. 4.43).

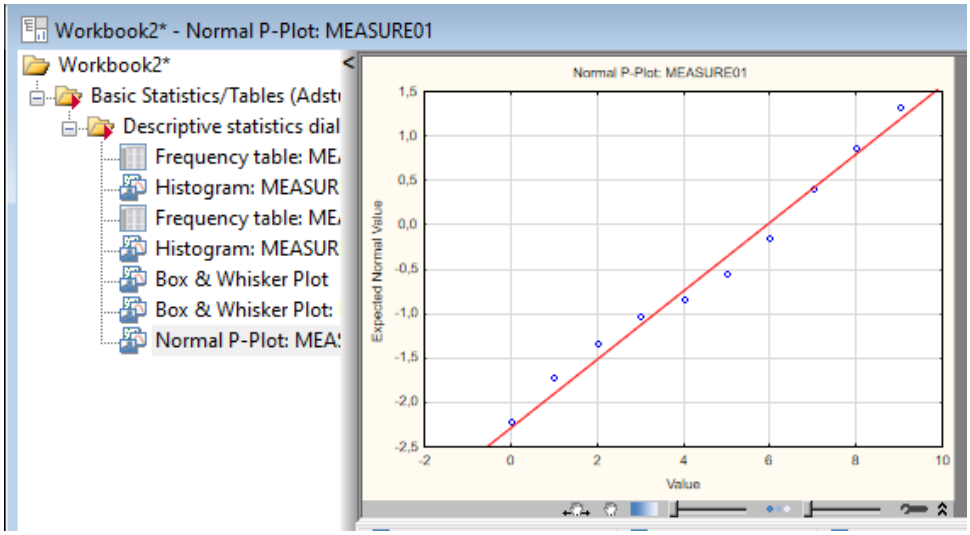


Рис. 4.43. Підтвердження нормального закону розподілу

4.2.2.2. Перевірка гіпотези на рівномірний розподіл

Розглянемо згенеровані випадкові значення, які за нульовою гіпотезою є рівномірно розподіленими на відрізку [0,1]. Створимо новий файл даних із 10000 випадками (Cases). Тоді послідовність дій: Edit → Fill/Standardize Blocks → Fill Random Values. Натиснемо Statistics → Distribution Fitting, виберемо рівномірний розподіл (рис. 4.44) і , відповідно, кнопка OK.

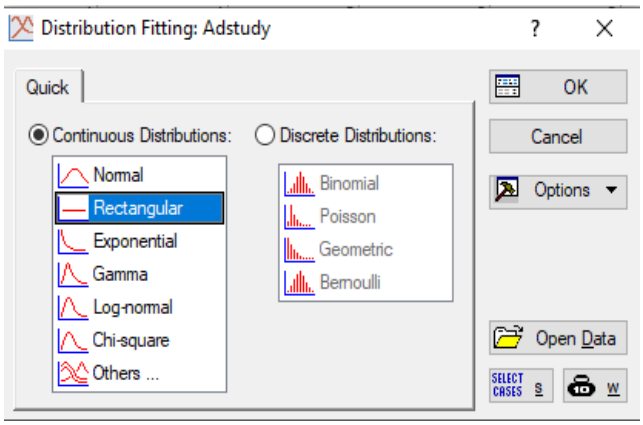


Рис. 4.44. Вибір кнопки Rectangular

У вікні, що з'явиться (рис. 4.45) як Variable обираємо створену змінну.

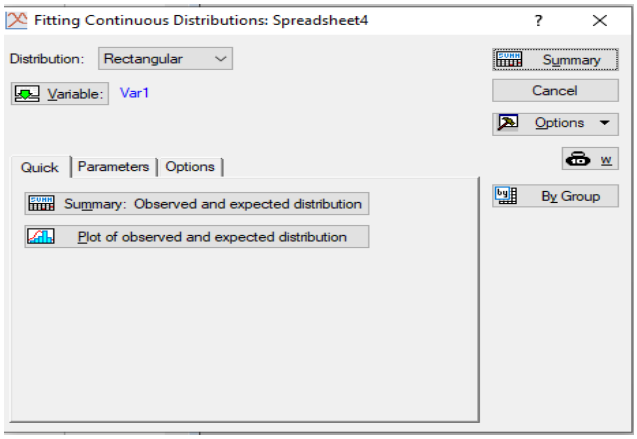


Рис. 4.45. Вибір змінної

Задання числових даних 0 та 1: у закладці Parameters як Lower limit та Min. range Parameter задають 0, а як Upper limit та Max. range Parameter, відповідно, значення 1 (рис. 4.46).

У закладці Options потрібно відмітити Yes (continuous) у розділі Kolmogorov-Smirnov test (рис. 4.47), далі натискання кнопки Summary приводить до результату.

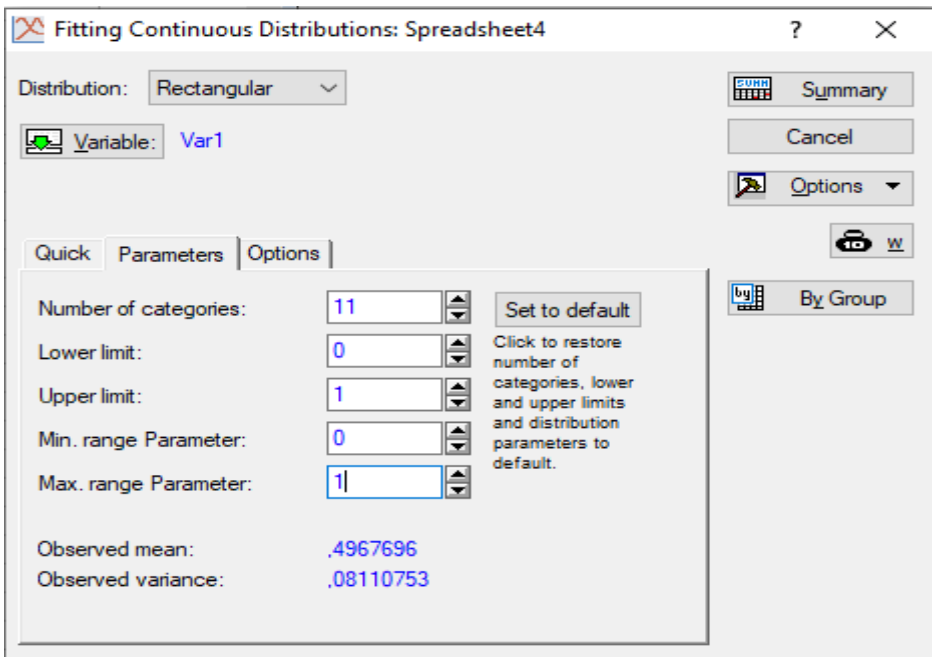


Рис. 4.46. Внесення даних

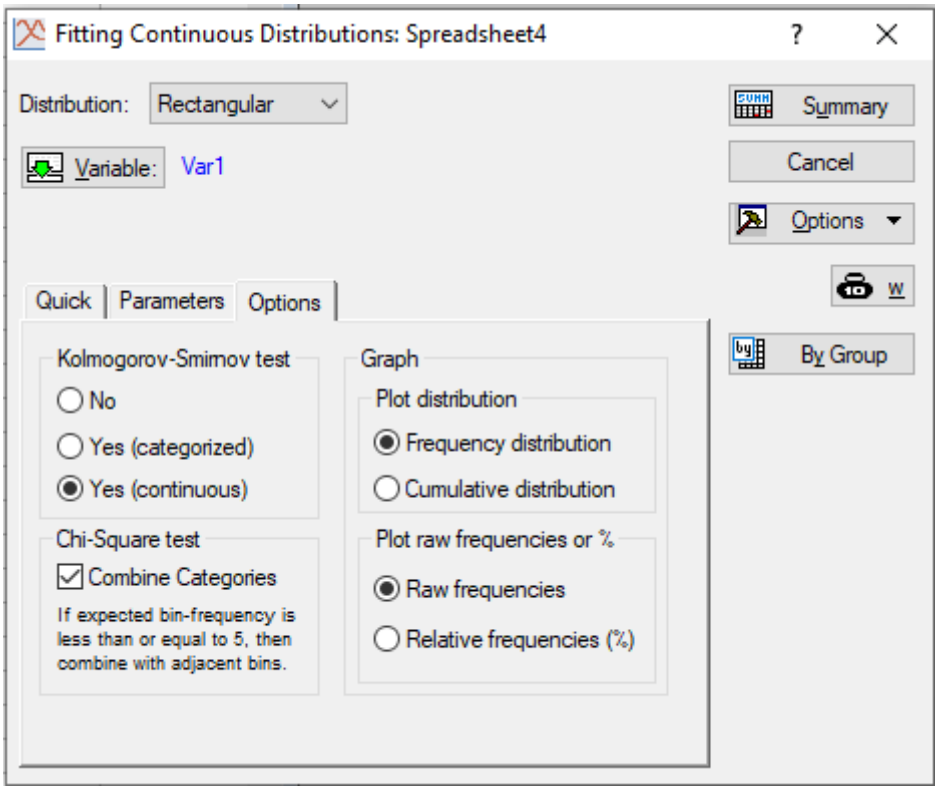


Рис. 4.47. Вибір тесту

У табличці, що з'явилась (рис. 4.48), вказана статистика для тестів Колмогорова-Смірнова та χ^2.

Variable: Var1, Distribution: Rectangular (Spreadsheet4)
Kolmogorov-Smirnov d = 0,02645, p = n.s.
Chi-Square = 13,23200, df = 8, p = 0,10411

Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumu Expe
<= 0,09091	98	98	9,80000	9,8000	90,90909	90,909	9,090909	9
0,18182	88	186	8,80000	18,6000	90,90909	181,818	9,090909	18
0,27273	87	273	8,70000	27,3000	90,90909	272,727	9,090909	27
0,36364	89	362	8,90000	36,2000	90,90909	363,636	9,090909	36
0,45455	76	438	7,60000	43,8000	90,90909	454,545	9,090909	45
0,54545	109	547	10,90000	54,7000	90,90909	545,455	9,090909	54
0,63636	92	639	9,20000	63,9000	90,90909	636,364	9,090909	63
0,72727	94	733	9,40000	73,3000	90,90909	727,273	9,090909	72
0,81818	96	829	9,60000	82,9000	90,90909	818,182	9,090909	81
0,90909	101	930	10,10000	93,0000	90,90909	909,091	9,090909	90
< Infinity	70	1000	7,00000	100,0000	90,90909	1000,000	9,090909	100

Рис. 4.48. Статистика опрацьованих даних

В даному прикладі гіпотеза про рівномірний розподіл приймається, оскільки Kolmogorov- Smirnov: p =n.s.; Chi-Square: p = 0.104.

Візуально підтвердженням висновку є побудова гістограми. Натиснувши кнопку Plot of observed and expected distribution на закладці Quick, можна графічно порівняти гістограму з щільністю обраного розподілу (рис. 4.49).

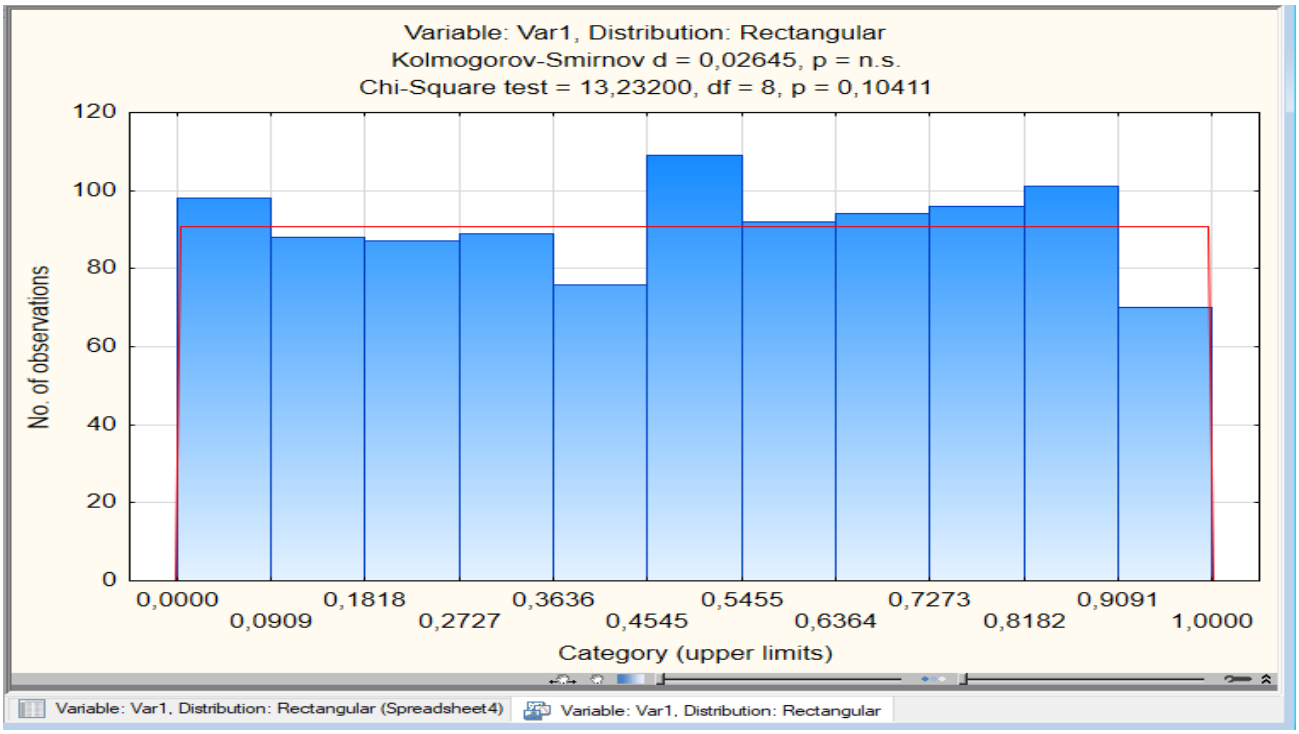


Рис. 4.49. Гістограма для даних прикладу

4.2.2.3. Перевірка гіпотез, використання t-test

Перевірка залежності двох вибірок

Для того, щоб перевірити гіпотезу про залежність двох показників виконують таку послідовність дій: Statistics → Nonparametrics →2×2 Table → ОК (рис. 4.50).

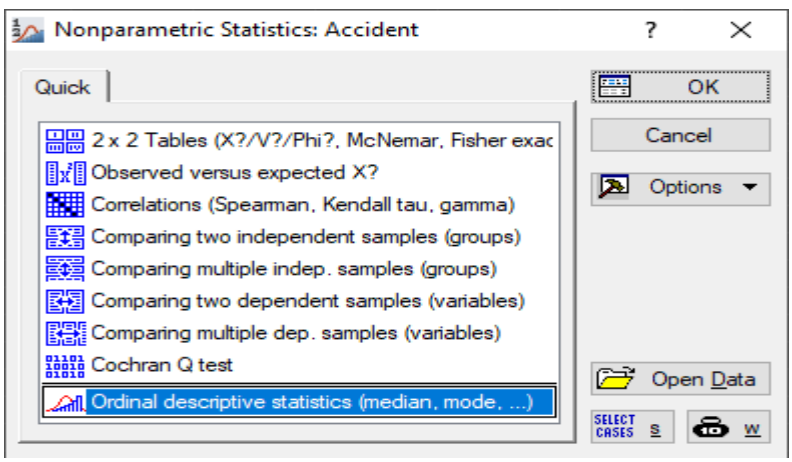


Рис. 4.50. Перевірка залежності між змінними

Таблицю використовують в разі потреби перевірки зв'язку між двома змінними. Наприклад, для перевірки, чи пов'язані між собою такі два показники: колір очей та колір волосся. Можна провести опитування в групі про колір очей та волосся і занести в таблицю ці дані (рис. 4.51).

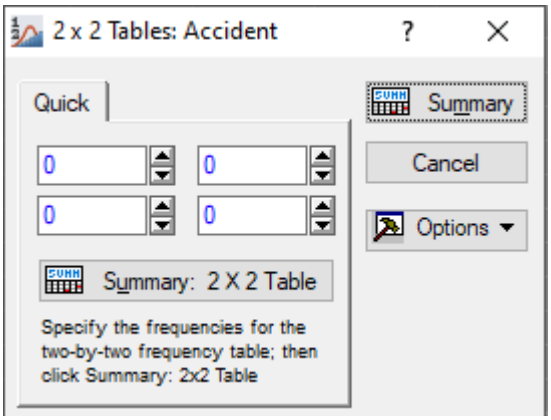


Рис. 4.51. Вигляд таблиці на екрані для занесення даних

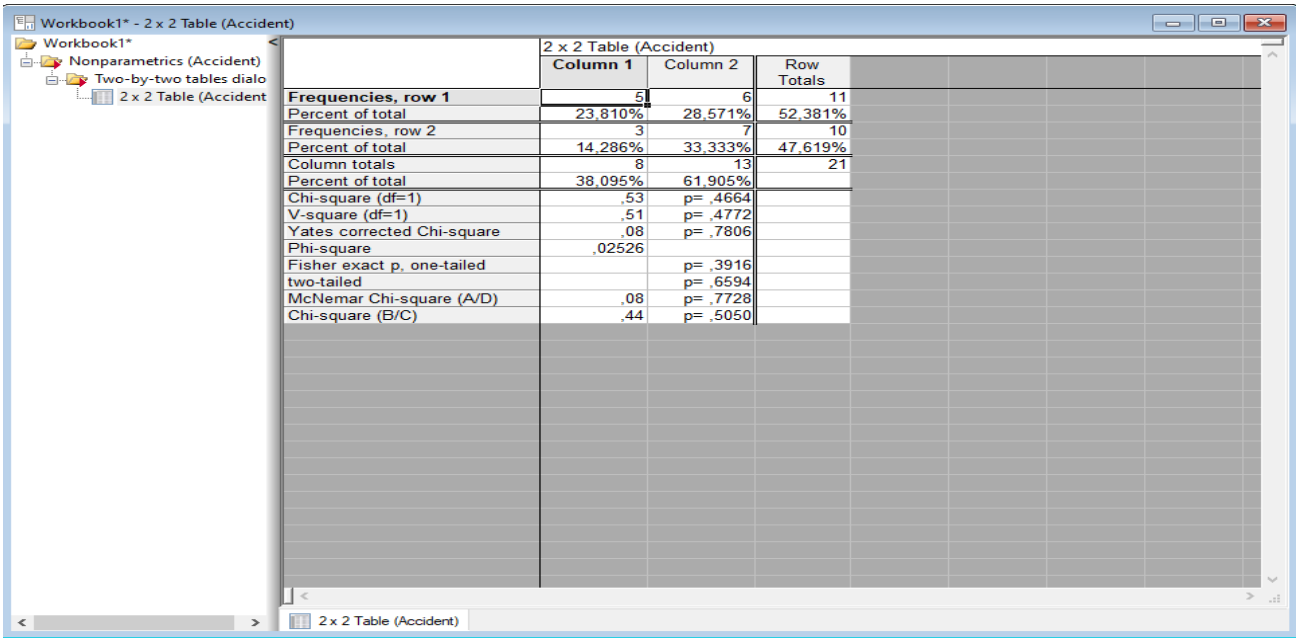
Нехай $p_1 = p_{11} + p_{12}$, $p_2 = p_{21} + p_{22}$, $q_1 = p_{11} + p_{21}$, $q_2 = p_{12} + p_{22}$.

	Очі		
Волосся	Темні	Світлі	Сума
Темне	p_{11}	p_{12}	p_1
Світле	p_{21}	p_{22}	p_2
Сума	q_1	q_2	

Зауваження. У разі незалежності кольору очей від кольору волосся дані для ймовірностей записують за формулою $p_{ij} = p_i q_j$, $i, j = 1, 2$.

За допомогою критерію (Chi-square) можна визначити ймовірність помилки, якщо, на основі конкретних даних, відхилено істинну гіпотезу про те, що колір очей та колір волосся−незалежні між собою.

Натискання кнопки Summary (рис. 4.51) показує ймовірність помилки, яка є на перетині Chi-square та Column 2 (рис. 4.52).



	Column 1	Column 2	Row Totals
Frequencies, row 1	5	6	11
Percent of total	23,810%	28,571%	52,381%
Frequencies, row 2	3	7	10
Percent of total	14,286%	33,333%	47,619%
Column totals	8	13	21
Percent of total	38,095%	61,905%	
Chi-square (df=1)	,53	p= ,4664	
V-square (df=1)	,51	p= ,4772	
Yates corrected Chi-square	,08	p= ,7806	
Phi-square	,02526		
Fisher exact p, one-tailed		p= ,3916	
two-tailed		p= ,6594	
McNemar Chi-square (A/D)	,08	p= ,7728	
Chi-square (B/C)	,44	p= ,5050	

Рис. 4.52. Імовірність помилки $p = 0,47$

Порівняння середніх двох вибірок (t-test)

t-test є одним з найпоширеніших методів порівняння середніх двох вибірок, дані яких можуть бути як залежні, так і незалежні. Теоретично t-test можна використовувати навіть для малих вибірок (приблизно 10 спостережень), якщо дані приблизно нормально розподілені (можна визначити візуально за гістограмою).

Відкриємо файл Adstudy.sta. Для того, щоб порівняти середні для двох вибірок натиснемо: Statistics → Basic Statistics/Tables → t-test, independent, by variables → OK (рис. 4.53).

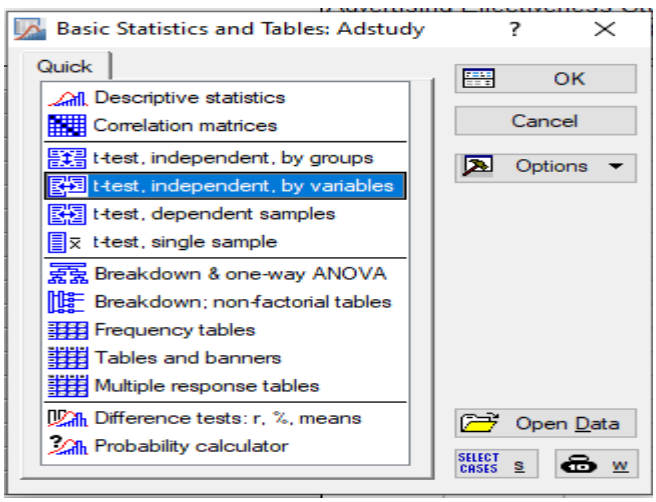


Рис. 4.53. Відкриття відповідного модуля

Як Variables (groups) вибирають MEASURE01−First list, та MEASURE02−second list, відповідно, далі Summary (рис. 4.54).

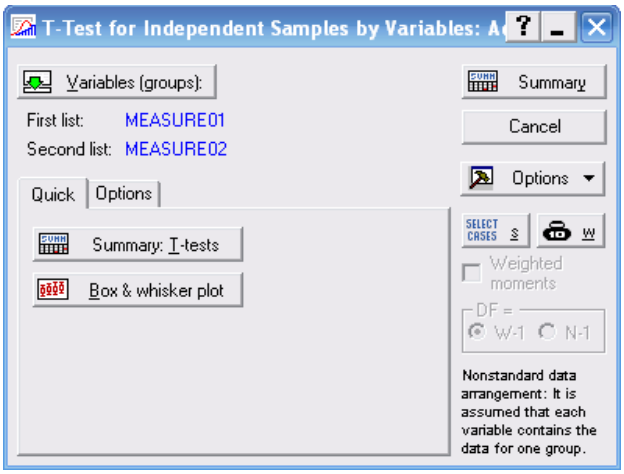


Рис. 4.54. Дані для порівняння

Після натискання кнопки Summary (рис. 4.55) обчислено середні для обох змінних. Також є ряд інших показників: t-статистика та F-статистика, показано кількість спостережень змінних, кількість степенів вільності df. У полі *p* вказано ймовірність помилки при відхиленні гіпотези про те, що середні обох вибірок збігаються.

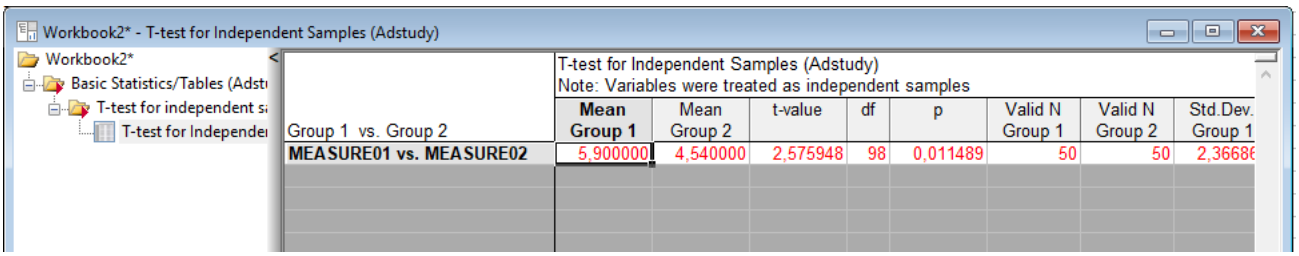


Рис. 4.55. Обчислені статистичні дані

Розглянемо роботу іншого модуля (рис. 4.53), тобто t-test, independent, by groups. Тоді як Grouping потрібно вибрати Gender, а як Dependent variables − MEASURE01 та MEASURE02 (рис. 4.56).

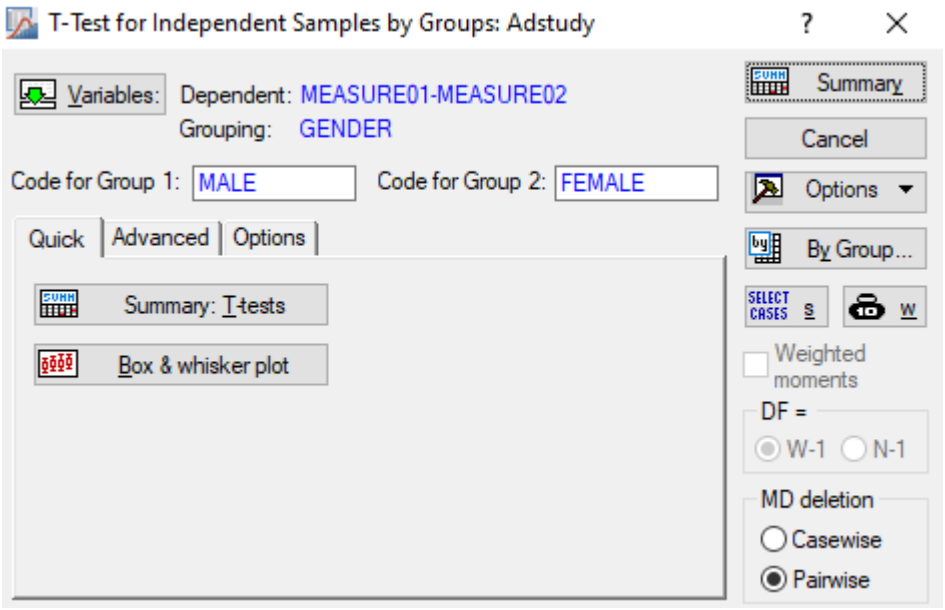
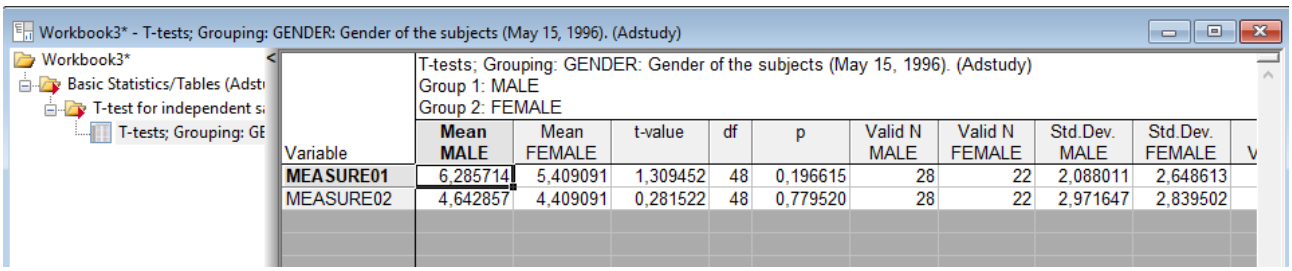


Рис. 4.56. Робота з модулем t-test, independent, by groups

Отримаємо порівняння середніх двох вибірок, які є згрупованими (рис. 4.57), тобто для кожної змінної буде пораховано значення середніх (окремо для чоловіків і для жінок), і будуть порівнюватися саме середні для чоловіків і жінок.



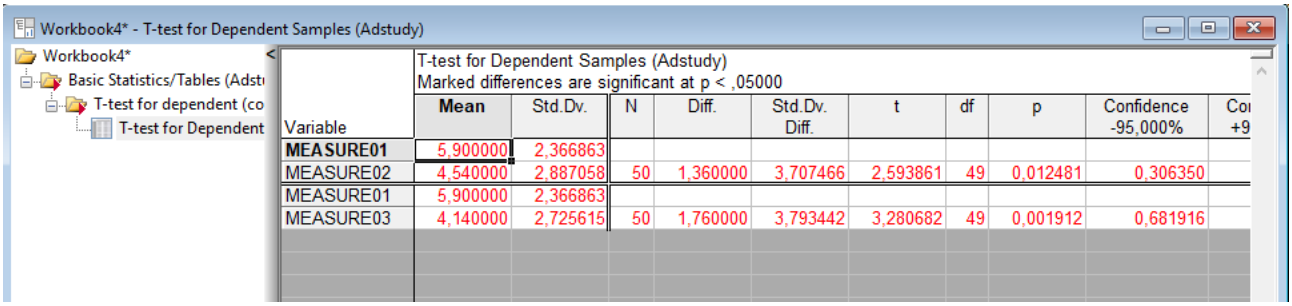
Variable	Mean MALE	Mean FEMALE	t-value	df	p	Valid N MALE	Valid N FEMALE	Std.Dev. MALE	Std.Dev. FEMALE
MEASURE01	6,285714	5,409091	1,309452	48	0,196615	28	22	2,088011	2,648613
MEASURE02	4,642857	4,409091	0,281522	48	0,779520	28	22	2,971647	2,839502

Рис. 4.57. Порівняння середніх

Отже, проведено аналіз гіпотези про рівність середніх в припущенні незалежності змінних. Проте, якщо врахувати, що відповідні значення для вибірок MEASURE01 та MEASURE02 є відповідями одних і тих же людей, то природніше вважати, що відповідні змінні все-таки залежні. Тому, доцільно проаналізувати гіпотезу, використовуючи тест для залежних вибірок.

Послідовність дій: Statistics → Basic Statistics/Tables → t-test, dependent samples → OK. Далі для Variables, як first variable, вибирають змінну MEASURE01, а як second variable– MEASURE02, MEASURE03 та кнопка Summary приводять до результату.

В отриманій таблиці (рис. 4.58) відображено результати порівняння середніх вибірок MEASUREO1 та MEASURE02 та вибірок MEASURE01 та MEASURE03, де в полі *p* вказано ймовірність помилки при відхиленні гіпотези про те, що середні обох вибірок збігаються.



T-test for Dependent Samples (Adstudy)
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Co +9
MEASURE01	5,900000	2,366863								
MEASURE02	4,540000	2,887058	50	1,360000	3,707466	2,593861	49	0,012481	0,306350	
MEASURE01	5,900000	2,366863								
MEASURE03	4,140000	2,725615	50	1,760000	3,793442	3,280682	49	0,001912	0,681916	

Рис. 4.58. Результати порівняння для трьох вибірок

Слід ще раз наголосити на тому, що t-test застосовують на практиці у тих випадках, коли:

1) достатньо невелика кількість спостережень;

2) можна вважати, що дані розподілені нормально (це можна візуально визначити за гістограмою або графіком на нормальному ймовірнісному папері − Normal probability plot).

Зазначимо, що з панелі t-тесту можна побудувати різноманітні графіки для візуального порівняння середніх та варіації в двох групах. Як приклад, (рис. 4.59) наведено графік з t-test, independent, by variables (у вікні аналізу натиснути Box & whisker plots). Такі графіки допомагають зробити попередні висновки про рівність середніх.

Тести порівняння для інших характеристик вибірок

Оцінити гіпотезу про рівність варіацій чи дисперсій змінних можна на базі F-test, який розміщений в таблиці результатів t-test.

Слід зазначити: якщо дві оцінки дисперсії близькі одна до одної, то значення F-статистики наближається до одиниці, адже F-статистику можна розглядати як відношення двох незалежних оцінок дисперсії. Також для оцінки гіпотези про рівність дисперсій можна використовувати тести Levene test та Brown-Forsythe.

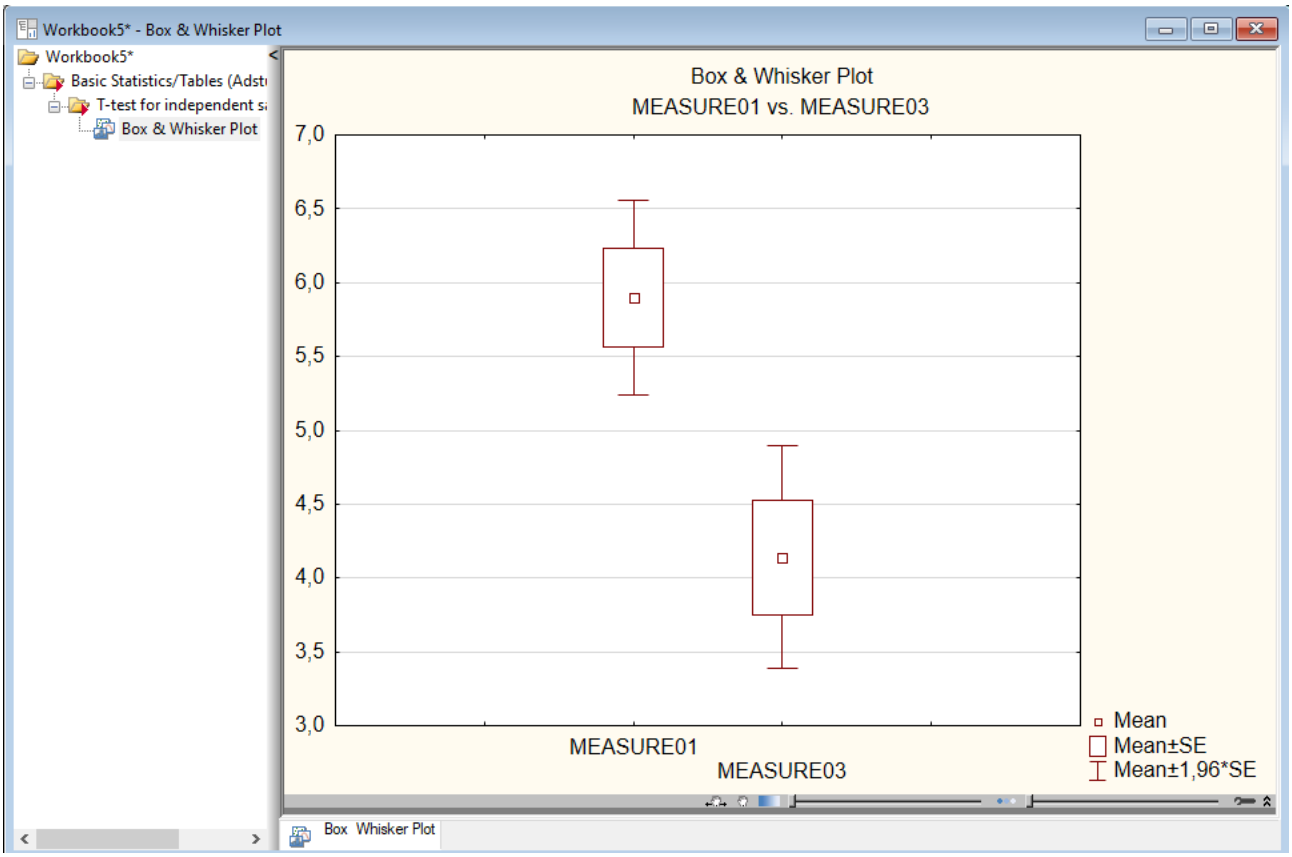


Рис. 4.59. Візуальне порівняння

Для цього вибирають: Statistics → Basic Statistics/Tables → Difference tests: r, %, means → OK.

У вікні (рис. 4.60) видно панелі трьох тестів, які дозволяють перевірити гіпотези про рівність таких характеристик вибірок: коефіцієнтів кореляції, математичних сподівань та ймовірностей.

Наприклад, можемо перевіряти гіпотезу про рівність математичних сподівань тоді, коли стандартні відхилення відомі.

У полях Difference between two correlation coefficients як значення r1 та r2 вводять обчислені попередньо значення кореляцій для двох вибірок, а як значення N1 та N2 вказують кількість елементів у двох вибірках, відповідно вибирають two-sided. У полі *p* після натискання Compute отримують ймовірність прийняття гіпотези про те, що кореляції обох вибірок рівні.

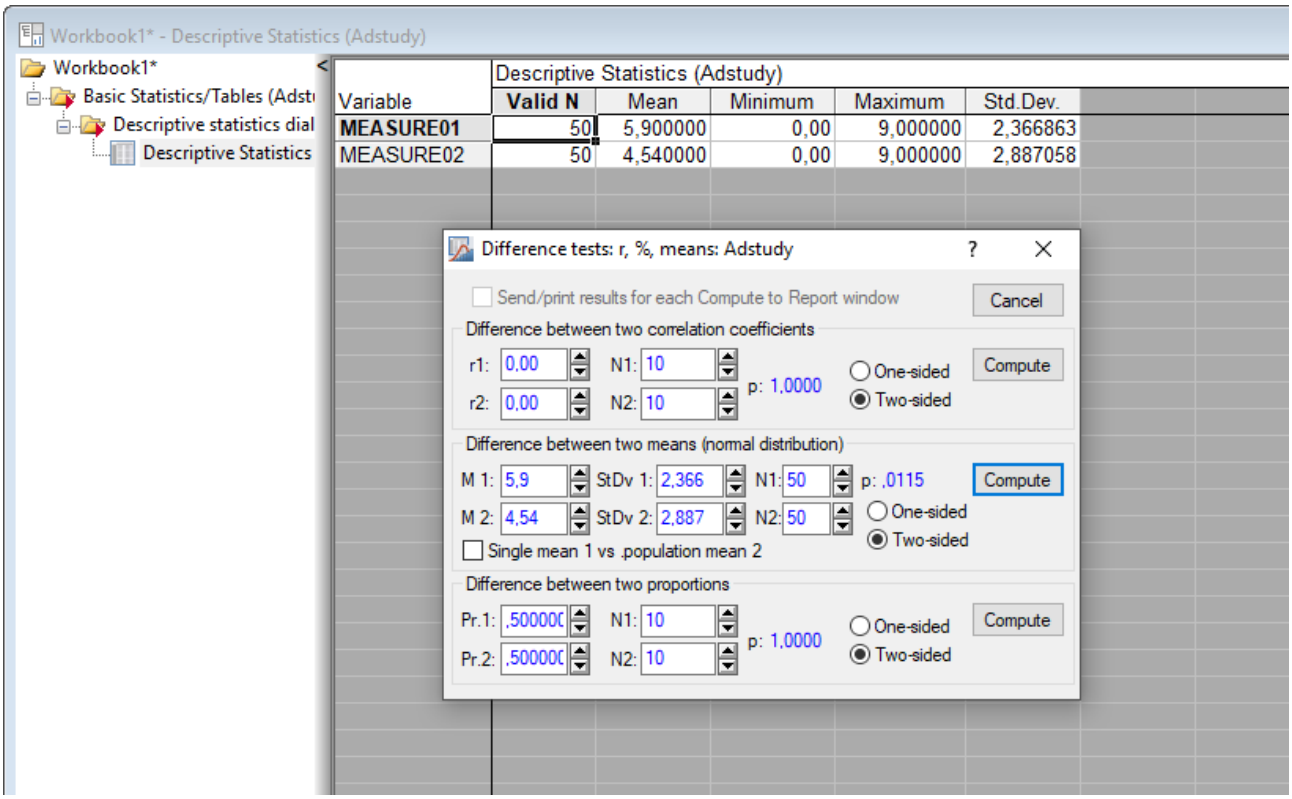


Рис. 4.60. Вигляд перевірки трьох тестів на екрані

Аналогічно, у полях Difference between two means вказують M 1 та M 2 – середні двох вибірок, а у полях StDv 1 та StDv 2 – стандартні квадратичні відхилення обох вибірок, відповідно. У поля N1 та N2 вводять кількість елементів в обох вибірках. В наведеному прикладі для вказаних значень, ймовірність помилки при відхиленні гіпотези про те, що середні обох вибірок збігаються, становить 0,0115 (рис. 4.60).

Якщо для наведених вище тестів потрібно попередньо підрахувати значення різних статистичних характеристик вибірок, то швидко зробити це можна так:

1) виділяємо ті значення змінної, для яких ми хочемо обчислити характеристику;

2) виконуємо Statistics → Statistics of Block Data → Columns → потрібна характеристика.

Завдання для самоконтролю

1. Пояснити початок роботи з пакетом STATISTICA.

2. Дати короткий опис модулів.

3. Пояснити введення змінних різними способами

4. Навести приклади створення файлів.

5. Пояснити використання на прикладі можливостей панелі Probability Distibutioп Calculator.

6. За обраними даними вибірки показати роботу в пакеті: побудувати інтервальний статистичний ряд відносних частот, показати вигляд гістограми, обрати гіпотезу про тип розподілу та перевірку її за критерієм згоди χ^2.

Вказівка. Дані взяти із варіантів завдань для індивідуальної роботи (завдання 2) після 3.3.

Список використаної літератури

1. Барковський В. В. Теорія ймовірностей та математична статистика / В. В. Барковський, Н. В. Барковська, О. К. Лопатін − Київ, ЦУЛ , 2002. − 448 с.

2. Гуткевич С. О. Політика ефективного розвитку підприємств: управлінський аспект : монографія / С.О. Гуткевич, Л.П. Шендерівська. – Київ: НТУУ «КПІ», 2016. – 211 с.

3. Іванюта І. Д. Елементи теорії ймовірностей та математичної статистики / І. Д. Іванюта, В. І. Рибалка, І. А. Рудоміно-Дусятська. – Київ: Слово, 2003.–272с.

4. Каніовська І. Ю. Теорія ймовірностей у прикладах і задачах / І. Ю. Каніовська. – Київ: ІВЦ "Видавництво «Політехніка»", ТОВ "Фірма «Періодика»", 2004. – 156 с.

5. Кушлик Б. Р. Стабілізація друкування малотиражної продукції офсетним друком: монографія / Б. Р. Кушлик, О. І. Кушлик-Дивульська ; за заг. ред. О. М. Величко. – Київ: КПІ ім. Ігоря Сікорського, Вид-во «Політехніка», 2017. – 162 с.

6. Кушлик-Дивульська О. І. Теорія ймовірностей та математична статистика: навч. посіб. / О. І. Кушлик-Дивульська, Н. В. Поліщук, Б. П. Орел, П.І. Штабалюк.− К.: НТУУ «КПІ», 2014. – 212 с.

7. Кушлик-Дивульська О. І. Вища математика: Елементи теорії ймовірності: Практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 186 «Видавництво та поліграфія» / КПІ ім. Ігоря Сікорського ; уклад.: О. І. Кушлик-Дивульська, Н. П. Селезньова. – Електронні текстові дані (1 файл: 2,4 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2022. – 105 с. – Назва з екрана. – Доступ: https://ela.kpi.ua/handle/123456789/46693.

8. Методичні вказівки до виконання лабораторних робіт (комп'ютерного практикуму) з дисципліни «Теорія ймовірностей і математична статистика» для студентів напряму підготовки 6.030601 «Менеджмент» студентів Видавничо-поліграфічного інституту [Електронний ресурс]: НТУУ «КПІ»; уклад.:

О. І. Кушлик-Дивульська, Б. Р. Кушлик. – Електронні текстові дані (1 файл: 4,60 Мбайт). – Київ : НТУУ «КПІ», 2016. – 205 с. – Назва з екрана. – Доступ : http://ela.kpi.ua/handle/123456789/17693.

9. Оленко А. Я. Компютерна статистика: навч. посіб. / А. Я. Оленко.−Київ: Вид.- поліграф. Центр «Київський університет», 2007. – 174 с.

10. Селезньова Н. П. Спеціальні питання вищої математики. Елементи теорії ймовірностей. Теорія і практикум. [Електронний ресурс] навч. посіб. для студ. спеціальності 151 «Автоматизація та комп'ютерно-інтегровані технології» КПІ ім. Ігоря Сікорського ; уклад. Н. П. Селезньова, Т. О. Рудик, та ін. – Електронні текстові дані (1 файл: 0,97 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 78 с. Назва з екрана. – Доступ: https://ela.kpi.ua/handle/123456789/42309.

11. Zolotukhina K. Researching the interaction of different printed materials types with liquids // K. Zolotukhina, S. Khadzhynova, O. Velychko, B. Kushlyk, O. Kushlyk-Dyvulska // Eastern-european journal of enterprise technologies. 2019. 3/1 (99). P. 26-35.

12. Хаджинова С. Є. Виявлення залежностей експериментальних досліджень крайового кута змочування. / С. Є. Хаджинова, К. І. Золотухіна, Б. Р. Кушлик, О. І. Кушлик-Дивульська // Технологія і техніка друкарства. – Київ: ВПІ КПІ ім. Ігоря Сікорського.–2018.−№4(62).−С.27-38. Назва з екрана: http://ttdruk.vpi.kpi.ua/article/view/167228/pdf_86.

Додатки

Таблиця Д.1

Значення функції $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2035	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0941	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0271	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0159	0155	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Таблиця Д.2

Значення функції Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43872	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48860	48899
2,3	48928	48966	48983	49010	49036	49061	49086	49110	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	0,49865		3,1	49903	3,2	49931	3,3	49952	3,4	49966
3,5	49977		3,6	49984	3,7	49989	3,8	49993	3,9	49995
4,0	499968									
4,5	499997									
5,0	49999997									

Таблиця Д.3

Значень χ^2 в залежності від k і рівня значущості α

Кількість степенів вільності k	α 0,95	α 0,90	α 0,50	α 0,30	α 0,20	α 0,10	α 0,05	α 0,025	α 0,01
1	0,004	0,016	0,455	1,074	1,642	2,71	3,84	5,0	6,6
2	0,103	0,211	1,386	2,41	3,22	4,60	5,99	7,4	9,2
3	0,352	0,584	2,37	3,67	4,64	6,25	7,82	9,4	11,3
4	0,711	1,064	3,36	4,88	5,99	7,78	9,49	11,1	13,3
5	1,145	1,61	4,35	6,06	7,29	9,24	11,07	12,8	15,1
6	1,635	2,20	5,35	7,23	8,56	10,64	12,59	14,4	16,8
7	2,17	2,83	6,35	8,38	9,80	12,02	14,07	16,0	18,5
8	2,73	3,49	7,34	9,52	11,03	13,36	15,51	17,5	20,1
9	3,32	4,17	8,34	10,66	12,24	14,68	16,92	19,0	21,7
10	3,94	4,86	9,34	11,78	13,44	15,99	18,31	20,5	23,2
11	4,58	5,58	10,34	12,90	14,63	17,28	19,68	21,9	24,7
12	5,23	6,30	11,34	14,01	15,84	18,55	21,0	23,3	26,2
13	5,89	7,04	12,34	15,12	16,98	19,81	22,4	24,7	27,7
14	6,57	7,79	13,34	16,22	18,15	21,1	23,7	26,1	29,1
15	7,26	8,55	14,34	17,32	19,31	22,3	25,0	27,5	30,6
16	7,96	9,31	15,34	18,42	20,5	23,5	26,3	28,8	32,0
17	8,67	10,08	16,34	19,51	21,6	24,8	27,6	30,2	33,4
18	9,39	10,86	17,34	20,6	22,8	26,0	28,9	31,5	34,8
19	10,11	11,65	18,34	21,7	23,9	27,2	30,1	32,9	36,2
20	10,85	12,44	19,34	22,8	25,0	28,4	31,4	34,2	37,6
21	11,59	13,24	20,3	23,9	26,2	29,6	32,7	35,5	38,9
22	12,34	14,04	21,3	24,9	27,3	30,8	33,9	36,8	40,3
23	13,09	14,85	22,3	26,0	28,4	32,0	35,2	38,1	41,6
24	13,85	15,66	23,3	27,0	29,6	33,2	36,4	39,4	43,0
25	14,61	16,47	24,3	28,2	30,7	34,4	37,7	40,6	44,3
26	15,38	17,29	25,3	29,2	31,8	35,6	38,9	41,9	45,6
27	16,15	18,11	26,3	30,3	32,9	36,7	40,1	43,2	47,0
28	16,93	18,94	27,3	31,4	34,0	37,9	41,3	44,5	48,3
29	17,71	19,77	28,3	32,5	35,1	39,1	42,6	45,7	49,6
30	18,49	20,60	29,3	33,5	36,2	40,3	43,8	47,0	50,9

Таблиця Д.4

Значення $t_\gamma = t(k, \gamma)$ (розподіл Стьюдента)

Кількість степенів вільності	$\gamma = 1 - \alpha$					
k	0,9	0,95	0,98	0,99	0,998	0,999
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,21	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,37	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,06	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,74	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,75
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,47	2,76	3,41	3,67
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	20,2	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29

Таблиця Д.5

Критичні точки розподілу *F* Фішера-Снедекора

Рівень значущості $\alpha = 0{,}01$

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	15,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Рівень значущості $\alpha = 0{,}05$

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,37	19,40	19,41
3	10,12	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,20	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Таблиця Д.6

Значення $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$

k	λ								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4		0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5				0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6							0,0001	0,0002	0,0003
k	λ								
	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050
3	0,0613	0,1804	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0573	0,0337
5	0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911
7	0,0001	0,0034	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1171
8		0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318
9		0,0002	0,0027	0,0132	0,0363	0,0688	0,1014	0,1241	0,1318
10			0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186
11			0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970
12			0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728
13				0,0002	0,0013	0,0052	0,0142	0,0296	0,0504
14				0,0001	0,0005	0,0022	0,0071	0,0169	0,0324
15					0,0002	0,0009	0,0033	0,0090	0,0194
16						0,0003	0,0014	0,0045	0,0109
17						0,0001	0,0006	0,0021	0,0058
18							0,0002	0,0009	0,0029
19							0,0001	0,0004	0,0014
20								0,0002	0,0006
21								0,0001	0,0003
22									0,0001

Таблиця Д.7

Значення $q = q(\gamma, n)$

$n \backslash \gamma$	0,95	0,99	0,999
5	1,37	2,67	5,64
6	1,09	2,01	3,88
7	0,92	1,62	2,98
8	0,80	1,38	2,42
9	0,71	1,20	2,06
10	0,65	1,08	1,80
11	0,59	0,98	1,60
12	0,55	0,90	1,45
13	0,52	0,83	1,33
14	0,48	0,78	1,23
15	0,46	0,73	1,15
16	0,44	0,70	1,07
17	0,42	0,66	1,01
18	0,40	0,63	0,96
19	0,39	0,60	0,92
20	0,37	0,58	0,88
25	0,32	0,49	0,73
30	0,28	0,43	0,63
35	0,26	0,38	0,56
40	0,24	0,35	0,50
45	0,22	0,32	0,46
50	0,21	0,30	0,43
60	0,188	0,269	0,38
70	0,174	0,245	0,34
80	0,161	0,226	0,31
90	0,151	0,211	0,29
100	0,143	0,198	0,27
150	0,115	0,160	0,211
200	0,099	0,136	0,185
250	0,089	0,120	0,162