

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Закарпатський угорський університет
імені Ференца Ракоці II
кафедра математики та інформатики

ВІКТОРІЯ ДЗЯМКО
Методичні вказівки до практичних занять

Теорія ймовірностей, математична статистика та економетрика
для студентів спеціальностей
D1 Облік і оподаткування
D2 Фінанси, банківська справа, страхування
та фондовий ринок (Фінансовий захист)



Берегове — 2025

Методичні вказівки до практичних занять з теорії ймовірностей, математичної статистики та економетрики для студентів спеціальностей D1 Облік і оподаткування, D2 Фінанси, банківська справа, страхування та фондовий ринок (Фінансовий захист). Розробник: Дзямко В.Й. Берегове: Закарпатський угорський університет ім. Ференца Ракоці II, 2025.- 82 с.

Затверджено до використання у навчальному процесі
на засіданні кафедри математики та інформатики ЗУУ ім. Ференца Ракоці II
(протокол № 3 від 15 жовтня 2025 року.)

Розглянуто та рекомендовано Радою із забезпечення якості вищої освіти
Закарпатського угорського університету ім. Ференца Ракоці II
(протокол № 10 від «24» листопада 2025)

Рекомендовано до видання в електронній формі (PDF)
рішенням Вченої ради Закарпатського угорського університету ім. Ференца Ракоці II
(протокол № 11 від «26» листопада 2025)

Розробник методичної вказівки:

Вікторія ДЗЯМКО – доцент кафедри математики та інформатики
ЗУУ імені Ференца Ракоці II, кандидат педагогічних наук, доцент

Рецензенти:

Тетяна БОЯРИЩЕВА – доцент кафедри теорії ймовірностей і математичного аналізу ДВНЗ
УжНУ, кандидат фізико-математичних наук, доцент (м. Ужгород)

Лівія МЕСАРОШ – доцент кафедри математики та інформатики
ЗУУ імені Ференца Ракоці II, кандидат фізик-математичних наук, доцент

За зміст методичної вказівки відповідальність несуть розробники

Відповідальний за випуск:

Мирослав Стойка – завідувач кафедри математики та інформатики ЗУУ ім. Ф.Ракоці II,
доцент, кандидат фізико-математичних наук

Видавництво: Закарпатський угорський університет імені Ференца Ракоці II (адреса: пл.
Кошута 6, м. Берегове, 90202. Електронна пошта: foiskola@kmf.uz.ua)

Курс лекцій з навчальної дисципліни «Теорія ймовірностей і математична статистика» для здобувачів вищої освіти першого (бакалаврського) рівня денної і заочної форми навчання для спеціальностей D1 Облік і оподаткування, D2 Фінанси, банківська справа, страхування та фондовий ринок (Фінансовий захист). У даному курсі лекцій коротко подано теоретичний матеріал, приклади розв'язування практичних завдань та пропонуються вправи для самостійної роботи.

ЗМІСТ

Вступ.....	4
Тема 1. Випадкові події. Типи подій.....	5
Тема 2. Класична ймовірність, статистична ймовірність, геометрична ймовірність.....	6
Тема 3. Теорема про ймовірність суми та добутку подій. Умовна ймовірність. Незалежні події.....	12
Тема 4. Схема повторних випробувань. Формули Бернуллі, Пуассона, локальна та інтегральна теорема Лапласа.....	15
Тема 5. Дискретні і неперервні випадкові величини. Основні характеристики. Властивості.....	18
Тема 6. Дискретні і неперервні випадкові величини. Основні характеристики. Властивості. Розподіли дискретних випадкових величин: біноміальний, Пуассона, геометричний.....	23
Тема 7. Функції випадкових величин.....	28
Тема 9. Граничні теореми теорії ймовірностей.....	31
Тема 10. Елементи математичної статистики. Вибірковий метод.....	35
Тема 11. Статистичні оцінки параметрів генеральної сукупності. Статистичні гіпотези.....	38
Тема 12. Елементи дисперсійного аналізу.....	40
Тема 13. Елементи теорії регресії і кореляції.....	45
Тема 14. Предмет економетрики. Методологія економетричного дослідження.....	58
Тема 15. Проста і множинна лінійна регресія. Метод найменших квадратів.....	59
Тема 16. Множинна лінійна регресія. Перевірка гіпотез.....	62
Тема 17. Мультиколінеарність та її вплив на оцінки параметрів моделі.....	65
Таблиці.....	74
Список використаних джерел.....	82

Вступ

Теорія ймовірностей та математична статистика є фундаментальними розділами математики, що мають широке застосування в науці, техніці, економіці, медицині та інших сферах людської діяльності. Вони дозволяють аналізувати випадкові явища, моделювати їх поведінку та робити висновки на основі даних. Теорія ймовірностей досліджує закономірності, що лежать в основі випадкових процесів, та надає методи для оцінки ймовірності різних подій. Математична статистика, в свою чергу, займається обробкою та аналізом даних, отриманих в результаті спостережень або експериментів, для виявлення значущих закономірностей, перевірки гіпотез та прогнозування.

Основною метою економетрики є побудова моделей, які допомагають:

- Оцінити економічні взаємозв'язки між змінними, такими як попит і пропозиція, інвестиції і економічне зростання, рівень безробіття та заробітна плата.
- Прогнозувати майбутні економічні тенденції на основі наявних даних, що має важливе значення для планування економічної політики та ведення бізнесу.
- Тестувати економічні гіпотези щодо того, як змінюються економічні змінні у відповідь на різні фактори (наприклад, як зміна податкової ставки може вплинути на споживання).

Метою цих методичних вказівок є допомогти студентам краще зрозуміти основні концепції теорії ймовірностей і математичної статистики, засвоїти ключові методи й інструменти, необхідні для роботи з випадковими явищами та даними, а також розвинути навички практичного застосування цих знань для вирішення різноманітних економічних задач.

Пропоновані рекомендації охоплюють теоретичні аспекти курсу, а також надають приклади розв'язання типових задач, що дозволяє закріпити матеріал на практиці.

Тема 1. Випадкові події. Типи подій.

Основними в теорії ймовірності є поняття **випадкового експерименту** – певного випробування, спостереження чи досліду, результати якого передбачити неможливо. Можливі наслідки випадкового експерименту називаються **випадковими подіями**. Випадкові події позначаються великими латинськими літерами $A, B, C, \dots$

Із кожною подією A можна пов'язати подію, яка полягає в тому, що A не настає. Цю подію називають **протилежною** до A і позначають $\bar{A}$.

Сумою подій B і C називається подія A (позначають $A = B + C$ або $A = B \cup C$), яка полягає в настанні принаймні однієї з цих подій.

Різницею подій B і C називається подія A (позначають $A = B - C$ або $A = B \setminus C$), яка настає тоді, коли настає подія B і не настає подія C .

Добутком подій B і C називається подія A (позначають $A = B \cdot C$ або $A = B \cap C$), яка полягає в тому, що в результаті експерименту відбувається як подія B , так і подія C .

Несумісними називаються події B і C , якщо відповідні їм множини елементарних подій не містять однакових елементів: $B \cap C = \emptyset$. Це означає, що коли одна з подій відбулась, друга подія відбутись не може.

Достовірна подія – подія, яка в результаті експерименту неодмінно відбудеться (позначається літерою U).

Неможлива подія – подія, яка в даному експерименті не може відбутись (позначається літерою V).

Подія **Аспричиняє** подію B (позначають $A \subset B$), якщо з того що, подія A настала випливає, що настає подія B .

Подія ω буде **елементарною**, якщо її не можна розкласти на простіші події. Множина $\Omega = \{\omega\}$ всіх елементарних подій називається **простором елементарних подій**.

Сукупність подій $A_1, A_2, \dots, A_n$ **утворюють повну групу**, якщо одна і тільки одна із цих подій в результаті експерименту обов'язково настає: $A_1 \cup A_2 \cup \dots \cup A_n = U, A_i \cap A_j = V, i \neq j$.

Розв'язання типових задач

Задача 1. Партія складається з деталей 1-го, 2-го і 3-го сорту, а також бракованих. Деталі ретельно перемішані. Із партії навмання беруть одну деталь. Застосувавши очевидні позначення, розглянемо події: $A_1 = \{1, 2\}, A_2 = \{1, 3\}, B_1 = \{2, 3\}, B_2 = \{2, \text{Бр}\}$. Яка з подій B_i утворює з подіями A_1 і A_2 повну групу?

Розв'язання. Згідно з означенням для повної групи подій має виконуватись співвідношення $A_1 \cup A_2 \cup B_i = U$. Достовірній події відповідає весь простір елементарних подій Ω .

У даному випробуванні простір елементарних подій Ω складається з чотирьох елементарних подій: деталь може бути 1-го, 2-го, 3-го сорту або бракованою, тобто $\Omega = \{1, 2, 3, \text{Бр}\}$. Знайдемо множину елементарних подій для суми подій A_1 і A_2 . Утворюючи об'єднання множин елементарних подій, до нього включають елементарні події відповідних подій, причому однакові елементарні події будуть один раз. Отже, $A_1 \cup A_2 = \{1, 2, 3\}$. Для того щоб доповнити цю множину до Ω потрібно додати до суми подій подію B_2 , бо множина її елементарних подій містить $\omega_i = \{\text{Бр}\}$.

Задача 2. Гральний кубик підкидають двічі. Випадкові події цього випробування: A – сума очок рівна 8, B – за другим разом випало 6 очок. Описати наслідки випробування та події $A, B, A + B, A \cdot B, A - B$.

Розв'язання. Наслідками випробування є наступні пари

$$\{(i, j), i, j = 1, 2, \dots, 6\}.$$

Тоді подія $A = \{(2,6), (5,3), (3,5), (4,4), (6,2)\}$, подія $B = \{(i, 6), i = 1, 2, \dots, 6\}$, $A + B = \{(5,3), (3,5), (4,4), (6,2), (i, 6), \text{де } i = 1, 2, \dots, 6\}$, $A \cdot B = \{(2,6)\}$, $A - B = \{(5,3), (3,5), (4,4), (6,2)\}$.

Задача 3. У квадрат на площині зі стороною $2a$ і з центром у початку координат, сторони якого паралельні до осей координат, навімання “кидають” точку. Описати простір елементарних подій.

Розв’язання. У цьому випадку можливими наслідками експерименту є попадання в будь-яку точку описаного квадрата. Якщо (x, y) – координатні точки на площині, то простір елементарних подій $\Omega : \{(x, y) : -a \leq x \leq a, -a \leq y \leq a\}$ є множиною нескінченною і незліченною (елементів цієї множини не можна ні полічити, ні занумерувати).

Задачі для самостійного розв’язання

1. Партія складається зі стандартних і нестандартних деталей, які ретельно перемішані. З неї навімання беруть дві деталі. Визначити множини: A – простір елементарних подій; B – множини елементарних подій для таких подій:

- а) A_1 – поява однієї стандартної і однієї нестандартної деталей;
- б) A_2 – поява не менш як однієї стандартної деталі;
- в) A_3 – поява не більш як однієї стандартної деталі.

2. Із ящика, в якому є бронзові, мідні, латунні та сталеві деталі, беруть одну деталь. Події A_1, A_2, A_3, A_4 означають відповідно, що взята деталь бронзова, латунна, мідна, сталева. Визначити подію $B = (A_1 \cup \overline{A_3}) \cap (\overline{A_4} \cup A_2)$.

3. У ящику містяться 3 латунні (Л), 3 сталеві (С) і 1 бронзова (Б) деталі. Беруть 2 деталі. Описати: а) простір елементарних подій Ω ; б) множини елементарних подій A_3 , яка утворює з подіями A_1 і A_2 повну групу, якщо $A_1 = \{СС, ЛС\}$, $A_2 = \{БЛ, ЛС\}$ – елементарні події, позначені згідно з матеріалом деталей.

4. Прилад складається з двох блоків 1-го типу і трьох блоків 2-го типу. Події A_k ($k = 1, 2$) означають, що працює k -й блок 1-го типу, а події B_j ($j = 1, 2, 3$) працює j -й блок 2-го типу. Прилад працює, якщо працює принаймні один блок першого типу і не менш як два блоки 2-го типу. Виразити подію C – «прилад працює» через події A_k і B_j .

5. Робітник виготовив n деталей. Нехай подія A_i ($i = 1, 2, \dots, n$) полягає в тому, що i -та деталь має дефект. Описати такі події:

- а) жодна з деталей не має дефектів;
- б) принаймні одна деталь має дефект;
- в) лише одна деталь має дефект

Тема 2. Класична ймовірність, статистична ймовірність, геометрична ймовірність.

Ймовірністю події A називається числова міра об’єктивної можливості настання цієї події в певному випробуванні. Позначається така ймовірність $P(A)$.

Властивості ймовірності:

- 1. Ймовірність достовірної події $P(\Omega) = 1$.
- 2. Ймовірність неможливої події $P(\emptyset) = 0$.
- 3. Ймовірність будь-якої випадкової події $0 \leq P(A) \leq 1$.

Класичне означення ймовірності. Ймовірністю випадкової події A називається відношення кількості елементарних подій m , які сприяють появі цієї події (становлять множину її елементарних подій), до загальної кількості n рівноможливих елементарних подій, що утворюють простір елементарних подій Ω :

$$P(A) = \frac{m}{n}. \quad (2.1)$$

Щоб обчислити ймовірність події A за цією формулою, потрібно знайти кількість елементарних подій у просторі Ω , а також кількість їх у множині, яка відповідає події A .

Приклад 2.1. Знайти ймовірність того, що число очок, що випаде на гральному кубіку при одному киданні, буде парним (подія A).

Розв'язання. Простір елементарних подій Ω містить $n = 6$ елементарних подій, подія $A = \{2, 4, 6\}$ містить $m = 3$ сприятливі події. За формулою (3.1) одержимо: $P(A) = \frac{3}{6} = \frac{1}{2}$.

Приклад 2.2. В урні 15 білих і 10 чорних кульок. З урни навмання виймають три кульки. Знайти ймовірність того, що 1) всі три кульки білі (подія A); 2) дві кулі – білі і одна чорна (подія B).

Розв'язання. 1) Число всіх елементарних подій дорівнює числу комбінацій з 25 по 3, тобто $n = C_{25}^3$. Знаходимо число сприятливих подій. Три білі можна вибрати з 15 білих кульок $m = C_{15}^3$ способами. Тому шукана ймовірність дорівнює

$$P(A) = \frac{m}{n} = \frac{C_{15}^3}{C_{25}^3} = \frac{15 \cdot 14 \cdot 13}{1 \cdot 2 \cdot 3} \cdot \frac{1 \cdot 2 \cdot 3}{25 \cdot 24 \cdot 23} = \frac{91}{460}.$$

2) Як і в 1) $n = C_{25}^3$. Знаходимо число сприятливих подій. Дві білі можна вибрати з 15 білих кульок C_{15}^2 способами, а 1 чорну з 10 чорних – C_{10}^1 способами. За правилом множення одержимо $m = C_{15}^2 C_{10}^1$. Маємо

$$P(B) = \frac{m}{n} = \frac{C_{15}^2 C_{10}^1}{C_{25}^3} = \frac{15 \cdot 14}{1 \cdot 2} \cdot \frac{10}{1} \cdot \frac{1 \cdot 2 \cdot 3}{25 \cdot 24 \cdot 23} = \frac{21}{46}.$$

Геометричне означення ймовірності. Якщо простір елементарних подій Ω можна подати у вигляді деякого геометричного образу, а множину елементарних подій для події A – як частину цього геометричного образу, то **ймовірність** події A визначається як відношення мір цих множин:

$$P(A) = \frac{\text{міра } A}{\text{міра } \Omega}. \quad (2.2.)$$

Під мірою розуміють довжину, площу, об'єм в одно-, дво- і тривимірному випадках відповідно.

Приклад 2.3. У середину круга радіуса r навмання кидають точку. Знайти ймовірність того, що точка виявиться всередині вписаного в круг квадрата?

Розв'язання. За формулою (2.2.) маємо $P(A) = \frac{S_{\text{кв}}}{S_{\text{кр}}}$. Нехай сторона квадрата дорівнює a , тоді його площа $S_{\text{кв}} = a^2$.

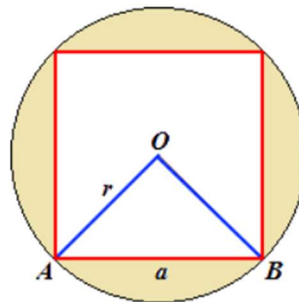


Рисунок.

Враховуючи, що у трикутнику AOB : $OA = OB = r$, $AB = a$, то сторона квадрата $a = \sqrt{2}r$ і $S_{\text{кв}} = a^2 = 2r^2$. Так як, площа круга рівна $S_{\text{кр}} = \pi r^2$, тоді шукана ймовірність дорівнює:

$$P(A) = \frac{S_{\text{кв}}}{S_{\text{кр}}} = \frac{2r^2}{\pi r^2} = \frac{2}{\pi}.$$

Приклад 2.4. Відрізок завдовжки l розділили на 3 частини, вибираючи дві точки поділу навмання. Знайти ймовірність того, що з утворених трьох відрізків можна скласти трикутник.

Розв'язання. Позначимо довжини частин відрізка через $x, y, l - x - y$. Очевидно, що $0 < x + y < l, x > 0, y > 0$. Якщо x, y розглядати як декартові координати на площині, то простір елементарних подій буде трикутник OAB (рисунок нижче).

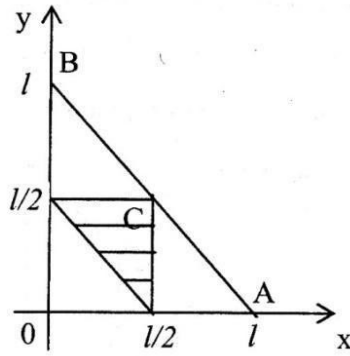


Рисунок.

Для того, щоб з відрізків $x, y, l - x - y$ можна було скласти трикутник, необхідно і достатньо, щоб кожен з цих відрізків був менший за суму двох інших, тобто $x < y + (l - x - y)$, $y < x + (l - x - y)$, $l - x - y < x + y$. Звідси $x < \frac{l}{2}$, $y < \frac{l}{2}$, $x + y > \frac{l}{2}$. Ця система нерівностей виділяє трикутник C, заштрихований на рисунку 3.2. Отже,

$$P(C) = \frac{\text{площа } C}{\text{площа } \Omega} = \frac{\frac{1}{8}l^2}{\frac{1}{2}l^2} = \frac{1}{4}.$$

Нехай в однакових умовах проводиться серія із n випадкових експериментів, у кожному з яких може статися деяка подія A . Якщо $\mu(A)$ – кількість експериментів, у яких подія A насталала, то відношення $v(A) = \frac{\mu(A)}{n}$ називається **відносною частотою** настання події.

Відносна частота має такі властивості:

1. $0 \leq v(A) \leq 1$;
2. $v(\Omega) = 1$;
3. $v(\emptyset) = 0$;
4. якщо A і B – несумісні події, то $v(A + B) = v(A) + v(B)$.

Статистичне означення ймовірності. Число, біля якого ґрунтуються відносні частоти випадкової події A при зростанні числа експериментів, називається **ймовірністю** події A і позначається $P(A)$. На практиці, при великій кількості експериментів, за ймовірність наближено приймають відносну частоту.

Приклад 2.5. При стрільбі була одержана відносна частота влучань 0,6. Скільки було зроблено пострілів, якщо одержано 12 промахів?

Розв'язання. Нехай було зроблено n пострілів. Тоді число влучань дорівнює $n - 12$, а відносна частота дорівнює $\frac{n-12}{n} = 0,6$. Звідси $n - 12 = 0,6n$, тоді $n = 30$.

Аксіоматичне означення ймовірностей. У 1933 році математиком А. Колмогоровим була запропонована система аксіом, яка лягла в основу сучасної теорії ймовірностей. Ця аксіоматика була визнана загальноприйнятим підходом до математичного опису події та ймовірності.

Кожній події поставлено у відповідність дійсне невід'ємне число $P(A)$. Це число називається **ймовірністю події** A . При цьому виконуються наступні *аксіоми*:

1. Ймовірність достовірної події дорівнює 1: $P(\Omega) = 1$.
2. Ймовірність неможливої події дорівнює 0: $P(\emptyset) = 0$.
3. Ймовірність настання суми двох несумісних подій дорівнює сумі ймовірностей подій: $P(A \cup B) = P(A) + P(B)$, $A \cap B = \emptyset$.
4. Ймовірність протилежної події: $P(\bar{A}) = 1 - P(A)$.

Теорема (додавання ймовірностей несумісних подій). Якщо події A і B несумісні, то

$$P(A + B) = P(A) + P(B) \quad (2.4)$$

Наслідок 3.1. Якщо події $A_1, A_2, \dots, A_n$ утворюють повну групу несумісних подій, то сума їх ймовірностей дорівнює 1:

$$\sum_{i=1}^n P(A_i) = 1.$$

Приклад 2.6. В ящику 10 червоних і 6 синіх гудзиків. Навмання виймають два гудзики. Яка ймовірність того, що гудзики будуть одного кольору?

Розв'язання. Введемо позначення: подія A – гудзики одного кольору, подія B – гудзики червоні, подія C – гудзики сині. Очевидно, $A = B + C$ і події B і C несумісні. За формулою (3.3) $P(A) = P(B) + P(C)$. Знайдемо $P(B)$ і $P(C)$. Число способів взяти 2 гудзики з 16 дорівнює C_{16}^2 . Число наслідків, сприятливих для події B , дорівнює C_{10}^2 , а для події C – C_6^2 .

Одержимо:

$$P(B) = \frac{C_{10}^2}{C_{16}^2} = \frac{10 \cdot 9}{1 \cdot 2} \cdot \frac{1 \cdot 2}{16 \cdot 15} = \frac{3}{8}; P(C) = \frac{C_6^2}{C_{16}^2} = \frac{6 \cdot 5}{1 \cdot 2} \cdot \frac{1 \cdot 2}{16 \cdot 15} = \frac{1}{8}.$$

Отже, $P(A) = \frac{3}{8} + \frac{1}{8} = \frac{1}{2}.$

Теорема (додавання ймовірностей сумісних подій). Якщо події A і B сумісні, тоді:

$$P(A + B) = P(A) + P(B) - P(A \cdot B). \quad (2.5)$$

Наслідок 3.3. Ймовірність суми трьох сумісних подій

$$P(A + B + C) = P(A) + P(B) + P(C) - P(A \cdot B) - P(A \cdot C) - P(B \cdot C) + P(A \cdot B \cdot C).$$

Приклад 2.7. Знайти ймовірність того, що навімання вибране двозначне число є кратним 2, або 5, або тому й іншому одночасно.

Розв'язання. Введемо позначення: подія A – вибране число кратне 2, або 5, або тому й іншому одночасно, подія B – число кратне 2, подія C – число кратне 5. Очевидно, що $A = B + C$. За формулою (3.4):

$$P(A) = P(B) + P(C) - P(BC).$$

Всього є $n = 90$ двозначних чисел, з них 45 кратних 2. Кількість двозначних чисел, кратних 5, за правилом добутку дорівнює $n = 9 \cdot 2 = 18$, з них 9 чисел кратних і 2 і 5. Маємо $P(B) = \frac{45}{90} = \frac{1}{2}, P(C) = \frac{18}{90} = \frac{1}{5}, P(BC) = \frac{9}{90} = \frac{1}{10}$. Отже,

$$P(A) = \frac{1}{2} + \frac{1}{5} - \frac{1}{10} = \frac{3}{5} = 0,6.$$

Вправи для аудиторної та самостійної роботи.

1. Задано множину цілих чисел $\{1, 2, 3, \dots, 45\}$. Навмання з цієї множини беруть одне число. Яка ймовірність того, що воно виявиться кратним 2? (подія A); воно виявиться кратним 3? (подія B). Знайти ймовірності наступних подій: $A + B, AB, B - A, \bar{A}B, \bar{A} + \bar{B}$. Підкидують 2 гральні кубики. Знайти ймовірність того, що: $A = \{\text{сума очок на гранях дорівнює } 5\}, B = \{\text{добуток очок на гранях дорівнює } 6\}$. Знайти ймовірність подій $A + B, AB, A - B$.

2. Коробка містить різнокольорові кульки: 10 червоних, 5 білих, 9 чорних і 6 зелених. Із коробки навімання виймають одну кульку. Знайти ймовірність наступних випадкових подій: кулька червоного кольору (подія A); кулька зелена (подія B); кулька не є зеленою (подія C).

3. Монету підкидують 3 рази. Описати простір елементарних подій. Знайти ймовірність того, що: $A = \{\text{випаде 2 рази герб}\}, B = \{\text{випаде 2 рази цифра}\}$. Знайти ймовірності подій: $A + B, AB$.

4. Яка ймовірність того, що взявши навімання 3 карти з колоди в 52 карти, ми дістанемо: а) карти однієї масті; б) одного значення; в) трьох тузів?

5. В ящику є 15 деталей, серед них 10 пофарбованих. Працівник навімання виймає 3 деталі. Яка ймовірність того, що вийняті деталі будуть пофарбовані? Яка ймовірність того, що вийняті будуть лише дві пофарбовані деталі?

Завданн для самостійної роботи.

1. Задано множину цілих чисел $\{1, 2, 3, \dots, 30\}$. Навмання з цієї множини беруть одне число. Яка ймовірність того, що воно виявиться кратним: а) числу 2 (подія A); б) числу 3 (подія B). Знайти ймовірності наступних подій: $A + B, \bar{A} \cdot B, B - A, A \cdot B$.

2. Для знищення цілі необхідно попасти в неї двічі. Ймовірність першого попадання 0,1 і не змінюється при промахах. Вона зростає у двічі після першого попадання. Яка ймовірність знищення цілі за перші два постріли?

3. Робиться два постріли по одній і тій самій мішені. Ймовірність влучання при першому пострілі дорівнює 0,6, при другому – 0,8. Знайдіть ймовірність того, що мішень матиме хоча б одну пробоїну.

4. Ймовірність перегорання за певний час кожної з двох електроламп дорівнює 0,5. Знайти ймовірність того, що за вказаний час: а) перегорять обидві лампочки; б) перегорить тільки одна з них; в) перегорить хоч би одна; г) ні одна лампочка не перегорить.

5. У двох урнах знаходиться по 6 білих і 8 чорних кульок. Не дивлячись дістають з кожної урни по одній кульці. Яка ймовірність того, що: а) обидві кульки білі; б) хоч би одна кулька була білою; в) одна кулька біла, а друга – чорна.

6. Кожен з трьох стрільців виконує один постріл по мішені. Ймовірність їх влучення в мішень відповідно дорівнює 0,6, 0,9 і 0,3. Знайти ймовірність того, що в результаті цих трьох пострілів у мішень буде: а) лише одна пробоїна; б) рівно дві пробоїни; в) три пробоїни?

7. В урні 30 куль, з них 5 – червоних, 10 – синіх, 14 – зелених і одна біла. З урни послідовно виймаємо кулі і назад їх не повертаємо. Яка ймовірність того, що перший раз вийняли червону кулю (подія А), у другий раз – синю (подія В) і третій раз – зелену (подія С).

8. Робітник обслуговує чотири верстати, що працюють незалежно один від одного. Ймовірність того, що протягом години увагу робітника потребуватиме 1-й верстат, дорівнює 0,1 ; 2-й – 0,2; 3-й – 0,15 і 4-й – 0,12. Яка ймовірність того, що протягом години: а) жоден верстат не потребує уваги робітника; б) усі верстати потребують уваги робітника; в) який-небудь один верстат потребує уваги робітника; г) хоча б один верстат потребує уваги робітника.

9. Екзаменаційний білет містить три питання. Ймовірність того, що студент знає відповідь на перше питання – 0,7, друге – 0,8, третє – 0,9. Знайти ймовірність, що студент відповів правильно на: а) всі питання; б) 2 з 3 питань; в) тільки одне питання?

10. У ящику міститься 20 однотипних деталей. Із них 12 стандартні, а решта – браковані. Деталі виймають по одній без повернення. Так було вийнято три деталі. Обчислити ймовірність наступних випадкових подій: А – три деталі виявляються стандартними; В – усі три деталі виявляються бракованими; С – дві стандартні й одна бракована; D – хоча б одна стандартна.

11. Готуючись до екзамену з теорії ймовірностей, студент з 25 екзаменаційних білетів вивчив лише 20. В якому випадку ймовірність витягти вивчений білет буде вищою: а) коли він тягне білет першим; б) коли він тягне білет другим.

12. Підкидають три гральних кубики. Знайти ймовірність того, що в сумі буде: а) 10 очок (подія А); б) 11 очок (подія В); в) 12 очок (подія С).

13. Два стрільці незалежно один від одного стріляють у ціль по одному разу. Ймовірність влучення в ціль для 1-го стрільця дорівнює 0,6, а для 2-го – 0,7. Яка ймовірність того, що: а) обидва стрільці влучають у ціль; б) перший влучить, а другий промахнеться; в) влучить тільки один із стрільців.

14. На тридцяти жетонах написані числа від 11 до 40. Яка ймовірність того, що навмання взятий жетон буде занумерований числом, сума цифр якого дорівнює 2 або 5?

14. Для руйнування моста досить влучення однієї авіаційної бомби. Знайти ймовірність того, що міст буде зруйнований, якщо на нього скинути 4 бомби, ймовірності влучення яких відповідно дорівнюють 0,3; 0,4; 0,6; 0,7.

15. Покупець може придбати акції двох компаній А та В. Надійність першої оцінюється експертами на рівні 90%, а другої – 80%. Чому рівна ймовірність того, що: а) обидві компанії протягом року не стануть банкрутами; б) хоча б одна з компаній збанкрутує?

16. На фірмі 550 співробітників, 380 з них мають вищу освіту, а 412 – середню спеціальну освіту, у 357 співробітників є вища і середня спеціальна освіта. Чому рівна

ймовірність того, що випадково вибраний співробітник має або середню спеціальну, або вищу освіту, або і ту і іншу?

17. Ймовірність того, що споживач побачить рекламу певного продукту по кожному з трьох центральних телевізійних каналів, рівна 0,05. Припустимо, що ці події незалежні в сукупності. Чому рівна ймовірність того, що споживач побачить рекламу: а) по всіх трьох каналах; б) хоча б по одному з каналів?

18. Студент розшукує потрібну йому формулу у трьох довідниках. Ймовірність того, що формула міститься у першому довіднику дорівнює 0,6, у другому – 0,7, в третьому – 0,8. Знайти ймовірність того, що формула міститься: а) тільки в одному довіднику; б) тільки в двох довідниках; в) хоча б в одному із довідників.

19. У грошово-речовій лотереї на кожен тисячу квитків припадає п'ятнадцять грошових і двадцять речових виграшів. Знайти ймовірність того, що при покупці трьох квитків: а) буде не менше двох виграшних; б) в одному з квитків буде речовий приз.

20. Для прийому партії, що складається з сорока стандартних виробів і десяти бракованих, застосовують вибірковий контроль. Партія бракується, якщо серед узятих навмання п'яти виробів виявиться: а) хоча б два бракованих; б) більше трьох бракованих.

21. Станція метрополітену обладнана трьома ескалаторами. Ймовірність безвідмовної роботи для першого ескалатора дорівнює 0,9; для другого – 0,85; для третього – 0,8. Знайти ймовірність того, що станеться поломка: а) не більше двох ескалаторів; б) третього ескалатора.

22. Пристрій складається з чотирьох елементів, які працюють незалежно. Ймовірності безвідмовної роботи цих елементів за час t відповідно дорівнюють 0,7; 0,6; 0,9; 0,8. Знайти ймовірність того, що за час t безвідмовно будуть працювати: а) три елементи; б) хоча б один елемент.

23. Літак долає три зони протиповітряної оборони. При доланні першої зони його може бути пошкоджено з ймовірністю 0,6; другої, за умови, що він пройшов першу, – 0,7; а третьої, за умови, що він пройшов перші дві, – 0,5. Знайти ймовірності наступних подій: а) проходження всіх зон; б) пошкодження літака під час проходження другої зони.

24. Для сигналізації аварії встановлено два сигналізатори, які працюють незалежно один від одного. Ймовірність того, що під час аварії спрацює перший сигналізатор дорівнює 0,95, а другий – 0,9. Знайти ймовірності того, що під час аварії спрацюють: а) принаймні один сигналізатор; б) тільки один сигналізатор.

25. Два спортсмени незалежно один від одного роблять по одному пострілу по одній мішені. Ймовірності влучення у мішень для першого та другого спортсменів відповідно дорівнюють 0,7 та 0,9. Знайти ймовірності того, що: а) обидва спортсмени не влучать у мішень; б) перший спортсмен влучить, а другий – ні.

26. Три студенти складають залік. Ймовірність того, що перший студент складе залік дорівнює 0,9, другий – 0,6 та третій – 0,7. Знайти ймовірність того, що: а) хоча б два студенти складуть залік; б) тільки один студент складе залік.

27. У бібліотеці підручників по вищій математиці 70% від кількості студентів, а підручників по економіці – 85%. У бібліотеку прийшов студент. Яка ймовірність, що: а) йому дістануться обидва підручники; б) йому дістанеться тільки один підручник?

28. Три стрілки, для яких ймовірності влучення в мішень дорівнюють відповідно 0,9, 0,7 і 0,8, роблять по два постріли. Визначити ймовірність хоча б одного влучення в мішень та ймовірність тільки одного влучення в мішень.

29. Для сигналізації про аварію встановлено три незалежно працюючих сигналізатори. Ймовірність того, що при аварії спрацює перший сигналізатор дорівнює 0,9; другий – 0,8; третій – 0,95. Знайти ймовірність, що при аварії спрацюють: а) хоча б один сигналізатор; б) рівно два сигналізатора.

30. З трьох гармат провели залп по цілі. Ймовірність влучення в ціль при одному пострілі з першої гармати дорівнює 0,9; з другої – 0,7; з третьої – 0,8. Знайти ймовірність подій: $A = \{\text{тільки два снаряди влучать у ціль}\}$ і $B = \{\text{хоча б один снаряд влучить у ціль}\}$.

Тема 3. Теореми про ймовірність суми та добутку подій. Умовна ймовірність. Незалежні події.

Теорема додавання ймовірностей. Нехай подія A є сумою двох подій B і C . Тоді:

а) якщо події B і C несумісні, то

$$P(A) = P(B \cup C) = P(B) + P(C);$$

б) якщо події B і C сумісні, то

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

Події B і C називаються **залежними**, якщо ймовірність однієї з них змінюється залежно від того, відбулась друга подія чи ні. У протилежному випадку події називаються **незалежними**.

Ймовірність події B , визначена за умови, що подія C відбулася, називається **умовною** і позначається $P(B / C)$.

Теорема множення ймовірностей. Нехай подія A є добутком двох подій B і C . Тоді:

а) якщо події B і C незалежні, то

$$P(A) = P(B \cap C) = P(B) \cdot P(C);$$

б) якщо події B і C залежні, то

$$P(A) = P(B \cap C) = P(B) \cdot P(C/B).$$

Ці теореми справджуються й для добутку n ($n > 2$) подій.

Ймовірність настання принаймні однієї події. Нехай у результаті випробування можуть відбутися n подій $A_1, A_2, \dots, A_n$. Потрібно знайти ймовірність того, що відбудеться принаймні одна з них. Позначимо цю подію літерою A . Тоді протилежною буде подія $\bar{A}$ яка полягає в тому, що в результаті випробування одночасно настали протилежні події:

$$\bar{A} = \bigcap_{i=1}^n \bar{A}_i.$$

Знайдемо ймовірність події A через ймовірність протилежної події:

$$P(A) = 1 - P(\bar{A}) = 1 - P\left(\bigcap_{i=1}^n \bar{A}_i\right).$$

Розв'язання типових задач

Задача 1. Партія містить 12 стандартних і чотири нестандартні деталі. Навмання беруть три деталі. Знайти ймовірність того, що серед узятих деталей:

- 1) не менш як дві стандартні;
- 2) усі три нестандартні;
- 3) принаймні одна стандартна.

Розв'язання. 1) Нехай подія A – «серед трьох узятих деталей не менш як дві стандартні». Тоді її можна подати як суму двох подій: A_1 – «серед трьох узятих деталей дві стандартні і одна нестандартна» і A_2 – «усі три узяті деталі стандартні». Події A_1 і A_2 несумісні, тому маємо

$$P(A) = P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

Ймовірності подій A_1 і A_2 знайдемо згідно з класичним означенням ймовірності:

$$n = C_{16}^3 = 560; m_1 = C_{12}^2 \cdot C_4^1 = 66 \cdot 4 = 264; m_2 = C_{12}^3 = 220;$$

Отже,

$$P(A) = \frac{484}{560} = 0,864.$$

2) Подія B – «усі три взяті деталі нестандартні». Цю подію можна подати як добуток трьох подій B_i ($i = 1, 2, 3$), де i -та деталь нестандартна, $B = \bigcap_{i=1}^3 B_i$.

Умовою задачі не задано, що деталі беруться з поверненням. Отже, взяти три деталі разом – означає те саме, що брати їх по одній без повернення, а тому події залежні. Згідно з цим ймовірність події B обчислюємо наступним чином:

$$P(B) = P(\cap_{i=1}^3 B_i) = P(B_1)P(B_2/B_1)P(B_3/B_1 \cap B_2) = \frac{4}{16} \cdot \frac{3}{13} \cdot \frac{2}{12} = 0,007.$$

3) Подія C – «із трьох деталей принаймні одна стандартна». Протилежна подія $\bar{C}$ – «усі три деталі нестандартні». Імовірність цієї події щойно знайдено: $P(\bar{C}) = P(B)$. Остаточного маємо:

$$P(C) = 1 - P(\bar{C}) = 1 - 0,007 = 0,993.$$

Задача 2. Нехай перша партія деталей складається з 10 стандартних і 3 нестандартних деталей, друга – із 15 стандартних і 4 нестандартних, третя – із 20 стандартних і 5 нестандартних деталей. Із кожної партії беруть по одній деталі. Знайти ймовірність того, що серед узятих деталей:

- 1) тільки одна стандартна;
- 2) тільки дві стандартні.

Розв'язання. Нехай згідно з умовою з кожної партії взято по одній деталі. При цьому можуть відбутися події A_1, A_2, A_3 , які полягають відповідно в тому, що деталь, яку взяли з першої, другої і третьої партії виявилась стандартною.

1) Подія A – «тільки одна із трьох деталей виявилась стандартною». Ця подія матиме вигляд:

$$A = A_1 \bar{A}_2 \bar{A}_3 \cup \bar{A}_1 A_2 \bar{A}_3 \cup \bar{A}_1 \bar{A}_2 A_3.$$

Групи подій, сумою яких є подія A , несумісні між собою, а події в кожній групі незалежні. Тому ймовірність події A обчислимо так:

$$P(A) = \frac{10}{13} \cdot \frac{4}{19} \cdot \frac{1}{5} + \frac{3}{13} \cdot \frac{15}{19} \cdot \frac{1}{5} + \frac{3}{13} \cdot \frac{4}{19} \cdot \frac{4}{5} = \frac{133}{1235} = 0,108.$$

2) Подія B – «тільки дві деталі із трьох виявились стандартними». Подамо цю подію через події A_1, A_2, A_3 та протилежні до них:

$$B = A_1 A_2 \bar{A}_3 \cup A_1 \bar{A}_2 A_3 \cup \bar{A}_1 A_2 A_3.$$

Подію B подано як суму несумісних груп подій. У кожній групі події незалежні. Знайдемо ймовірність події B :

$$P(B) = \frac{10}{13} \cdot \frac{15}{19} \cdot \frac{1}{5} + \frac{10}{13} \cdot \frac{4}{19} \cdot \frac{4}{5} + \frac{3}{13} \cdot \frac{15}{19} \cdot \frac{4}{5}.$$

Задача 3. Перевезення вантажів для підприємства забезпечують дві бригади, які з цієї метою щодня в першу зміну мають виділяти по одному автомобілю. Імовірність виходу автомобіля на лінію в першій бригаді дорівнює 0,7, а в другій – 0,6. Знайти ймовірність того, що в першу зміну на підприємстві перевозитимуться вантажі.

Розв'язання. Розглянемо події: A – «на підприємстві в першу зміну перевозитимуться вантажі»; A_1 – «для перевезення вантажів прибув автомобіль із першої бригади»; A_2 – «для перевезення вантажів прибув автомобіль із другої бригади». Тоді $A = A_1 \cup A_2$. Події A_1 і A_2 сумісні, тому

$$P(A) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

Очевидно, що події A_1 і A_2 незалежні і $P(A_1 \cap A_2) = P(A_1)P(A_2)$. Остаточного дістаємо:

$$P(A) = 0,7 + 0,6 - 0,7 \cdot 0,6 = 0,88.$$

Задача 4. Прилад складається із трьох вузлів, які працюють незалежно один від одного, причому другий і третій вузли взаємозамінювані. Ймовірності виходу з ладу вузлів на заданому часовому проміжку становлять відповідно 0,2; 0,3 і 0,4. Знайти ймовірність того, що протягом заданого часу прилад працюватиме.

Розв'язання. Розглянемо події: A – «прилад працює протягом заданого часу»; B_1 – «перший вузол працює»; B_2 – «другий вузол працює»; B_3 – «третій вузол працює». Подія A настає, якщо працюють перший та другий вузли, або перший та третій вузли, або всі три вузли разом. Звідси:

$$A = B_1 \cap (B_2 \cup B_3).$$

За умовою задачі маємо, що події B_1 і $B_2 \cup B_3$ незалежні, а події B_2 і B_3 – сумісні.

Тому

$$A = P(B_1 \cap (B_2 \cup B_3)) = P(B_1)P(B_2 \cup B_3) = P(B_1)(P(B_2)P(B_3) - P(B_2 \cap B_3)) = 0,8(0,7 + 0,6 - 0,7 \cdot 0,6) = 0,704.$$

Під час обчислення враховано, що умовою задачі задано ймовірності протилежних подій.

Задача 5. Ймовірність влучення стрілкою у першу область мішені дорівнює 0,45, у другу область – 0,35, у третю – 0,15. Знайти ймовірність того, що при одному пострілі стрілок влучить у першу або другу області мішені.

Розв'язання. Позначимо за подію A — влучення у першу область мішені; за подію B — влучення у другу область мішені.

При одному пострілі події A та B несумісні. Тому ймовірність влучення в першу або другу області мішені буде

$$P(A + B) = 0.45 + 0.35 = 0.8.$$

Задача 6. Студент знає 20 з 25 питань програми іспиту. Знайти ймовірність того, що він дасть правильну відповідь на три питання, які він отримав на іспиті.

Розв'язання. Нехай подія A_1 полягає в тому, що студент знає відповідь на перше питання, A_2 — на друге, A_3 — на третє. Тоді подія B — «він дасть правильну відповідь на три питання» має вигляд

$$B = A_1 A_2 A_3.$$

Використавши теорему добутку ймовірностей залежних подій, знайдемо

$$P(B) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) = \frac{20}{25} \cdot \frac{19}{24} \cdot \frac{18}{23} = \frac{57}{115}.$$

Задача 7. Підприємець має акції двох компаній. Ймовірність отримання дивідендів по акціях тільки однієї з двох компаній дорівнює 0,38, причому для першої компанії вона дорівнює 0,8. Знайти ймовірність того, що він отримає дивіденди по акціях другої компанії.

Розв'язання. Нехай подія A_1 полягає в отриманні дивідендів по акціях першої компанії, A_2 — другої. Тоді

$$P(A_1) = 0,8 \text{ і } P(A_2) = x.$$

Подія «отримано дивіденди по акціях тільки однієї з двох компаній» має вигляд

$$P(A_1 \bar{A}_2) + P(A_2 \bar{A}_1) = P(A_1)P(\bar{A}_2) + P(A_2)P(\bar{A}_1).$$

Звідки маємо рівняння

$$0,8(1 - x) + x(1 - 0,8) = 0,36,$$

розв'язавши яке, знайдемо $x = 0,7$.

Приклади для самостійного розв'язання

1. У разі масового виготовлення виробів брак становить у середньому 1,5% загальної кількості всіх виробів. 3-поміжпридатних виробів 85,3 % становлять вироби 1-го сорту. Знайти ймовірність того, що навмання взятий виріб належить до 1-го сорту.

2. Контролер перевіряє партію з двох деталей 1-го сорту, двох 2-го сорту і трьох – 3-го сорту. Деталі беруть по одній навмання без повернення. Знайти ймовірність того, що деталь 1-го сорту з'явиться раніше за деталь 3-го сорту.

3. У цеху є три резервні двигуни, для кожного з яких ймовірність бути ввімкненим у даний момент дорівнює 0,3. Знайти ймовірність того, що в даний момент ввімкнено:

- 1) принаймні два двигуни;
- 2) принаймні один двигун.

4. Із партії суконь дівчина має намір вибрати дорогі. Ймовірність того, що навмання взята сукня дорога, дорівнює 0,7. Знайти ймовірність того, що із трьох суконь тільки дві будуть дорогими.

5. Від вокзалу відправились два автобуси. Ймовірність своєчасного прибуття кожного з них дорівнює 0,92. Знайти ймовірність такої події:

A – обидва автобуси прибудуть своєчасно;

B – обидва автобуси запізняться;
 C – тільки один автобус прибуде своєчасно

Тема 4. Схема повторних випробувань. Формули Бернуллі, Пуассона, локальна та інтегральна теореми Лапласа.

Незалежні випробування. Нехай проводяться n випробувань, у кожному з яких подія A може відбутись або не відбутись. Якщо ця ймовірність у кожному випробуванні не залежить від того, відбулась вона в інших випробуваннях чи ні, то такі випробування називаються незалежними щодо події A . Згідно з означенням випробування також незалежні, якщо в кожному з них ймовірність настання події A однакова. Ймовірність того, що подія A відбудеться в кожному з незалежних випробувань, позначають $P(A) = p$, а ймовірність настання протилежної події

$$P(\bar{A}) = 1 - p = q.$$

Для розв'язання задач на повторні незалежні випробування застосовують такі формули і теореми.

Формула Бернуллі. Ймовірність того, що в n незалежних випробуваннях, у кожному з яких ймовірність $P(A) = p$, подія A відбудеться m раз, подається так:

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

Формула застосовується, якщо $n \leq 10$.

Найімовірніша кількість. Частота m_0 настання події A в n незалежних повторних випробуваннях називається **найімовірнішою кількістю** (появи цієї події), якщо їй відповідає найбільша ймовірність. Вона визначається за формулою:

$$np - q \leq m_0 \leq np + q.$$

Розподіл може мати одне або два найімовірніші числа.

Локальна теорема Лапласа. Якщо в кожному з n незалежних випробувань ймовірність появи події A стала і не дорівнює 0 і 1, то ймовірність того, що в цих n випробуваннях подія A відбудеться рівно m раз, наближено дорівнює

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x),$$

де $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x = \frac{m-np}{\sqrt{npq}}$. Функцію $\varphi(x)$ протабульовано. При користуванні таблицею її значень потрібно мати на увазі, що ця функція парна, тобто $\varphi(-x) = \varphi(x)$, а при $x \geq 4$ її значення практично дорівнюють нулю.

Локальна теорема Лапласа дає змогу обчислювати ймовірності $P_n(m)$, якщо $n > 10$ і $p > 0,1$.

Формула Пуассона. Якщо число n проведених випробувань велике, а ймовірність появи події A в кожному випробуванні мала ($p \leq 0,01$), то ймовірність появи події A в цих n випробуваннях m раз обчислюють за формулою Пуассона:

$$P_n(m) \approx \frac{a^m}{m!} e^{-a},$$

де $a = np$ – середнє число появ події A в n випробуваннях.

Інтегральна теорема Лапласа. Якщо в кожному з n незалежних випробувань ймовірність p появи події A стала і не дорівнює 0 і 1, то ймовірність того, що в n випробуваннях подія A відбудеться від m_1 до m_2 раз, наближено дорівнює:

$$P_n(m_1, m_2) \approx \Phi(x_2) - \Phi(x_1),$$

де

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt - \text{функція Лапласа};$$
$$x_1 = \frac{m_1 - np}{\sqrt{npq}}; \quad x_2 = \frac{m_2 - np}{\sqrt{npq}}.$$

Функцію $\Phi(x)$ протабульовано. При користуванні таблицею її значень потрібно мати на увазі, що функція $\Phi(x)$ непарна, тобто $\Phi(-x) = -\Phi(x)$, а при $x \geq 5$ $\Phi(x) = 0,5$.

Відхилення відносної частоти від ймовірності. Ймовірність того, що при проведенні n незалежних випробувань відхилення відносної частоти події A від її ймовірності за модулем не перевищить $\varepsilon > 0$, визначається за формулою:

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 2\Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right).$$

Розв'язання типових задач

Задача 1. Із партії, в якій 12 стандартних і 4 нестандартні деталі, навмання беруться 3 деталі з поверненням. Знайти ймовірність того, що серед узятих деталей:

- 1) усі три стандартні;
- 2) не більш як одна нестандартна;
- 3) принаймні одна нестандартна.

Розв'язання. Маємо схему трьох незалежних випробувань. Нехай подія A – «узята щоразу деталь стандартна», тоді $P(A) = p = \frac{12}{16} = 0,75$.

- 1) Ймовірності обчислюватимемо за формулою Бернуллі:

$$P_3(3) = C_3^3 p^3 q^0 = 0,75^3 = 0,421875.$$

2) Подію «із трьох деталей не більш як одна нестандартна» можна розглядати так: узято 3 стандартні деталі або 2 стандартні і одну нестандартну деталь. У позначеннях формули Бернуллі

$$P_3(x \geq 2) = P_3(3) + P_3(2) = 0,421875 + C_3^2 \cdot 0,75^2 \cdot 0,25 = 0,84375.$$

3) Протилежною для даної буде подія «усі три деталі стандартні». Їй рівносільна подія $P_3(x < 3)$ Обчислимо цю ймовірність:

$$P_3(x < 3) = 1 - P_3(3) = 1 - 0,421875 = 0,578125.$$

Задача 2. Частка довгих волокон у партії бавовни становить у середньому 0,6 загальної кількості волокон. Скільки потрібно взяти волокон, щоб найімовірніше число довгих волокон серед них дорівнювало 40?

Розв'язання. Скористаємося формулою, за якою визначається найімовірніше число: $np - q \leq m_0 \leq np + q$. Підставимо значення відомих величин:

$$0,6n - 0,4 \leq 40 \leq 0,6n + 0,6; \quad -0,4 - 40 \leq -0,6n \leq 0,6 - 40,$$

$$\frac{39,4}{0,6} \leq n \leq \frac{40,4}{0,6}.$$

Задача має два розв'язки: $n = 66$ і $n = 67$.

Задача 3. На кожні 40 відштампованих виробів у середньому припадає 4 дефектних. Із усієї продукції навмання узято 400 виробів. Знайти ймовірність того, що серед них 350 виробів будуть без дефектів.

Розв'язання. Подія A – «узято виріб без дефекту». За умовою $P(A) = p = 0,9$. Проведено $n = 400$ незалежних випробувань. Розв'яжемо задачу за формулою локальної теореми Лапласа:

$$P_n(m) = \frac{1}{\sqrt{npq}} \cdot \varphi(x), \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x = \frac{m - np}{\sqrt{npq}}.$$

Підставляючи дані за умовою задачі, дістаємо: $x = \frac{350 - 400 \cdot 0,9}{\sqrt{400 \cdot 0,9 \cdot 0,1}} = -1,67$. За таблицями знаходимо $\varphi(-1,67) = 0,098$ беручи до уваги, що $\varphi(x)$ – парна функція.

$$\text{Отже, } P_{400}(350) = \frac{0,0989}{6} = 0,0165.$$

Задача 4. Завод відправив на базу 1000 доброякісних виробів. За час перебування в дорозі кожний виріб може бути пошкоджено з імовірністю 0,003. Знайти ймовірність того, що на базу придуть 3 пошкоджені вироби.

Розв'язання. Якщо подія A – «виріб пошкоджено», то її ймовірність $p = 0,003$. Розглядається схема незалежних випробувань, $n = 1000$. Імовірність події A досить мала, тому задачу розв'яжемо за формулою Пуассона: $P_n(m) = \frac{a^m}{m!} e^{-a}$, $a = np$.

Виконуючи обчислення, знаходимо:

$$a = np = 1000 \cdot 0,003 = 3; P_{1000}(3) = \frac{3^3}{3!} e^{-3} = 0,229$$

Задача 5. Зерна пшениці проростають з ймовірністю 0,95. Знайти ймовірність того, що із 2000 посіяних зерен зійде від 1880 до 1920.

Розв'язання. Подія A – «зерно пшениці зійшло». Її ймовірність $p = 0,95$, кількість незалежних випробувань $n = 2000$. Застосуємо формулу інтегральної теореми Лапласа:

$$P_n(m_1, m_2) = \Phi(x_2) - \Phi(x_1), \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt - \text{функція Лапласа};$$

$$x_1 = \frac{m_1 - np}{\sqrt{npq}}; x_2 = \frac{m_2 - n}{\sqrt{npq}}.$$

Далі виконаємо обчислення:

$$x_1 = \frac{1880 - 2000 \cdot 0,95}{\sqrt{2000 \cdot 0,95 \cdot 0,05}} = -2,0519567; x_2 = \frac{1920 - 1900}{\sqrt{95}} = 2,0519567$$

$$P_{2000}(1880; 1920) = \Phi(2,06) - \Phi(-1,03) = \Phi(2,06) + \Phi(1,03) = 0,4803 + 0,3485 = 0,8288.$$

Значення функції Лапласа беруться з відповідної таблиці.

Задача 6. В урні 15 кульок: 10 чорних і 5 білих. Випробування виконують так: виймають одну кульку, фіксують її колір і кладуть назад; перемішують кульки і виймають ще одну кульку. Таку процедуру повторюють кілька разів. Яка ймовірність того, що під час такого випробування з п'яти вийнятих кульок буде чотири білих?

Розв'язання. Ймовірність вийняти білу кульку під час кожного випробування дорівнює $p = \frac{5}{15} = \frac{1}{3}$, а ймовірність того, що білої кульки не виймуть, дорівнює $1 - p = \frac{2}{3}$. За формулою Бернуллі знаходимо ймовірність появи чотирьох білих кульок серед п'яти вийнятих:

$$P_5(4) = C_5^4 \cdot \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^{5-4} = 5 \cdot \frac{2}{768} = \frac{10}{768}.$$

Приклади для самостійного розв'язання

1. На кожні 30 штампованих виробів у середньому припадає 6 виробів з дефектом. Знайти ймовірність того, що з 5 навмання взятих виробів 3 виявляться без дефекту.

2. Імовірність виграшу облігації за весь період позики становить 0,6. Куплено 5 облігацій. Знайти ймовірність такої події:

- 1) виграють дві облігації;
- 2) виграш випаде принаймні на одну облігацію;
- 3) виграють не більш як дві облігації.

3. Для забезпечення роботи на деякому будівельному об'єкті автопідприємство має 6 автомобілів. Імовірність виходу кожного автомобіля на лінію в першу зміну дорівнює 0,8. Знайти ймовірність нормальної роботи автопідприємства, якщо для цього в першу зміну потрібно мати на лінії не менш як 4 автомобілі.

4. Яку частку (у відсотках) виробів 1-го сорту має виробляти автомат, щоб у партії із 100 навмання взятих виробів найімовірніша кількість виробів 1-го сорту дорівнювала 80?

5. Частка 2-го сорту деякої масової продукції в середньому становить 20 %. Навмання взято 100 примірників цієї продукції. Яка кількість виробів 2-го сорту в утвореній групі найімовірніша і яка ймовірність того, що в цій групі буде саме така кількість виробів 2-го сорту?

6. Ймовірність несплати податків для кожного із 400 підприємств дорівнює 0,2. Яка ймовірність того, що податки несплатять 80 підприємств?

Тема 5. Дискретні і неперервні випадкові величини. Основні характеристики. Властивості.

Випадковою називають величину, яка в результаті випробування набуває одне і лише одне з можливих значень, наперед не відоме і залежне від випадкових причин, які заздалегідь не можуть бути враховані.

Випадкові величини прийнято позначати великими латинськими літерами: $X, Y, Z, \dots$, а їх можливі значення – відповідними малими: $x, y, z, \dots$.

Розрізняють дискретні і неперервні випадкові величини. Якщо множина можливих значень випадкової величини зчисленна, то таку величину називають **дискретною**. **Неперервною випадковою величиною** називається величина, яка набуває будь-яких значень з деякого скінченного або нескінченного інтервалу (області).

Співвідношення, що встановлює зв'язок між можливими значеннями випадкової величини та відповідними їй ймовірностями, називають **законом розподілу випадкової величини**.

Якщо у результаті експерименту дискретна випадкова величина X набуває одного із значень $x_1, x_2, \dots, x_n$, тобто відбулась одна з повної групи несумісних подій: $X = x_1, X = x_2, \dots, X = x_n$ з ймовірностями цих подій $p_1 = P(X = x_1), p_2 = P(X = x_2), \dots, p_n = P(X = x_n)$, то

$$\sum_{i=1}^n P(X = x_i) = \sum_{i=1}^n p_i = 1.$$

Якщо ж величина набуває зліченної множини значень, то $\sum_{i=1}^{\infty} p_i = 1$.

Закони розподілу дискретних випадкових величин задаються в табличній формі, аналітичній (наводиться формула, за якою обчислюються ймовірності для заданих значень випадкової величини), графічній (у прямокутній системі координат задається набір точок $(x_i; p_i)$ сполучивши точки відрізками прямих, дістанемо многокутник розподілу ймовірностей).

При табличному способі задання закону розподілу перший рядок таблиці містить можливі значення випадкової величини, а другий – відповідні ймовірності

X	x_1	x_2	x_3	...	x_n
p	p_1	p_2	p_3	...	p_n

Ця таблиця називається **таблицею розподілу** ймовірностей дискретної випадкової величини X або **рядом розподілу**.

Функцією розподілу випадкової величини X називається функція $F(x)$, яка для кожного значення x дорівнює ймовірності того, що випадкова величина X набуде значення, меншого за x , тобто $F(x) = P(X < x)$.

Для дискретної випадкової величини, заданої рядом розподілу, функцію розподілу задають наступним чином

$$F(x) = \sum_{x_i < x} p(x_i).$$

Функція розподілу – неспадна, неперервна зліва, $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$. Для довільних α, β , $P(\alpha \leq X \leq \beta) = F(\beta) - F(\alpha)$.

Якщо X – неперервна випадкова величина, то $F(x)$ – функція неперервна й диференційовна; її похідна $F'(x) = f(x)$ називається **щільністю розподілу ймовірностей**.

Властивості щільності розподілу ймовірностей:

1. $f(x) \geq 0$ для всіх $x \in R$;

2. $\int_{-\infty}^{+\infty} f(x)dx = 1;$
3. $P\{a \leq X < b\} = \int_a^b f(x)dx.$
4. $F(x) = \int_{-\infty}^x f(t)dt.$

Числові характеристики випадкових величин

1. Математичне сподівання або середнє значення дискретної випадкової величини X , називається число, яке дорівнює сумі добутків усіх можливих значень X на відповідні імовірності

$$MX = \sum_{i=1}^n x_i p_i.$$

Якщо X приймає нескінчену кількість значень, то

$$MX = \sum_{i=1}^{\infty} x_i p_i,$$

причому математичне сподівання існує, якщо ряд у правій частині рівності збігається абсолютно.

Якщо неперервна випадкова величина приймає значення на відрізку $[a, b]$ та має щільність імовірності $f(x)$, то її математичне сподівання знаходиться за формулою

$$MX = \int_a^b x f(x) dx.$$

Якщо ж можливі значення неперервної випадкової величини X належать нескінченному інтервалу $(-\infty; +\infty)$, то її математичне сподівання знаходиться за формулою

$$MX = \int_{-\infty}^{+\infty} x f(x) dx,$$

за умови, що інтеграл збігається абсолютно.

Основні властивості математичного сподівання:

- 1) Математичне сподівання сталої дорівнює самій сталій, тобто $MC = C$;
- 2) Сталий множник можна виносити за знак математичного сподівання, тобто $M(CX) = C \cdot MX$;
- 3) Математичне сподівання суми випадкових величин дорівнює сумі їх математичних сподівань, тобто $M(X + Y) = MX + MY$;
- 4) Математичне сподівання добутку незалежних дискретних випадкових величин дорівнює добутку їх математичних сподівань, тобто $M(XY) = MX \cdot MY$.

2. Дисперсія або розсіювання DX випадкової величини X називають число, яке дорівнює математичному сподіванню квадрата відхилення випадкової величини X від її математичного сподівання, тобто

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

Основні властивості дисперсії:

- 1) Дисперсія будь-якої випадкової величини X невід'ємна, тобто $DX \geq 0$;
- 2) Дисперсія сталої C дорівнює нулеві, тобто $DC = 0$;
- 3) Сталий множник C можна виносити за знак дисперсії, при цьому сталий множник треба піднести у квадрат, тобто $D(CX) = C^2 DX$;
- 4) Дисперсія суми двох незалежних випадкових величин X та Y дорівнює сумі їх дисперсій, тобто $D(X + Y) = DX + DY$.

3. Корінь квадратний з дисперсії називається **середнім квадратичним відхиленням** і позначається:

$$\sigma(X) = \sqrt{D(X)}, \text{ або } \sigma_x = \sqrt{D_x}.$$

Якщо від випадкової величини віднімемо її математичне сподівання, то дістанемо центровану випадкову величину, математичне сподівання якої дорівнює нулю. Ділення випадкової величини на її середнє квадратичне відхилення називається **нормуванням** цієї випадкової величини.

4. Початковий, центральний і абсолютний початковий моменти порядку величини X визначають відповідно за такими формулами:

$$\nu_k = MX^k, \mu_k = M(X - MX)^k, \alpha_k = M|X|^k.$$

Якщо існує початковий абсолютний момент порядку k , то існують усі моменти нижчих порядків.

5. Мода і медіана. Крім математичного сподівання, яке є основною характеристикою положення випадкової величини, у теорії ймовірностей знаходять застосування інші характеристики положення – мода і медіана випадкової величини.

Модою $Mo(X)$ випадкової величини X називається її найбільш імовірне можливе значення. Для дискретної випадкової величини X , яка має ряд розподілу $Mo(X)$ дорівнює тому можливому значенню, якого величина набуває з найбільшою ймовірністю.

Для неперервної випадкової величини X , заданої щільністю ймовірності $f(x)$ на проміжку $[a; b]$, мода дорівнює тому значенню x , при якому $f(x)$ досягає максимуму або (у разі відсутності максимуму) набуває найбільшого значення. Якщо $f(x)$ набуває найбільшого значення в одній точці x з проміжку $[a; b]$, то розподіл випадкової величини X називається *унімодальним*, а якщо в кількох точках — *полімодальним*.

Медіаною $Me(X)$ випадкової величини X називається те значення x , для якого

$$P\{X < Me(X)\} = P\{X > Me(X)\} = \frac{1}{2}.$$

Медіана, як правило, застосовується для неперервної випадкової величини і геометрично подає абсцису точки, в якій пряма $x = Me(X)$ поділяє навпіл площу, обмежену кривою розподілу $f(x)$ і віссю Ox .

6. Асиметрія й ексцес. Крім розглянутих характеристик випадкової величини застосовуються й інші, які характеризують той чи інший бік розподілу випадкової величини.

Коефіцієнт асиметрії або **асиметрія** – нормований центральний момент 3-го порядку. Цей коефіцієнт характеризує асиметрію графіка $f(x)$ щодо точки математичного сподівання випадкової величини та визначається за формулою:

$$A_s = \frac{\mu_3}{\sigma^3}.$$

Коефіцієнт ексцесу або **ексцес** – нормований центральний момент 4-го порядку:

$$E_s = \frac{\mu_4}{\sigma^4} - 3.$$

Він характеризує ступінь відхилення закону розподілу випадкової величини X від нормального (гауссового).

Розв'язання типових задач

Задача 1. Проводиться випробування надійності системи, яка складається з трьох приладів, що працюють незалежно один від одного. Надійність (імовірність безвідказної роботи) першого приладу дорівнює 0,9, другого — 0,8, третього — 0,7. Побудувати ряд розподілу випадкової величини X — кількості надійних приладів у системі. Побудувати функцію розподілу випадкової величини X і знайти ймовірності того, що X набуде можливого значення з проміжку $(0,5; 2)$.

Розв'язання. Випадкова величина X набуває можливих значень 0, 1, 2, 3. Позначимо відповідно через g_1, g_2, g_3 імовірності безвідказної роботи першого, другого і третього приладів. Тоді за умовою задачі $g_1 = 0,9$; $g_2 = 0,8$; $g_3 = 0,7$, а отже, імовірність виходу з ладу цих приладів становить відповідно $\bar{g}_1 = 0,1$; $\bar{g}_2 = 0,2$; $\bar{g}_3 = 0,3$. За теоремами додавання і множення ймовірностей обчислимо ймовірності, з якими величина X набуває можливих значень:

$$\begin{aligned} p_0 &= P\{X = 0\} = \bar{g}_1 \bar{g}_2 \bar{g}_3 = 0,006; \\ p_1 &= P\{X = 1\} = g_1 \bar{g}_2 \bar{g}_3 + \bar{g}_1 g_2 \bar{g}_3 + \bar{g}_1 \bar{g}_2 g_3 = 0,092; \\ p_2 &= P\{X = 2\} = g_1 g_2 \bar{g}_3 + g_1 \bar{g}_2 g_3 + \bar{g}_1 g_2 g_3 = 0,398; \\ p_3 &= P\{X = 3\} = g_1 g_2 g_3 = 0,504. \end{aligned}$$

Ряд розподілу величини X має вигляд:

X	0	1	2	3
-----	---	---	---	---

P	0,006	0,092	0,398	0,504
-----	-------	-------	-------	-------

Для нього виконується умова: $p_0 + p_1 + p_2 + p_3 = 1$.

За рядом розподілу та за означенням маємо функцію розподілу $F(x)$:

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 0,006, & 0 < x \leq 1, \\ 0,006 + 0,0092 = 0,098, & 1 < x \leq 2, \\ 0,006 + 0,0092 + 0,398 = 0,496, & 2 < x \leq 3, \\ 0,006 + 0,0092 + 0,398 + 0,504 = 1, & x > 3. \end{cases}$$

З властивостей функції розподілу отримаємо, що

$$P\{0,5 < X < 2\} = F(2) - F(0,5) = 0,098 - 0,006 = 0,092.$$

Задача 2. Маємо 4 заготовки для виготовлення деталей. Імовірність виготовлення придатної деталі дорівнює 0,75. Знайти закон розподілу випадкової величини X – кількість заготовок, що їх буде використано для виготовлення придатної деталі. Знайти математичне сподівання та дисперсію випадкової величини X , а також імовірність того, що із цих заготовок буде виготовлено стандартну деталь.

Розв'язання. Подамо закон розподілу для випадкової величини X у табличній формі. Очевидно, що випадкова величина може набувати значень 1, 2, 3, 4. Значення $X = 1$ буде тоді, коли з першої заготовки виготовлено стандартну деталь, а ймовірність цього дорівнює 0,75. Випадкова величина набуває значення 2, якщо з першої заготовки виготовлено браковану деталь, а з другої – придатну. За теоремою множення ймовірностей імовірність цієї події:

$$P(X = 2) = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}.$$

Аналогічно, $X = 3$, якщо деталі, виготовлені з першої та другої заготовок, браковані, а деталь, яку виготовлено з третьої заготовки – придатна:

$$P(X = 3) = \left(\frac{1}{4}\right)^2 \cdot \frac{3}{4} = \frac{3}{64}.$$

Нарешті, $X = 4$, якщо деталі, виготовлені з перших трьох заготовок, браковані:

$$P(X = 4) = \left(\frac{1}{4}\right)^3 = \frac{1}{64}.$$

Запишемо закон розподілу:

x_i	1	2	3	4
$p(x_i)$	0,75	0,1875	0,046875	0,015625

Легко перевірити, що сума ймовірностей у законі розподілу дорівнює 1. Знайдемо математичне сподівання та дисперсію випадкової величини за наведеними вище формулами.

$$\begin{aligned} MX &= 1 \cdot \frac{3}{4} + 2 \cdot \frac{3}{64} + 4 \cdot \frac{1}{64} = \frac{58}{64}, \\ MX^2 &= 1 \cdot \frac{3}{4} + 4 \cdot \frac{3}{16} + 9 \cdot \frac{1}{64} = \frac{105}{64}, \\ DX &= \frac{105}{64} - \left(\frac{58}{64}\right)^2 = \frac{3008}{4096}. \end{aligned}$$

Якщо подія A – «із чотирьох заготовок виготовлено одну придатну деталь», то

$$P(A) = 1 - \left(\frac{1}{4}\right)^4 = \frac{255}{256}.$$

Задача 3. Задано функцію

$$F(x) = \begin{cases} 0, & x \leq 1; \\ ax^2 + bx, & 1 < x \leq 4; \\ 1, & x > 4. \end{cases}$$

Довести, що можна дібрати такі значення a і b , при яких $F(x)$ буде функцією розподілу ймовірностей випадкової величини X . Знайти $P(2 \leq X \leq 3)$.

Розв'язання. Щоб знайти a і b скористаємося неперервністю функції розподілу в точках $x = 1$ і $x = 4$. Дістанемо систему рівнянь:

$$\begin{cases} a + b = 0; \\ 16a + 4b = 1. \end{cases} \begin{cases} a = -b; \\ 16b + 4b = 1. \end{cases} \begin{cases} b = -\frac{1}{12}; \\ a = \frac{1}{12}. \end{cases}$$

Отже,

$$F(x) = \frac{1}{12}x^2 - \frac{1}{12}x.$$

Доведемо, що на цьому проміжку функція монотонно зростає. Відшукуємо похідну функції: $F'(x) = \frac{1}{6}x - \frac{1}{12}$. Похідна дорівнює нулю при $x = \frac{1}{2}$. На проміжку $(0; 4)$ похідна функції $F(x)$ додатна, а отже, ця функція зростає. Отже, $F(x)$ задає закон розподілу випадкової величини X .

Обчислюємо ймовірність:

$$P(2 \leq X < 3) = F(3) - F(2) = \frac{1}{12} \cdot 9 - \frac{1}{12} \cdot 4 + \frac{1}{12} \cdot 2 = \frac{1}{3}.$$

Задача 4. Випадкова величина X задана законом розподілу:

X	2	3	10
P	0,1	0,4	0,5

Знайти середнє квадратичне відхилення.

Розв'язання. Математичне сподівання знайдемо

$$MX = 2 \cdot 0,1 + 3 \cdot 0,4 + 10 \cdot 0,5 = 6,4$$

Далі знайдемо математичне сподівання X^2 :

$$MX^2 = 2^2 \cdot 0,1 + 3^2 \cdot 0,4 + 10^2 \cdot 0,5 = 54,$$

$$DX = 54 - 6,4^2 = 13,04.$$

Середнє квадратичне відхилення обчислимо

$$\sigma(X) = \sqrt{DX} = \sqrt{13,04} = 3,61.$$

Задача 5. У грошовій лотереї випущено 100 білетів. Розігрується один виграш у 50 гривень і 10 виграшів по 1 гривні. Знайти закон розподілу, математичне сподівання і дисперсію дискретної випадкової величини X – вартості можливого виграшу для власника одного лотерейного білету.

Розв'язання. Випишемо можливі значення X :

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = 50.$$

Ймовірність цих можливих значень дорівнюють:

$$p_1 = \frac{89}{100}, \quad p_2 = \frac{10}{100} = 0,1, \quad p_3 = \frac{1}{100} = 0,01.$$

Закон розподілу приймає вигляд:

X	0	1	50
P	0,89	0,1	0,01

Перевіримо контрольну суму:

$$0,89 + 0,1 + 0,01 = 1$$

Математичне сподівання знайдемо

$$MX = x_1 p_1 + x_2 p_2 + x_3 p_3 = 50 \cdot 0,01 + 1 \cdot 0,1 + 0 \cdot 0,89 = 0,6,$$

а для знаходження дисперсії попередньо побудуємо закон розподілу величини X^2 :

X	0	1	2500
P	0,89	0,1	0,01

Далі знайдемо її математичне сподівання

$$MX^2 = 2500 \cdot 0,01 + 1 \cdot 0,1 + 0 \cdot 0,89 = 25,1.$$

Тоді

$$DX = 25,1 - 0,6^2 = 24,74.$$

Задачі для самостійного розв'язання

1. Дискретну випадкову величину X задано функцією розподілу:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 0,3, & 0 < x \leq 2 \\ 0,4, & 2 < x \leq 3 \\ 0,7, & 3 < x \leq 4 \\ 1, & x > 4 \end{cases}$$

Знайти закон розподілу випадкової величини X у табличній формі.

2. Випадкова величина X задається функцією розподілу:

$$F(x) = \begin{cases} 0, & x \leq 1 \\ a + b \cdot \operatorname{arctg} x, & x > 1 \end{cases}$$

Знайти коефіцієнти a, b та щільність $f(x)$.

3. Проводяться чотири незалежних постріли в однакових умовах по деякій цілі. Ймовірність влучення при одному пострілі дорівнює 0,25. Знайти закон розподілу для числа влучень у ціль.

4. Партія товару складається з 25 виробів, серед яких 6 бракованих. Навмання вибирають 3 вироби для перевірки їх якості. Знайти закон розподілу випадкової величини X – числа вибраних бракованих виробів.

5. Проводяться чотири незалежних постріли в однакових умовах по деякій мішені. Ймовірність влучення при одному пострілі дорівнює 0,25. Знайти закон розподілу для числа влучень у мішень.

Тема 6. Дискретні і неперервні випадкові величини. Основні характеристики. Властивості. Розподіли дискретних випадкових величин: біноміальний, Пуассона, геометричний.

1. Біноміальний закон розподілу

Біноміальний закон розподілу виникає у схемі Бернуллі, коли проводиться n незалежних випробувань, у кожному з яких випадкова подія A може або відбутися зі сталою ймовірністю p , або невідбутися з ймовірністю $q = 1 - p$, а випадкова величина X – кількість появ події A в цих n випробуваннях.

Дискретна величина X набуває цілих невід'ємних значень m ($m = 0, 1, \dots, n$) з ймовірностями, які обчислюються за формулою Бернуллі

$$P_n(m) = P(X = m) = C_n^m p^m (1 - p)^{n-m}.$$

Числові характеристики розподілу:

$$MX = np, \quad DX = np(1 - p).$$

2. Закон розподілу Пуассона

Розподіл Пуассона виникає в схемі Бернуллі, коли кількість n виконуваних незалежних випробувань велика, ймовірність p появи події A в кожному випробуванні мала, а випадкова величина X – кількість появ події A в цих n випробуваннях.

Дискретна випадкова величина має розподіл Пуассона, якщо вона набуває зліченної множини значень $m = 0, 1, \dots$ з ймовірностями

$$P_n(m) = P(X = m) = \frac{a^m}{m!} e^{-a}, \quad (a > 0).$$

Цей розподіл застосовується в задачах статистичного контролю якості, у теорії надійності, теорії масового обслуговування. Математичне сподівання і дисперсія в цьому розподілі однакові і дорівнюють a . Для цього розподілу складено таблиці щодо різних значень $a(0,1 - 20)$. У таблицях для відповідних значень a наведено ймовірності

$$P(X = m) \text{ і } P(X \geq m).$$

Якщо у схемі незалежних повторних випробувань n велике і p або $1 - p$ прямують до нуля, то біноміальний розподіл апроксимується розподілом Пуассона, коли $a = np$.

3. Геометричний розподіл

Геометричний розподіл виникає, коли незалежні випробування, у кожному з яких подія A може відбутися зі сталою ймовірністю p , проводяться до першого «невдалого» випробування (подія A не відбулась) і далі припиняються, а випадкова величина X – кількість проведених «вдалих» випробувань.

Закон подається формулою:

$$P(X = m) = p(1 - p)^{m-1}, m = 1, 2, \dots$$

Геометричний закон розподілу застосовується у задачах статистичного контролю якості і теорії надійності. Числові характеристики розподілу:

$$MX = \frac{1}{p}, \quad DX = \frac{1-p}{p^2}.$$

4. Гіпергеометричний розподіл

Гіпергеометричний розподіл описує ймовірність настання m успішних результатів у n випробуваннях, якщо значення n мале порівняно з обсягом сукупності N :

$$P(X = m) = \frac{C_k^n C_{N-k}^{n-m}}{C_N^n}, m = 0, 1, 2, \dots, n; k \leq n.$$

Наприклад, імовірність того, що з деталей, які випадково вибрано з партії обсягом N , m виявляться дефектними, має гіпергеометричний закон розподілу (k – кількість дефектних деталей у партії). Цей закон розподілу застосовується в задачах статистичного контролю якості та в суміжних галузях. Числові характеристики розподілу:

$$MX = \frac{kn}{N}, \quad DX = \frac{nk(N-k)(N-n)}{N^2(N-1)}.$$

Зі зменшенням відношення $\frac{n}{N}$ гіпергеометричний розподіл наближається до біноміального з параметрами $np = \frac{k}{N}$. Дуже часто гіпергеометричний розподіл апроксимується розподілом Пуассона, якщо $a = \frac{nk}{N}$.

5. Рівномірний закон розподілу

Рівномірний розподіл неперервної випадкової величини X виникає у випробуваннях типу кидання намання точки на відрізок $[a; b]$ (X – відстань точки від межі a відрізка), або у випробуваннях, пов'язаних з округленням вимірювань фізичних величин за допомогою приладів (X – похибка округлення).

Якщо ймовірність потрапляння випадкової величини на інтервал пропорційна до довжини інтервалу і не залежить від розташування інтервалу на осі, то вона має **рівномірний закон розподілу**.

Щільність рівномірного закону розподілу має вид:

$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x > b. \end{cases}$$

Функція розподілу рівномірного закону розподілу має вид:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

Математичне сподівання та дисперсія рівномірного закону розподілу дорівнюють

$$MX = \frac{a+b}{2}, DX = \frac{(b-a)^2}{12}.$$

Рівномірний закон розподілу легко моделювати. За допомогою функціональних перетворень із величин, розподілених рівномірно, можна діставати величини з довільним законом розподілу.

6. Показниковий закон розподілу

Показниковий, або **експоненціальний**, розподіл неперервної випадкової величини має широке застосування в теорії надійності технічного обладнання для характеристики терміну безвідказної роботи елементів та пристроїв і в теорії масового обслуговування для характеристики тривалості обслуговування.

Неперервна випадкова величина X розподілена за **показниковим законом** з параметром $a > 0$ якщо її щільність розподілу має вигляд:

$$f(x) = \begin{cases} 0, & x < 0, \\ ae^{-ax}, & x \geq 0. \end{cases}$$

Функція розподілу показникового закону розподілу задається формулою:

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

Показниковому розподілу задовольняють: час телефонної розмови, час ремонту техніки, час безвідмовної роботи комп'ютера.

Математичне сподівання та дисперсія величини X , розподіленої за показниковим законом розподілу можна знайти за формулами:

$$MX = \frac{1}{a}, DX = \frac{1}{a^2}.$$

Розв'язання типових задач

Задача 1. У цеху є 5 верстатів. Імовірність того, що верстат працює, дорівнює 0,8. Знайти ймовірність того, що працюватимуть не менш як 3 верстати.

Розв'язання. Імовірність того, що працює будь-який верстат, дорівнює 0,8. Тому справджується біноміальний закон розподілу:

$$P(X = m) = C_n^m p^m (1-p)^{n-m}, \quad m = 0, 1, 2, \dots, n$$

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5).$$

Зазначені ймовірності знайдемо за наведеною щойно формулою.

$$P(X \geq 3) = C_5^3 0,8^3 0,2^2 + C_5^4 0,8^4 0,2 + 0,8^5 =$$

$$= 0,2048 + 0,4096 + 0,32768 = 0,94208.$$

Задача 2. При виготовленні довільного виробу інструмент з імовірністю $p = 0,2$ може бути пошкодженим і потребуватиме заміни. Знайти математичне сподівання і дисперсію кількості виробів, які будуть виготовлені цим інструментом.

Розв'язання. Нехай випадкова величина X – кількість деталей, виготовлених до заміни цим інструментом. Ця випадкова величина може набувати значень $0, 1, 2, \dots$. Побудуємо закон розподілу цієї величини. Вона набуває значення, що дорівнює нулю, якщо при виготовленні першого виробу інструмент буде пошкоджено: $P(X = 0) = p = 0,2$. Якщо інструмент буде пошкоджено при виготовленні другого виробу, то $X = 1$; $P(X = 1) = p(1 - p)$. Аналогічно

$$P(X = 2) = p(1 - p)^2, P(X = 3) = p(1 - p)^3, \dots, \\ P(X = k) = p(1 - p)^k, \dots$$

Для обчислення математичного сподівання і дисперсії зіставимо здобутий закон розподілу з геометричним законом розподілу $P(Y = m) = p(1 - p)^{m-1}, m = 1, 2, \dots$. Очевидно, що $X = Y - 1$. Скориставшись властивостями математичного сподівання та дисперсії, дістанемо:

$$MX = M(Y - 1) = MY - 1 = \frac{1}{p} - 1 = 5 - 1 = 4,$$

$$DX = D(Y - 1) = DY = \frac{1-p}{p^2} = 20.$$

Задача 3. На автоматичну телефонну станцію надходить простий потік викликів, інтенсивність (щільність) якого $k = 0,8$ (викликів/хв.). Знайти ймовірність того, що за дві хвилини: а) не надійде жодного виклику; б) надійде рівно один виклик; в) надійде хоча б один виклик.

Розв'язання. Нехай X – випадкова величина (число викликів за 2 хвилини), яка розподілена за законом Пуассона з параметром $kt = 0,8 \cdot 2 = 1,6$. Маємо:

$$\text{а) } P_2(0) = P(X = 0) = \frac{(kt)^0}{0!} e^{-kt} = e^{-kt} = 0,2024;$$

$$\text{б) } P_2(1) = P(X = 1) = \frac{(1,6)^1}{1!} e^{-kt} = 0,323;$$

$$\text{в) } P(X \geq 1) = 1 - P(X = 0) = 1 - P_2(0) = 0,798.$$

Задача 4. Випадкова величина X розподілена рівномірно. Знайти щільність її розподілу, якщо $P(X \geq 3) = 0,4$, а математичне сподівання рівне $MX = 2$.

Розв'язання. Щільність рівномірного розподілу $f(x) = \frac{1}{b-a}$. Отже, потрібно визначити область зміни випадкової величини. Складаємо систему рівнянь:

$$\begin{cases} \int_3^b \frac{1}{b-a} dx = 0,4; \\ \int_a^b \frac{x}{b-a} dx = 2; \end{cases} \begin{cases} \frac{b-3}{b-a} = 0,4; \\ \frac{a+b}{2} = 2; \end{cases} \begin{cases} 0,6b + 0,4a = 3; \\ b = 4 - a; \end{cases} \begin{cases} b = 7, \\ a = -3. \end{cases}$$

Тому

$$f(x) = \begin{cases} 0, & x \leq -3; \\ 0,1 - 3 < x \leq 7; \\ 0, & x > 7. \end{cases}$$

Задачі для самостійного розв'язання

1. Брак при виготовленні деталей становить у середньому 5 %. Знайти ймовірність того, що серед 5 навмання взятих деталей: а) не виявиться жодної бракованої; б) будуть дві браковані деталі.

2. Робітник обслуговує 6 однотипних верстатів. Імовірність того, що верстат потрібно обслуговувати протягом часу T , дорівнює $\frac{1}{3}$.

Знайти ймовірність того, що за час T :

а) потрібно буде обслуговувати 3 верстати;

б) кількість вимог на обслуговування за час T буде від 2 до 5.

3. У деякій місцевості на кожні 100 кавунів припадає в середньому один, маса якого не менша за 10 кг. Знайти ймовірність того, що в партії із 400 кавунів буде:

- а) 3 кавуни масою не менш як 10 кг;
- б) не менш ніж 2 такі кавуни.

4. Технічними умовами передбачено, що довжина заготовки деякої деталі має бути між 24 і 25 см. Якщо довжина деталі розподіляється нормально з параметрами $a = 24,6$ см і $\sigma = 0,4$ см, то яка частка заготовок матиме довжину, що виходить за межі, задані технічними умовами?

5. Розмір деталі – нормально розподілена величина з $MX = 10$. Деталь вважається стандартною, якщо її відхилення від математичного сподівання не перевищує 0,8. Побудувати графік залежності між значеннями σ і часткою (у відсотках) бракованих деталей.

Нормальний закон розподілу

Нормальний (гауссівський) закон розподілу неперервної величини є одним із найчастіше застосовуваних на практиці розподілів. Йому підлягають похибки вимірювань різних фізичних величин, розміри або маса виробів, які сходять з поточної лінії, тощо.

Неперервна випадкова величина X називається **розподіленою за нормальним законом** з параметрами a і σ , якщо її щільність розподілу має вигляд

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

для всіх $x \in (-\infty; \infty)$.

Параметри a і σ , які входять до виразу щільності розподілу, є відповідно математичним сподіванням та середнім квадратичним відхиленням випадкової величини. Функція розподілу нормально розподіленої випадкової величини X має вигляд

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

Нормальний закон розподілу широко застосовується в математичній статистиці. Головна особливість нормального закону полягає в тому, що він є граничним законом, до якого наближаються інші закони розподілу за типових умов.

Для обчислення ймовірності потрапляння випадкової величини, розподіленої нормально, на проміжок використовується функція Лапласа:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt,$$
$$P(\alpha \leq X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Часто застосовується також формула

$$P(|X - a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Задача 1. Висунуто гіпотезу про те, що відхилення розміру деталі від номіналу є випадковою нормально розподіленою величиною з $MX = 0$ і $DX = 25$ мкм². Чи відповідає заданій гіпотезі те, що в перевірених 6 деталях відхилення належить проміжку [5,13]? Рівень значущості $\alpha = 0,0005$.

Розв'язання. Розглянемо подію A – «відхилення в 6 деталей належить проміжку [5,13]». Обчислимо ймовірність цієї події і зіставимо її з рівнем значущості α . Якщо ймовірність буде меншою за α , то результат випробування не відповідатиме висунутій гіпотезі. Ймовірність події A знайдемо за теоремою множення ймовірностей:

$$P(A) = p^6, \text{ де } p = P(5 \leq X < 13).$$

Обчислимо цю ймовірність:

$$P(5 \leq X < 13) = \Phi\left(\frac{13-0}{5}\right) - \Phi\left(\frac{5-0}{5}\right) = \Phi(2,6) - \Phi(1) = 0,4953 - 0,3413 = 0,154.$$

Тоді

$$P(A) = 0,154^6 < 0,2^6 = 0,000064.$$

Імовірність події A менша від рівня значущості. Отже, гіпотеза про закон розподілу не відповідає значенням випадкової величини у випробуваннях.

Задача 2. Похибка спостереження X при вимірюванні довжини розподілена нормально з $\alpha = 5$ мм і $\sigma = 4$ мм. Знайти ймовірність того, що виміряне значення відхилиться від істинного більш ніж на 10 мм.

Розв'язання. Згідно з умовою потрібно знайти $P(|X| \geq 10)$. Виразимо цю ймовірність через ймовірність протилежної події:

$$\begin{aligned} P(|X| \geq 10) &= 1 - P(|X| < 10) = 1 - P(-10 < X < 10) = \\ &= 1 - \Phi\left(\frac{10-5}{4}\right) + \Phi\left(\frac{-10-5}{4}\right) = 1 - \Phi(1,25) - \Phi(3,75) = \\ &= 1 - 0,3944 - 0,4999 = 0,1057. \end{aligned}$$

Задачі для самостійного розв'язання

1. За метеорологічних умов, що склалися в аеропорту, ймовірність своєчасного відправлення кожного рейсу дорівнює 0,6. Випадкова величина X – кількість своєчасно відправлених рейсів з трьох передбачених розкладом. Скласти ряд розподілу і математичне сподівання та дисперсію цієї випадкової величини.

2. Випадкова точка A підлягає нормальному розподілу у колі радіусом R . Знайти математичне сподівання $M\xi$ і дисперсію $D\xi$ випадкової величини ξ , яка є відстанню від точки A до центра кола.

3. Випадкова величина X підлягає нормальному закону розподілу з математичним сподіванням $MX = 30$ і середнім квадратичним відхиленням $\sigma_X = 10$. Знайти інтервал, в який з ймовірністю 0,95 потрапляє випадкова величина у результаті випробувань.

4. Проводять вимірювання діаметра деталі. Похибки вимірювання беруть як випадкову величину X , яка підлягає нормальному закону розподілу з параметрами: математичне сподівання $MX = 0$, середнє квадратичне відхилення $\sigma_X = 1$. Знайти ймовірність того, що вимірювання буде зроблено з похибкою, яка не перевищує за абсолютним значенням величини 2 мм

Тема 7. Функції випадкових величин

Нехай X – дискретна випадкова величина, яку задано табличним законом розподілу:

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Відомо, що $Y = \varphi(X)$, тоді закон розподілу Y має такий вигляд:

x_i	$\varphi(x_1)$	$\varphi(x_2)$	...	$\varphi(x_n)$
p_i	p_1	p_2	...	p_n

Числові характеристики функції можна знайти за її законом розподілу або за формулами:

$$MY = \sum_{i=1}^n \varphi(x_i) p_i; \quad DY = \sum_{i=1}^n (\varphi(x_i))^2 p_i - \left(\sum_{i=1}^n \varphi(x_i) p_i \right)^2.$$

Нехай X – неперервна випадкова, яку задано щільністю розподілу $f_1(x)$. Якщо $Y = \varphi(X)$ і φ — диференційована функція, монотонна в області значень X , то щільність розподілу цієї функції подається у вигляді

$$f(y) = f_1(\psi(y))|\psi'(y)|,$$

де ψ — функція, обернена до φ . Якщо φ — не монотонна функція в області зміни аргументу, то обернена функція неоднозначна і щільність розподілу $f(y)$ визначається як сума стількох доданків, скільки значень має обернена функція:

$$f(y) = \sum_{i=1}^n f_1(\psi_i(y))|\psi'_i(y)|,$$

де $\psi_i(y)$ — функції, обернені до φ .

Розв'язання типових задач

Задача 1. Робітник обслуговує 4 верстати, які розміщено в одну лінію на відстані a один від одного. Знайти закон розподілу для випадкової величини Z – відстані, яку проходить робітник між двома обслуговуваннями, вважаючи, що події – «робітник перебуває біля будь-якого з верстатів» і «довільний верстат потребує обслуговування» – рівноможливі.

Розв'язання. Розглянемо випадкові величини: X – номер верстата, біля якого перебуває робітник; Y – номер верстата, який потребує обслуговування. Якщо верстати перенумерувати у тій послідовності, в якій їх розміщено (1, 2, 3, 4), то закони розподілу X і Y наберуть такого вигляду:

x_i	1	2	3	4
p_i	0,25	0,25	0,25	0,25

y_i	1	2	3	4
p_i	0,25	0,25	0,25	0,25

При цьому $Z = a|X - Y|$. Складаючи закон розподілу Z , обчислимо значення функції для всіх можливих комбінацій значень X і Y . Таких комбінацій буде 16. Імовірності для всіх цих комбінацій однакові і дорівнюють $\frac{1}{16}$ (за теоремою множення ймовірностей):

z_i	0	a	$2a$	$3a$	a	0	a	$2a$	$2a$	a	0	a	$3a$	$2a$	a	0
p_i	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

У таблиці значення z_i повторюються. Складемо нову таблицю, в якій такі значення запишемо один раз, а їхні ймовірності додамо.

z_i	0	a	$2a$	$3a$
p_i	0,25	0,375	0,25	0,125

Задача 2. Випадкову величину X задано щільністю розподілу:

$$f_1(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{2}, & 0 < x \leq 1 \\ \frac{1}{3}x, & 1 < x \leq 2 \\ 0, & x > 2 \end{cases}$$

Знайти щільність розподілу $f(y)$, якщо $Y = X^2$.

Розв'язання. Дослідимо задану функцію на монотонність: $y' = 2x$, $2x = 0, x = 0$. Отже, критична точка міститься на мережі зміни X . Функція монотонно зростає в області зміни аргументу x . Тому можна застосувати формулу:

$$f(y) = f_1(\psi(y))|\psi'(y)|, \text{ де } x = \psi(y) = \sqrt{y}, \psi'(y) = \frac{1}{2\sqrt{y}}.$$

Щільність розподілу для X має два відмінні від нуля аналітичні вирази. Стільки ж виразів матиме і щільність розподілу Y .

Якщо $x \in (0; 1]$, $y \in (0; 1]$, то $f_1(x) = \frac{1}{2}$, $f(y) = \frac{1}{4\sqrt{y}}$. Якщо $x \in (0; 2]$, $y \in (1; 4]$, то $f_1(x) = \frac{1}{3}x$, $f(y) = \frac{1}{6}$. Звідси маємо:

$$f(y) = \begin{cases} 0, & y \leq 0; \\ \frac{1}{4\sqrt{y}}, & 0 < y \leq 1; \\ \frac{1}{6}, & 1 < y \leq 4; \\ 0, & y > 4. \end{cases}$$

Задача 3. Швидкість обробки – випадкова величина X з напівнормальним законом розподілу:

$$f_1(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad x > 0.$$

Знайти закон розподілу випадкової величини Y – часу, потрібного для обробки деталі, якщо сумарний час її обробки дорівнює A .

Розв'язання. З умови задачі випливає, що $Y = \frac{A}{X}$. Ця функція монотонна при додатних значеннях X . Скориставшись тією самою формулою, яка застосовувалась у попередніх прикладах, дістанемо:

$$f(y) = f_1(\psi(y))|\psi'(y)|, \quad x = \psi(y) = \frac{A}{y}, \quad \psi'(y) = -\frac{A}{y^2};$$

$$f(y) = \frac{2}{\sqrt{2\pi}\sigma} e^{-\frac{A^2}{2y^2\sigma^2}} \cdot \frac{A}{y^2} = \frac{2A}{\sigma\sqrt{2\pi}y^2} e^{-\frac{A^2}{2y^2\sigma^2}}.$$

Оскільки область зміни Y – множина додатних чисел, то маємо:

$$f(y) = \begin{cases} \frac{2A}{\sigma\sqrt{2\pi}y^2} e^{-\frac{A^2}{2y^2\sigma^2}}, & y > 0; \\ 0, & y \leq 0. \end{cases}$$

Задачі для самостійного розв'язання

1. Кількість відказів телевізора протягом гарантійного строку розподілена за законом Пуассона за $\lambda = 0,5$. У разі i -го відказу витрати на ремонт $y_i = (i^2 + 2i)c$. Знайти математичне сподівання і дисперсію витрат упродовж гарантійного строку.

2. Закон розподілу похибок при вимірюванні радіуса R кола – нормальний з параметрами $a = 100$, $\sigma = 0,25$. Знайти закон розподілу і числові характеристики похибок при обчисленні довжини кола та площі круга.

3. Залишок матеріалу A на початок місяця становив 300 одиниць. Витрати матеріалу за день роботи – випадкова величина, яка рівномірно розподілена на проміжку $(10; 15]$. Знайти закон розподілу та математичне сподівання:

а) для тривалості періоду Y на який вистачить матеріалу;

б) для залишку матеріалу Z після 20 днів роботи.

4. Урожайність зернових, ц/га випадкова величина, рівномірно розподілена на проміжку $(15; 45]$. Знайти закон розподілу та математичне сподівання випадкової величини Y – собівартості виробництва 1 ц зерна, якщо витрати з виробництва зерна на 1 га становлять b грн.

5. Нехай маса виловленої риби є неперервною нормально розподіленою випадковою величиною X з параметрами $MX = 375$ г, $DX = 25$ г. Знайти ймовірність того, що маса однієї виловленої риби буде знаходитись у межах від 300 до 425 г.

6. Закон розподілу випадкової величини X заданий функцією щільності ймовірності $p(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x-2)^2}{18}}$. Знайти:

а) числові характеристики випадкової величини MX і DX ;

б) функцію розподілу;

в) інтервал, в який з імовірністю 0,96 потрапляє випадкова величина у результаті випробувань.

Тема 9. Граничні теореми теорії ймовірностей

Доведено, що при деяких відносно широких умовах сукупна поведінка достатньо великої кількості випадкових величин втрачає випадковий характер і приводить до результату, що майже не має елементів випадковості і дозволяє передбачити їх поведінку. Для практики, зокрема для задач математичної статистики, важливо знати умови, за якими це відбувається. Ці умови вказуються у відповідних граничних теоремах, які об'єднані загальною назвою *закон великих чисел*.

Основою закону великих чисел є нерівність Чебишова. Вона є частковим випадком більш загальної **нерівності**, яка носить ім'я іншого вченого – **Маркова**:

$$P(|X| \geq \varepsilon) \leq \frac{M(|X|)}{\varepsilon}.$$

Суть цієї нерівності полягає у тому, що для $\forall \varepsilon > 0$ і будь-якої випадкової величини X , яка має скінченне математичне сподівання, ймовірність події $\{\omega: |X| \geq \varepsilon\}$ не перевищує добутку частки $1/\varepsilon$ на математичне сподівання модуля випадкової величини $M|X|$.

отримуємо іншу нерівність

$$P(|X| < \varepsilon) \geq 1 - \frac{M(|X|)}{\varepsilon},$$

яка є протилежною до неї.

Використовуючи нерівність вище, можна довести **нерівність Чебишова**: Якщо випадкова величина X має скінченну дисперсію, то для $\forall \varepsilon > 0$

$$P(|X - MX| < \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2}.$$

З останньої нерівності одержимо $\forall \varepsilon > 0$

$$P(|X - MX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}.$$

На підставі нерівності вище також можна знайти ймовірність того, що відхилення випадкової величини X від її математичного сподівання не перевищить трьох своїх середніх квадратичних відхилень ($\varepsilon = 3\sigma$):

$$P(|X - MX| < 3\sigma) \geq 1 - \frac{\sigma^2}{(3\sigma)^2} = \frac{8}{9} \cong 0,889.$$

Аналогічно

$$P(|X - MX| \geq 3\sigma) \leq \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9} \cong 0,111.$$

Тобто, ймовірність відхилення будь-якої випадкової величини X від свого математичного сподівання на величину, що перевищує 3σ , не більше 1/9. Згадаємо, що для нормального закону ця ймовірність дорівнює 0,0027 (див. розділ 7).

Крім формул (5), (6), з нерівності Чебишова випливають такі **наслідки**:

1. Якщо $X_1, X_2, \dots, X_n, \dots$ – послідовність попарно незалежних випадкових величин, що мають скінченну дисперсію $\sigma^2 = DX_i$ і математичне сподівання $a = MX_i, i = 1, 2, \dots$, то для $\forall \varepsilon > 0$

$$P\left|\left|\frac{1}{n} \sum_{i=1}^n X_i - a\right| < \varepsilon\right| \geq 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

2. Якщо $X_1, X_2, \dots, X_n, \dots$ – послідовність попарно незалежних випадкових величин, причому їхні дисперсії обмежені згори числом C і $\sigma_i^2 = DX_i, a_i = MX_i, i = 1, 2, \dots$, то $\forall \varepsilon > 0$

$$P\left|\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n a_i\right| < \varepsilon\right| \geq 1 - \frac{C}{n\varepsilon^2}.$$

3. Нехай $\frac{m}{n}$ – відносна частота появи події A в n незалежних випробуваннях, p – ймовірність появи події A в кожному з цих випробувань. Тоді для $\forall \varepsilon > 0$

$$P\left|\left|\frac{m}{n} - p\right| < \varepsilon\right| \geq 1 - \frac{pq}{n\varepsilon^2}.$$

Теорема Чебишова (закон великих чисел). Якщо випадкові величини в послідовності $X_1, X_2, \dots, X_n, \dots$ попарно незалежні, а їхні дисперсії обмежені згори одним і тим же числом C , тобто не перевищують постійного числа C

$$DX_i \leq C,$$

то $\forall \varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n MX_i \right| < \varepsilon \right] = 1.$$

Звідси видно, що, якщо дисперсії попарно незалежних випадкових величин $X_1, X_2, \dots, X_n, \dots$ рівномірно обмежені, то $\forall \varepsilon > 0$ справедлива дана рівність

Тобто, середнє арифметичне досить великої кількості незалежних випадкових величин, дисперсії яких рівномірно обмежені, втрачає характер випадкової величини і з великою ймовірністю приймає значення близьке до середнього арифметичного їх математичних сподівань.

На практиці досить часто використовується **наслідок** з цієї теореми (**закон великих чисел для однаково розподілених випадкових величин, що мають однакову дисперсію**): Якщо $X_1, X_2, \dots, X_n, \dots$ послідовність попарно незалежних однаково розподілених випадкових величин з математичним сподіванням a і дисперсією σ^2 , то для будь-якого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - a \right| < \varepsilon \right] = 1.$$

Теорема Чебишова є підґрунтям вибіркового методу, який широко використовується у статистиці. Суть його полягає у тому, що за відносно невеликою випадковою вибіркою судять про всю (генеральну) сукупність досліджуваних об'єктів.

З рівності не випливає, що $\lim_{n \rightarrow \infty} f(n) = \lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = a$. В теоремі лише говориться про те, що при досить великій кількості спроб середнє арифметичне випадкових величин втрачає випадковий характер і буде як завгодно мало відрізнятися від їхнього математичного сподівання a . В цьому випадку кажуть, що $f(n)$ прямує до a за ймовірністю і записують так:

$$f(n) = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\text{ймовірністю}} a$$

Зрозуміло, що збіжність за ймовірністю відрізняються від збіжності в розумінні математичного аналізу. Адже, якщо $f(n) \xrightarrow[n \rightarrow \infty]{} a$ в розумінні математичного аналізу, то починаючи з деякого N , для всіх наступних n обов'язково виконується нерівність

$$|f(n) - a| < \varepsilon.$$

В той же час, якщо $f(n)$ при $n \rightarrow \infty$ прямує за ймовірністю до a , то для окремих значень n нерівність може не виконуватися.

Закон великих чисел для кількості успіхів у випробуваннях Бернуллі називають **теоремою Бернуллі**: Відносна частота успіхів в n незалежних випробуваннях Бернуллі при $n \rightarrow \infty$ збігається за ймовірністю до ймовірності успіху в одному випробуванні, тобто для будь-якого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1,$$

де m – кількість успіхів в n випробуваннях Бернуллі, p – ймовірність успіху в одному випробуванні.

Теорема Бернуллі є теоретичним підґрунтям статистичного означення ймовірності події.

Оскільки з теореми слідує, що для $\forall \varepsilon > 0$ нерівність $\left|\frac{m}{n} - p\right| < \varepsilon$ виконується для великих n з ймовірністю, близькою до одиниці, то можна вважати, що при великих n $p \cong \frac{m}{n}$.

Центральна гранична теорема (у формулюванні Ляпунова): Якщо $X_1, X_2, \dots, X_n, \dots$ – незалежні однаково розподілені випадкові величини, що мають скінченну дисперсію $\sigma^2 = DX_i$ і математичне сподівання $a = MX_i, i = 1, 2, \dots$, то для довільного дійсного x має місце рівність

$$\lim_{n \rightarrow \infty} P\left\{\frac{\frac{1}{n} \sum_{i=1}^n X_i - a}{\frac{\sigma}{\sqrt{n}}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Тобто, теорема Ляпунова твердить, що розподіл стандартизованого середнього

$$Y_n = \frac{\frac{1}{n} \sum_{i=1}^n X_i - MX}{\frac{\sigma_X}{\sqrt{n}}}$$

арифметичного n випадкових величин прямує при досить великих n до нормального нормованого розподілу. Вирішення питання, при яких значеннях n використовувати нормальне наближення залежить від потрібної точності обчислення ймовірності. Переважно центральну граничну теорему використовують, якщо $n > 10$.

На практиці збіжність розподілу ймовірностей суми випадкових величин до нормального закону виконується доволі часто. Цим можна пояснити те, що нормальний розподіл часто зустрічається в практичних задачах. Зокрема в економіці, деякі показники також являють собою суму великої кількості доданків, кожний з яких має малий внесок в сумарний показник. В цьому випадку досліджуваний показник має приблизно нормальний розподіл.

Слід зауважити, що локальна та інтегральна теореми Муавра-Лапласа є наслідками центральної граничної теореми.

Математична статистика.

Тема 10. Елементи математичної статистики. Вибірковий метод

10.1. Задачі математичної статистики

Математична статистика вивчає методи, які дають змогу за результатами випробувань роботи певні ймовірнісні висновки.

Нехай подія A має ймовірність, але її значення $p = P(A)$ нам невідоме; необхідно оцінити p за допомогою випробувань. Наприклад, треба оцінити число бракованих виробів в достатньо великій партії з N виробів. Якщо партія містить M бракованих виробів, то $\frac{M}{N}$ — це ймовірність того, що навмання взятий виріб є бракованим. Зрозуміло, що задача оцінки числа бракованих виробів зводиться до задачі оцінки невідомої ймовірності.

Є цілий ряд важливих задач, пов'язаних з невідомими функціями розподілу. Так може статись, що нам нічого невідомо про функцію розподілу $F(x)$ випадкової величини X . Тоді може бути поставлене питання про наближене подання функції $F(x)$ за допомогою даних, здобутих в результаті випробувань.

Проте може бути і так, що нам відомо аналітичний вигляд функції розподілу $F(x_1, \alpha_1, \dots, \alpha_n)$, але невідомо значення параметрів $\alpha_1, \dots, \alpha_n$, від яких залежить ця функція; треба знайти оцінки цих параметрів.

Важливою задачею є перевірка статистичних гіпотез про вид невідомого розподілу або про величину параметрів розподілу, вид якого відомий.

Отже, задача математичної статистики полягає у створенні методів збору в обробки статистичних даних для одержання наукових і практичних висновків.

10.2. Вибірка з генеральної сукупності

Генеральною сукупністю в математичній статистиці називається множина однотипних об'єктів, кількісна чи якісна ознака яких підлягає вивченню. Підмножина об'єктів, дібраних у відповідний спосіб із генеральної сукупності, називається **вибірковою сукупністю**. Вважаємо, що ознака, яка вивчається, є випадковою величиною X із функцією розподілу $F(x)$. Результати вибірки розглядатимемо як послідовність незалежних однаково розподілених випадкових величин $X_1, X_2, \dots, X_n$. Закон розподілу для всіх X_i визначається функцією $F(x)$. Результати вибірки — реалізації випадкових величин — позначатимемо відповідно через $x_1, \dots, x_n$. Розмістивши ці числа в порядку зростання і записавши частоти n_i , з якими зустрічаються ці значення, дістанемо варіаційний, або статистичний, ряд:

x_i	x_1	x_2	$\dots$	x_k
Частоти	n_1	n_2	$\dots$	n_k

На підставі такого ряду можна побудувати статистичну функцію розподілу $F_n^*(x) = \sum_{x_i < x} \frac{n(x_i)}{n}$. Якщо $n \rightarrow \infty$, то статистична функція розподілу збігається до теоретичної функції розподілу.

Статистичний ряд графічно подається полігоном розподілу. Щоб побудувати його, на осі абсцис відкладають значення реалізацій, а на осі ординат — відповідні їм частоти (відносні частоти). Здобуті точки сполучають відрізками прямих.

У разі, коли X — неперервна величина і обсяг вибірки великий, результати вибірки подають інтервальним рядом. Для цього область реалізацій розбивають на k інтервалів і для кожного інтервалу визначають частоти. Здобутий ряд геометрично подається гістограмою. Для побудови її на осі абсцис відкладають інтервали, а на них як на основах будують прямокутники, висота яких пропорційна до частоти (відносної частоти) інтервалу. Гістограма дає певне уявлення про графік щільності розподілу.

10.3. Обчислення вибіркових характеристик

Для вибіркової сукупності обчислюють числові характеристики — вибіркові випадкові функції: вибірку середню $\bar{x}$, вибірку дисперсію s^2 , статистичні моменти розподілу тощо. Реалізації цих вибіркових функцій знаходять за формулами, вигляд яких залежить від того, в якій формі подано вибіркові дані. Якщо вибіркові дані не згруповано, то

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Якщо вибіркові дані зведено у статистичний ряд, то:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i n_i,$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 n_i$$

Якщо дані подаються інтервальним рядом, то перехід до статистичного ряду виконують, обчислюючи для кожного інтервалу його середину.

Якщо число варіант досить велике, то для спрощення обчислень застосовують групування емпіричних даних. Розіб'ємо інтервал, в якому лежать вибіркові дані, на деяке число s рівних частин довжини h . Якщо x_i — середина i -ї частини, n_i — число варіант, що потрапили в i -у частину, то $\bar{x}$ і s^2 обчислюються за формулами вище.

Для дальшого спрощення обчислень застосовують так званий метод умовних варіант. Умовні варіанти визначаються рівністю:

$$u_i = \frac{x_i - c}{h},$$

де c — умовний нуль (новий початок відліку).

Найбільша простота обчислень досягається тоді, коли за умовний нуль взяти варіанту, яка розміщена приблизно посередині варіаційного ряду і, яка має найбільшу частоту (хоча це і не обов'язково). Тоді:

$$\bar{x} = h\bar{u} + c$$

$$s^2 = h^2(\overline{u^2} - (\bar{u})^2)$$

Приклад 10.1. Із генеральної сукупності зроблено вибірку обсягом $n = 32$. Здобуто такі реалізації випадкової величини: 2,2; 7,1; 6,3; 3,9; 5,9; 5,6; 5,6; 4,7; 7,9; 3,2; 6,1; 5,5; 6,4; 6,0; 6,9; 4,7; 6,4; 6,9; 6,7; 7,9; 4,2; 6,7; 6,0; 9,2; 5,5; 6,5; 3,5; 4,9; 7,2; 4,9; 8,9; 5,7. Скласти інтервальний ряд і побудувати гістограму. За допомогою умовних моментів розподілу знайти $\bar{x}$, s^2 .

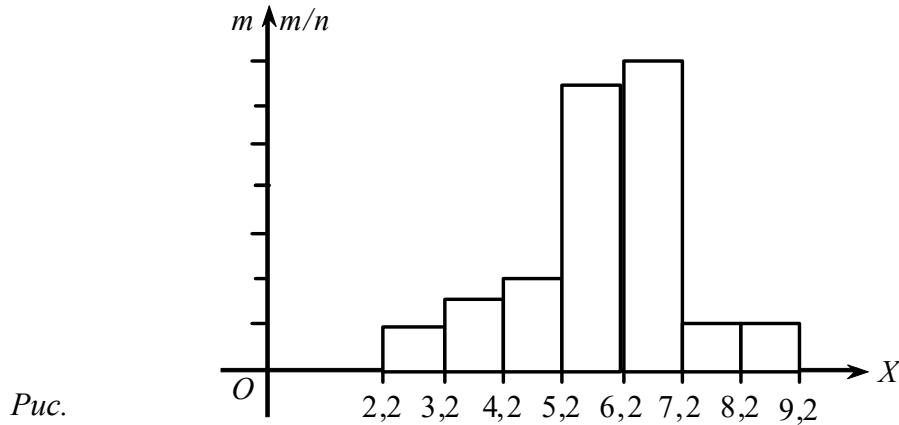
Розв'язання. Для побудови інтервального ряду розбиваємо область реалізацій на 7 інтервалів з однаковими довжинами інтервалів:

$$\Delta x = \frac{\max(x_i) - \min(x_i)}{7}; \Delta x = \frac{9,2 - 2,2}{7} = 1.$$

Частоти кожного інтервалу знайдемо, визначивши для кожного значення інтервал. Якщо значення x_i потрапляє на межу, то збільшуємо на 1 частоту нижнього інтервалу.

Інтервал	2,2—3,2	3,2—4,2	4,2—5,2	5,2—6,2	6,2—7,2	7,2—8,2	8,2—9,2
Частота	2	3	4	9	10	2	2

Згідно зі знайденим рядом будуюмо гістограму (рис.).



На підставі побудованої гістограми можна висунути гіпотезу про нормальний закон розподілу в сукупності.

Для обчислення умовних моментів розподілу складемо таблицю, в якій запишемо середини інтервалів x_i , їхні частоти n_i і нові змінні u_i . Візьмемо c , що дорівнює $x_6 = 6,7$. У наступних стовпцях обчислені значення $n_i u_i$, $n_i u_i^2$, а в останньому рядку таблиці — їхні суми.

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$
2,7	2	-4	-8	32
3,7	3	-3	-9	27
4,7	4	-2	-8	16
5,7	9	-1	-9	9
6,7	10	0	0	0
7,7	2	1	2	2
8,7	2	2	4	8
Сума	32	—	-28	94

$$c = 6,7; h = 1;$$

$$\bar{u} = \frac{1}{n} \sum_{i=1}^k n_i u_i = \frac{-28}{32} = 0,875; \quad \overline{u^2} = \frac{1}{n} \sum_{i=1}^k n_i u_i^2 = \frac{94}{32} = 2,9;$$

$$s^2 = h^2 (\overline{u^2} - (\bar{u})^2) = 2,9 - 0,875^2 = 2,025;$$

$$\bar{x} = h \bar{u} + c = 0,875 + 6,7 = 7,575.$$

Тема 11. Статистичні оцінки параметрів генеральної сукупності. Статистичні гіпотези

Точкові оцінювання невідомих параметрів розподілу.

11.1. Класифікація статистичних оцінок

Оцінка параметра розподілу сукупності θ у загальному випадку є випадковою величиною, яка визначається за даними вибірки і використовується замість невідомого значення параметра, який потрібно оцінити.

При точковому оцінюванні параметра необхідно знайти функцію $h = h(X_1, X_2, \dots, X_n)$, значення якої при заданій реалізації $x_1, x_2, \dots, x_n$ вибірки $X_1, X_2, \dots, X_n$ приймають за наближене значення параметра θ_0 . Для того, щоб порівнювати різні оцінки і вибрати кращу, існує така їх класифікація.

Оцінка $\hat{\theta} = h(X_1, X_2, \dots, X_n)$ параметра θ називається **незміщеною**, якщо $M\hat{\theta} = \theta$ її математичне сподівання збігається зі значенням параметра. Якщо $M\hat{\theta} = \theta + b(\theta)$, то $b(\theta)$ називають **зсувом** оцінки $\hat{\theta}$.

Послідовність оцінок $\hat{\theta}_n = h(X_1, X_2, \dots, X_n), n = 1, 2, \dots$ параметра θ називається **спроможною**, якщо $\hat{\theta}_n \xrightarrow{p} \theta, n \rightarrow \infty$.

Послідовність оцінок $\hat{\theta}_n = h(X_1, X_2, \dots, X_n), n = 1, 2, \dots$ параметра θ називається **сильно спроможною**, якщо $\hat{\theta}_n \rightarrow \theta$ з імовірністю одиниця при $n \rightarrow \infty$.

Найчастіше застосовується критерій, який полягає у виборі оцінки, що має найменшу можливу дисперсію при даному об'єму вибірки. Така оцінка називається **ефективною**.

Незміщеною сильно спроможною оцінкою математичного сподівання a випадкової величини X є вибіркове середнє: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Реалізацію цієї оцінки позначають :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Вибіркове середнє для дискретного статистичного ряду знаходиться за формулою:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n n_i x_i,$$

де $\sum_{i=1}^k n_i = n$. Вибіркове середнє для інтервального статистичного ряду знаходиться за формулою:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s n_i z_i, (14.3)$$

де z_i – середина i -го інтервалу, $\sum_{i=1}^s n_i = n$.

Незміщеність оцінки $\bar{X}$ математичного сподівання a випадкової величини X випливає з того, що кожне X_i ($i = 1, \dots, n$) є копією X , тому має математичне сподівання a і дисперсію σ^2 . Отже, $M(\bar{X}) = \frac{1}{n} \sum_{i=1}^n M(X_i) = \frac{1}{n} na = a$.

Оскільки для довільного $nM(\bar{X}) = a$, то оцінка $\bar{X}$ математичного сподівання a також сильно спроможна.

Оцінкою дисперсії σ^2 випадкової величини X є вибіркова дисперсія:

$$\bar{s}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Вибіркова дисперсія $\bar{s}^2$ є зміщеною оцінкою для σ^2 . Можна показати, що:

$$M(\bar{s}^2) = \frac{n-1}{n} \sigma^2,$$

тобто $M(\bar{s}^2) \neq \sigma^2$. Нехай $s^2 = \frac{n}{n-1} \bar{s}^2$, тоді $M(s^2) = \frac{n}{n-1} M(\bar{s}^2) = \sigma^2$. Отже,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

є незміщеною оцінкою дисперсії σ^2 випадкової величини X .

Якщо математичне сподівання a відоме, то незміщеною сильно спроможною оцінкою дисперсії σ^2 є оцінка: $s_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2$.

Вибіркову дисперсію для статистичного ряду знаходять за формулою:

$$\bar{s}^2 = \frac{1}{n} \sum_{i=1}^k (n_i x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2 \quad (14.5)$$

Відповідно $s^2 = \frac{1}{n-1} \sum_{i=1}^k (n_i x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2$.

Вибіркову дисперсію для інтервального ряду знаходять за формулою:

$$\bar{s}^2 = \frac{1}{n} \sum_{i=1}^s (n_i z_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^s n_i z_i^2 - \bar{x}^2$$

відповідно

$$s^2 = \frac{1}{n-1} \sum_{i=1}^s (n_i z_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^s n_i z_i^2 - \bar{x}^2$$

Величина $\bar{s} = \sqrt{\bar{s}^2}$ називається **вибірковим середнім квадратичним відхиленням**, $s = \sqrt{s^2}$ – **вибірковим виправленим середнім квадратичним відхиленням** (s^2 – вибірка виправлена дисперсія).

11.2. Означення медіани та моди

Медіаною M_e називають варіанту, яка ділить варіаційний ряд на дві частини, рівні по числу варіант. Для дискретних статистичних рядів:

$$M_e = \begin{cases} x_m, & \text{при } n = 2m - 1, \\ \frac{x_m + x_{m+1}}{2}, & \text{при } n = 2m. \end{cases}$$

Для інтервальних статистичних рядів:

$$M_e = x_i + h_i \frac{\frac{n}{2} - \sum_{j=1}^{i-1} n_j}{n_i},$$

де x_i – початок медіанного інтервалу, тобто такого, якому відповідає перша з нагромаджених частот, що перевищує половину всіх спостережень, h_i – довжина i -го інтервалу, n_i – частота медіанного інтервалу.

Мода – це елемент, який найчастіше трапляється у вибірці. Для дискретних статистичних рядів: $M_0 = x_j$, якщо $n_j = \max n_i$.

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 з найбільшою частотою, n_i – частота i -го інтервалу.

Приклад. Обчислити незміщені оцінки математичного сподівання та дисперсії, вибірку медіану та моду вибірки:

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

Розв'язання. Об'єм вибірки $n = 5 + 15 + 50 + 16 + 10 + 4 = 100$. Вибіркове середнє:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n n_i x_i = \frac{1}{100} (5 \cdot 12 + 15 \cdot 14 + 50 \cdot 16 + 16 \cdot 18 + 10 \cdot 20 + 4 \cdot 22) = 16,46;$$

$$\frac{1}{n} \sum_{i=1}^n n_i x_i^2 = \frac{1}{100} (5 \cdot 12^2 + 15 \cdot 14^2 + 50 \cdot 16^2 + 16 \cdot 18^2 + 10 \cdot 20^2 + 4 \cdot 22^2) = 275,8;$$

$$\bar{s}^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2 = 275,8 - 16,46^2 = 4,8684;$$

$$s^2 = \frac{n}{n-1} \bar{s}^2 = \frac{100}{99} \cdot 4,8684 = 4,92; \quad M_e = \frac{16 + 16}{2} = 16; \quad M_0 = 16;$$

Тема 12. Елементи дисперсійного аналізу

У математичній статистиці (а це основний математичний апарат обробки результатів моделювання) використовується поняття гіпотези.

Гіпотезою називається припущення:

- • про закони розподілу ймовірностей випадкових величин;
- • значеннях характеристик випадкових величин;
- • збігу законів розподілу двох і більше випадкових величин і ін.

Зазвичай вихідну гіпотезу називають нульовою і позначають H_0 . Протилежне твердження називають конкуруючою гіпотезою і позначають H_1 .

Гіпотеза піддається перевірці. Сенс цієї перевірки в тому, щоб прийняти або відхилити її з допустимим мінімальним ризиком. При цьому можливі помилки:

- • *забракувати* перевіряється гіпотезу, якщо вона вірна, що відповідає так званій *помилці першого роду*;
- • *прийняти* перевіряється гіпотезу, коли вона не вірна, значить, зробити *помилку другого роду*.

Правило, за яким приймається судження про істинність або хибність основної гіпотези H_0 , називають *критерієм перевірки*, або *критерієм згоди*.

У практиці моделювання та обробки експериментальних даних дуже часто необхідно вирішувати проблему підтвердження або спростування гіпотези про приналежність двох або більше вибірок однієї генеральної сукупності.

До цієї проблеми призводять такі завдання:

- • порівняльна оцінка різних технологічних процесів по їх продуктивності, точності, економічності;
- • порівняння конструктивних особливостей приладів, машин, засобів озброєння і ін.

Ознаки, за якими проводиться порівняльна оцінка, часто не є детермінованими, мають розсіюванням. Наприклад, точність ніколи не може бути абсолютною, так як вимірювальні прилади завжди несуть в собі помилку.

Найбільш загальним і часто вживаним на практиці методом порівняння якостей об'єктів є дисперсійний аналіз.

Сутність *дисперсійного аналізу* полягає в перевірці гіпотези про тотожність вибірових дисперсій одній і тій же генеральній дисперсії.

Чому дослідників цікавить порівняння саме дисперсій, а не будь-яких інших характеристик? Зауважимо, що є методики порівняння, наприклад, математичних очікувань і ін., Але вони не володіють такою спільністю, як дисперсійний аналіз.

А справа в тому, що дисперсія характеризує важливі конструкторські та технологічні показники, такі як:

- • точність приладів;

- розсіювання точок влучення при стрільбі і ін.

І ще дисперсійний аналіз одночасно вирішує проблему перевірки гіпотези про рівність середніх значень вибірок.

Завдання порівняння дисперсій зводиться до перевірки вихідної гіпотези (нульової гіпотези H_0) про приналежність двох вибірок однієї і тієї ж генеральної сукупності.

Для перевірки гіпотези про рівність дисперсій потрібно мати незалежну функцію, яка обчислюється за даними експерименту.

Такою функцією є функція Фішера (розподіл Фішера, F-розподіл), що визначається так:

$$F = \frac{\frac{U}{k_1}}{\frac{V}{k_2}},$$

де U і V - випадкові величини, що мають розподіл χ^2 ; k_1 і k_2 - ступеня свободи випадкових величин U і V відповідно, $k_1 = N_1 - 1$, $k_2 = N_2 - 1$, $i=1, \dots, N_1$ і $j=1, \dots, N_2$ - кількість випробувань (обсяги вибірок).

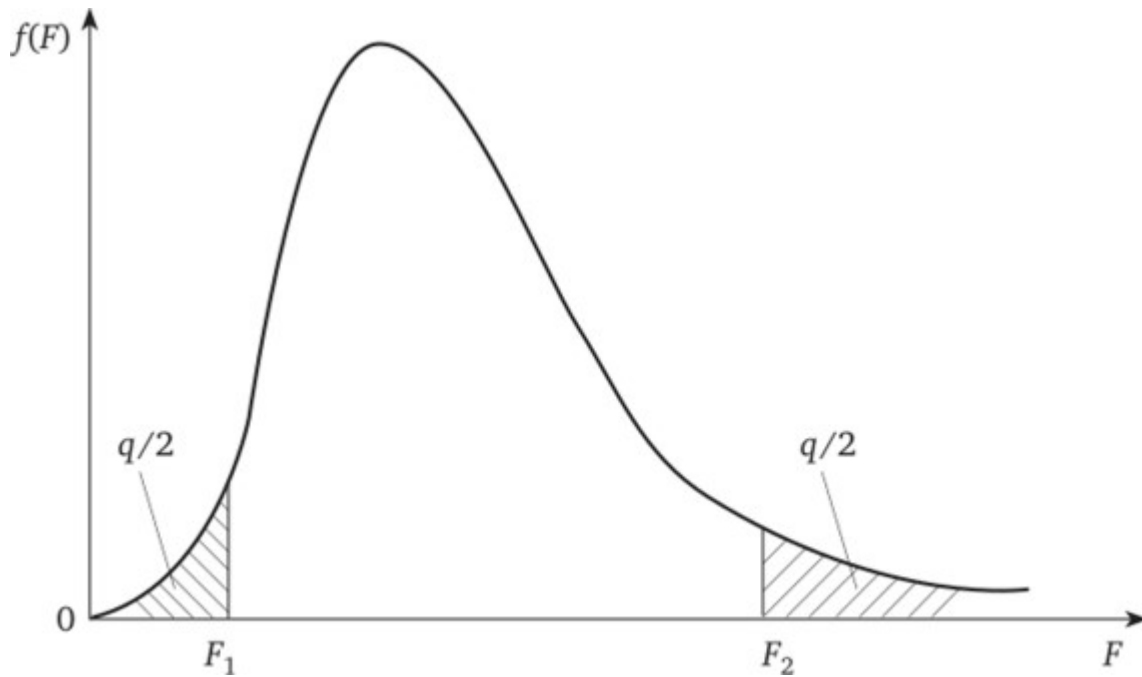
Чому χ^2 є мірою порівняння дисперсій? Тому що дисперсії, будучи сумою квадратів помилок, мають розподіл χ^2 .

Розподіл χ^2 -квадрат визначається наступним чином:

$$f(x) = \frac{1}{2^{v/2} \Gamma(v/2)} x^{(v/2)-1} e^{-x/2}, \quad v = 1, 2, \dots, \quad x > 0,$$

де v - число ступенів свободи; e - число Ейлера (2,71 ...); Γ - гамма- функція.

Графік щільності F-розподілу показаний на рис. 5.2.



Мал. Графік щільності F-розподілу

Отже, випадкова величина

$$F = \frac{S_1^2}{S_2^2},$$

де S_1^2 і S_2^2 - незсунені оцінки дисперсій, отриманих з незалежних вибірок, взятих з нормальних сукупностей, має розподіл Фішера (F-розподіл).

Величина F випадкова, судити однозначно по її величині про підтвердження або спростування гіпотези про однорідність досліджуваних вибірок можна.

Тому вводиться $1 - \alpha$ -вий рівень значущості, чисельно рівний ймовірності *неприйнятних* відхилень від прийнятої гіпотези. Області неприйнятних значень F показані на рис. 5.2 штрихуванням. Граничні точки допустимих значень F визначаються точками F_1 і F_2 , відповідними можливостям $\alpha/2$.

Якщо обчислене за даними експерименту значення F потрапляє в область між точками F_1 і F_2 :

$$F_1 \leq F \leq F_2,$$

то прийнята гіпотеза не спростовується. Зауважимо, що випадкова величина

$$F^* = \frac{1}{F} = \frac{S_2^2}{S_1^2}$$

також має F-розподіл із ступенями свободи $do_x = N_1 - 1$ і $do_2 = N_2 - 1$ відповідно. Отже, ймовірність попадання числа F в ліву критичну область дорівнює

$$P(F < F_1) = P\left(\frac{1}{F} > \frac{1}{F_1}\right) = P\left(F^* > \frac{1}{F_1}\right).$$

Звідси випливає, що *ліва* критична точка F-розподілу відповідає *правій* критичній точці F'-розподілу, тобто праві точки розподілів F і F* визначають ліву і праву точки і F₂. Тому в таблицях представлені тільки праві (F₂) критичні точки F-розподілу.

У таблицях значення F₂ наведені в залежності від $q/2$, числа ступенів свободи $do_x = N_x - 1$ і $do_2 = N_2 - 1$.

Зазвичай при обчисленні F в чисельник відносини ставлять значення більшої дисперсії.

Отже, при $F < F_2$ прийнята гіпотеза не спростовується, при $F > F_2$ - не підтверджується.

До організації надійшли дві буссоли. Перша з них при вимірюванні п'ять разів одного і того ж кута показала дисперсію $S_f = 0,1$ град². За результатами семи вимірів другий буссолью того ж кута отримана дисперсія $S_f = 0,15$ град².

Однотипні чи буссоли? Чи однакові вони по точності вимірювання кутів? Висунемо і перевіримо гіпотезу про їх однотипності для рівня значущості $q = 10\%$.

Розв'язання

$$F = \frac{S_2^2}{S_1^2} = \frac{0,15}{0,1} = 1,5; k_1 = 5 - 1, k_2 = 7 - 1 = 6.$$

По таблиці F-розподілу для ступенів свободи $do_2 = 6$, що відповідає більшій дисперсії, і $do_x = 4$, що відповідає меншій дисперсії, і рівня значущості $q/2$ знаходимо $P_2 = 6,16$.

Так як $F = 1,5 < P_2 = 6,16$, то для рівня значущості $q = 10\%$ гіпотеза про однаковість бусоль спростовується.

Отже: чим менше рівень значимості q , тим менше ймовірність забракувати перевіряється гіпотезу, коли вона вірна, тобто зробити помилку першого роду.

Але зі зменшенням рівня значущості (збільшенням F_2) розширюється область допустимих помилок, що призводить до збільшення ймовірності прийняття неправильного рішення, тобто здійснення помилки другого роду.

На закінчення викладеного відзначимо, що як би не була велика обсяг статистичного матеріалу (N_1, N_2), критерій Фішера (втім, як і будь-який інший) не може дати абсолютно достовірну відповідь про справедливість або несправедливість перевіряється гіпотези, так як

ми оперуємо випадковими числами, тобто спростування гіпотези при $F > F_2$ не є категоричним доказом її несправедливості, так само як і підтвердження гіпотези при $F < F_2$ не означає категоричної докази її справедливості. Не виключено, що в тому і в іншому випадках рішення може виявитися помилковим.

Судження про підтвердження або відхилення висунутої гіпотези висловлюється з певним ступенем вірогідності.

Приклад. Нехай проведено чотири види дослідів на кожному з трьох рівнів фактору F. Результати дослідів занесені в таблицю. Необхідно на рівні значущості 0,05 перевірити нульову гіпотезу про рівність групових середніх (нехай вибірки взято з нормальних сукупностей з однаковими дисперсіями).

Таблиця

Номер дослідів	Рівні фактору		
	F ₁	F ₂	F ₃
1	38	20	21
2	36	24	22
3	35	26	31
4	31	30	34
$y_i^{сеп}$	35	25	27

Для нашого прикладу $n = 12$, $I = 3$.

Знайдемо відповідні групові середні $y_i^{сеп}$ та між групову середню $y^{сеп} = 29$ і занесемо їх до таблиці.

Відшукаємо S_R^2 , S_A^2 .

$$S_A^2 = \sum_{i=1}^I n_i (y_i^{сеп} - y^{сеп})^2 = (35 - 29)^2 + (25 - 29)^2 + (27 - 29)^2 = 224$$

$$S_R^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - y_i^{сеп})^2 = (38 - 35)^2 + (36 - 35)^2 + (35 - 35)^2 + (31 - 35)^2 + \\ + (20 - 25)^2 + (24 - 25)^2 + (26 - 25)^2 + (30 - 25)^2 + \\ + (21 - 27)^2 + (22 - 27)^2 + (31 - 27)^2 + (34 - 27)^2 = 204$$

Складемо вибірову характеристику F.

$$F = \frac{S_A^2 / (I - 1)}{S_R^2 / (n - I)} = \frac{112}{22,67} = 4,94$$

Враховуючи кількість степенів вільності 2 і 9, та рівень значущості 0,05, за таблицею розподілу Фішера знайдемо критичну точку $F_{кр}(0,05;2;9)=4,26$. Оскільки $F > F_{кр}$ – нульову гіпотезу відкидаємо. Іншими словами, групові середні різняться суттєво.

Завдання для самостійного розв'язування

Приклад. Нехай проведено чотири види дослідів на кожному з трьох рівнів фактору F. Результати дослідів занесені в таблицю. Необхідно на рівні значущості 0,01 перевірити нульову гіпотезу про рівність групових середніх (нехай вибірки взято з нормальних сукупностей з однаковими дисперсіями).

Таблиця

Номер дослідів	Рівні фактору		
I	F ₁	F ₂	F ₃
1	35	24	20
2	32	28	26
3	38	24	30
4	39	32	36
$\bar{y}_i^{сер}$	35	26	28

Тема 13. Елементи теорії регресії і кореляції

Велику роль в пізнанні навколишньої дійсності, постійному уточненні й поглибленні знань людини про все більш нові її сторони і властивості відіграє кількісний аналіз, що базується на приматі якісного дослідження причинно-наслідкових зв'язків між явищами.

Розрізняють два види залежності між економічними явищами і процесами:

- функціональні, коли зміна однієї змінної X (аргументу, чинника) на одиницю свого вимірювання зумовлює зростання або зменшення іншої змінної Y (функції) на певну величину, тобто для кожного можливого (допустимого) значення X відповідає цілком певне значення Y;
- стохастичні (вірогідність), при яких зміна однієї випадкової величини X викликає зміну ряду розподілу (середнього значення) іншої Y, тобто кожному функціональному значенню аргументу відповідає статистичний розподіл функції і навпаки.

Функціональну (жорсткі, однозначні) залежність не слід змішувати з формулами, що використовуються для обчислення різних статистичних показників, оскільки перші характеризують одну з багатоманітних (об'єктивних) форм зв'язку між явищами і процесами в природі і суспільстві, а другі – формальний (арифметичний) підхід до оцінки отриманих результатів, часто у вигляді ланцюга співмножників.

Стохастична залежність може бути суттєвою, тобто обумовлена внутрішньо властивими даному явищу причинами, і несуттєвими, які викликані дією зовнішніх (випадкових) причин (середовищем); безпосередніми і опосередкованими, стійкими і нестійкими, сильними і слабкими, простими (між двома змінними) і складними (між залежною змінною Y і декількома чинниками-аргументами $x_1, x_2, \dots, x_n$).

Використання методів кореляції і регресії базується на математико-статистичній обробці вибірових даних. Завдання дослідника полягає у виявленні закономірностей, що приховані за похибками вимірювання, помилками спостереження, випадковими причинами, зробити ці закономірності більш очевидними, абстрагувати від всього другорядного, незначного і концентруючись на найважливішому, суттєвому.

Регресія – це одностороння стохастична залежність між випадковими величинами, в якій кожному значенню X відповідає ряд значень Y і, навпаки; кожному значенню Y – безліч значень X . На відміну від кореляційної, функція регресії необоротна. Це обумовлено наступними обставинами:

- Слід розрізняти: причинні зв'язки між явищами, коли X – причина, а Y – наслідок, або $X > Y$ (врожайність і кількість опадів, добрив);

X $\rightarrow$ V_2 i $Y_1 \leftrightarrow Y_2$

(вік чоловіків і жінок, які беруть шлюб);

помилкові (абсурдні) зв'язки, які виходять при формальному підході до досліджуваних явищ, без з'ясування причин, що обумовлюють даний зв'язок (число пожеж у країні – причиною є засуха). Ці зв'язки частіш за все обумовлені відсутністю попереднього їх аналізу, неправомірною (некоректною) розмірністю досліджуваних змінних (абсолютні величини) – продуктивність праці і чисельність трудящих.

Регресійну модель слід розглядати як математичний вид реального закономірного зв'язку. В економічних дослідженнях представляє інтерес не просто вивчення взаємозв'язків процесів і явищ, а кількісний вид цих взаємозв'язків. Тому до моделі перш за все ставиться вимога найбільшої відповідності характеру досліджуваного процесу, можливості економічної інтерпретації всіх його параметрів і наближення розрахункових результатів до досліджених даних. Звідси значне підвищення вимог до точності, надійності й адекватності кожного параметра моделі в цілому.

Процес кореляційного і регресійного аналізу складається з таких послідовних етапів:

- попередні угруповання статистичних даних і виявлення форми зв'язку;
- складання рівнянь парної регресії за кожним чинником;
- оцінки тісноти зв'язку, надійності й достовірності отриманої залежності;
- розробки регресійної багатофакторної моделі явища, що вивчається, оцінки її точності й визначення сили впливу врахованих чинників;
- аналіз досліджуваного явища (показників) за допомогою складеного тижня.

Дані подібного аналізу можна розглядати як основу при оцінці протікання досліджуваного процесу.

Обґрунтування форми зв'язку змінних і розрахунок параметрів теоретичної лінії регресії

Після попередньої обробки статичних даних приступають до складання рівняння регресії Y за кожним аргументом X . Загальна схема послідовних дій полягає в наступному:

- за дослідженими даними складатися кореляційна таблиця і кореляційне поле в звичайній декартовій координатній сітці;
- обчислюються середні $\bar{Y}_{xi}$ і будується емпірична лінія регресії $\bar{Y}_{cp}$ по X_i , що характеризує зміну середнього значення функції під впливом чинника аргументу.

Середні інтервальні значення $\bar{Y}_{cp}$ визначаються за допомогою співвідношення

$$\bar{Y} = \frac{\sum m_i y_i}{\sum m_i}.$$

Доцільно використовувати спрощений спосіб обчислень шляхом переходу до нових значень функції

$$\bar{Y}'_i = \frac{Y_i - C_{yi}}{\Delta Y},$$

де C_{yi} – вибраний центр переходу до нових значень. Перетворені змінні переводяться в натуральні:

$$\bar{Y}_{cp} = \bar{Y}'_{cp} \cdot \Delta Y + C_y,$$

де ΔY – величина інтервалу.

Середні інтервальні значення функції $\bar{Y}_{cp}$ відображаються на кореляційному полі у вигляді точок із середини інтервалів зміни X_i . Потім ці точки з'єднуються і отримана ламана лінія має назву емпіричної. Її зигзаги сильніше виявляються в інтервалах з малою кількістю спостережень. За законом великих чисел можна стверджувати, що емпірична лінія регресії все більше згладжуватиметься при зростанні числа спостережень.

Граничне положення емпіричної лінії регресії, до якого вона прагне при необмеженому збільшенні числа спостережень має назву теоретичної лінії регресії, а процес її знаходження – вирівнюванням емпіричної лінії регресії.

Суттєву роль в кореляційному аналізі відіграє підбір форми математичного рівняння, що найкращим чином описує досліджуваний процес.

Математичні функції $\bar{Y} = f(x_i)$ для опису залежності можуть бути найрізноманітнішими. Найширше розповсюдження знайшли лінійні рівняння регресії виду

$$\bar{Y} = kx_i \pm b.$$

Параметри лінійного рівняння легко визначаються, рівняння зручне в застосуванні, теорія лінійної кореляції вивчена найбільш докладно. Проте лінійна залежність має свої специфічні властивості, які не можуть належати всьому різноманіттю економічних явищ.

Основні властивості лінійної функції:

- швидкість зміни Y із зростанням X постійна, не залежить від величини X , тобто $\frac{\Delta y}{\Delta x} = const$;
- абсолютні максимум і мінімум Y в будь-якому діапазоні зміни X досягаються на його кінцях.

З цього виходить, по-перше, що різноманітні процеси, швидкість перебігу яких змінна і залежить від досягнутого рівня, тільки грубо приблизно можуть описуватися за допомогою лінійної функції; по-друге, якщо абсолютний максимум або мінімум величину, яка вивчається, досягається всередині даного діапазону зміни X , то за допомогою лінійної функції він

виявлений бути не може. Ця остання обставина особливо суттєва при вирішенні задач оптимізації.

Далі лінійна регресія Y по X припускає виконання наступних строго певних умов:

- дисперсія величини Y , відповідна даному значенню X , постійна, тобто $\sigma_{yx}^2 = \sigma_x^2$, або пропорційна даній функції від X , тобто $\sigma_{yx}^2 = \sigma_x^2 h(x)$.

Вказані умови на практиці виконуються рідко. Отже складність реальних процесів, нерівномірність їх протікання, наявність екстремальних значень і різнонаправленість дії окремих взаємозв'язаних чинників у більшості випадків не дозволяють достатньо обґрунтовано застосовувати лінійні методи. Це призводить до необхідності розрахунку нелінійних рівнянь.

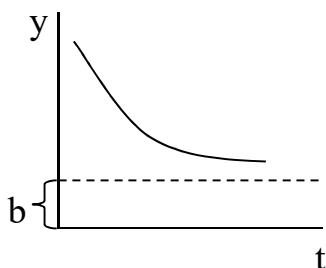
Виходячи з фізичної, технологічної, економічної сутності дослідженого процесу і враховуючи вид емпіричної лінії регресії, робиться гіпотеза про форму залежності функціональної ознаки Y від аргументу X по суті, забезпечуючи найкраще наближення розрахункових результатів до початкових даних.

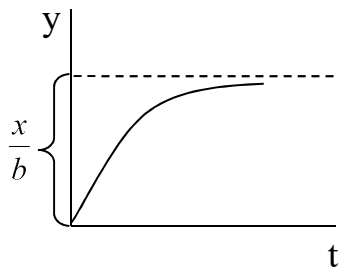
Для раціонального вибору форми зв'язку дослідник повинен мати в розпорядженні достатній вибір інструментів для оцінки залежності.

Доцільно користуватися дво- і трипараметричними рівняннями. Для моделювання показників, що монотонно зростають, можна отримати добре наближення двопараметричними значеннями, а моделі трипараметричні при коректному їх підборі виконуються для досить точного опису більшості процесів, що допускають єдиний екстремум. Монотонність або екстремальність залежності повинна бути обумовлена характеристиками процесу і підтверджуватися виглядом емпіричної лінії регресії.

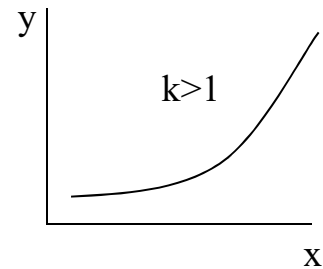
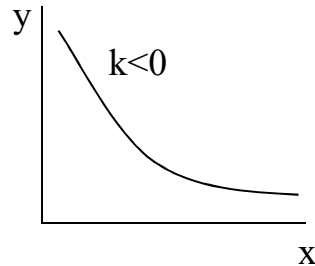
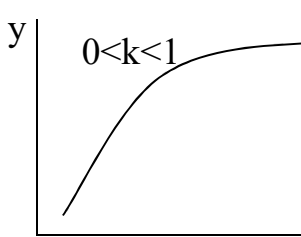
Використання багатопараметричних рівнянь недоцільне, хоча за допомогою максимуму функції можна отримати близьке наближення до початкових даних. Проте таким чином описується не стільки велика тенденція, скільки випадкові відхилення. Такі функції можуть мати максимуми і мінімуми, не виправдані по суті. Крім того, як складання таких функцій, так і їх застосування для практичних розрахунків різко ускладнюється.

Більшість математичних функцій, що використовуються для опису техніко-економічних показників, шляхом функціональних перетворень Y по X (роздільно або одночасно) можуть бути зведені до лінійного вигляду. При цьому метод перетворень залежить від форми зв'язку.





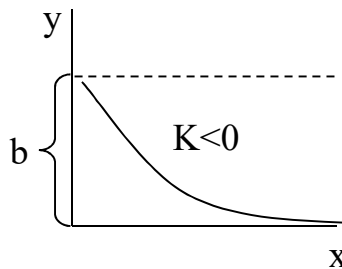
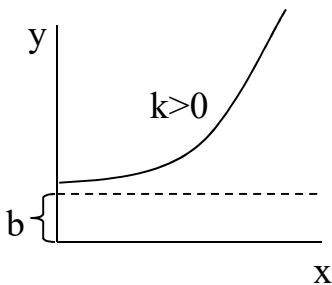
Гіпербола вигляду $\bar{y} = \frac{k}{x} + b$ перетвориться в лінійну шляхом заміни. Статична функція вигляду $\bar{y} = bk^x$ перетвориться в лінійну шляхом логарифмуванням. У результаті маємо $\lg \bar{y} = \lg b + k \lg x$. Позначимо $\bar{y}' + \lg \bar{y}, b' = \lg b, x' = \lg x$.



У результаті маємо $\bar{y}' = kx' + b'$.

Показова функція виду $y = be^{kx}$ перетвориться в лінійну логарифмуванням $\lg \bar{y} = \lg b + kx \lg e$.

Позначимо $\bar{y}' = \lg y', b' = \lg b, k' = k \lg e$, при цьому $\lg e = 0.4343$. У результаті маємо $y_1 x + b_1$.



Теоретична лінія регресії може бути подана у вигляді плавної кривої яка кількісно виражає зв'язок між середніми інтервальними значеннями $\bar{y}_{cp}$ і відповідними значеннями X (аргументами). Процес знаходження невідомих параметрів теоретичної залежності є однією з важливих проблем теорії кореляції і регресії.

Існує кілька способів встановлення аналітичної залежності, точніше, знаходження невідомих параметрів рівняння: графічний, методи середніх, найменших квадратів та інші.

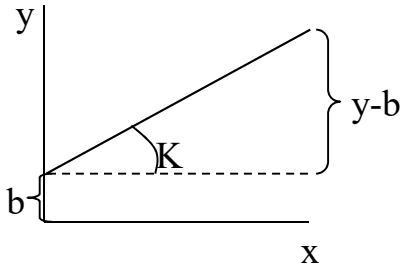
При графічному способі залежність визначають за допомогою лінійки (протягнутої нитки) так, щоб по обох сторонах проведеної потім лінії розташовувалося приблизно однакова кількість точок. Якщо характер розташування точок вказує на наявність криволінійної

залежності, то використовується лекало, або шляхом функціональних перетворень змінних нелінійна залежність зводиться до лінійного вигляду.

Числове значення коефіцієнта залежності функції $\bar{y} = kx + b$ визначається за формулою

$$k = \frac{y - b}{x},$$

де y і x - похідні числові значення за графіком.



Відшукавши як мінімум дві точки, по них проводять пряму.

Якщо лінійну пряму розташувати так, щоб точки приблизно однаково розташовувалися по обидві сторони від неї, то параметри лінійного рівняння можна визначити з рівняння

$$\frac{y - y_1}{y_2 - y_3} = \frac{x - x_1}{x_2 - x_3},$$

де y_1, x_1, y_2, x_2 - значення координат двох точок, які найкращим чином "вписуються" у пряму.

Сутність середньоарифметичного способу полягає в знаходженні середньоарифметичного значення від всіх n значень Y і X , а також додаткових середніх для точок, чисельні значення яких менше x_{cp}^b і y_{cp}^b і більше x_{cp}^n і y_{cp}^n . У цьому випадку сума відхилень відстаней точок по обидві сторони від теоретичної лінії, визначених на осі ординат, дорівнює нулю.

Кутовий коефіцієнт теоретичної лінії знаходять за формулою

$$k = \frac{y_{cp}^b - y_{cp}^n}{x_{cp}^b - x_{cp}^n},$$

де y_{cp}^b, x_{cp}^b - верхня середня (координати точок); y_{cp}^n, x_{cp}^n - нижня середня.

Значення вільного члена b знаходять за графіком.

Основними недоліками графічного методу визначення математичної залежності слід вважати суб'єктивність «якнайкращого розташування» лінійної прямої. Якщо на тривалість виконання роботи впливають одночасно кілька чинників, то графічний метод можна застосовувати тоді, коли є можливість заздалегідь визначити залежність роздільно по кожному чиннику за умови, що значення решти чинників залишаються постійними.

Для обчислення параметрів рівняння виду $\bar{y} = kx + b$ частіш за все користуються методом найменших квадратів. При цьому ставиться умова, щоб сума квадратів відхилень (відстаней) всіх досліджених точок від ординат, обчислених за рівнянням прямої ε_i , була мінімальною. Іншими словами, пряма повинна проходити якомога ближче до вершин емпіричної лінії регресії. Це означає, що параметри K і b управління регресії треба визначити з рівняння:

$$\varepsilon_i = \sum_{i=1}^n (y_i - \tilde{y}_i) = \min ,$$

де y_i – ординати досліджуваних точок;

$\tilde{y}_i$ – ординати розрахункових точок, визначені за рівнянням регресії $\bar{y} = kx + b$ таким чином $\varepsilon_i = \sum_{i=1}^n [y_i - (kx + b)]^2 = F(k, b) \min$.

Умовою екстремуму даної функції слід вважати рівність нулю часткових виробничих, узятих за параметрами K і b

$$\frac{dF}{dk} = 0 \text{ у } \frac{dF}{db} = 0, [F(u)] = F_u(u) * u' \text{ звідси } \begin{aligned} \frac{dF}{dk} &= -2 \sum_{i=1}^n [y_i - (kx_i + b)] x_i = 0 \\ \frac{dF}{db} &= -2 \sum_{i=1}^n [y_i - (kx_i + b)] = 0. \end{aligned}$$

Скоротивши на (-2) і розкривши квадратні дужки, отримаємо систему лінійних рівнянь

$$\begin{aligned} \sum_{i=1}^n y_i x_i &= k \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i x_i &= k \sum_{i=1}^n x_i + bn, \end{aligned}$$

підставивши сюди чисельні значення відповідних величин, знайдемо параметри K і b .

У разі лінійної залежності геометричне і алгебраїчне значення коефіцієнта регресії полягає в тому, що він кількісно характеризує на скільки в середньому змінюється y при зміні X на одиницю свого вимірювання. Чим більше чисельні значення коефіцієнта регресії, тим більше відносний приріст функції при зміні аргументу.

При знаходженні параметрів параболу виду $\bar{y} = ax^2 + bx + c$ необхідно скласти і вирішувати систему з трьох нормальних рівнянь, яке розв'язується, виходячи з вимоги методу

найменших квадратів, тобто $\varepsilon_i = \sum_{i=1}^n (y_i - \tilde{y}_i)^2 = \min$.

Підставляючи $\tilde{y} = ax_i^2 + bx + c$, маємо

$$\varepsilon_i = \sum_{i=1}^n [y_i - (ax_i^2 + bx_i + c)]^2 = \min = F(a, b, c).$$

Знаходимо часткові похідні $\frac{dF}{da}, \frac{dF}{db}, \frac{dF}{dc}$ і прирівнюємо їх до нуля

$$\frac{dF}{da} = -2 \sum_{i=1}^n [y_i - (ax_i^2 + bx_i + c)] x_i^2 = 0$$

$$\frac{dF}{db} = -2 \sum_{i=1}^n [y_i - (ax_i^2 + bx_i + c)] x_i = 0$$

$$\frac{dF}{dc} = -2 \sum_{i=1}^n [y_i - (ax_i^2 + bx_i + c)] = 0$$

Після відповідного перетворення маємо

$$\sum_{i=1}^n y_i x_i^2 = a \sum_{i=1}^n x_i^4 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^2$$

$$\sum_{i=1}^n y_i x_i = a \sum_{i=1}^n x_i^3 + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n y_i = a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i + cn$$

Не важко помітити, за яким правилом складається система нормальних рівнянь для знаходження невідомих параметрів шуканої функції.

Приклад. За даними таблиці спостережень

x_i	2	4	6
y_i	5	3	7

скласти рівняння регресії Y на X та X на Y .

Складемо таблицю

Таблиця

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
2	5	4	25	10
4	3	16	9	12
6	7	36	49	42
$\sum_{i=1}^3 x_i = 12$	$\sum_{i=1}^3 y_i = 15$	$\sum_{i=1}^3 x_i^2 = 56$	$\sum_{i=1}^3 y_i^2 = 83$	$\sum_{i=1}^3 x_i y_i = 64$

За формулами (3.102) для $n = 3$ отримуємо

$$\alpha = \frac{56 \cdot 15 - 12 \cdot 64}{3 \cdot 56 - 12^2} = 3, \quad \beta = \frac{3 \cdot 64 - 12 \cdot 15}{3 \cdot 56 - 12^2} = \frac{1}{2}.$$

Отже, рівняння регресії Y на X буде

$$y = \alpha + \beta x = 3 + \frac{x}{2}.$$

Аналогічно за формулами (3.103) знаходимо:

$$\alpha' = \frac{83 \cdot 12 - 15 \cdot 64}{3 \cdot 83 - 15^2} = \frac{3}{2}; \quad \beta' = \frac{3 \cdot 64 - 15 \cdot 12}{3 \cdot 83 - 15^2} = \frac{1}{2}.$$

$$x = \alpha' + \beta'x = \frac{3}{2} + \frac{y}{2}.$$

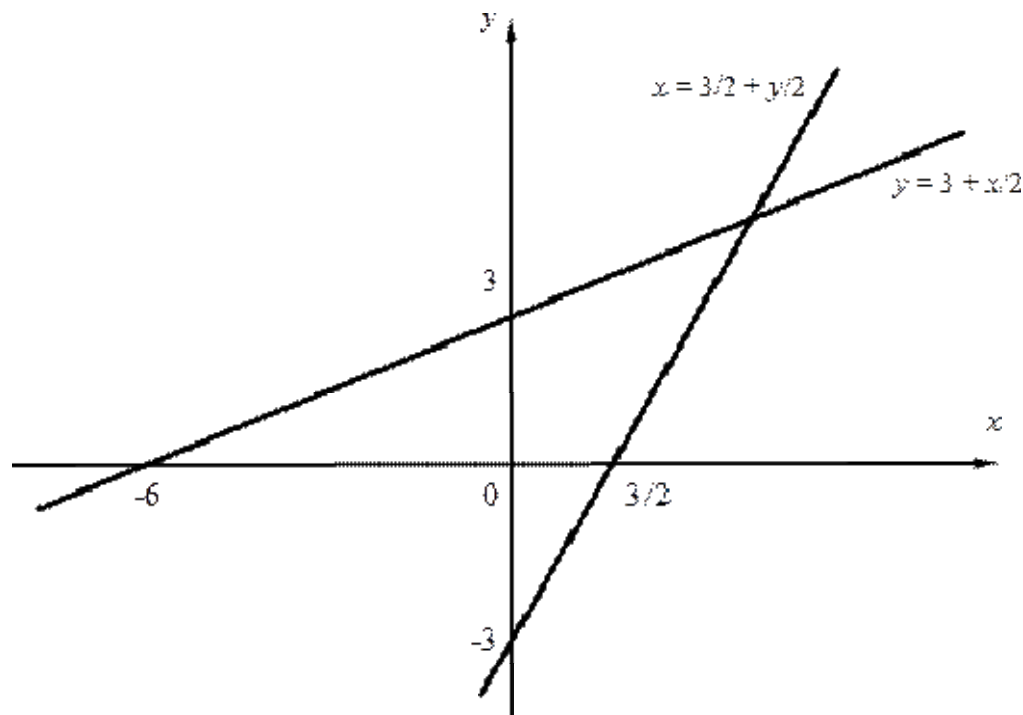


Рис.

Вправи для аудиторної та самостійної роботи.

1. Два арбітри оцінили майстерство 10 спортсменів. Записати рівняння лінійної регресії Y на X . Реалізувати графічно.

X	3	10	7	2	8	5	6	9	1	4
Y	6	2	1	3	9	4	5	7	10	8

2. Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведена

$Y = y_i$, шт.	10	35	68	146	152
$X = x_i$	44	43	42	41	40

в таблиці:

Побудувати рівняння лінійної регресії Y на X і знайти значення Y при $X = 39$.

3. Вимірювання температури в грудні, здійснені у двох містах, що умовно позначені А і В, наведено в таблиці:

Місто А $Y = y_i, C^0$	-10,2	-11,0	-12,4	-12,8	-13,0
Місто В $X = x_i, C^0$	-20,2	-20,8	-21,4	-21,8	-22,0

Побудувати рівняння лінійної регресії Y на X і знайти значення Y при $X = -23$.

4. Залежність урожайності пшениці Y від глибини зволоження X наведено в таблиці:

$Y = y_i$, ц/га	10	12	14	16	18
$X = x_i$, см	2	3	8	11	12

Побудувати рівняння лінійної регресії Y на X і знайти значення Y при $X = 13$.

5. Залежність пружності Y сталевих болтів від вмісту в них нікелю X наведена в таблиці:

$Y = y_i, \%$	34,8	35,0	35,8	36,2	37,7
$X = x_i, \%$	2,20	2,35	2,42	2,58	2,65

Побудувати рівняння лінійної регресії Y на X і знайти значення Y при $X = 2,7$.

Індивідуальні завдання.

1. Залежність між собівартістю X та кількістю Y виготовлених виробів наведено у таблиці:

$Y = y_i$, тис. шт.	5,2	4,3	3,8	3,6	3,5	2,2
$X = x_i$, тис. грн.	0,8	0,9	1,1	1,2	1,4	1,5

Побудувати рівняння лінійної регресії Y на X .

2. Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведена в таблиці:

$Y = y_i$, шт.	15	25	68	156	152
$X = x_i$	44	43	42	41	40

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 39$.

3. Вимірювання температури в грудні, здійснені у двох містах, що умовно позначені А і В, наведено в таблиці:

Місто А $Y = y_i, C^0$	-10,8	-11,4	-12,2	-12,8	-13,6
Місто В $X = x_i, C^0$	-20,2	-20,8	-21,4	-21,8	-22,0

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = -23$.

4. Залежність урожайності пшениці Y від глибини зволоження X наведено в таблиці:

$Y = y_i$, ц/га	12	14	15	16	18
$X = x_i$, см	2	3	8	11	12

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 13$.

5. Залежність пружності Y сталевих болтів від вмісту в них нікелю X наведена в таблиці:

$Y = y_i, \%$	35,2	35,6	35,8	36,6	37,5
$X = x_i, \%$	2,20	2,35	2,42	2,58	2,65

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 2,7$.

6. Залежність між собівартістю X та кількістю Y виготовлених виробів наведено у таблиці:

$Y = y_i$, тис. шт.	5,7	4,5	3,8	3,7	3,5	2,2
$X = x_i$, тис. грн.	0,6	0,8	1,3	1,4	1,5	1,6

Побудувати рівняння лінійної регресії Y на X .

7. Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведена в таблиці:

$Y = y_i$, шт.	10	35	68	146	152
$X = x_i$	43	42	41	40	39

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 38$.

8. Вимірювання температури в грудні, здійснені у двох містах, що умовно позначені А і В, наведено в таблиці:

Місто А $Y = y_i, C^0$	-10,2	-11,0	-12,4	-12,8	-13,0
Місто В $X = x_i, C^0$	-20,6	-20,8	-21,6	-21,8	-22,5

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = -24$.

9. Залежність урожайності пшениці Y від глибини зволоження X наведено в таблиці:

$Y = y_i$, ц/га	10	12	14	16	18
$X = x_i$, см	4	8	11	12	14

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 15$.

10. Залежність пружності Y сталевих болтів від вмісту в них нікелю X наведена в таблиці:

$Y = y_i$, %	35,4	35,0	35,8	36,2	37,7
$X = x_i$, %	2,25	2,35	2,45	2,56	2,63

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 2,7$.

11. Залежність між собівартістю X та кількістю Y виготовлених виробів наведено у таблиці:

$Y = y_i$, тис. шт.	5,4	4,5	3,9	3,7	3,5	2,8
$X = x_i$, тис. грн.	0,8	0,9	1,1	1,2	1,4	1,5

Побудувати рівняння лінійної регресії Y на X .

12. Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведена в таблиці:

$Y = y_i$, шт.	10	28	78	106	132
$X = x_i$	44	43	42	41	40

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 39$.

13. Вимірювання температури в грудні, здійснені у двох містах, що умовно позначені А і В, наведено в таблиці:

Місто А $Y = y_i$, $^{\circ}\text{C}$	-10,0	-11,5	-12,0	-12,6	-13,2
Місто В $X = x_i$, $^{\circ}\text{C}$	-20,2	-20,8	-21,4	-21,8	-22,0

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = -23$.

14. Залежність урожайності пшениці Y від глибини зволоження X наведено в таблиці:

$Y = y_i$, ц/га	12	15	18	22	25
$X = x_i$, см	2	3	8	11	12

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 13$.

15. Залежність пружності Y сталевих болтів від вмісту в них нікелю X наведена в таблиці:

$Y = y_i$, %	34,4	35,0	35,6	36,8	37,9
$X = x_i$, %	2,20	2,35	2,42	2,58	2,65

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 2,7$.

16. Залежність між собівартістю X та кількістю Y виготовлених виробів наведено у таблиці:

$Y = y_i$, тис. шт.	5,7	4,5	3,8	3,7	3,5	2,2
$X = x_i$, тис. грн.	0,6	0,8	1,1	1,3	1,4	1,6

Побудувати рівняння лінійної регресії Y на X .

17. Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведена в таблиці:

$Y = y_i$, шт.	10	35	68	146	152
$X = x_i$	45	44	43	42	41

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 39$.

18. Вимірювання температури в грудні, здійснені у двох містах, що умовно позначені А і В, наведено в таблиці:

Місто А $Y = y_i$, $^{\circ}\text{C}$	-10,2	-11,0	-12,4	-12,8	-13,0
Місто В $X = x_i$, $^{\circ}\text{C}$	-20,0	-20,6	-21,2	-21,5	-22,4

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = -24$.

19. Залежність урожайності пшениці Y від глибини зволоження X наведено в таблиці:

$Y = y_i$, ц/га	10	12	14	16	18
$X = x_i$, см	3	6	8	10	12

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 13$.

20. Залежність пружності Y сталевих болтів від вмісту в них нікелю X наведена в таблиці:

$Y = y_i$, %	35,4	35,0	35,8	36,2	37,7
$X = x_i$, %	2,26	2,38	2,46	2,54	2,62

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 2,7$.

21. Залежність між собівартістю X та кількістю Y виготовлених виробів наведено у таблиці:

$Y = y_i$, тис. шт.	5,2	4,4	3,9	3,7	3,5	2,8
$X = x_i$, тис. грн.	0,8	0,9	1,1	1,2	1,4	1,5

Побудувати рівняння лінійної регресії Y на X .

22. Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведена в таблиці:

$Y = y_i$, шт.	10	28	48	116	132
$X = x_i$	44	43	42	41	40

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 39$.

23. Вимірювання температури в грудні, здійснені у двох містах, що умовно позначені А і В, наведено в таблиці:

Місто А $Y = y_i$, $^{\circ}\text{C}$	-10,4	-11,2	-11,8	-12,4	-13,4
Місто В $X = x_i$, $^{\circ}\text{C}$	-20,2	-20,8	-21,4	-21,8	-22,0

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = -23$.

24. Залежність урожайності пшениці Y від глибини зволоження X наведено в таблиці:

$Y = y_i$, ц/га	15	18	20	22	25
$X = x_i$, см	2	3	8	11	12

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 13$.

25. Залежність пружності Y сталевих болтів від вмісту в них нікелю X наведена в таблиці:

$Y = y_i$, %	34,8	35,6	35,8	36,6	37,2
$X = x_i$, %	2,20	2,35	2,42	2,58	2,65

Побудувати рівняння лінійної регресії Y на X і знайти значення Y при $X = 2,7$.

26. Залежність між собівартістю X та кількістю Y виготовлених виробів наведено у таблиці:

$Y = y_i$, тис. шт.	5,7	4,5	3,8	3,7	3,5	2,2
$X = x_i$, тис. грн.	0,6	0,8	1,2	1,4	1,5	1,7

Побудувати рівняння лінійної регресії Y на X .

27. Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведена в таблиці:

$Y = y_i$, шт.	10	35	68	146	152
$X = x_i$	42	41	40	39	38

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 37$.

28. Вимірювання температури в грудні, здійснені у двох містах, що умовно позначені А і В, наведено в таблиці:

Місто А $Y = y_i$, $^{\circ}\text{C}$	-10,2	-11,0	-12,4	-12,8	-13,0
Місто В $X = x_i$, $^{\circ}\text{C}$	-20,0	-20,4	-20,9	-21,7	-22,0

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = -23$.

29. Залежність урожайності пшениці Y від глибини зволоження X наведено в таблиці:

$Y = y_i$, ц/га	10	12	14	16	18
$X = x_i$, см	4	6	8	11	12

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 13$.

30. Залежність пружності Y сталевих болтів від вмісту в них нікелю X наведена в таблиці:

$Y = y_i$, %	35,4	35,0	35,8	36,2	37,7
$X = x_i$, %	2,26	2,38	2,48	2,58	2,65

Побудувати рівняння лінійної регресії Y на X і знайти Y при $X = 2,7$.

Тема 14. Предмет економетрики. Методологія економетричного дослідження.

Економетрія посідає одне з чільних місць серед дисциплін, які є основою сучасної економічної науки. Вона є однією з основних дисциплін фундаментальної підготовки бакалаврів з економічних спеціальностей і за загальною, компетентною думкою разом з такими дисциплінами як мікроекономіка і макроекономіка створює теоретичний базис для аналітичних досліджень в економіці. Таке місце економетрії визначається тією роллю, яку вона відіграє в сучасних економічних дослідженнях. За допомогою економетричних методів вирішують наступні важливі питання :

- 1) підтвердження або відхилення економічних гіпотез і законів ;
- 2) емпіричний вивід економічних законів ;
- 3) прогнозування різноманітних економічних показників ;
- 4) економіко-математичний аналіз явища або процесу, що досліджується .

Економетрія як наукова дисципліна утворилася на основі економічної теорії, математичної економіки, економічної і математичної статистики. Таким чином, економетрія використовує основні базові поняття, термінологію і методи зазначених дисциплін. Але вона не є механічним поєднанням зазначених дисциплін, а є окремою, оскільки відрізняється від своїх складових.

Однією з основних задач **економічної статистики** є збір, обробка і представлення економічних даних у спеціальній формі – у вигляді таблиць, графіків, діаграм тощо. Економетрія також користується цим інструментарієм, але йде далі, застосовуючи отримані статистичні дані для аналізу економічних взаємозв'язків і прогнозування.

Внаслідок того, що взаємозв'язки між економічними показниками носять стохастичний характер в економетрії широко застосовуються методи **математичної статистики**. Однак в силу специфіки отримання статистичних даних в економіці економетрія розробляє і використовує спеціальні методи і прийоми, які у математичній статистиці не зустрічаються.

Економетричні моделі представляють собою окремий клас економіко-математичних моделей і характеризуються наступними особливостями:

- 1) економетричні моделі - **моделі функціональні** ;
- 2) економетричні моделі - **моделі прикладні (емпіричні)** ;
- 3) економетричні моделі - **моделі дескриптивні** ;
- 4) економетричні моделі - **моделі стохастичні**.

Задача : Оцініть параметри лінійної однофакторної моделі залежності за даними табл.

- Звітні дані будівельного підприємства, що характеризують кількість кв. метрів зданого житла (у) і використаних будівельних матеріалів (х)

Місяць	Рік									
	2000		2001		2002		2003		2004	
	у, тис. кв. м	х, млн. грн.	у, тис. кв. м	х, млн. грн.	у, тис. кв. м	х, млн. грн.	у, тис. кв. м	х, млн. грн.	у, тис. кв. м	х, млн. грн.
Січень	5,6	20,1	4,5	23,4	8,1	22,4	7,6	25,7	7,8	26,4
Лютий	5,2	22,3	4,9	22,1	5,4	27,6	6,4	25,9	8,7	28,1
Березень	6,4	22,7	5,8	24,8	5,5	28,5	7,3	26,2	9,3	28,7
Квітень	5,7	22,5	5,3	24,3	5,2	27,1	6,9	26,1	8,9	28,3
Травень	5,6	22,3	5,2	24,1	5,1	27,0	6,7	25,9	8,7	28,1
Червень	5,9	22,6	5,7	24,5	5,4	27,6	7,1	26,0	8,8	28,2
Липень	6,0	22,7	5,9	24,9	5,6	28,7	7,4	26,8	9,1	28,5
Серпень	6,5	22,8	6,0	25,0	5,7	28,8	7,6	26,9	9,3	28,7
Вересень	6,2	22,6	5,8	24,6	5,5	28,4	7,2	26,2	9,2	28,6
Жовтень	6,4	22,7	5,9	24,9	5,8	28,9	7,7	27,1	9,5	28,9
Листопад	5,1	22,1	4,6	23,6	4,9	26,9	6,1	25,3	8,0	27,6
Грудень	5,2	22,5	4,7	23,7	5,0	27,0	6,2	25,5	8,2	27,8
Середнє	5,4	22,7	4,8	23,9	5,2	27,2	6,3	25,7	8,3	27,9

Тема 15. Проста і множинна лінійна регресія. Метод найменших квадратів

Загальна лінійна економетрична модель- це регресійна модель, яка встановлює лінійну залежність між економічними показниками, один з яких є залежною (пояснюваною) змінною, а всі інші – незалежними(пояснюючими) змінними моделі.

Залежна змінна для такої моделі розглядається, як ендогенна змінна, а незалежні змінні – як екзогенні.

Теоретична (“канонічна”) загальна лінійна економетрична модель може бути специфікована в наступній формі :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon,$$

де y – залежна (пояснювана) змінна моделі, $x_1, x_2, \dots, x_m$ – незалежні (пояснюючі) змінні моделі або фактори, $\beta_0, \beta_1, \dots, \beta_m$ – параметри моделі, ε – стохастична складова моделі, m – кількість пояснюючих змінних моделі. Зазначимо, що параметри $\beta_1, \beta_2, \dots, \beta_m$ ще прийнято називати **коефіцієнтами регресії**,

Теоретична модель є гіпотетичною конструкцією і дійсна (як це відмічалось у попередній темі) для всієї генеральної сукупності спостережень за змінними моделі. Невідомі параметри $\beta_j, (j = \overline{0, m})$ цієї моделі є константами, а випадкова величина ε – взагалі неспостережувана і ми можемо зробити тільки припущення щодо закону її розподілу.

Вибіркова (емпірична) загальна лінійна економетрична модель має наступний вигляд :

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m + e,$$

де y – залежна (пояснювана) змінна моделі, $x_1, x_2, \dots, x_m$ – незалежні (пояснюючі) змінні моделі (фактори), b_0, b_1, b_m – параметри вибіркової моделі, e – залишки моделі.

Вибіркова модель є реальною конструкцією і будується на основі певної статистичної вибірки з генеральної сукупності. На відміну від моделі (2.1) параметри вибіркової моделі $b_0, b_1, \dots, b_m$ є оцінками (наближеними значеннями) параметрів $\beta_0, \beta_1, \beta_m$ і випадковими величинами, а залишки « e » можна оцінити (розрахувати) на основі статистичних даних. Таким чином, вибіркова модель завжди є тільки оцінкою (вдалою або невдалою) реальної але невідомої теоретичної моделі.

Вибіркова (емпірична) функція регресії для загальної лінійної економетричної моделі має наступний вигляд :

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m,$$

де $\hat{y}$ – оцінка математичного сподівання залежної (пояснюваної) змінної моделі, $x_1, x_2, \dots, x_m$ – незалежні (пояснюючі) змінні моделі (фактори), b_0, b_1, b_m – параметри вибіркової регресії.

Для побудови і аналізу загальної лінійної економетричної моделі широко застосовується апарат матричної алгебри. Тому для подальших викладок подамо загальну лінійну економетричну модель у матричній формі. Оскільки теоретична модель використовується для канонічного подання деякого економічного явища або процесу, а реально оперуємо в процесі дослідження цього явища (процесу) тільки вибірковою моделлю, саме вибірккову модель подамо в матричному вигляді :

$$Y = XB + e,$$

де

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \dim Y = n \times 1 - \text{вектор спостережень за залежною змінною моделі};$$

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{m1} \\ 1 & x_{12} & x_{22} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{mn} \end{pmatrix}, \dim X = n \times k - \text{матриця спостережень за пояснюючими}$$

змінними моделі, яку ще називають **регресійною матрицею**;

$$B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}, \dim B = k \times 1 - \text{вектор оцінок параметрів моделі (вектор параметрів вибіркової}$$

моделі);

$$e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}, \dim e = n \times 1 - \text{вектор залишків моделі.}$$

Для всіх наведених вище форм представлення загальної лінійної моделі прийняті наступні загальні позначення, які будуть у подальшому постійно використовуватися:

n – розмір статистичної вибірки (кількість спостережень в статистичній вибірці);

m – число незалежних (пояснюючих) змінних моделі;

k = m + 1 – число параметрів моделі.

Питання для розгляду:

1. Що таке кореляційний і регресійний аналіз? Основні напрямки їх використання в економетричних дослідженнях?
2. Розкрийте сутність і особливості використання методу найменших квадратів?

Задача: Виявіть кореляційні зв'язки між двома показниками за значеннями спостережень на основі даних в таблиці. Показники характеризують виробничу ситуацію на підприємствах будівельної галузі.

- Звітні дані будівельного підприємства, що характеризують кількість кв. метрів зданого житла (y) і використаних будівельних матеріалів (x)

Місяць	Рік
--------	-----

	2000		2001		2002		2003		2004	
	у, тис. кв. м	х, млн. грн.	у, тис. кв. м	х, млн. грн.	у, тис. кв. м	х, млн. грн.	у, тис. кв. м	х, млн. грн.	у, тис. кв. м	х, млн. грн.
Січень	5,6	20,1	4,5	23,4	8,1	22,4	7,6	25,7	7,8	26,4
Лютий	5,2	22,3	4,9	22,1	5,4	27,6	6,4	25,9	8,7	28,1
Березень	6,4	22,7	5,8	24,8	5,5	28,5	7,3	26,2	9,3	28,7
Квітень	5,7	22,5	5,3	24,3	5,2	27,1	6,9	26,1	8,9	28,3
Травень	5,6	22,3	5,2	24,1	5,1	27,0	6,7	25,9	8,7	28,1
Червень	5,9	22,6	5,7	24,5	5,4	27,6	7,1	26,0	8,8	28,2
Липень	6,0	22,7	5,9	24,9	5,6	28,7	7,4	26,8	9,1	28,5
Серпень	6,5	22,8	6,0	25,0	5,7	28,8	7,6	26,9	9,3	28,7
Вересень	6,2	22,6	5,8	24,6	5,5	28,4	7,2	26,2	9,2	28,6
Жовтень	6,4	22,7	5,9	24,9	5,8	28,9	7,7	27,1	9,5	28,9
Листопад	5,1	22,1	4,6	23,6	4,9	26,9	6,1	25,3	8,0	27,6
Грудень	5,2	22,5	4,7	23,7	5,0	27,0	6,2	25,5	8,2	27,8
Середнє	5,4	22,7	4,8	23,9	5,2	27,2	6,3	25,7	8,3	27,9

Згрупуйте подані дані, побудуйте кореляційне поле, кореляційну таблицю і лінію кореляції

Тема 16. Множинна лінійна регресія. Перевірка гіпотез

Найпростішою серед лінійних економетричних моделей є модель парної лінійної регресії (або проста лінійна модель), яка описує зв'язок всього між двома економічними змінними - показниками.

Економетричною моделлю парної лінійної регресії (простою лінійною моделлю) називається регресійна модель, яка описує лінійний зв'язок між двома економічними показниками, один з яких є залежною (пояснюваною), а другий – незалежною (пояснюючою) змінною.

Виходячи з вищезазначених позначень для простої лінійної регресії маємо:

- **теоретичну (“канонічну”)** модель парної лінійної регресії

$$y = \beta_0 + \beta_1 x + \varepsilon;$$

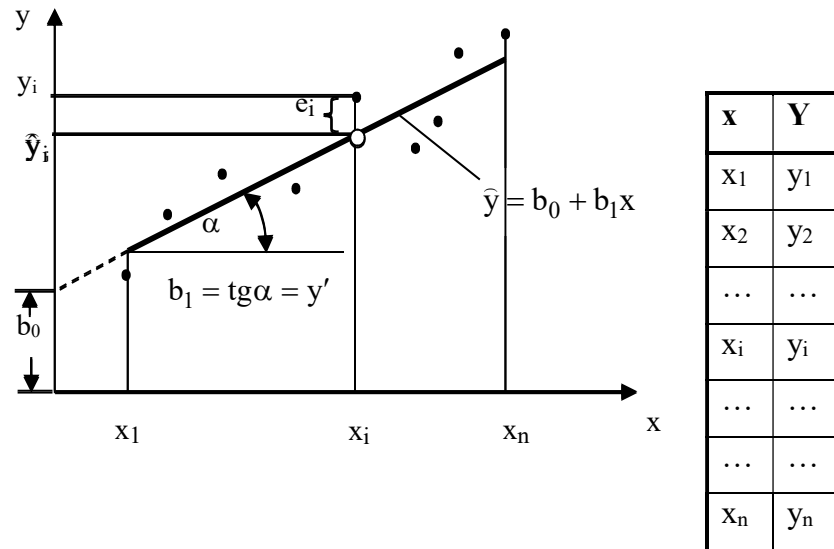
- **вибіркову (емпіричну)** модель парної лінійної регресії

$$y = b_0 + b_1 x + e;$$

- **вибіркову функцію** парної лінійної регресії

$$\hat{y} = b_0 + b_1x .$$

Рівняння (2.7) представляє собою параметричне рівняння прямої, тому на площині xOy вибірковій функції парної лінійної регресії відповідає вибіркова (емпірична) пряма регресії. Графічно вибіркова функція регресії і пряма регресії для деякої вибірки представлені на рисунку



Парна лінійна регресія

Параметри моделі парної лінійної регресії мають спеціальну назву. Параметр b_0 називається **перетином**, а b_1 – **нахилом**. Математична інтерпретація цих параметрів зрозуміла з наведеного рисунку.

Для побудови загальної лінійної моделі використовують статистичну інформацію щодо діяльності підприємства і здійснюють такі етапи: математико-статистичний аналіз, побудову лінійної регресійної моделі, перевірку побудованої моделі на адекватність, аналіз отриманих результатів.

На етапі математико-статистичного аналізу проводять перевірку основних припущень класичного регресійного аналізу, крім того, здійснюють найважливішу процедуру багатофакторного аналізу – перевірку факторів на мультиколінеарність. Термін “мультиколінеарність” означає, що в багатофакторній регресійній моделі дві або більше незалежних змінних (факторів) пов’язані між собою лінійною залежністю або, іншими словами, мають високий ступінь кореляції ($r_{xixj} \rightarrow 1, i \neq j$).

Для здійснення математико-статистичного аналізу будують матрицю коефіцієнтів парної кореляції. Потім аналізують коефіцієнти парної кореляції між факторами. Результатом етапу математико-статистичного аналізу є знаходження множини основних незалежних між собою факторів, що є базою для побудови регресійної моделі.

На другому етапі для побудови лінійної моделі широке використання отримали «покроковий» метод і метод «виключень». Сутність «покрокового» методу полягає в тому,

що фактори по черзі включаються в модель доти, доки вона не стане задовільною. Порядок включення вибирають за допомогою коефіцієнта кореляції як міри важливості факторів (незалежних змінних), які ще не включені в модель. Цей метод передбачає розрахунок часткових F-критеріїв для факторів, що здійснювали значний вплив на результативний показник. Далі визначають показники, які здійснювали найбільший вплив на результативний показник і значення часткових F-критеріїв перевищують нормативні значення.

Метод “виключень” полягає в тому, що вибирається низка факторів, які ймовірно можуть впливати на результативний показник. Потім, по черзі виключаються ті фактори, в яких найменший коефіцієнт кореляції (згідно з матрицею статистики), а значення часткових F-критеріїв не перевищують нормативні значення. Таким чином, залишаються лише ті змінні, які відповідають розглянутим вище умовам.

На наступному етапі аналізу перевіряється адекватність моделі з використанням F-критерію Фішера і t-критерію Стюдента. Статистична оцінка надійності коефіцієнта регресії здійснюється за допомогою t-критерію Стюдента. Він застосовується для оцінки тісноти зв'язку між незалежною змінною x і залежною y . При використанні цього критерію формулюється нульова гіпотеза. Потім отримане значення t-розподілу Стюдента порівнюється з критичним. Якщо фактичне значення t-розподілу Стюдента перевищує критичне, то спростовується нульова гіпотеза й зв'язок між змінними x і y вважається щільним. Якщо ні, то приймається нульова гіпотеза, а фактори моделі вважаються статистично неадекватними і виключаються з моделі при встановленому рівні значущості в 5% і 1%.

F-тест використовується для оцінки того, чи важливе пояснення, яке дає рівняння в цілому. Якщо фактичне значення F-критерію вище нормативного, то модель адекватна, а її фактори залишаються у рівнянні.

Для перевірки адекватності економетричної моделі використовується тест Дарбіна-Уотсона, який спрямований для перевірки кореляції між залишками використовується. Він включає такі етапи:

1. Розраховуються d-статистики для аналізованої вибірки даних. Як відомо з теорії, значення d-статистики лежать у межах від 0 до 4.
2. Порівнюються отримані d-статистики з табличними d-статистиками при рівні значущості $\alpha = 0,05$, кількості факторів k , що присутні в моделі, і кількості спостережень n . Якщо розраховане значення d-статистики знаходиться в проміжку від 0 до d_L ($0 < d < d_L$), то це свідчить про наявність позитивної автокореляції. Якщо значення d потрапляє в зону невизначеності, тобто набуває значення $d_L \leq d \leq d_U$, або $4 - d_U \leq d \leq 4 - d_L$, то ми можемо зробити висновки ні про наявність, ні про відсутність автокореляції. Якщо $4 - d_L < d < 4$, то маємо негативну автокореляцію. Нарешті, якщо $d_U < d < 4 - d_U$, то автокореляції немає.

Для оцінки адекватності моделі важливе значення має перевірка її на гомо або гетероскедастичність. Суть цього явища полягає в тому, що варіація кожної ε_i навколо її математичного сподівання не залежить від значення x . Дисперсія кожної ε_i зберігається сталою незалежно від малих чи великих значень факторів: σ_ε^2 не є функцією x_{ij} , тобто $\sigma_\varepsilon^2 \neq f(x_{1i}, x_{2i}, \dots, x_{pi})$.

Якщо σ_ε^2 не є сталою, а її значення залежать від значень x , можемо записати $\sigma_\varepsilon^2 = f(x_{1i}, x_{2i}, \dots, x_{pi})$. У цьому разі маємо справу з гетероскедастичністю. Оцінка моделі на наявність гетероскедастичності полягає в тому, що на першому етапі здійснюється тестування

моделі на наявність гетероскедастичності. І якщо підтверджується гіпотеза про її наявність, то на другому етапі модель виключається.

Тестування моделі на гетероскедастичність здійснюється на підставі тесту рангової кореляції Спірмена. Значущість отриманого коефіцієнта рангової кореляції Спірмена перевіряється за допомогою t-критерія Стюдента при $(n-2)$ кількості ступенів свободи.

Фактичне значення t-критерію Стюдента зіставляється з $t_{кр}$. Якщо $t_f > t_{кр}$, то підтверджується гіпотеза про наявність гетероскедастичності. А коли $t_f < t_{кр}$, то приймається гіпотеза про гомоскедастичність.

На етапі аналізу отриманих результатів здійснюється економічна інтерпретація отриманої економетричної моделі. На цьому етапі обґрунтовується економічна доцільність отриманих результатів.

Питання для розгляду:

1. Назвіть основні складові елементи моделі парної лінійної регресії?
2. Що таке випадковий член? Назвіть основні причини існування випадкового члена?
3. Як визначається залишки в моделі парної регресії?
4. Назвіть основні напрямки інтерпретації рівняння регресії?
5. Проінтерпретуйте будь-яке лінійне рівняння регресії?

Задача: Проведіть регресійний аналіз, коли відомо, що координати двох спостережень наступні: $y = 2$, коли $x = 1$, і $y = 4$ при $x = 3$. Відомо, що рівняння регресії має вигляд

$$y = a + bx.$$

Задача: Використовуючи дані задачі 4.1, додатково добавимо третє спостереження: $y = 6$ при $x = 3$. Проведіть регресійний аналіз методом найменших квадратів.

Задача: Проінтерпретуйте отриману регресійну залежність витрат на оплату житла і отриманого власного доходу, якщо вона відображена наступним чином:

$$y = -27,6 + 0,178 x.$$

Задача: Проінтерпретуйте отриману регресійну залежність витрат на оплату житла від часу, якщо вона відображена наступним чином:

$$y = 48,9 + 4,84 t.$$

Тема 17. Мультиколінеарність та її вплив на оцінки параметрів моделі

Мультиколінеарність – це поняття, яке використовується для опису проблеми, коли нестрога лінійна залежність між пояснювальними змінними призводить до отримання ненадійних оцінок регресії. Проте, така залежність зовсім необов'язково дає незадовільні оцінки. Якщо

всі інші умови задовільні, тобто якщо кількість спостережень і вибіркві дисперсії пояснювальних змінних великі, а дисперсія випадкого члена – мала, то в результаті можна отримати досить позитивні оцінки.

Мультиколінеарність повинна виникати за рахунок сполучення нестрогої залежності однією (або більше) незадовільних умов, це – питання ступеня визначеності явища, а не його виду. Оцінки регресії будуть неадекватні від неї у відповідній мірі значущості. Розгляд цієї проблеми починається тільки тоді, коли вона досить суттєво впливає на результати оцінки регресії.

Ця проблема є звичною для регресії часових рядів, тобто коли значення показників складаються із рядів спостережень протягом визначеного періоду часу. Якщо дві або більше незалежних змінних мають часовий тренд, то між ними буде існувати кореляція, і це може призвести до мультиколінеарності.

Існують різні методи для зменшення мультиколінеарності. Вони діляться на дві категорії: до першої категорії відносяться методи, спрямовані на виконання умов, що забезпечують надійність оцінок регресії; до другої – відносяться використання зовнішньої інформації. Якщо з початку використовувати можливі значення показників, то, звичайно, було б важливим збільшити кількість спостережень. Якщо, наприклад, використовуються часові ряди, то це можна зробити шляхом скорочення терміну кожного періоду часу.

Якщо використовуються дані перехресної виборки і дослідник знаходиться на стадії планування дослідження, то можна збільшити точність оцінок регресії і послабити проблему мультиколінеарності за рахунок більших витрат коштів на збільшення розміру виборки та інші методи.

Найбільш поширеним способом виявлення мультиколінеарності є побудова матриці статистики, в якій відображені коефіцієнти кореляції. Незалежні показники, між якими коефіцієнт кореляції перевищує 0,8 або 80%, розглядають на наявність мультиколінеарності.

У випадку виявлення наявності мультиколінеарності існує декілька простих шляхів її усунення. Основними серед них є наступні.

1. Вилучення змінної (або змінних) з моделі. При цьому з моделі вилучається одна із змінних колінеарної пари. Слід зазначити, що таке вилучення змінних можливе тільки у випадку, коли це не суперечить логіці економічних зв'язків. У протилежному випадку це може призвести до помилки специфікації.

2. Зміна аналітичної форми економетричної моделі. Іноді заміна однієї функції регресії іншою (наприклад, лінійної нелінійною) , якщо це не суперечить апріорній інформації, дає змогу уникнути явища мультиколінеарності.

3. Збільшення спостережень. З точки зору теорії, мультиколінеарність та невелика кількість спостережень у вибірці – це одна і та ж проблема. Тому збільшення спостережень у статистичній вибірці або використання іншої статистичної вибірки може усунути, або принаймні зменшити вплив мультиколінеарності.

4. Перетворення статистичних даних. Позбутися мультиколінеарності можна і шляхом наступних перетворень вихідних даних стосовно пояснюючих змінних :

- а) замість самих даних узяти їхні відхилення від середніх;
- б) замість абсолютних значень даних взяти відносні значення ;
- в) стандартизувати змінні.

5. Використання додаткової первинної інформації. Аналіз і використання первинної додаткової інформації інколи дозволяє зняти проблему мультиколінеарності.

Узагальнений метод найменших квадратів

Узагальнений метод найменших квадратів (УНК) враховує інформацію про неоднаковість дисперсії. Щоб проілюструвати це, розглянемо модель лінійної регресії:

$$Y=XA+u.(4.1)$$

- де A – параметри (вектор) економетричної моделі;
 Y – залежна змінна;
 X – незалежна змінна;
 u – випадковий член.

Задача полягає в знаходженні оцінок елементів вектора A в моделі. Для цього використовується матриця S , за допомогою якої коригується вхідна інформація.

Оскільки S - додатно визначена матриця, то вона може бути зображена як добуток PP^T , де матриця P є невиродженою, тобто:

$$S=PP^T,(4.2)$$

коли

$$P^{-1}S(P^{-1})^T=E,$$

$$((P^{-1})^T)^{-1}=S^{-1},$$

Помноживши рівняння ліворуч на матрицю P^{-1} , дістанемо:

$$P^{-1}Y=P^{-1}XA+P^{-1}u.$$

Позначимо $Y^*=P^{-1}Y$; $X^*=P^{-1}X$, $u^*=P^{-1}u$

Тоді модель матиме вигляд:

$$Y^*=X^*A+u^*.$$

Використовуючи дану модель, неважко показати, що $M(u^*u^{*T})=\sigma^2E$, тобто модель задовольняє умові, коли параметри моделі можна оцінити на основі МНК, яка може бути виражена:

$$\hat{A} = (X^* X^*)^{-1} X^{*'} Y^* = (X^T S^{-1} X)^{-1} X^T S^{-1} Y.$$

Ця оцінка є незміщеною лінійною оцінкою вектора A , який має найменшу дисперсію і матрицю коваріацій:

$$\text{var}(\hat{A}) = \sigma_u^2 (X^{*T} X^*)^{-1} = \sigma_u^2 (X^T S^{-1} X)^{-1}.$$

Незміщену оцінку для дисперсії σ_u^2 можна знайти так:

$$(Y^* - X^* \hat{A})^T (Y^* - X^* \hat{A}) / (n-k-1) = (Y - X \hat{A})^T S^{-1} (Y - X \hat{A}) / (n-k-1) = u^T S^{-1} u / (n-k-1).$$

При заданій матриці S оцінку параметрів моделі можна обчислити згідно із моделлю (4.7), а стандартну помилку - згідно з (4.8). Тому можна сконструювати звичайні критерії значущості і довірчі інтервали для параметрів a .

Визначивши залишки $u = Y - X \hat{A}$ і помноживши ліворуч на матрицю P^{-1} , отримаємо:

$$P^{-1} Y - P^{-1} X \hat{A} = P^{-1} u; \text{ або } u^* = Y^* - X^* \hat{A}$$

Звідси $Y^* = X^* A + u^*$.

Тоді $Y^{*T} Y^* = (X^* \hat{A} + u^*)^T (X^* \hat{A} + u^*)$.

Оскільки $\hat{B} = (X^{*T} X^*)^{-1} X^{*'} Y^* = (X^T S^{-1} X)^{-1} X^T S^{-1} Y$, то

$$Y^T S^{-1} Y = \hat{A}^T X^T S^{-1} Y + u^T S^{-1} u. \quad (4.10)$$

Отже, ми розбили загальну суму квадратів для Y^* на суму квадратів регресії і залишкову. Згідно з цими даними дисперсійний аналіз буде виконаний для перетворених вхідних даних. Крім того, коли незалежна змінна Y^* вимірювана відносно початку відліку, а не у формі відхилення від середньої, то необхідно визначити її середнє значення і скористатись ним для корекції загальної суми квадратів і суми квадратів регресії.

Щоб оцінити параметри моделі, коли дисперсії залишків визначаються $M(uu') = \sigma_u^2 S$, потрібно визначити матрицю S . Спинимось на визначенні матриці S .

Оскільки явище гетероскедастичності пов'язане лише з тим, що змінюються дисперсії залишків, а коваріація між ними відсутня, то матриця S має бути діагональною, а саме:

$$S = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\lambda_3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{1}{\lambda_n} \end{pmatrix}, \quad S = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Щоб пояснити, чому саме такий вигляд має ця матриця, потрібно визначити: за наявності гетероскедастичності для певних вихідних даних одна (або кілька) пояснювальних змінних можуть різко змінюватись від одного спостереження до іншого, тоді як залежна змінна має такі самі коливання, як і для попередніх спостережень.

Але це означає, що дисперсія залишків, яка змінюватиметься від одного спостереження до іншого (чи для групи спостережень), може бути пропорційною до величини пояснювальної змінної X (або до її квадрата), яка зумовлює гетероскедастичність, або пропорційною до квадрата залишків.

Звідси в матриці S значення σ_i^2 можна обчислити, користуючись гіпотезами:

1. $M(uu') = \sigma_u^2 x_{ij}$, тобто дисперсія залишків пропорційна до зміни пояснюючої змінної x_{ij} ;
2. $M(uu') = \sigma_u^2 x_{ij}^2$, тобто зміна дисперсії пропорційна до зміни квадрата пояснюючої змінної x_{ij}^2 ;
3. $M(uu') = \sigma_u^2 \{IuI\}^2$, тобто дисперсія залишків пропорційна до зміни квадрата залишків за модулем.

Для першої гіпотези: $\sigma_i^2 = 1/x_{ij}$

Для другої гіпотези: $\sigma_i^2 = 1/x_{ij}^2$

Для третьої гіпотези: $\sigma_i^2 = \{IuI\}^2$, або $\sigma_i^2 = (a_0 - a_1 x_{ij})^2$, або $\sigma_i^2 = (a_0 - a_1 x_{ij}^{-1})^2$

Оскільки матриця S — симетрична і додатно визначена, то при $S = P'P$, матриця P має вигляд:

$$S = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix}, \quad P = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sqrt{\lambda_3}} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}.$$

Розглянемо дослідження впливу на економічний показник Y — реальне споживання країни (у млрд. грн.) трьох факторів: x_1 — купівлі та оплати товарів і послуг (у млрд.

грн.), x_2 – усіх заощаджень від загального грошового доходу (у % від загальної суми доходу), x_3 – рівня ставки ПДВ (у %). Необхідно перевірити фактори на мультиколінеарність.

Таблиця

Дані до задачі

№ п/п	y_i	x_{1i}	x_{2i}	x_{3i}
	25,74	4,69	11,97	29,23
	25,34	5,64	13,43	29,35
	31,26	6,26	12,92	33,40
	33,50	6,99	14,74	30,97
	32,30	6,36	14,64	32,92
	38,90	7,60	17,10	37,27
	41,58	7,12	15,63	30,97

Таблиця (продовження)

48,02	6,81	15,35	33,58
43,30	8,67	15,85	35,62
51,78	7,83	18,05	34,99
52,14	7,84	17,24	39,34
54,94	8,85	20,52	41,50
59,18	9,61	19,18	45,58
62,22	10,67	19,03	41,08
63,62	11,04	21,45	40,54

65,01	11,85	22,25	42,75
67,78	12,94	24,75	43,89
71,45	14,24	25,03	41,95
75,24	15,67	27,87	44,06
77,38	16,33	30,48	46,77

Розв'язання

1-й крок:

нормалізуємо змінні x_1, x_2, x_3 економетричної моделі за формулою:

$$x_j^* = \frac{(x_j - \bar{x}_j)}{\sqrt{n \cdot \sigma_j}},$$

де $n = 20$ – кількість спостережень; $m = 3$ – кількість незалежних

змінних; $(j = 1, m)$; $\bar{x}_j$ – середня арифметична незалежної змінної x_j ; σ_j –

середнє квадратичне відхилення незалежної змінної x_j .

Обчислюємо середні арифметичні значення:

$$\bar{x}_1 = 9,3505; \quad \bar{x}_2 = 18,8740; \quad \bar{x}_3 = 37,7880.$$

Обчислюємо середні квадратичні відхилення:

$$\sigma_1 = 3,2841; \quad \sigma_2 = 4,9763; \quad \sigma_3 = 5,5017.$$

Будуємо матрицю нормалізованих змінних X^* .

2-й крок:

Будуємо кореляційну матрицю (формула 4.2):

$$R = X^{*T} \cdot X^* = \begin{pmatrix} 1 & 0,9777 & 0,8627 \\ 0,9777 & 1 & 0,8750 \\ 0,8627 & 0,8750 & 1 \end{pmatrix}$$

3-й крок:

Обчислюємо визначник кореляційної матриці R :

$$|R| = 0,010276$$

Обчислимо експериментальне значення χ^2 — критерію (формула 4.3):

$$\chi_{експ}^2 = - \left[n - 1 - \frac{1}{6}(2m + 5) \right] \ln |R| = - \left(20 - 1 - \frac{2 \cdot 3 + 5}{6} \right) \cdot \ln 0,010276 = 78,588$$

Порівняємо значення $\chi_{експ}^2$ з табличним при $\frac{1}{2}m(m-1) = 3$ ступенях свободи та рівні значущості $\alpha = 0,05$: $\chi_{табл}^2(0,05; 3) = 7,8147$.

Так як $\chi_{експ}^2 > \chi_{табл}^2$, то в масиві незалежних змінних існує мультиколінеарність у сукупності.

4-й крок:

визначимо матрицю похибок (формула 4.4):

$$C = R^{-1} = \begin{pmatrix} 22,8036 & -21,6877 & -0,6946 \\ -21,6877 & 24,8940 & -3,0737 \\ -0,6946 & -3,0737 & 4,2888 \end{pmatrix}$$

5-й крок:

Розрахуємо експериментальні значення F — критеріїв (формула 4.5):

$$F_1 = 185,33; \quad F_2 = 203,10; \quad F_3 = 27,95$$

Розраховані значення критеріїв порівняємо з табличним при рівні значущості $\alpha = 0,05$ та ступенях свободи $n - m = 20 - 3 = 17$; $m - 1 = 3 - 1 = 2$:

$$F_{табл} = F(0,05; 17; 2) = 19,44$$

Оскільки кожне експериментальне значення більше від

табличного: $F_1 > F_{табл}$; $F_2 > F_{табл}$; $F_3 > F_{табл}$; робимо висновок, що кожна незалежна змінна мультиколінеарна з іншими. Визначимо коефіцієнти детермінації для кожної змінної

$$R_1^2 = 0,9562; \quad R_2^2 = 0,9598; \quad R_3^2 = 0,7668$$

6-й крок: знайдемо часткові коефіцієнти кореляції (формула 4.7), які характеризують щільність зв'язку між двома змінними за умови, що інші змінні не впливають на цей зв'язок (існування парної мультиколінеарності):

$$r_{12} = 0,9103; \quad r_{13} = 0,0702; \quad r_{23} = 0,2975.$$

Якщо порівняти абсолютні значення часткових і парних коефіцієнтів, то можна побачити, що перші значно менші, ніж останні. Тому на основі знання часткових коефіцієнтів кореляції висновок про мультиколінеарність робити неможливо. Для цього необхідно ще виконати 7-й крок.

7-й крок:

Розрахуємо експериментальні значення t — критеріїв:

$$t_{12} = 9,065; \quad t_{13} = 0,290; \quad t_{23} = 1,285$$

Розраховані значення критеріїв t_{kj} порівняємо з табличним при рівні значущості $\alpha = 0,05$ та $n - m = 17$ ступенях свободи:

$$t_{табл} = t(0,05; 17) = 2,110.$$

Оскільки $t_{12} > t_{табл}$, $t_{13} < t_{табл}$, $t_{23} < t_{табл}$ то між першою та другою незалежними змінними існує мультиколінеарність.

Висновок: у масиві незалежних змінних існує мультиколінеарність. Доцільно виключити з масиву незалежних змінних або змінну x_1 або змінну x_2 та будувати двохфакторну лінійну модель.

Завдання для самостійної роботи

1. Поняття про мультиколінеарність та її вплив на оцінку параметрів моделі
2. Тестування наявності мультиколінеарності
3. Алгоритм Феррара-Глобе

Таблиці

Таблиця 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	3969	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	2036	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	0940	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	0277	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	0158	0154	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

x	4,0	4,2	4,4	4,6	4,8	5,0
$\varphi(x)$	0,00013	0,0000589	0,0000249	0,0000101	0,0000040	0,0000015

Таблиця 2. Значення функції Лапласа $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$.

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	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	41308	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	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	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	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	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997

x	4,0	4,2	4,4	4,6	4,8	5,0
$\Phi_0(x)$	0,4999683	0,4999867	0,4999946	0,4999979	0,4999992	0,4999997

Таблиця 3. Квантилі хі-квадрат розподілу $\chi^2_{p,k}$

k	P										
	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995
1	0,044	0,032	0,001	0,004	0,016	0,46	2,71	3,84	5,02	6,63	7,88
2	0,010	0,020	0,051	0,103	0,211	1,39	4,61	5,99	7,38	9,21	10,6
3	0,072	0,115	0,216	0,352	0,584	2,37	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	3,36	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	4,35	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	5,35	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	6,35	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	7,34	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	8,34	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	9,34	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	10,3	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	11,3	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	12,3	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	13,3	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	14,3	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	15,3	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	16,3	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	17,3	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,4	11,7	18,3	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	19,3	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	20,3	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	21,3	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	14,8	22,3	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	23,3	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	24,3	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	25,3	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	26,3	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	27,3	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	28,3	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	29,3	40,3	43,8	47,0	50,9	53,7
35	17,2	18,5	20,6	22,5	24,8	34,3	46,1	49,8	53,2	57,3	60,3
40	20,7	22,2	24,4	26,5	29,1	39,3	51,8	55,8	59,3	63,7	66,8
45	24,3	25,9	28,4	30,6	33,4	44,3	57,5	61,7	65,4	70,0	73,2
50	28,0	29,7	32,4	34,8	37,7	49,3	63,2	67,5	71,4	76,2	79,5
75	47,2	49,5	52,9	56,1	59,8	74,3	91,1	96,2	100,8	106,4	110,3
100	67,3	70,1	74,2	77,9	82,4	99,3	118,5	124,3	129,6	135,6	140,2

Таблиця 4. Квантилі розподілу Стьюдента $t_{y,k}$.

k	P						
	0,750	0,900	0,950	0,975	0,990	0,995	0,999
1	1,000	3,078	6,314	12,71	31,82	63,66	318
2	0,816	1,886	2,920	4,303	6,965	9,925	22,3
3	0,765	1,638	2,353	3,182	4,541	5,841	10,2
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,090

Таблиця 5. Квантилі розподілу Фішера-Снедекора $F_p(k_1, k_2)$

(k_1 — число ступенів вільності більшої дисперсії, k_2 — число ступенів вільності меншої дисперсії)

$$p = 0,95$$

k_2	k_1								
	1	2	3	4	5	6	7	8	9
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	16,13	9,55	9,78	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,75	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,85	3,98	3,59	3,36	3,20	3,09	3,01	2,93	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,30	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,37	2,30
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

$(p = 0,95)$

k_2	k_1								
	10	12	15	20	24	30	40	60	120
1	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3
2	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49
3	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55
4	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66
5	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40
6	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70
7	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27
8	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97
9	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75
10	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58
11	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45
12	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34
13	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25
14	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18
15	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11
16	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06
17	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01
18	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97
19	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93
20	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90
21	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87
22	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84
23	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81
24	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79
25	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77
26	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75
27	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73
28	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71
29	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70
30	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68
40	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58
60	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47
120	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35
∞	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22

$$p = 0,99$$

k_2	k_1								
	1	2	3	4	5	6	7	8	9
1	4052	4999,5	5403	5625	5764	5859	5928	5982	6022
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07
40	7,31	5,18	4,31	3,53	3,51	3,29	3,12	2,99	2,89
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41

$(p = 0,99)$

k_2	k_1								
	10	12	15	20	24	30	40	60	120
1	6056	6106	6157	6209	6235	6261	6287	6313	6339
2	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49
3	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22
4	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56
5	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11
6	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97
7	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74
8	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95
9	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40
10	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00
11	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69
12	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45
13	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25
14	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09
15	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96
16	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84
17	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75
18	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66
19	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58
20	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52
21	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46
22	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40
23	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35
24	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31
25	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27
26	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23
27	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20
28	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17
29	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14
30	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11
40	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92
60	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73
120	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53
∞	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32

Список використаних джерел

Основна

1. Васильків І.М. В Основи теорії ймовірностей і математичної статистики : навч. посібник. – Львів : ЛНУ імені Івана Франка, 2020. – 184 с. ISBN 978–617–10–0354–5
2. Теорія ймовірностей та математична статистика (конспект лекцій + тести) : навчальний посібник. Вид. 2-ге, допов. / Я.Т.Соловко, П.Г.Остафійчук, О.З.Гарпуль, С.А.Войтик. – Івано-Франківськ: Репозитарій / ЗВО «Університет Короля Данила», 2021. – 150 с.
3. Теорія ймовірностей і математична статистика: Навчальний посібник / М. К.Бугір – Тернопіль, «Підручники й посібники», 2020 р. — 404 с.
4. Дюженкова О.Ю., Дудкін М.Є., Степахо І.В. Теорія ймовірностей. Навчальний посібник. К.: НТУУ «КПІ ім. Ігоря Сікорського», 2021.- 103 с

Допоміжна

1. Теорія ймовірностей і математична статистика: Навчальний посібник / Барковський В.В., Барковська Н.В., Лопатін О.К. – Київ: Центр навчальної літератури, 2019. — 424 с.
2. Теорія ймовірностей і математична статистика: Посібник з розв'язування задач / Г.І. Кармелюк – Київ: Центр навчальної літератури, 2019. — 576 с.
3. Теорія ймовірностей і математична статистика: Навч.-метод. посібник. у 2 ч. Ч. І. Теорія ймовірностей / В. І.Жлуктенко , С. І. Наконечний – К.: КНЕУ, 2019. — 304 с
4. Синявська О.О. Методичні вказівки та варіанти типово-розрахункових робіт з теорії ймовірностей для студентів математичного факультету. Ужгород: ДВНЗ «УжНУ», 2019. 58 с.
5. О.В.Козьменко, О.В.Кузьменко. Економіко-математичні методи та моделі (економетрика). Навчальний посібник, 2019. Видавництво Університетська книга, 406 с. ISBN 978-966-680-722-2
6. Дзямко Вікторія. ПРИКЛАДНА СПРЯМОВАНІСТЬ ВИКЛАДАННЯ ВИЩОЇ МАТЕМАТИКИ, МАТЕМАТИЧНОЇ СТАТИСТИКИ СТУДЕНТАМ ЕКОНОМІЧНОГО ПРОФІЛЮ ЗВО. Наука і техніка сьогодні. Серія: право, економіка, педагогіка, техніка, фізико-математичні науки- №8(36).-С.594-603

[https://doi.org/10.52058/2786-6025-2024-8\(36\)-594-603](https://doi.org/10.52058/2786-6025-2024-8(36)-594-603)