Міністерство освіти і науки України

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ
УНІВЕРСИТЕТ
БУДІВНИЦТВА ТА АРХІТЕКТУРИ

О.М. Стасенко

ТЕКСТИ ЛЕКЦІЙ

З КУРСУ «Математика для економістів

(теорія ймовірності і математична статистика)**»**

для студентів спеціальності 051 «Економіка»

Харків 2016

Міністерство освіти і науки України

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА ТА АРХІТЕКТУРИ**

О.М. Стасенко

МАТЕМАТИКА ДЛЯ ЕКОНОМІСТІВ
(теорія ймовірності і математична статистика)

Рекомендовано
науково-методичною радою університету
в якості текстів лекцій
для студентів спеціальності 051 «Економіка»

Харків 2016

УДК -

Рецензенти:
А.А. Янцевич, доктор фіз.-мат. наук, професор кафедри математичних методів в економіці Харківського національного університету ім. В.Н. Каразіна
Є.В. Поклонський, канд. фіз.-мат. наук, доцент кафедри вищої математики Харківського національного університету будівництва та архітектури (ХНУБА)

Рекомендовано кафедрою вищої математики,
протокол № 3 від 24.10.2016
Затверджено науково- методичною радою університету,
протокол № 3 від 23.11.2016

Автори: О.М. Стасенко

Стасенко О.М. Математика для економістів (теорія ймовірності і математична статистика): Тексти лекцій. – Х.: ХНУБА, 2016. – 90с.

Викладено теоретичні та практичні основи таких розділів вищої математики: лінійна алгебра, аналітична геометрія, вступ до аналізу, диференціальна та інтегральне числення, функції однієї змінної, диференціальні рівняння, ряди.

Теоретичний матеріал супроводжується великою кількістю прикладів, які подаються з розв’язанням. Запропоновано питання для самоперевірки.

Видання розроблено відповідно діючої програми з «Математики для економістів(вищої математики)» для студентів спеціальності «Економіка».

Призначається для студентів технічних вишів.

Іл. 76; бібліогр.: 19 назв.

© О.М. Стасенко, 2016

ЛЕКЦІЯ 1

1.1 Елементи комбінаторики

1.2 Класифікація подій. Основні поняття та визначення

1.1 Елементи комбінаторики

На практиці часто доводиться мати справу з задачами, в яких потрібно підрахувати число всіх можливих способів розташування деяких предметів або число всіх можливих способів здійснення деякої дії. Задачі такого типу називаються комбінаторними. Зокрема, комбінаторні міркування лежать в основі розв'язання багатьох задач теорії ймовірностей – важливого розділу математики, присвяченого вивченню випадкових явищ. В основі розв'язання таких задач лежать наступні два правила.

1. Правило додавання. Якщо деякий елемент A можна вибрати m способами, а другий елемент B – n способами, які виключають один одного, то вибір будь-якого одного з цих елементів (або A, або B) можна здійснити $m + n$ способами.

2. Правило множення. Якщо елемент A можна вибрати m способами і якщо після кожного такого вибору елемент B можна вибрати n способами, то вибір пари елементів (A,B) в указаному порядку можна здійснити $m \cdot n$ способами.

Слід відзначити, що ці правила легко узагальнюються на випадок довільної скінченої кількості елементів.

Приклад. У розіграшу першості країни з футболу приймають участь 16 команд. Скількома способами можуть бути розподілені золота та срібна медалі?

Розв'язання. Золоту медаль може отримати одна з 16 команд. Після того як виявлено власника золотої медалі, срібну медаль може мати одна із 15 команд. Отже, загальне число способів, якими можуть бути розподілені золота та срібна медалі, дорівнює $15 \cdot 16 = 240$.

Різні групи, що складені з яких-небудь предметів і відрізняються одна від одної або порядком цих предметів, або самими предметами, називаються взагалі сполуками.

Предмети, з яких складаються сполуки, називаються *елементами***.**

Серед можливих сполук виділимо наступні: *розміщення, перестановки, комбінації (сполучення).*

Розміщеннями з n *різних елементів по* k називаються усілякі групи, які містять k елементів, узятих з даних n елементів, і які відрізняються одна від одної або складом елементів, або їх порядком.

Число розміщень із n різних елементів по k без повторень позначається символом A_n^k і обчислюється за формулою

$$A_n^k = n(n-1)(n-2)...(n-k+1) = \frac{n!}{(n-k)!}. \qquad (1.1)$$

Число розміщень із n різних елементів по k з повтореннями позначається символом $\overline{A}_n^k$ і обчислюється за формулою

$$\overline{A}_n^k = n^k. \qquad (1.2)$$

Приклад. Скільки двозначних чисел можна скласти з цифр 1, 2, 3, 4 (без повторення цифр)?

Розв'язання. Складемо двозначні числа, які будуть відрізнятися одне від одного лише цифрами: 12, 13, 14, 23, 24, 34. Отримано шість чисел. Змінюючи цифри місцями в кожному числі, отримаємо нові числа: 21, 31, 41, 32, 42, 43. Таким чином, з чотирьох цифр (без повторень) складено 12 різних двозначних чисел. Інакше, з чотирьох елементів по два складено всього 12 розміщень: A_4^2, за формулою (1.1) отримаємо той самий результат: $A_4^2 = 4 \cdot 3 = 12$.

Приклад. До профкому обрано дев'ять осіб. Із них потрібно обрати голову та його замісника. Скількома способами це можна зробити?

Розв'язання. Задача зводиться до знаходження числа елементів по два, бо тут істотно й те, хто буде обраний до керівництва профкому, і те, як розподіляться обов'язки між обраними. Таким чином, за формулою (1.1) шукана кількість способів

$$A_9^2 = 9 \cdot 8 = 71.$$

Розміщення з n елементів по n, тобто всілякі групи з n різних елементів, які відрізняються одна від одної лише порядком елементів, називаються *перестановками.*

Число перестановок з n різних елементів без повторень позначається символом P_n і обчислюється за формулою

$$P_n = A_n^n = n!, \qquad (1.3)$$

де

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n.$$

Зауваження. За означенням прийнято вважати,що $0! = 1$.

Число перестановок з n різних елементів з повторенням, які можна скласти з k_1 елементів першого типу, k_2 елементів другого типу, ..., k_m елементів m-го типу, обчислюється за формулою

$$P_n(k_1, k_2, \ldots, k_m) = \frac{n!}{k_1! k_2! \ldots k_m!}, \qquad (1.4)$$

де

$$k_1 + k_2 + \ldots + k_m = n.$$

Приклад.Скільки різних чотиризначних чисел можна скласти з цифр 0, 1, 2, 3, якщо кожна цифра в зображенні числа зустрічається один раз?

Розв'язання. Чотиризначне число може бути подане як деяка перестановка з цифр 0, 1, 2, 3, в якій перша цифра відмінна від нуля. Оскільки число перестановок з чотирьох цифр дорівнює $P_4 = 4!$ і з них $P_3 = 3!$ перестановок починаються з нуля, то шукана кількість чисел дорівнює

$$P_4 - P_3 = 4! - 3! = 4 \cdot 3! - 3! = (4-1) \cdot 3! = 3 \cdot 3! = 3 \cdot 2 \cdot 1 \cdot 3 = 18.$$

Комбінаціями (сполученнями) з n різних елементів по k називаються усілякі групи, які містять k елементів, узятих з даних n елементів, і які відрізняються одна від одної принаймні одним елементом (порядок при цьому не враховується).

Число комбінацій із n різних елементів по k без повторень позначається символом C_n^k і обчислюється за формулою

$$C_n^k = \frac{A_n^k}{P_k} = \frac{n(n-1)(n-2)...(n-k+1)}{1\cdot 2\cdot 3\cdot ...\cdot k} = \frac{n!}{k!(n-k)!}. \qquad (1.5)$$

Число комбінацій без повторень із n різних елементів по k дорівнює числу комбінацій із цих самих n різних елементів по k, тобто

$$C_n^k = C_n^{n-k}. \qquad (1.6)$$

Число комбінацій із n різних елементів по k з повторенням позначається символом $\overline{C}_n^k$ і обчислюється за формулою

$$\overline{C}_n^k = C_{n+k-1}^k = \frac{(n+k-1)!}{k!(n-1)!}. \qquad (1.7)$$

Приклад. Скільки різних правильних дробів можна скласти з чисел 1, 2, 3, 5, 7 ,11, 13, беручи їх попарно?

Розв'язання. Різних пар з даних чисел, перший елемент яких менший за другий, буде, очевидно, стільки, скільки можна скласти комбінацій з семи елементів по два. Звідси за формулою (1.5) отримаємо шукане число:

$$C_7^2 = \frac{7\cdot 6}{1\cdot 2} = 21.$$

Приклад. Скількома способами можна розсадити чотирьох студентів на 25 місцях?

Розв'язання. Шукане число способів дорівнює числу розміщень з 25 по чотири, тобто

$$A_{25}^4 = 25\cdot 24\cdot 23\cdot 22 = 303600 \ .$$

Приклад.Скількома способами можна розмістити на полиці чотири книжки?

Розв'язання. Мова йде про всілякі групи з чотирьох різних елементів, які відрізняються одна від одної лише порядком. Тому шукане число способів можна знайти за формулою (1.3)

$$P_4 = 4! = 1\cdot 2\cdot 3\cdot 4 = 24.$$

Приклад. У ящику шість білих і вісім чорних кульок. Із нього одночасно виймають дві кульки одного кольору. Скількома способами можна це зробити?

Розв'язання. Дві білих кульки можна вийняти з ящика $C_6^2 = \frac{6\cdot 5}{1\cdot 2} = 15$ способами, а дві чорні кульки - $C_8^2 = \frac{8\cdot 7}{1\cdot 2} = 28$ способами. Оскільки з ящика виймають одночасно дві кульки одного кольору, то це можуть бути або дві білі, або дві чорні кульки, тому за правилом додавання шуканих способів.

1.2 Класифікація подій. Основні поняття та визначення

Основні поняття теорії ймовірностей - подія і ймовірність.

Подія – це результат випробування.

Ймовірність події – це числова міра степеня об'єктивної можливості цієї події.

Подія називається *випадковою*, якщо при здійсненні певної сукупності умов вона може або настати, або не настати. Події прийнято позначати буквами *A, B, C* і т.ін., а їх ймовірності – буквою *P*, наприклад, ймовірність події *A* позначається *P(A).*

Вірогідною називається подія, яка в результаті випробування обов'язково повинна відбутися (настати).

Неможливою називається подія, яка в результаті випробування не може відбутися.

Події називаються *несумісними*, якщо поява однієї з них виключає можливості появи інших подій в одному і тому самому випробуванні.

Дві несумісних події називаються *протилежними*, якщо одна з них повинна обов'язково відбутися.

Подія протилежна події *A* позначається $\overline{A}$.

Події називаються *сумісними*, якщо поява однієї з них не виключає можливості появи інших подій в одному і тому самому випробуванні.

Декілька подій утворюють *повну групу*, якщо в результаті випробування настане хоча б одна з них. Інакше кажучи, поява хоча б однієї з подій повної групи є вірогідною подією.

Зокрема, якщо події, які утворюють повну групу, попарно несумісні, то в результаті випробування з'явиться одна і тільки одна з цих подій.

Події називаються *рівноможливими*, якщо є основа вважати, що ні одна з них не є більш можливою, ніж інша.

Елементарні наслідки, в яких подія, яка нас цікавить, настає, називаються *сприятливими цій події.*

Питання для самоперевірки

1 Що називають сполуками?

2 Які види сполук розрізняють?

3 Які сполуки називають розміщеннями, перестановками, комбінаціями?

4 Наведіть формули для обчислення числа розміщень, перестановок та комбінацій з повторенням і без повторення елементів.

ЛЕКЦІЯ 2

2.1 Класична та статистична ймовірність, її властивості

2.2 Відносна частота появи події. Геометрична ймовірність

2.1 Класична та статистична ймовірність, її властивості

Імовірність події A – це відношення числа сприятливих цій події наслідків до загального числа всіх рівно можливих несумісних елементарних наслідків, які утворюють повну групу.

Ймовірність події *A* обчислюється за формулою

$$P(A) = \frac{m}{n}, \tag{2.1}$$

де

n – загальне число наслідків,

m – число наслідків, сприятливих появі події A.

Властивості ймовірності:

1 Ймовірність вірогідної події дорівнює одиниці.

2 Ймовірність неможливої події дорівнює нулю.

3 Ймовірність випадкової події

$$0 < P(A) < 1.$$

Отже, ймовірність будь-якої події задовольняє подвійній нерівності $0 \le P(A) \le 1$.

Приклад. Кинуто два гральних кубика. Знайти ймовірність того, що сума очок на гранях, що випали, дорівнює семи.

Розв'язання. Введемо подію A – сума очок на гранях, що випали, дорівнює семи. Тоді за формулою (2.1)

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

На грані "першого" грального кубика, що випала, може з'явитися одне очко, два очка, ..., шість очок. Аналогічні шість елементарних наслідків можливі при киданні "другого" кубика. Кожний із наслідків кидання "першого" кубика може поєднуватися з кожним із наслідків кидання "другого". Тому загальне число можливих елементарних наслідків випробування складатиме

$$n = 6 \cdot 6 = 36.$$

Ці наслідки єдино можливі і, в силу симетрії кубиків, рівно можливі.

Сприятливими події A, яка нас цікавить, будуть наступні шість наслідків:

1) 1, 6; 1 + 6 = 7,
2) 2, 5; 2 + 5 = 7,
3) 3, 4; 3 + 4 = 7,
4) 6, 1; 6 + 1 = 7,
5) 5, 2; 5 + 2 = 7,
6) 4, 3; 4 + 3 = 7.

Таким чином, шукана ймовірність дорівнює

$$P(A) = \frac{6}{36} = \frac{1}{6}.$$

Приклад. Кинуто два гральних кубика. Знайти ймовірність наступних подій:

а) сума очок, що випали, дорівнює восьми, а різниця – чотирьом;

б) сума очок, що випали, дорівнює восьми, якщо відомо, що їх різниця дорівнює чотирьом.

Розв'язання.

а) Введемо подію A – сума очок, що випали, дорівнює восьми, а різниця – чотирьом.

Розмірковуючи далі таким самим чином, як і в попередній задачі, знаходимо $n = \overline{A}_6^2 = 6^2 = 36$. Наслідків, сприятливих події A, є два (m = 2):

1) 2, 6; 2 + 6 = 8, 6 - 2 = 4,

2) 6, 2; 6 + 2 = 8, 6 - 2 = 4.

Тоді за формулою (4.8) шукана ймовірність

$$P(A) = \frac{2}{36} = \frac{1}{18}.$$

б) Введемо подію B – сума очок, що випали, дорівнює восьми, якщо відомо, що їх різниця дорівнює чотирьом.

Звертаємо увагу на те, що за умовою відомо – різниця очок, що випали, дорівнює чотирьом. Тому загальне число наслідків n = 4:

1) 1, 5; 5 – 1 = 4,

2) 2, 6; 6 – 2 = 4,

3) 5, 1; 5 – 1 = 4,

4) 6, 2; 6 – 2 = 4.

Наслідків, сприятливих події B,є два (m = 2):

1) 2, 6; 2 + 6 = 8,

2) 6, 2; 6 + 2 = 8.

Таким чином, шукана ймовірність

$$P(A) = \frac{2}{4} = \frac{1}{2}.$$

2.2 Відносна частота появи події. Геометрична ймовірність

Статистична ймовірність (відносна частота) події A – це відношення числа випробувань, в яких подія A настала, до загального числа фактично проведених випробувань.

Відносна частота події A обчислюється за формулою

$$W(A) = \frac{m}{n}, \qquad (2.2)$$

де

n – загальне число випробувань,

m – число випробувань, в яких подія A настала.

Відносна частота має такі самі властивості, що і ймовірність.

Зауваження:

1 Відносна частота може бути обчислена лише після проведення випробовувань. а ймовірність – до проведення випробовувань.

2 Відносна частота зміцнюється мало (тим менше, чим більше проведено випробувань), коливаючись навколо деякого сталого числа (ймовірності).

3 Недоліком статистичного означення є неоднозначність статистичної ймовірності.

Для подолання недоліку класичного означення ймовірності, який полягає в тому, що воно не може бути застосовано для випробувань з нескінченним числом наслідків, вводять геометричні ймовірності – ймовірності попадання точки в область (відрізок, частину площини і т.ін.).

Якщо в область G на площині (рис. 2.1) кидається навмання точка (під цим розуміють, що точка може попасти в будь-яку частину області G з ймовірністю, пропорційною лише її площі S_g і не залежній ні від її розташування, ні від її форми), то ймовірність попадання точки в область $\underline{g}$, яка знаходиться всередині G, визначається рівністю

$$P = \frac{S_g}{S_G}. \qquad (2.3)$$

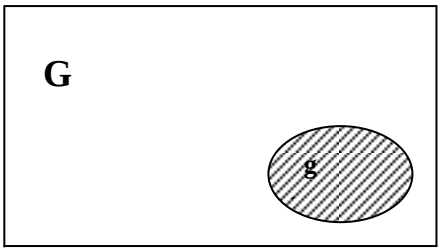


Рисунок 2.1

Аналогічно визначається ймовірність попадання точки у тривимірний об’єм і на відрізок.

Приклад. Під час випробування партії приладів відносна частота непридатних приладів виявилася рівною 0,1. Знайти число придатних приладів, якщо всього було перевірено 200 штук.

Розв’язання. Скориставшись формулою (2.2), знаходимо кількість непридатних приладів, які містяться у випробуваній партії:

$$m = m \cdot w(A) = 200 \cdot 0{,}1 = 20$$

Тоді число придатних приладів в партії 200 – 20 = 180.

Приклад. Усередину круга радіуса R навмання кинута точка. Знайти ймовірність того, що точка попаде усередину вписаного в круг квадрата. Припускається, що ймовірність попадання точки в квадрат пропорційна площі квадрата і не залежить від його розташування відносно круга.

Розв'язання. Площа круга $S_{кр} = \pi R^2$, площа квадрата $S_{кв} = 2R^2$ (рис.20.2). Отже, за формулою (2.3) шукана ймовірність $P = \frac{S_{кв}}{S_{кр}} = \frac{2R^2}{\pi R^2} = \frac{2}{\pi}$.

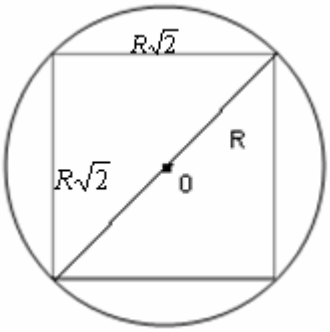


Рисунок 2.2

Приклад. Відомо, що телефонний дзвінок повинен відбутися від 11 год. до 11 год. 30 хв. Яка ймовірність того, що дзвінок відбудеться в останні 10 хв. указаного проміжку часу, якщо момент дзвінка випадковий?

Розв'язання. Скористаємося геометричною схемою. Для цього проміжок часу від 11 год. до 11 год. 30 хв. подамо у вигляді відрізка *AB* довжиною в 30 одиниць, а проміжок часу від 11 год. 20 хв. до 11 год. 30 хв. – у вигляді відрізка *CB* довжиною в 10 одиниць (рис.2.3).

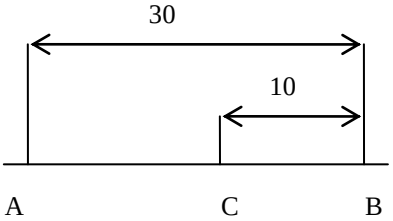


Рисунок 2.3

Випадковий дзвінок у деякий момент розглянутої півгодини зображується точкою, навмання взятою на відрізку *AB*. Тоді ймовірність того, що дзвінок відбудеться в інтервалі від 11 год. 20 хв. до 11 год. 30 хв. на отриманій схемі означає ймовірність того, що точка, навмання взята на відрізку *AB*, виявиться такою, що належить належною відрізку *CB*.

Ця ймовірність, очевидно, дорівнює

$$P = \frac{10}{30} = \frac{1}{3}.$$

Питання для самоперевірки

1 Що вивчає теорія ймовірностей?

2 Перерахуйте основні поняття теорії ймовірностей.

3 Що розуміють під випробуванням?

4 Що називають подією? Ймовірністю?

5 Яку подію називають випадковою, вірогідною, неможливою? Наведіть приклади.

6 Які події називають несумісними? Сумісними? Наведіть приклади.

7 Які події називають рівно можливими? Наведіть приклади.

8 Що являє собою повна група подій? Наведіть приклади.

9 Дайте означення класичної, статистичної та геометричної ймовірностей.

10 Сформулюйте властивості ймовірності та наведіть формулу для її обчислення.

ЛЕКЦІЯ 3

3.1 Теореми додавання ймовірностей

3.2 Теореми множення ймовірностей

До основних теорем теорії ймовірностей відносяться теореми додавання і множення ймовірностей.

3.1 Теореми додавання ймовірностей

Сумою $A+B$ двох сумісних подій A і B називають подію, яка полягає в появі або події A, або події B, або обох цих подій.

Сумою декількох сумісних подій називають подію, яка полягає в появі хоча б однієї з цих подій.

Зокрема, якщо дві події A і B – несумісні, то $A+B$ – подія, яка полягає в появі однієї з цих подій, байдуже якої.

1. Ймовірність появи однієї з двох несумісних подій, байдуже якої, дорівнює сумі ймовірностей цих подій:

$$P(A+B)=P(A)+P(B). \qquad (3.1)$$

Ця теорема справджується для будь-якого скінченого числа попарно несумісних подій, тобто

$$P(A_1+A_2+...+A_n)=P(A_1)+P(A_2)+...+P(A_n). \qquad (3.2)$$

2. Сума ймовірності подій $A_1, A_2,..., A_n$, які утворюють повну групу, дорівнює одиниці:

$$P(A_1)+P(A_2)+...+P(A_n)=1. \qquad (3.3)$$

3. Сума ймовірностей протилежних подій дорівнює одиниці:

$$P(A)+P(\overline{A})=1. \qquad (3.4)$$

Прийнято $P(A)$ і $P(\overline{A})$ позначати відповідно через p і q. Таким чином, якщо $P(A) = p$, а $P(A) = q$, то згідно з формулою (3.4)

$$p+q=1. \qquad (3.5)$$

4. Імовірність появи хоча б однієї з двох сумісних подій дорівнює сумі ймовірностей цих подій без ймовірності їх сумісної появи:

$$P(A+B)=P(A)+P(B)-P(AB). \qquad (3.6)$$

Використовуючи формулу (3.6) треба мати на увазі, що події A і B можуть бути як незалежними, так і залежними. Якщо події A і B несумісні, то їх суміщення є подією неможливою і, отже, $P(AB)=0$. Формула (3.6) для несумісних подій набуває вигляду

$$P(A+B)=P(A)+P(B),$$

тобто отримана вже відома формула (3.1).

Таким чином, формула (3.6) вірна як для сумісних так і для несумісних подій.

3.2 Теореми множення ймовірностей

Добутком двох подій A і B називають подію AB, яка полягає у сумісній появі (суміщенні) цих подій.

Добутком декількох подій називають подію, яка полягає у сумісній появі всіх цих подій.

Умовною ймовірністю $P_A(B)$ називають ймовірність події B, обчислену за припущенням, що подія A уже відбулася.

1. Імовірність сумісної появи двох подій дорівнює добутку ймовірності однієї з них на умовну ймовірність другої, обчислену за припущенням, що перша подія вже настала:

$$P(AB)=P(A)\cdot P_A(B)=P(B)\cdot P_B(A). \qquad (3.7)$$

Наслідок:

$$P(A_1A_2...A_n)=P(A_1)P_{A_1}(A_2)P_{A_1A_2}(A_3)...P_{A_1A_2...A_{n-1}}(A_n).$$

Зауважимо, що порядок, у якому розташовані події, може бути обрано будь-яким. Зокрема, для трьох подій

$$P(ABC)=P(A)P_A(B)P_{AB}(C) \qquad (3.8)$$

Подію B називають незалежною від події A, якщо поява події A не змінює ймовірності події B, тобто якщо умовна ймовірність події B дорівнює її безумовній ймовірності:

$$P_A(B)=P(B)$$

Декілька подій називають попарно незалежними, якщо кожні дві з них незалежні.

Декілька подій називають незалежними в сукупності (або просто незалежними), якщо незалежні кожні дві з них і незалежні кожна подія і всі можливі добутки решти.

2. Якщо події A і B є незалежними, то

$$P(AB)=P(A)P(B) \qquad (3.9)$$

3. Імовірність сумісної появи декількох подій, незалежних у сукупності, дорівнює добутку ймовірностей цих подій:

$$P(A_1A_2...A_n)=P(A_1)P(A_2)...P(A_n). \qquad (3.10)$$

Зауважимо, що на практиці про незалежність подій, як правило, за змістом задачі.

4. Імовірність появи хоча б однієї з подій $A_1, A_2, ..., A_n$, незалежних у сукупності, дорівнює різниці між одиницею і добутком ймовірностей протилежних подій $\overline{A_1}, \overline{A_2}, ..., \overline{A_n}$:

$$P(A) = 1 - q_1 q_2 ... q_n . \qquad (3.11)$$

Якщо події $A_1, A_2, ..., A_n$ мають однакову ймовірність появи хоча б однієї з цих подій, то

$$P(A) = 1 - q^n . \qquad (3.12)$$

Приклад. Імовірність того, що стрілець під час стрільби по мішені виб'є 10 очок, дорівнює 0,4; 9 очок – 0,2; 8 очок – 0,18; 7 очок – 0,1; 6 очок або менше – 0,12. Знайти ймовірність того, що стрілець в результаті одного пострілу виб'є не менше 7 очок (подія *A*).

Розв'язання:

1 спосіб. Подія *A* відбудеться, якщо стрілець виб'є або 7 очок (подія A_1), або 8 очок (подія A_2), або 9 очок (подія A_3), або 10 очок (подія A_4). Оскільки події A_1, A_2, A_3, A_4 несумісні, то за формулою (3.2)

$$P(A) = P(A_1) + P(A_2) + P(A_3) + P(A_4) .$$

За умовою задачі $P(A_1) = 0{,}1$, $P(A_2) = 0{,}18$, $P(A_3) = 0{,}2$ і $P(A_4) = 0{,}4$, тому шукана ймовірність

$$P(A) = 0{,}1 + 0{,}18 + 0{,}2 + 0{,}4 = 0{,}88 .$$

2 спосіб. Очевидно, що подія *A* є протилежного події $\overline{A}$ (стрілець виб'є 6 очок або менше). Оскільки, за умовою задачі $P(\overline{A}) = 0{,}12$, то за формулою (3.4) будемо мати:

$$P(A) = 1 - P(\overline{A}) = 1 - 0{,}12 = 0{,}88 .$$

Як бачимо, одержали такий самий результат.

Приклад. Знайти ймовірність того, що навмання взяте двозначне число виявиться кратним або 2, або 5, або і 2, і 5 одночасно (подія *A*).

Розв'язання. Введемо події: A_1 – навмання взяте двозначне число кратне 2, A_2 – навмання взяте двозначне число кратне 5. Тоді подію А можна розглядати як суму подій A_1 і A_2, тобто $A = A_1 + A_2$. Оскільки події A_1 і A_2 сумісні, то за формулою (3.6)

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

Двозначні числа – це 10, 11, ..., 98, 99. Усіх їх 90. Очевидно 45 із них кратні 2 (сприятливі події A_1), 18 кратні 5 (сприятливі події A_2) і, нарешті, 9 кратні і 2, і 5 одночасно (сприятливі події $A_1 A_2$). Тому за формулою (3.1)

$$P(A_1) = \frac{45}{90} = 0{,}5; \;\; P(A_2) = \frac{18}{90} = 0{,}2; \;\; P(A_1 A_2) = \frac{9}{90} = 0{,}1$$

і отже, шукана ймовірність

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

Приклад. У разі збільшення напруги може відбутися розрив електричного кола внаслідок виходу з ладу одного з трьох послідовно з'єднаних елементів. Ймовірності відмови елементів відповідно дорівнюють 0,2, 0,3 і 0,4. Обчислити ймовірність того, що розриву кола не відбудеться.

Розв'язання. Нехай події $\overline{A_1}$, $\overline{A_2}$, $\overline{A_3}$ означають вихід із ладу відповідно першого, другого і третього елементів. Їх ймовірність за умовою відповідно дорівнює:

$$P(\overline{A_1}) = q_1 = 0{,}2; P(\overline{A_2}) = q_2 = 0{,}3; P(\overline{A_3}) = q_3 = 0{,}4.$$

Тоді ймовірність протилежних подій A_1, A_2, A_3 (відповідно перший, другий і третій елементи не виходять із ладу, тобто працюють) за формулою (3.5) дорівнює:

$$p_1 = 1 - q_1 = 0{,}8; p_2 = 1 - q_2 = 0{,}7; p_3 = 1 - q_3 = 0{,}6.$$

Подія *A*, яка полягає в тому, що розриву кола не відбудеться, є суміщення незалежних подій A_1, A_2, A_3, тобто $A = A_1 A_2 A_3$.

Отже, за формулою (21.10) шукана ймовірність

$$P(A) = P(A_1)P(A_2)P(A_3) = p_1 p_2 p_3 = 0{,}8 \cdot 0{,}7 \cdot 0{,}6 = 0{,}336.$$

Приклад. Яка ймовірність того, що взятий навмання виріб виявиться першосортним, якщо відомо, що 3% усієї продукції складають нестандартні вироби, а 75% стандартних виробів задовольняють вимогам першого сорту?

Розв'язання. Позначимо через A_1 подію, яка полягає у тому, що виріб виявиться стандартним, а через A_2 – подію, яка полягає у тому, що стандартний виріб задовольняє вимогам першого сорту. Тоді подія *A* (взятий навмання виріб виявиться першосортним) складається з суміщення подій A_1 і A_2, тобто $A = A_1 \cdot A_2$. Оскільки події A_1 і A_2 залежні (виріб буде першосортним за умови, що він стандартний), то за формулою (3.7)

$$P(A) = P(A_1 \cdot A_2) = P(A_1) \cdot P_{A_1}(A_2).$$

За умовою задачі відомо, що 75% стандартних виробів задовольняють вимогам першого сорту, отже, $P_{A_1}(A_2) = 75\% = 0{,}75$. Крім того, відомо, що 3% усієї продукції складають нестандартні вироби. Тому стандартні вироби складають 100% - 3%=97% усієї продукції. Отже, ймовірність того, що взятий виріб виявиться стандартним, буде такою:

$$P(A_1) = 97\% = 0{,}97.$$

Таким чином, шукана ймовірність

$$P(A) = 0{,}97 \cdot 0{,}75 = 0{,}7275.$$

Як бачимо з розглянутих задач, для знаходження шуканої ймовірності в кожній конкретній задачі застосовувалася лише одна з теорем: або теорема додавання, або теорема множення. Однак на практиці часто зустрічаються задачі, для розв'язання яких доводиться застосовувати одночасно ці теореми та їх наслідки. Покажемо це на прикладах.

Приклад. Три станки працюють незалежно один від одного. Ймовірність безперебійної роботи першого станка протягом деякого часу *T* дорівнює

$p_1 = 0{,}9$, другого – $p_2 = 0{,}8$, третього – $p_3 = 0{,}7$. Яка ймовірність того, що протягом указаного проміжку часу

а) усі три станки працюють безперебійно (подія *A*);

б) лише два станки працюють безперебійно (подія *B*);

в) лише один станок працює безперебійно (подія *C*);

г) хоча б один з станків працює безперебійно (подія *D*);

д) жоден станок не працює (подія *E*)?

Розв'язання. Введемо наступні події:

A_1 – працює перший станок,

A_2 – працює другий станок;

A_3 – працює третій станок.

Тоді протилежні їм події:

$\overline{A_1}$ – не працює перший станок;

$\overline{A_2}$ – не працює другий станок;

$\overline{A_3}$ – не працює третій станок.

За умовою задачі $P(A_1) = p_1 = 0{,}9; P(A_2) = p_2 = 0{,}8; P(A_3) = p_3 = 0{,}7$, отже,

$$P(\overline{A_1}) = q_1 = 1 - 0{,}9 = 0{,}1; P(\overline{A_2}) = q_2 = 1 - 0{,}8 = 0{,}2; P(\overline{A_3}) = q_3 = 1 - 0{,}7 = 0{,}3.$$

а) очевидно, що подія $A = A_1 A_2 A_3$. Оскільки за умовою задачі події A_1, A_2, A_3 незалежні (станки працюють незалежно один від одного), то за формулою (3.10)

$$P(A) = P(A_1)P(A_2)P(A_3) = p_1 p_2 p_3.$$

Виконуючи обчислення далі, знаходимо:

$$P(A) = 0{,}9 \cdot 0{,}8 \cdot 0{,}7 = 0{,}504.$$

б) Подія *B* полягає у тому, що працюють лише два станки з трьох. Це можливо в разі появи однієї з наступних подій: або $A_1 A_2 \overline{A_3}$, або $A_1 \overline{A_2} A_3$, або $\overline{A_1} A_2 A_3$. Отже,

$$B = A_1 A_2 \overline{A_3} + A_1 \overline{A_2} A_3 + \overline{A_1} A_2 A_3.$$

Оскільки події A_1, A_2, A_3, $\overline{A_1}$, $\overline{A_2}$, і $\overline{A_3}$ незалежні в сукупності, а події $A_1 A_2 \overline{A_3}$, $A_1 \overline{A_2} A_3$ і $\overline{A_1} A_2 A_3$ несумісні, то використовуючи послідовно формули (21.2) і (21.10), знаходимо:

$$P(B) = P(A_1 A_2 \overline{A_3} + A_1 \overline{A_2} A_3 + \overline{A_1} A_2 A_3) = P(A_1 A_2 \overline{A_3}) + P(A_1 \overline{A_2} A_3) + P(\overline{A_1} A_2 A_3) =$$

$$= P(A_1)P(A_2)P(\overline{A_3}) + P(A_1)P(\overline{A_2})P(A_3) + P(\overline{A_1})P(A_2)P(A_3) = p_1 p_2 q_3 + p_1 q_2 p_3 + q_1 p_2 p_3.$$

Підставляючи в отриману формулу відповідні значення ймовірностей, остаточно будемо мати:

$$P(B) = 0{,}9 \cdot 0{,}8 \cdot 0{,}3 + 0{,}9 \cdot 0{,}2 \cdot 0{,}7 + 0{,}1 \cdot 0{,}8 \cdot 0{,}7 = 0{,}216 + 0{,}126 + 0{,}056 = 0{,}398.$$

в) Аналогічні міркування приводять до того, що

$$P(C) = P(A_1 \overline{A_2}\,\overline{A_3} + \overline{A_1} A_2 \overline{A_3} + \overline{A_1}\,\overline{A_2} A_3) = p_1 q_2 q_3 + q_1 p_2 q_3 + q_1 q_2 p_3.$$

Після підстановки до цієї формули відповідних значень ймовірностей, знаходимо:

$$P(C) = 0{,}9 \cdot 0{,}2 \cdot 0{,}3 + 0{,}1 \cdot 0{,}8 \cdot 0{,}3 + 0{,}1 \cdot 0{,}2 \cdot 0{,}7 = 0{,}054 + 0{,}024 + 0{,}014 = 0{,}092.$$

г) Імовірність події D знаходимо за формулою (3.11):

$$P(D) = 1 - q_1 q_2 q_3.$$

Після обчислень, отримаємо:

$$P(D) = 1 - 0{,}1 \cdot 0{,}2 \cdot 0{,}3 = 1 - 0{,}006 = 0{,}994.$$

Цю саму ймовірність можна знайти інакше. Очевидно, що $D = A + B + C$. Тоді

$$P(D) = P(A) + P(B) + P(C) = 0{,}504 + 0{,}398 + 0{,}092 = 0{,}994.$$

Відзначимо, що цей спосіб розв'язання задачі є нераціональним. Дійсно, якби ймовірності $P(A)$, $P(B)$ і $P(C)$ не були знайдені раніше, то обчислення $P(D)$ значно б ускладнилося.

д) Подія E полягає із суміщення незалежних подій $\overline{A_1}$, $\overline{A_2}$ і $\overline{A_3}$. Тому

$$P(E) = P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}) = P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot P(\overline{A_3}) = q_1 q_2 q_3.$$

Підставляючи числові дані, отримаємо:

$$P(E) = 0{,}1 \cdot 0{,}2 \cdot 0{,}3 = 0{,}006.$$

З іншого боку, враховуючи, що події D і E являються незалежними і, отже, $P(E) + P(D) = 1$, а $P(D)$ уже відома і дорівнює 0,994, знаходимо:

$$P(E) = 1 - 0{,}994 = 0{,}006.$$

Як бачимо, отримали такий самий результат. Зауважимо, що тут було б помилковим вважати, що протилежними являються події A і E. Дійсно,

$$P(A) + P(E) = 0{,}504 + 0{,}006 = 0{,}510 \neq 1.$$

Питання для самоперевірки

1 Які теореми вважать основними теоремами теорії ймовірностей?

2 Що розуміють під сумою подій A і B ?

3 Сформулюйте теорему додавання для несумісних подій. Наведіть формулу для обчислення .

4 Сформулюйте теорему додавання для сумісних подій. Наведіть формулу для обчислення .

5 Чому дорівнює сума ймовірностей подій, які утворюють повну групу ?

6 Чому дорівнює сума ймовірностей протилежних подій ?

7 Що розуміють під добутком подій A і B ?

8 Що називають умовною ймовірністю?

9 Сформулюйте теорему множення для незалежних подій. Наведіть формулу для обчислення .

10 Сформулюйте теорему множення для залежних подій. Наведіть формулу для обчислення.

11 Чому дорівнює ймовірність появи хоча б однієї із n незалежних подій?

12 Чому дорівнює ймовірність появи хоча б однієї із n незалежних подій $A_1, A_2, ..., A_n$, якщо ці події мають однакову ймовірність.

ЛЕКЦІЯ 4

4.1 Формули повної ймовірності та Байєса

4.1 Формула повної ймовірності. Формула Байєса

Імовірність події A, яка може настати лише за умови появи однієї з несумісних подій (гіпотез) $B_1, B_2, ..., B_n$, що утворюють повну групу, дорівнює сумі добутків ймовірностей кожної з гіпотез на відповідну умовну ймовірність події A:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + ... + P(B_n) \cdot P_{B_n}(A), \qquad (4.1)$$

де $P(B_1) + P(B_2) + ... + P(B_n) = 1.$

Рівність (4.13) називають формулою повної ймовірності.

Нехай подія A може настати лише за умови появи однієї з несумісних подій (гіпотез) $B_1, B_2, ..., B_n$, які утворюють повну групу подій. Якщо подія A вже відбулася, тоді ймовірності гіпотез можуть бути переоцінені за формулами Байєса

$$P_A(B_i) = \frac{P(B_i) \cdot P_{B_i}(A)}{P(A)}, \qquad i = 1,2,...n, \qquad (4.2)$$

де $P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + ... + P(B_n) \cdot P_{B_n}(A).$

Приклад. Імовірність одного влучення в ціль в результаті одного залпу з двох гармат дорівнює 0,38. Знайти ймовірність влучення в ціль в результаті одного пострілу з першої гармати, якщо відомо, що для другої гармати ця ймовірність дорівнює 0,8.

Розв'язання. Введемо події:

A– влучення в ціль в результаті одного залпу з двох гармат;

A_1 – влучення в ціль в результаті одного пострілу з першої гармати;

A_2 – влучення в ціль в результаті одного пострілу з другої гармати.

Тоді $A = A_1\overline{A_2} + \overline{A_1}A_2$. Оскільки події $A_1\overline{A_2}$ і $\overline{A_1}A_2$ є несумісними, а події A_1, A_2, $\overline{A_1}$, $\overline{A_2}$ - незалежними у сукупності, то

$$P(A) = P(A_1) \cdot P(\overline{A_2}) + P(\overline{A_1})P(A_2) = p_1 q_2 + q_1 p_2 .$$

За умовою задачі $P(A_2) = p_2 = 0{,}8$, тоді $P(\overline{A_2}) = q_2 = 1 - 0{,}8 = 0{,}2$. Враховуючи також, що *P(A)* = 38, а $q_1 = 1 - p_1$, будемо мати :

$$0{,}38 = 0{,}2\,p_1 + 0{,}8(1 - p_1),$$

звідки шукана ймовірність:

$$p_1 = \frac{0{,}42}{0{,}6} = 0{,}7 .$$

Приклад. Імовірність хоча б одного влучення в ціль в результаті чотирьох пострілів (подія *A*) дорівнює 0,9984. Знайти ймовірність влучення в ціль в результаті одного пострілу (подія A_1).

Розв'язання. Згідно з формулою (3.12)

$$P(A) = 1 - q^n ;$$

За умовою відомо, що $P(A)$ = 0,9984, n = 4 (кількість пострілів), тому

$$0{,}9984 = 1 - q^4 \text{ або } q^4 = 1 - 0{,}9984 = 0{,}0016,$$

звідки

$$q = 0{,}2.$$

Таким чином, $q = P(\overline{A_1}) = 0{,}2$, отже, шукана ймовірність:

$$P(A_1) = p = 1 - 0{,}2 = 0{,}8.$$

Приклад. В урну, яка містить дві кулі, опущений біла куля, після чого з неї навмання витягнули одну кулю. Знайти ймовірність того, що витягнута куля виявиться білою, якщо рівноможливі всі можливі припущення про початковий склад куль (за кольором).

Розв'язання. Позначимо через A подію – витягнули білу кулю. Можливі наступні припущення (гіпотези) про початковий склад куль: B_1 – білих куль немає, B_2 – одна біла куля, B_3 – дві білі кулі.

Отже, всього є три гіпотези, причому за умовою вони рівноймовірні, і сума ймовірностей гіпотез дорівнює одиниці (оскільки вони утворюють повну групу подій), тоді ймовірність кожної з гіпотез дорівнює $\frac{1}{3}$, тобто

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}.$$

Умовна ймовірність того, що буде витягнута біла куля, за умови, що спочатку в урні не було білих куль, $P_{B_1}(A) = \frac{1}{3}$.

Умовна ймовірність того, що буде витягнута біла куля, за умови, що спочатку в урні була одна біла куля, $P_{B_2}(A) = \frac{2}{3}$.

Умовна ймовірність того, що буде витягнута біла куля, за умови, що спочатку в урні було дві білих кулі, $P_{B_3}(A) = 1$.

Імовірність того, що буде витягнута біла куля, знаходимо за формулою повної ймовірності:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + P(B_3) \cdot P_{B_3}(A) = \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot 1 = \frac{2}{3}.$$

Приклад. Два автомати виготовляють однакові деталі, які надходять на загальний конвеєр. Продуктивність першого автомата вдвічі більше другого. Перший автомат виготовляє 70% деталей відмінної якості, другий – 85%. Навмання взята з конвеєра деталь виявилась відмінної якості. Знайти ймовірність того, що ця деталь виготовлена першим автоматом.

Розв'язання. Позначимо: подія A – деталь відмінної якості. Можна зробити два припущення: B_1 – деталь виготовлена першим автоматом, причому за умовою задачі, перший автомат виготовляє вдвічі більше деталей, ніж

другий, тому $P(B_1)=\frac{2}{3}$; B_2 – деталь виготовлена другим автоматом, причому $P(B_2)=\frac{1}{3}$.

Умовна ймовірність того, що деталь відмінної якості виготовлена першим автоматом, $P_{B_1}(A)=0{,}7$. Для другого автомату ця ймовірність дорівнює $P_{B_2}(A)=0{,}85$. Імовірність того, що взята навмання деталь виявиться відмінної якості, обчислюється за формулою повної ймовірності:

$$P(A)=P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)=\frac{2}{3}\cdot 0{,}7+\frac{1}{3}\cdot 0{,}85=0{,}75.$$

Шукана ймовірність того, що взята деталь відмінної якості, виготовлена першим автоматом, за формулою Байєса дорівнює:

$$P_A(B_1)=\frac{P(B_1)\cdot P_{B_1}(A)}{P(A)}=\frac{2/3\cdot 0{,}7}{0{,}75}=0{,}62.$$

Питання для самоперевірки

1 Запишіть формулу повної ймовірності.
2 Сформулюйте теорему Байєса.

ЛЕКЦІЯ 5

5.1 Повторні випробування. Формула Бернуллі

5.1 Повторні випробування. Формула Бернуллі

Якщо проводяться декілька випробувань, при чому ймовірність події *A* в кожному випробуванні не залежить від наслідків інших випробувань, то такі випробування називають незалежними відносно події *A*.

Будемо далі розглядати лише такі незалежні випробування, в яких подія *A* має одну й ту ж саму ймовірність.

Розглядається серія із *n* незалежних випробувань з двома можливими наслідками, в кожному з яких подія *A* може відбуватись з імовірністю *p* (випробування незалежні, якщо ймовірність будь-якого наслідку будь-якого випробування не залежить від того, які були наслідки інших випробувань). Нехай A_j ($j=1,2,\cdots,n$) позначає подію, що означає наставання події *A* у *j*-му випробуванні. Тоді кожну з 2^n елементарних подій серії можна зобразити у вигляді добутку *n* множників, кожен з яких дорівнює A_j або $\overline{A}_j$.

Теорема. Ймовірність $p_n(k)$ того, що у серії з *n* випробовувань подія настає *k* раз, задається рівністю

$$p_n(k)=C_n^k p^k(1-p)^{n-k}\quad (k=0,1,...,n).\quad (1)$$

Доведення. Події, що нас цікавить, сприяють ті елементарні події, у яких події A_j спостерігаються k раз, а події $\overline{A}_j$ – $(n-k)$ раз (наприклад, $A_1 \cdot A_2 \cdot ... \cdot A_k \cdot \overline{A}_{k+1} \cdot ... \cdot \overline{A}_n$, $A_1 \cdot A_2 \cdot ... \cdot A_{k-1} \cdot \overline{A}_k \cdot A_{k+1} \cdot \overline{A}_{k+2} \cdot ... \cdot \overline{A}_n$ і т.п.). В силу незалежності подій A_j ймовірність кожної такої елементарної події на підставі теореми множення ймовірностей дорівнює $p^k(1-p)^{n-k}$. Оскільки подібних елементарних подій буде C_n^k, то з урахуванням їх несумісності і теореми додавання ймовірностей остаточно одержимо

$$p_n(k) = C_n^k p^k (1-p)^{n-k}.$$

Позначивши через $q = 1 - p$ отримаємо $P_n(k) = C_n^k p^k q^{n-k}$.

Останній вираз називають **формулою Бернуллі**. Імовірність того, що в n незалежних випробуваннях, у кожному з яких імовірність появи події A дорівнює p ($0<p<1$), подія настане рівно $к$ раз (байдуже в якій послідовності), дорівнює

$$P_n(k) = C_n^k p^k q^{n-k} \tag{5.1}$$

де

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

Слід зауважити, що формулу Бернуллі використовують тоді, коли значення n мале, оскільки при великих значеннях n формула потребує виконання дій над дуже великими числами.

Приклад. Імовірність того, що витрата електроенергії протягом доби не перевищує установленої норми, дорівнює 0,8. Знайти ймовірність того, що в найближчі 5 діб витрата електроенергії протягом 3 діб не перевищить норму.

Розв'язання. Ймовірність нормальної витрати електроенергії протягом кожної із 5 діб є сталою і дорівнює 0,8. Отже, ймовірність перевитрати електроенергії за кожну добу також є сталою і дорівнює $q = 1 - p = 1 - 0{,}8 = 0{,}2$. Оскільки в розглянутій задачі n мале (n = 5), то шукану ймовірність знаходимо за формулою (5.1):

$$P_5(3) = C_5^3 p^3 q^2 = C_5^2 p^3 q^2.$$

Після обчислень остаточно отримаємо $P_5(3) = \frac{5 \cdot 4}{1 \cdot 2} \cdot (0{,}8)^3 \cdot (0{,}2)^2 = 0{,}2048.$

Приклад. Тридцять відсотків виробів даного підприємства є продукцією вищого ґатунку сорту. Дехто придбав 6 виробів, виготовлених на цьому підприємстві. Чому дорівнює ймовірність того, що 4 із них вищого ґатунку?

Розв'язання. Ймовірність того, що виріб, виготовлений на даному підприємстві, є вищого ґатунку, є сталою і дорівнює 0,3. Тоді ймовірність того, що виріб є виробом вищого ґатунку, також є сталою і дорівнює q = 1 - 0,3 = 0,7. Отже, ймовірність того, що з 6 придбаних виробів чотири будуть виробами вищого ґатунку, знаходимо за формулою Бернуллі:

$$P_6(4) = C_6^4 p^4 q^2 = C_6^2 p^4 q^2 = \frac{6 \cdot 5}{1 \cdot 2} \cdot (0{,}3)^4 \cdot (0{,}7)^2 \approx 0{,}06.$$

Приклад. На кожному з двох крил літака установлені по два двигуни, кожен з яких може вийти з ладу під час польоту незалежно один від одного з імовірністю p=0,1. Яка ймовірність того, що політ закінчиться нормально, якщо:

а) літак може летіти на будь-яких двох двигунах;

б) літак може летіти при умові, що на кожному крилі працює хоча б один двигун.

Розв'язання. Позначимо через A подію, яка полягає у тому, що політ закінчиться нормально.

а) Нехай D_k – подія, яка полягає у тому, що під час польоту вийдуть із ладу лише k (k= 0,1,2,3,4) двигуни. Тоді $\overline{A} = D_3 \cup D_4$, де події D_3 і D_4 несумісні. Таким чином, $P(\overline{A}) = P(D_3) + P(D_4)$. При обчисленні $P(D_3)$ і $P(D_4)$ скористаємося схемою Бернуллі при n=4 і p=0.1:

$$P(D_3)= p_4(3)= C_4^3 \cdot p^3 \cdot (1-p)= 4\cdot(0.1)^3\cdot 0.9= 0.0036;$$

$$P(D_4)= p_4(4)= C_4^4 \cdot p^4 \cdot (1-p)^0=4\cdot(0.1)^4=0.0001.$$

Тому $P(A)=1-P(\overline{A})=1-0.0036-0.0001=0.996$.

б) Нехай $П_k$ ($Л_k$) – подія, яка полягає у тому, що під час польоту на правому (лівому) крилі вийдуть із ладу лише k (k= 0,1,2) двигуни.

Тоді подію A можна представити у вигляді суми несумісних подій: $A=П_1\cdot Л_0+П_0\cdot Л_1+П_1\cdot Л_1+П_0\cdot Л_0$. При обчисленні $P(П_0)=P(Л_0)$ і $P(П_1)=P(Л_1)$ скористаємося схемою Бернуллі при n=2 і p=0.1:

$$P(П_0)=p_2(0)= C_2^0 \cdot p^0 \cdot (1-p)^2=(0.9)^2=0.81;$$

$$P(П_1)=p_2(1)= C_2^1 \cdot p\cdot (1-p)=2\cdot 0.1\cdot 0.9=0.18.$$

Таким чином, з урахуванням незалежності подій $П_k$ і $Л_j$ одержимо:

$$P(A)=P(П_1)\cdot P(Л_0)+P(П_0)\cdot P(Л_1)+P(П_1)\cdot P(Л_1)+P(П_0)\cdot P(Л_0)=$$
$$= 0.18\cdot 0.81+0.81\cdot 0.18+0.18\cdot 0.18+0.81\cdot 0.81= 0.98.$$

Питання для самоперевірки

1 Які події називають незалежними?

2 Що включає поняття "повторні випробування" ?

3 Як знайти ймовірність того, що в результаті n незалежних випробувань подія настане рівно $к$ раз?

4 Формула Бернуллі та її застосування.

ЛЕКЦІЯ 6

6.1 Граничні теореми. Локальна та інтегральна теореми Лапласа

6.1 Граничні теореми. Локальна та інтегральна теореми Лапласа

Граничні теореми. Якщо число випробувань n велике, то застосовування формули Бернуллі приводить до дуже громіздких обчислень, тому в таких випадках використовують наближені формули (формулу Пуассона, а також формули, які виражають зміст локальної та інтегральної теореми Лапласа).

1. Локальна теорема Лапласа. Якщо ймовірність p настання події A в кожному з n незалежних випробувань стала, причому $0 < p < 1$, а число випробувань досить велике, то ймовірність $P_n(k)$ того, що в цих випробуваннях подія A настане $к$ раз, наближено дорівнює

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi\,(x), \tag{6.1}$$

де

$$x = \frac{k - np}{\sqrt{npq}},$$

а функція $\varphi\,(x)$ визначається рівністю

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

Таблиця значень функції $\varphi\,(x)$ наведена в додатку А, причому лише для додатних значень φ, бо ця функція парна, тобто $\varphi\,(-x) = \varphi\,(x)$. Для значень $x > 5$ потрібно вважати, що $\varphi\,(x) \approx 0$.

На рис.6.1 зображено графік функції $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

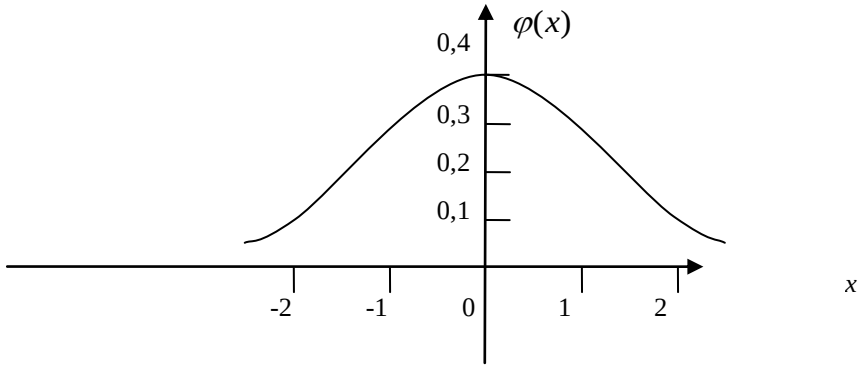


Рисунок 6.1

Для більшої наочності прийняті різні масштаби по осях. Побудована крива (крива ймовірностей) дозволяє для кожного значення $x = \frac{k - np}{\sqrt{npq}}$ знайти найближче значення $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, а звідси й $P_n(k) \approx \frac{1}{\sqrt{npq}}$ $\varphi(x)$. Ця формула дає тим більш близькі до точного значення ймовірності $P_n(k)$ результати, чим більше значення $\sqrt{npq}$. Причому в даному випадку має значення не лише n, а й значення pq. При одних і тих самих значеннях n обчислення ймовірності за формулою Лапласа більш близьке до її значення, обчисленого за формулою Бернуллі, чим більше pq до 0,25 (це значення є для pq найбільшим). У задачах зі значенням p і q, близькими до нуля, застосування формули (6.1) дає більш помітні відхилення від точних значень $P_n(k)$, які отримують за формулою Бернуллі.

У зв'язку з цим для так званих рідкісних подій (p близьке до 0) з успіхом застосовується наближена формула Пуассона.

2. Формула Пуассона. Якщо в кожному випробуванні ймовірність p настання події A є сталою і малою, а число випробувань n є досить великим, то ймовірність того, що подія A настане $к$ раз, наближено дорівнює

$$P_n(k) \approx \frac{\lambda^k e^{-\lambda}}{k!}, \qquad (6.2)$$

де $\lambda = np$.

Слід зазначити, що наближення тим краще, чим більше n і менше p. Застосування цієї формули припустиме за умови $\lambda = np \le 10$.

Для спрощення розрахунків практично замість безпосереднього обчислення ймовірності $P_n(k)$ за формулою Пуассона необхідно використовувати таблицю значень функції Пуассона. Ця функція є функцією двох змінних $к$ і $\lambda = np$.

Застосування у подібних випадках локальної теореми Лапласа приводить до неточних значень ймовірностей $P_n(k)$.

Доцільність застосування обчислень за формулою Пуассона, наприклад, для $n = 100$ і p = 0,01 показує наступна таблиця значень $P_{100}(к)$:

Таблиця 6.1

$к$	0	1	2	5	9
За формулами:					
Бернуллі	0,366	0,370	0,185	0,003	0,000001
Лапласа	0,242	0,411	0,242	0,001	0,000000
Пуассона	0,368	0,368	0,184	0,003	0,000001

3. Інтегральна теорема Лапласа. Якщо ймовірність p настання деякої події A в кожному випробувані є сталою, причому $0<p<1$, а число випробувань n досить велике, то ймовірність того, що ця подія відбудеться не менше $к_1$ раз і не більше $к_2$ рази, наближено дорівнює

$$P_n(k_1, k_2) \approx \Phi(x_2) - \Phi(x_1), \qquad (6.3)$$

де

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}},$$

а функція Лапласа $\Phi(x)$ визначається рівністю

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

Функція $\Phi(x)$ - непарна, тобто $\Phi(-x) = -\Phi(x)$ і монотонно зростаюча, а $\lim\limits_{x\to\infty} \Phi(x) = 0{,}5$. До цієї границі вона наближається досить швидко.

Вже при $x = 5$ функція $\Phi(x)$ дуже мало відрізняється від неї. Дійсно, $\Phi(5) = 0{,}499997$, тому для всіх $x > 5$ приймається $\Phi(x) \approx 0{,}5$.

Точність ймовірностей, які отримують за формулою (6.3), цілком задовольняє потребам практики, якщо добуток npq має порядок декількох

сотень. Якщо задача не потребує більшої точності відповіді, то цю формулу можна використовувати й тоді, коли добуток npq невеликий, але більший за 20. У противному випадку треба користуватися не інтегральною теоремою Лапласа, а формулою Бернуллі та теоремою додавання ймовірностей.

Частинні випадки:

1. Формула (6.3) спрощується, якщо $к_1$ менше, а $к_2$ більше np на одне й те саме число ε:

$$P_n(np-\varepsilon \le m \le np+\varepsilon) = P_n(|m-np| \le \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right).$$

2. Імовірність того, що в n незалежних випробувань, у кожному з яких імовірність настання події дорівнює p ($0<p<1$), абсолютна величина відносної частоти появи події від імовірності появи події не перевищує додатного числа ε, наближено дорівнює подвоєній функції Лапласа при $x = \varepsilon\sqrt{\frac{n}{pq}}$:

$$P\left(\left|\frac{m}{n}-p\right| \le \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right). \qquad (6.4)$$

Приклад. Імовірність влучення в рухому мішень приймається рівною 0,7. Яка ймовірність того, що з 20 пострілів 15 виявиться вдалими?

Розв'язання. За умовою: $n = 20$, $к = 15$, $p = 0{,}7$ і $q = 0{,}3$. Знаходимо шукану ймовірність за наближеною формулою Лапласа, для чого обчислимо спочатку значення x:

$$x = \frac{к-np}{\sqrt{npq}} = \frac{15-20\cdot 0{,}7}{\sqrt{20\cdot 0{,}7\cdot 0{,}3}} \approx \frac{1}{2{,}0494} \approx 0{,}49\,.$$

Далі за формулою (22.2) отримаємо:

$$P_{20}(15) \approx \frac{1}{2{,}0494}\cdot\varphi(0{,}49) = \frac{1}{2{,}0494}\cdot 0{,}3538 \approx 0{,}1726\,.$$

Обчислення за точною формулою, тобто за формулою Бернуллі, дає наступний результат:

$$P_{20}(15) = C_{20}^{15}p^{15}q^5 = C_{20}^{5}p^{15}q^5 = \frac{20!}{15!5!}\cdot(0{,}7)^{15}\cdot(0{,}3)^5 = 0{,}1789.$$

Легко підрахувати що відносна похибка наближеного результату складає 3,5 %. Зі зростанням значення n цей відсоток знижується.

Нехай за умовою попередньої задачі потрібно знайти ймовірність того, що 5 пострілів виявляться вдалими. Застосування формули Бернуллі дає:

$$P_{20}(5) = C_{20}^{5}\cdot((0{,}7)^5\cdot(0{,}3)^{15} \approx 0{,}00003739\,.$$

Отримане значення ймовірності вказує на те, що очікувана подія практично не можлива.

Застосування наближеної формули Лапласа дає:

$$P_{20}(5) \approx \frac{1}{2{,}0494}\cdot\varphi(4{,}39) \approx 0\,.$$

Відзначимо, що значення φ(4,39) не може бути знайденим, оскільки в додатку А наведені значення φ (x) для $x \le 3{,}99$. Значенням $x > 3{,}99$ відповідають дуже малі значення φ (x), що вказує на практичну неможливість події.

Приклад. Імовірність виготовлення виробів вищого ґатунку на даному підприємстві дорівнює 0,4. Знайти ймовірність того, що серед 26 виробів, навмання взятих, половина матиме вищий ґатунок.

Розв'язання. За умовою : $n = 26$, $к = 13$, $p = 0{,}4$ і $q = 0{,}6$. Оскільки $n = 26$ - велике число, скористаємося локальною теоремою Лапласа:

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x), \text{ де } x = \frac{к - np}{\sqrt{npq}}.$$

Знаходимо значення x:

$$x = \frac{13 - 26 \cdot 0{,}4}{\sqrt{26 \cdot 0{,}4 \cdot 0{,}6}} \approx \frac{2{,}6}{2{,}5} = 1{,}04 .$$

За додатком А знаходимо $\varphi(x) = \varphi(1{,}04) = 0{,}2323$. Шукана ймовірність

$$P_{26}(13) \approx \frac{0{,}2323}{2{,}5} \approx 0{,}093 .$$

Приклад. За установленого технологічного процесу фабрика випускає у середньому 70 % продукції першого ґатунку. Чому дорівнює ймовірність того, що у партії з 2100 виробів першосортних міститься між 1450 і 1500?

Розв'язання. Відомо число незалежних випробувань - n = 2100. Імовірність настання події у кожному окремому випробувані - p = 0,7. Треба знайти ймовірність того, що число появи події міститься між $к_1$=1450 і $к_2$=1500. Шукану ймовірність знаходимо за формулою (6.3):

$$P_n(k_1;k_2) \approx \Phi(x_2) - \Phi(x_1), \text{ де } x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

Знаходимо значення x_1 і x_2:

$$x_1 = \frac{1450 - 2100 \cdot 0{,}7}{\sqrt{2100 \cdot 0{,}7 \cdot 0{,}3}} = -\frac{20}{21} \approx -0{,}95; \quad x_2 = \frac{1500 - 2100 \cdot 0{,}7}{\sqrt{2100 \cdot 0{,}7 \cdot 0{,}3}} = \frac{30}{21} \approx 1{,}43;$$

Враховуючи, що функція Лапласа непарна, тобто Ф(-x)=-Ф(x), отримаємо

P_{2100} (1450; 1500) ≈ Ф(1,43) - Ф(-0,95) = Ф(1,43) + Ф(0,95) =

= 0,4236 + 0,3289 = 0,7525.

Значення функції Лапласа Ф(1,43) і Ф(0,95) знайдені за додатком Б.

Приклад. Завод відправив на базу 500 виробів. Імовірність пошкодження виробу у дорозі дорівнює 0,002. Знайти ймовірність того, що у дорозі буде пошкоджено виробів: а) рівно 3; б) менше 3; в) більше 3; г) хоча б один.

Розв'язання. Число n = 500 велике, ймовірність p = 0,002 мала і події "пошкодження виробів" незалежні, тому має місце формула Пуассона:

$$P_n(к) \approx \frac{\lambda^k \cdot e^{-\lambda}}{к!}, \text{ де } \lambda = np$$

а) знаходимо λ: $\lambda = np = 500 \cdot 0{,}002 = 1$.

Знаходимо ймовірність того, що буде пошкоджено рівно три ($к$ = 30 вироби: $P_{500}(3) \approx \frac{e^{-1}}{3!} \approx \frac{0{,}36788}{6} \approx 0{,}0613$

б) Знаходимо ймовірність того, що буде пошкоджено менше трьох виробів ($к < 3$):

$$P_{500}(k<3)=P_{500}(0)+P_{500}(1)+P_{500}(2)=e^{-1}+e^{-1}+\frac{e^{-1}}{2}=\frac{5e^{-1}}{2}\approx\frac{5}{2}\cdot 0{,}36788\ =0{,}9197\,.$$

в) Знаходимо ймовірність того, що буде пошкоджено більше трьох виробів. Події "пошкоджено більше трьох виробів" і "пошкодженні не більше трьох виробів" протилежні тому,

$P_{500}(к > 3) + P_{500}(к \le 3) = 1$.

Звідси $P_{500}(к > 3) = 1 - P_{500}(к \le 3) = 1 - (P_{500}(к < 3) + P_{500}(к = 3))$.

Використовуючи результати отримані вище, маємо

$P_{500}(к > 3) = 1 - (0{,}9197 + 0{,}0613) = 0{,}019$.

г) Знаходимо ймовірність того, що буде пошкоджено хоча б один виріб. Події A (пошкоджено хоча б один виріб) і $\bar{A}$ (жоден виріб не пошкоджено) є протилежними, отже, $P(A) + P(\bar{A}) = 1$.

Звідси шукана ймовірність

$P(A) = 1 - P(\bar{A}) = 1 - P_{500}(0) = 1- e^{-1} = 1 - 0{,}36788 \approx 0{,}632$..

Приклад. Імовірність появи події у кожному із 900 незалежних випробувань дорівнює 0,5. Знайти ймовірність того, що відносна частота появи події відхиляється від її ймовірності за абсолютною величиною не більше ніж на 0,02.

Розв'язання. За умовою $n = 900$, $p = 0{,}5$, $q = 0{,}5$, $\varepsilon = 0{,}02$.

Треба знайти ймовірність

$$P\left(\left|\frac{m}{900}-0{,}5\right|\le 0{,}02\right).$$

Скористаємося формулою (6.4):

$$P(\left|\frac{m}{n}-P\right|\le\varepsilon)\approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$$

Маємо

$$P(\left|\frac{m}{900}-0{,}5\right|\le 0{,}02)\approx 2\Phi\left(0{,}02\cdot\sqrt{\frac{900}{0{,}5\cdot 0{,}5}}\right)=2\Phi(1{,}2)\,.$$

Далі за додатком Б знаходимо $\Phi(1{,}2)=0{,}3849$

Отже, шукана ймовірність

$$P(\left|\frac{m}{900}-0{,}5\right|\le 0{,}02)\approx 2\cdot 0{,}3849\ =0{,}7698\,.$$

Питання для самоперевірки

1 Сформулюйте локальну теорему Лапласа. У яких випадках вона застосовується?

2 Як знайти значення $\varphi(x)$, коли $x > 5$?

3 Наведіть формулу Пуассона для знаходження $P_n(к)$. В яких випадках вона застосовується?

4 Сформулюйте інтегральну теорему Лапласа.

5 Який вигляд має функція Лапласа $\varphi(x)$?

6 Властивості функції Лапласа $\varphi(x)$. Як знайти значення $\varphi(x)$, коли $x > 5$?

7 Наведіть формулу для обчислення $P(\left|\frac{m}{n} - P\right| \le \varepsilon)$.

ЛЕКЦІЯ 7

7.1 Дискретні випадкові величини та закони їх розподілу

7.2 Біноміальний закон та закон Пуассона

7.1 Дискретні випадкові величини та закони їх розподілу

Випадковою називають величину, яка в результаті випробовування набуває одне і тільки одне можливе значення, наперед відомо, яке саме і яке залежить від випадкових причин, які заздалегідь не можуть бути враховані.

Випадкові величини прийнято позначати великими буквами *X, Y, Z*, а їх можливі значення - відповідними *x, y, z*. Наприклад, якщо випадкова величина *X* має три можливих значення, то вони будуть позначені x_1, x_2, x_3. Випадкові величини бувають дискретними і неперервними.

Дискретною (перервною) називають випадкову величину, яка набуває окремі, ізольовані можливі значення з певними ймовірностями.

Число можливих значень дискретної випадкової величини може бути скінченим або нескінченим (в останньому випадку множину всіх можливих значень називають лічильною).

Законом розподілу (рядом розподілу) дискретної величини називають перелік її можливих значень і відповідних їм ймовірностей.

Закон розподілу дискретної випадкової величини *X* може бути заданий у вигляді таблиці, перший рядок якої містить можливі значення x_i, а другий їх імовірності p_i:

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

(7.1)

де $\sum_{i-1}^{n} p_i = p_1 + p_2 + ... + p_n = 1$.

Якщо множина можливих значень *X* нескінчена (лічильна), то ряд $\sum_{i-1}^{n} p_i$ збігається і його сума дорівнює одиниці.

Закон розподілу дискретної випадкової величини *X* може бути також заданий аналітичною формулою:

$P(X = x_i) = \varphi(x_i)$.

Або з допомогою інтегральної функції (про цю функцію мова піде у матеріалі наступного заняття).

Закон розподілу дискретної випадкової величини можна зобразити графічно, для чого в прямокутній декартовій системі координат відкладають точки $M_1(x_1 ; p_1)$, $M_2(x_2 ; x_2)$, ... , $M_n(x_n ; p_n)$ (x_i можливі значення *X*, p_i - відповідні їм ймовірності) і з'єднують їх відрізками прямих. Отриману фігуру називають прямокутником розподілу.

Приклад. У партії з шести деталей міститься чотири стандартних. Навмання відібрано три деталі (без повернення).

Скласти закон розподілу дискретної випадкової величини X - числа стандартних деталей серед відібраних.

Побудувати многокутник отриманого розподілу.

Розв'язання. Випадкова величина X - число стандартних деталей серед відібраних має наступні можливі значення: x_2 = 1 (одна деталь стандартна), x_3=2 (дві деталі стандартні), x_4 = 3 (усі три деталі стандартні). Ймовірності цих значень відповідно дорівнюють:

$$P(X=1)=\frac{C_4^1\cdot C_2^2}{C_6^3}=\frac{4\cdot1\cdot1\cdot2\cdot3}{6\cdot5\cdot4}=\frac{1}{5};$$

$$P(X=2)=\frac{C_4^2\cdot C_2^1}{C_6^3}=\frac{4\cdot3\cdot2\cdot1\cdot2\cdot3}{1\cdot2\cdot1\cdot6\cdot5\cdot4}=\frac{3}{5};$$

$$P(X=3)=\frac{C_4^3\cdot C_2^0}{C_6^3}=\frac{4\cdot1\cdot1\cdot2\cdot3}{1\cdot6\cdot5\cdot4}=\frac{1}{5};$$

Шуканий закон розподілу X має вигляд:

X	1	2	3
P	1/5	3/5	1/5

Контроль: $\sum_{i=1}^{4} p_i=\frac{1}{5}+\frac{3}{5}+\frac{1}{5}=1.$

Будуємо багатокутник розподілу випадкової величини X (рис.7.1)

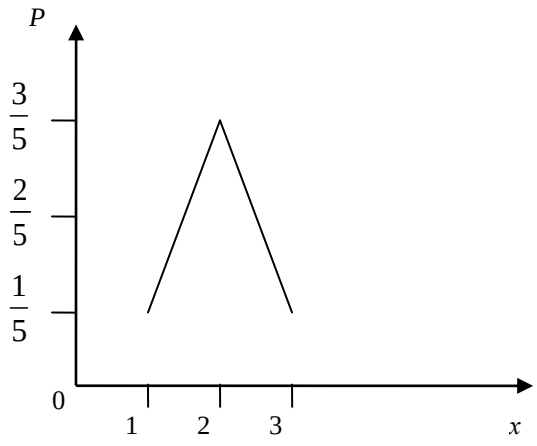


Рисунок 7.1

7.2 Біноміальний закон та закон Пуассона

Закон розподілу дискретної випадкової величини X - числа появи події у n незалежних випробуваннях, в кожному із яких ймовірність появи події є сталою і рівною p, називають біноміальним.

Ймовірність можливого значення $X = к$ (числа $к$ появи події в n випробуваннях) обчислюють за формулою Бернуллі:

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

Якщо число випробувань велике, а ймовірність появи події у кожному випробуванні мала, то використовують формулу Пуассона:

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

де $\lambda = пр$.

У цьому випадку говорять, що випадкова величина розподілена за законом Пуассона.

Приклад. Ймовірність виходу мікросхеми із ладу протягом заданого проміжку часу дорівнює 0,003. Яка ймовірність того, що із 200 мікросхем:

а) жодна не вийде з ладу;

б) вийдуть з ладу не більше двох?

Розв'язання. Точний розв’язок прикладу одержуємо за формулою (1) при n=200 і p=0,003:

à) $p_{200}(0) = 0{,}997^{200} = 0{,}5483$;

á) $p_{200}(0) + p_{200}(1) + p_{200}(2) = 0{,}997^{200} + 200 \cdot 0{,}997^{199} \cdot 0{,}003 +$

$+19900 \cdot 0{,}997^{198} \cdot 0{,}003^2 = 0{,}9771.$

Наближений розв’язок знаходимо за формулою (2) при λ=0.6:

à) $p_{200}(0) \approx e^{-0.6} = 0{,}5488;$

á) $p_{200}(0) + p_{200}(1) + p_{200}(2) \approx e^{-0.6}\left(1 + 0{,}6 + \frac{0{,}6^2}{2}\right) = 0{,}9769.$

Питання для самоперевірки

1 Яка величина називається випадковою?

2 Як прийнято позначати випадкові величини та їх значення?

3 Що являє собою закон розподілу випадкової величини?

4 Як можна задати закон розподілу випадкової величини?

5 Яка випадкова величина називається дискретною?

6 Як прийнято задавати закон розподілу дискретної випадкової величини? Чому?

7 У чому полягає біноміальний закон та закон Пуассона дискретної випадкової величини?

ЛЕКЦІЯ 8

8.1 Числові характеристики дискретних випадкових величин та їх властивості

8.2 Інтегральна та диференціальна функції розподілу і їх властивості

8.1 Числові характеристики дискретних випадкових величин та їх властивості

До числових характеристик дискретної випадкової величини відносяться математичне сподівання, дисперсія і середнє квадратичне відхилення.

Математичним сподіванням дискретної випадкової величини називають суму добутків усіх її можливих значень на їх імовірності:

$$M(X) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + ... + x_n p_n. \qquad (8.1)$$

Якщо випадкова величина приймає лічильну множину можливих значень, то

$$M(X) = \sum_{i=1}^{\infty} x_i p_i,$$

причому припускається, що ряд, який знаходиться у правій частині рівності, збігається абсолютно і сума всіх ймовірностей p_i дорівнює одиниці. При повторних випробуваннях математичне сподівання обчислюється за формулою

$$M(X) = np, \qquad (8.2)$$

де n - число випробувань; p - ймовірність появи події в одному випробуванні.

Математичне сподівання X наближено дорівнює середньому арифметичному значень випадкової величини, які спостерігались (тим точніше, чим більше число випробувань):

$$M(X) \approx X_{ср.}$$

Слід підкреслити, що математичне сподівання можна обчислити до проведення випробувань, а середнє значення - лише після їх проведення.

Властивості математичного сподівання

1. $M(C) = C$, де $C = const$.
2. $M(X_1 + X_2 + ... + X_n) = M(X_1) + M(X_2) + ... + M(X_n)$.
3. $M(X_1 X_2 ... X_n) = M(X_1) \cdot M(X_2) \cdot ... \cdot M(X_n)$.
4. $M(CX) = C \cdot M(X)$.

Дисперсією (розсіянням) дискретної випадкової величини X називають математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання:

$$D(X) = M(X - M(X))^2. \qquad (8.3)$$

Зауважимо, що відхиленням називають різницю між випадковою величиною і її математичним сподіванням.

Математичне сподівання відхилення дорівнює нулю: $M[X - M(X)] = 0$.

Нарівні з терміном "відхилення" використовують термін "центрована величина". Центрованою випадковою величиною $\overset{\circ}{X}$ називають різницю між випадковою величиною і її математичним сподіванням:

$$\overset{\circ}{X} = X - M(X) \ .$$

Назва "центрована величина" пов'язана з тим, що математичне сподівання є центром розподілу *X*.

Зауважимо також, що на практиці дисперсію доцільно обчислювати за формулою

$$D(X) = M(X^2) - M^2(X). \qquad (8.4)$$

Для обчислення дисперсії у разі повторних випробувань використовують формулу

$$D(X) = npq, \qquad (8.5)$$

де n - число випробувань; p - імовірність появи події в одному випробуванні; q - імовірність появи події в одному випробуванні.

Властивості дисперсії

1. $D(C) = 0$, де $C = const$.
2. $D(X \pm Y) = D(X) + D(Y)$.
3. $D(CX) = C^2 \cdot D(X)$.

Середнім квадратичним відхиленням дискретної випадкової величини називають квадратний корінь з дисперсії:

$$\sigma(X) = \sqrt{D(X)} \ . \qquad (8.6)$$

Приклад. Дискретно випадкова величина *X* набуває лише два можливих значення: x_1 і x_2, причому $x_1 < x_2$. Ймовірність того, що *X* набуває значення x_1, дорівнює 0,2. Зайти закон розподілу *X*, якщо математичне сподівання $M(X)$ = 2,6 і середнє квадратичне відхилення $\sigma(X) = 0{,}8$.

Розв'язання. Сума ймовірностей усіх можливих значень дискретної випадкової величини дорівнює одиниці, тому ймовірність того, що *X* набуває значення x_2, буде такою: $p_2 = 1 - 0{,}2 = 0{,}8$.

Запишемо закон розподілу *X*:

X	x_1	x_2
P	0,2	0,8

Для знаходження x_1 і x_2 за формулами (8.1) і (8.4) виразимо відоме математичне сподівання і дисперсію через x_1 і x_2:

$M(X) = 0{,}2x_1 + 0{,}8\,x_2$,

$D(X) = 0{,}2 \cdot x_1^2 + 0{,}8 \cdot x_2^2 - M^2(X)$.

Підставляючи до цих рівнянь значення $M(X) = 2{,}6$ і $D(X) = \sigma^2(X) = (0{,}8)^2 = 0{,}64$, після елементарних перетворень отримаємо систему двох рівнянь з двома невідомими:

$$\begin{cases} 0{,}2x_1 + 0{,}8x_2 = 2{,}6; \\ 0{,}2x_1^2 + 0{,}8x_2^2 = 7{,}4. \end{cases}$$

Помножимо обидві частини кожного рівняння системи на п’ять. Тоді

$$\begin{cases} x_1 + 4x_2 = 13; \\ x_1^2 + 4x_2^2 = 37. \end{cases}$$

Виражаючи далі з першого рівняння системи x_1 і підставляючи його вираз у друге рівняння, отримаємо:

$(13 - 4x_2)^2 + 4x_2^2 = 37.$

Після нескладних перетворень це рівняння набуває вигляду

$5x_2^2 - 26x_2 + 33 = 0,$

а його розв'язок -

$$x_2 = \frac{13 \pm \sqrt{169 - 165}}{5} = \frac{3 \pm 2}{5}.$$

Отже, маємо два розв'язки системи:

$x_2^1 = 3, \quad x_1^1 = 13 - 4x_2^1 = 1;$

$x_2^{11} = 2{,}2, \quad x_1^{11} = 13 - 4x_2^{11} = 4{,}2.$

За умовою задачі $x_1 < x_2$, тому даній задачі задовольняє лише перший розв'язок, тобто $x_1 = 1$; $x_2 = 3$.

Таким чином, шуканий закон розподілу:

X	1	3
P	0,2	0,8

Приклад. Знайти дисперсію дискретної випадкової величини X - числа появи події A у двох незалежних випробуваннях, якщо ймовірності появи події у цих випробування однакові і відомо, що $M(X) = 0{,}9$.

Розв'язання.

І спосіб. Можливі значення величини такі: $x_1 = 0$ (подія не настала), $x_2 = 1$ (подія настала один раз) і $x_3 = 2$ (подія настала двічі).

Знайти ймовірність можливих значень за формулою Бернуллі:

$P_2(0) = C_2^0 p^0 q^2 = q^2;$

$P_2(1) = C_2^1 p^1 q^1 = 2pq;$

$P_2(2) = C_2^2 p^2 q^0 = p^2.$

Запишемо закон розподілу X:

Можливі значення	0	1	2
Імовірності	q^2	$2pq$	p^2

Знайдемо $M(X)$:

$M(X) = 0 \cdot q^2 + 1 \cdot 2qp + 2 \cdot p^2 = 2p(p + q) = 2p \cdot 1 = 2p.$

Згідно з умовою $M(X) = 0{,}9$, тобто $2p = 0{,}9$.

Звідси $p = 0{,}45$, отже, $q = 1 - 0{,}45 = 0{,}55$.

Шукана дисперсія

$D(X) = M(X^2) - M^2(X) = 0 \cdot q^2 + 1 \cdot 2pq + 4p^2 - 0{,}81 =$

$= 2 \cdot 0{,}45 \cdot 0{,}55 + 4 \cdot 0{,}2025 - 0{,}81 = 0{,}495.$

ІІ спосіб. За формулою (8.2) $M(X) = np$.

Оскільки, за умовою $M(X) = 0{,}9$, а $n = 2$, то $2p = 0{,}9$. Звідси $p = 0{,}45$ і, отже, $q = 0{,}55$. Шукану дисперсію знаходимо за формулою (8.5):

$D(X) = npq = 2 \cdot 0{,}45 \cdot 0{,}55 = 0{,}495.$

Перевага другого способу є наочною.

8.2 Неперервні випадкові величини. Інтегральна та диференціальна функції розподілу і їх властивості

Неперервною називають випадкову величину, можливі значення якої цілком заповнюють деякий проміжок (скінчений або нескінчений).

Очевидно, що число можливих значень неперервної випадкової величини нескінчене.

Слід відзначити, що дане означення неперервної випадкової величини не є точним. Оскільки, на відміну від дискретної випадкової величини, яка може бути задана переліком усіх її можливих значень і їх ймовірностей, неперервну випадкову величину не можна задати таким способом, то виникає необхідність в тому, щоб дати загальний спосіб завдання будь-яких типів випадкових величин. З цією метою вводять функції розподілу ймовірностей випадкової величини.

Функцією розподілу (інтегральною функцією) називають функцію $F(x)$, визначальну ймовірність того, що випадкова величина X у результаті випробування набуває значення, меншого за x, тобто

$$F(x) = P(X < x). \qquad (8.7)$$

Тепер можна дати точніше означення неперервної випадкової величини. Випадкову величину називають неперервною, якщо її функція розподілу є неперервною, кусково-диференційованою функцією з неперервною похідною.

Властивостями функції розподілу є:

Властивість 1. Значення функції розподілу належать відрізку [0; 1], тобто $0 \le F(x) \le 1$.

Властивість 2. $F(x)$ - неспадна функція, тобто $F(x_2) \ge F(x_1)$, якщо $x_2 > x_1$.

Наслідок 1. Ймовірність того, що випадкова величина набуде значення, яке належить інтервалу (а; в) дорівнює приросту функції розподілу на цьому інтервалі:

$$P(a < X < в) = F(в) - F(a).$$

Наслідок 2. Ймовірність того, що неперервна випадкова величина X набуде одного певного значення, дорівнює 0.

Властивість 3. Якщо можливі значення випадкової величини належать інтервалу (*а; в*), то:

1) $F(x) = 0$ при $x \le a$;
2) $F(x) = 1$ при $x \ge в$.

Наслідок 3. Якщо можливі значенні неперервної випадкової величини заповнюють усю числову вісь *0x*, то вірними є наступні граничні співвідношення:

$$\lim_{x \to -\infty} F(x) = 0; \quad \lim_{x \to +\infty} F(x) = 1$$

Графік цієї функції подано на рис.8.1:

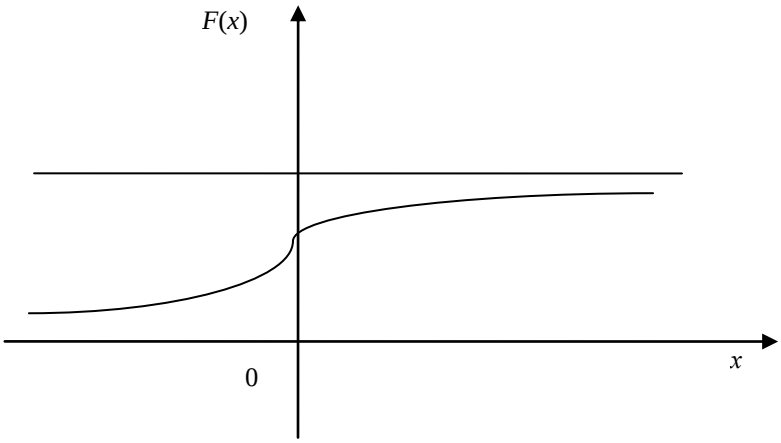


Рисунок 8.1

Спосіб завдання неперервної величини за допомогою функції розподілу не є єдиним. Неперервну випадкову величину можна також задати, використовуючи іншу функцію, яку називають щільністю розподілу, або щільністю ймовірності (іноді її називають диференціальною функцією) і позначають $f(x)$.

За означенням

$$f(x) = F'(x), \tag{8.8}$$

тобто щільність розподілу є першою похідною від функції розподілу.

Із цього означення випливає, що

$$F(x) = \int_{-\infty}^{x} f(x)dx. \tag{8.9}$$

Властивостями щільності розподілу є:

Властивість 1. Щільність розподілу - невід'ємна функція, тобто $f(x) \geq 0$.

Властивість 2. Невласний інтеграл від щільності розподілу в межах від $-\infty$ до $+\infty$ дорівнює одиниці, тобто

$$\int_{-\infty}^{+\infty} f(x)dx = 1. \tag{8.10}$$

Імовірність того, що неперервна випадкова величина X набуває значення, яке належить інтервалу $(a; в)$

$$P(a < X < в) = \int_{a}^{в} f(x)dx = F(в) - F(a), \tag{8.11}$$

де $f(x)$ – диференціальна функція (щільність розподілу); $F(x)$ – інтегральна функція (функція розподілу).

Приклад. Випадкова величина X задана інтегральною функцією

$$F(x) = \begin{cases} 0 & при \quad x \leq 2, \\ 0{,}5x - 1 & при \quad 2 < x \leq 4, \\ 1 & при \quad x > 4. \end{cases}$$

Знайти ймовірність того, що в результаті випробування X набуває значення: а) меншого за 0,2; б) меншого за 3; в) не меншого за 3; г) не меншого за 5.

Розв'язання.

а) За означенням функції розподілу маємо:

$F(x) = P(X < x)$.

Отже, $P(X < 0{,}2) = F(0{,}2) = 0$, бо при $x \le 2$ функція $F(x) = 0$.

б) $P(X < 3) = F(3) = 0{,}5 \cdot 3 - 1 = 1{,}5 - 1 = 0{,}5$,

бо значення $x = 3$ належить інтервалу (2,4], на якому функція $F(x) = 0{,}5x - 1$.

в) Події $X \ge 3$ і $X < 3$ є протилежними, тому

$P(X \ge 3) + P(X < 3) = 1$.

Звідси, враховуючи, що $P(X < 3) = 0{,}5$ (див. п. б), отримаємо

$P(X \ge 3) = 1 - P(X < 3) = 1 - 0{,}5 = 0{,}5$.

г) Сума ймовірностей протилежних подій дорівнює одиниці, тому

$P(X \ge 5) + P(X < 5) = 1$.

Звідси, використовуючи умову, в силу якої при $x > 4$ функція $F(x) = 1$, отримаємо

$P(X \ge 5) = 1 - P(X < 5) = 1 - F(5) = 1 - 1 = 0$.

Приклад. Дискретна випадкова величина задана законом розподілу

X	3	4	7	10
p	0,2	0,1	0,4	0,3

Знайти інтегральну функцію і побудувати її графік.

Розв'язання.

1. Якщо $x \le 3$, то $F(x)=0$. Дійсно, значень, менших за число 3, величина X не набуває. Отже, при $X \le 3$ функція $F(x) = P(X < x) = 0$.

2. Якщо $3 < x \le 4$, то $F(x) = 0{,}2$. Дійсно, X може набути значення 3 з ймовірністю 0,2.

3. Якщо $4 < x \le 7$, то $F(x) = 0{,}3$. Дійсно, X може набути значення 3 з ймовірністю 0,2 і значення 4 з імовірністю 0,1. Отже, одного з цих значень, байдуже яке саме, X може набути (за теоремою додавання ймовірностей несумісних подій) з імовірністю 0,2 + 0,1 = 0,3.

4. Якщо $7 < x \le 10$, то $F(x) = 0{,}7$. Дійсно, X може набути значення 3 з імовірністю 0,2, значення 4 з імовірністю 0,1 і значення 7 з імовірністю 0,4, отже, одне з цих значень, байдуже яке саме, X може набути (за теоремою додавання ймовірностей несумісних подій) з імовірністю 0,2 + 0,1 + 0,4 = 0,7.

5. Якщо $x > 10$, то $F(x) = 1$. Дійсно, подія $X \le 10$ вірогідна і тому її ймовірність дорівнює одиниці.

Таким чином, шукана інтегральна функція аналітично може бути записана таким чином:

$$F(x)=\begin{cases} 0 & при \quad x\le 3, \\ 0{,}2 & при \quad 3<x\le 4, \\ 0{,}3 & при \quad 4<x\le 7, \\ 0{,}7 & при \quad 7<x\le 10, \\ 1 & при \quad x>10. \end{cases}$$

Графік інтегральної функції зображений на рис. 8.2

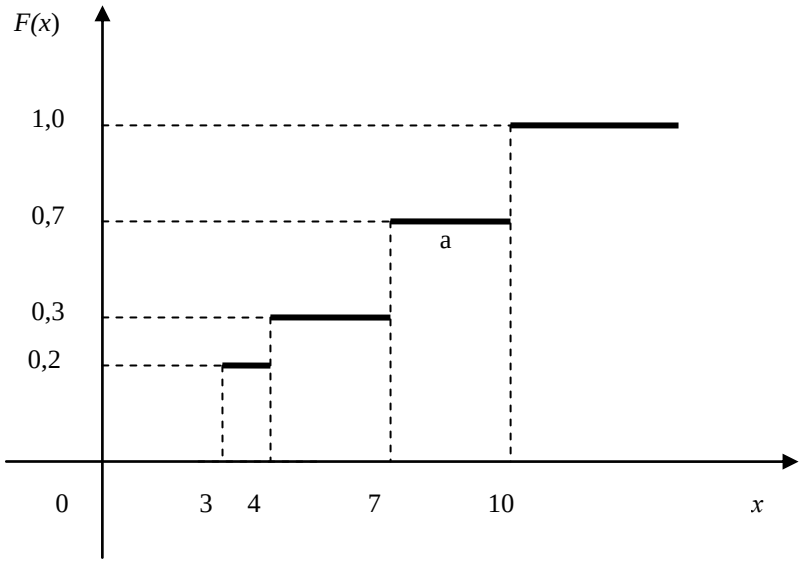


Рисунок 8.2

Із рис.8.2 видно, що графік інтегральної функції дисперсної випадкової величини має східчастий вигляд.

Приклад. Дана диференціальна функція неперервної випадкової величини X:

$$f(x)=\begin{cases} 0 & при \quad x\le 1, \\ x-0{,}5 & при \quad 1<x\le 2, \\ 0 & при \quad x>2. \end{cases}$$

Знайти інтегральну функцію $F(x)$.

Розв'язання. За формулою (8.9)

$$F(x)=\int_{-\infty}^{x} f(x)dx.$$

Якщо $x\le 1$, то $f(x)=0$, отже, $F(x)=\int_{-\infty}^{\infty} 0\cdot dx=0.$

Якщо $1<x\le 2$, то

$$F(x)=\int_{-\infty}^{\infty} 0dx+\int_{1}^{x}(x-0{,}5)dx=(\frac{x^2}{2}-0{,}5x)\,|_1^x=\frac{x^2}{2}-\frac{x}{2}-\frac{1}{2}+\frac{1}{2}=\frac{1}{2}(x^2-x).$$

Якщо $x>2$, то

$$F(x)=\int_{-\infty}^{1} 0dx+\int_{1}^{2}(x-0{,}5)dx+\int_{2}^{x} 0dx=(\frac{x^2}{2}-0{,}5x)\,|_1^2=2-1-\frac{1}{2}+\frac{1}{2}=1.$$

$$F(x)=\begin{cases} 0 & при \quad x\le 1, \\ \frac{1}{2}(x^2-x) & при \quad 1<x\le 2, \\ 1 & при \quad x>2. \end{cases}$$

Питання для самоперевірки

1 Назвіть числові характеристики дискретної випадкової величини.

2 Що називають математичним сподіванням дискретної випадкової величини? Яким чином воно її характеризує?

3 Перерахуйте властивості математичного сподівання.

4 Що називають дисперсією дискретної випадкової величини? Яким чином вона її характеризує?

5 Якою формулою для обчислення дисперсії користуються на практиці?

6 Перерахуйте властивості дисперсії випадкової величини.

7 Що називають середнім квадратичним відхиленням дискретної випадкової величини? Для чого його вводять?

8 Яку випадкову величину називають неперервною?

9 Що називають інтегральною функцією? Сформулюйте її властивості.

10 Дайте означення диференціальної функції (щільності розподілу).

11 Сформулюйте властивості диференціальної функції.

ЛЕКЦІЯ 9

9.1 Числові характеристики неперервних випадкових величин, їх закони розподілу

9.2 Рівномірний розподіл, нормальний розподіл, їх числові характеристики

9.1 Числові характеристики неперервних випадкових величин, їх закони розподілу

Неперервна випадкова величина має такі самі числові характеристики, як і дискретна. Ці характеристики мають ті ж самі властивості, що і випадку дискретної випадкової величини

Математичне сподівання неперервної випадкової величини X, можливі значення якої належать усій числовій осі $O\,x$, визначаються рівністю:

$$M(X)=\int_{-\infty}^{+\infty} xf(x)dx, \qquad (9.1)$$

де $f(x)$ – диференціальна функція.

Припускається, що невласний інтеграл збігається абсолютно.

Зокрема, якщо всі можливі значення X належать до інтервалу $(a;\ в)$, то

$$M(X)=\int_{a}^{в} xf(x)dx. \qquad (9.2)$$

Дисперсією неперервної випадкової величини називають математичним сподіванням квадрата її відхилення.

Якщо можливі значення неперервної величини X належать до осі $0x$, то дисперсія визначається рівністю

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 f(x)dx, \qquad (9.3)$$

що рівносильне

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - M^2(X), \qquad (9.4)$$

Зокрема, якщо всі можливі значення X належать до інтервалу ($a;в$), то

$$D(X) = \int_{a}^{в} (x - M(X))^2 f(x)dx, \text{ або } D(X) = \int_{a}^{в} x^2 f(x)dx - M^2(X). \qquad (9.5)$$

Зауважимо, що всі властивості математичного сподівання і дисперсії, указані вище для дискретних величин, зберігаються і для неперервних величин.

Середнє квадратичне відхилення $\sigma(X)$ неперервної випадкової величини визначається так само, як і величини дискретної - рівністю:

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

Приклад. Випадкова величина X задана інтегральною функцією

$$F(x) = \begin{cases} 0, & при \quad x \le 0, \\ \dfrac{x^2}{4}, & при \quad 0 < x \le 2, \\ 1, & при \quad x > 2. \end{cases}$$

Знайти :

а) диференціальну функцію;

б) математичне сподівання, дисперсію і середнє квадратичне відхилення;

в) ймовірність того, що в результаті випробовування X набуде значення, яке належить інтервалу $(1; 4)$ (двома способами).

Побудувати графіки функцій $F(x)$ і $f(x)$.

Розв'язання:

а) Знаходимо диференціальну функцію $f(x)$, для чого продиференціюємо по x інтегральну функцію $F(x)$:

$$f(x) = F'(x) = \begin{cases} 0, & при \quad x \le 0, \\ \dfrac{x}{2}, & при \quad 0 < x \le 2, \\ 0, & при \quad x > 2. \end{cases}$$

б) Застосовуючи послідовно формули (9.2) і (9.4), обчислюємо відповідно математичне сподівання, дисперсію і середнє квадратичне відхилення випадкової величини X:

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{0} x \cdot 0 \cdot dx + \int_{0}^{2} x\frac{x}{2}dx + \int_{2}^{\infty} x \cdot 0 \cdot dx = \frac{1}{2}\int_{0}^{2} x^2 dx = \frac{1}{2}\frac{x^3}{3}\Bigg|_{0}^{2} = \frac{4}{3};$$

$$D(X)=\int_{-\infty}^{\infty}x^2 f(x)dx-M^2(X)=\int_{-\infty}^{0}x\cdot 0\cdot dx+\int_{0}^{2}x^2\cdot\frac{x}{2}dx+\int_{2}^{\infty}x^2\cdot 0\cdot dx-\left(\frac{4}{3}\right)^2=$$

$$=\frac{1}{2}\int_{0}^{2}x^3dx-\frac{16}{9}=\frac{1}{2}\cdot\frac{x^4}{4}\Bigg|_{0}^{2}-\frac{16}{9}=2-\frac{16}{9}=\frac{2}{9};$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{2}{9}}=\frac{\sqrt{2}}{3}.$$

Зауважимо, що обчислення дисперсії за формулою (9.3) є більш трудомістким. Пропонуємо переконатись в цьому самостійно.

в) Вказану ймовірність знайдемо двома способами:

І спосіб. Шукана ймовірність дорівнює приросту інтегральної функції на заданому інтервалі:

$$P(1<X<4)=F(4)-F(1)=1-\frac{1}{4}=\frac{3}{4}.$$

ІІ спосіб. Шукана ймовірність дорівнює визначеному інтегралу в межах від 1 до 4 від диференціальної функції:

$$P(1<X<4)=\int_{1}^{2}\frac{x}{2}dx+\int_{2}^{4}0\cdot dx=\frac{1}{2}\cdot\frac{x^2}{2}\Bigg|_{1}^{2}=1-\frac{1}{4}=\frac{3}{4}.$$

Зауваження. Для обчислення ймовірності того, що в результаті випробовування X набуде значення, яке належить інтервалу $(a;\ в)$, доцільно скористатися функцією, яка задана умовою задачі.

Графіки функцій $F(x)$ і $f(x)$ зображені на рис. 9.1 і 9.2 відповідно.

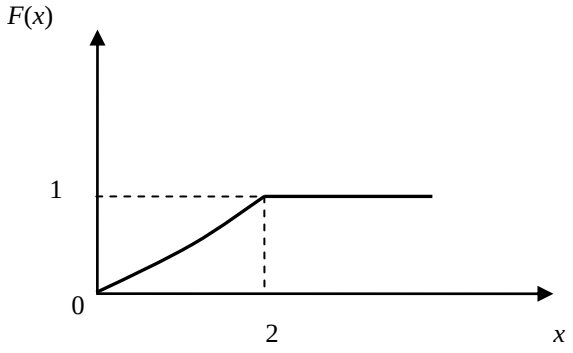


Рисунок 9.1

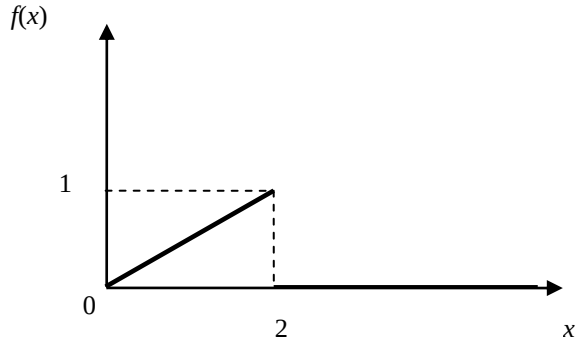


Рисунок 9.2

9.2 Рівномірний розподіл, нормальний розподіл, їх числові характеристики

Розподіл ймовірностей називають рівномірним, якщо на інтервалі, до якого належать усі можливі значення випадкової величини, щільність розподілу зберігає старе значення.

Щільність ймовірності рівномірного розподілу має вигляд:

$$f(x)=\begin{cases} 0, & якщо\ x \le a; \\ \dfrac{1}{в-a}, & якщо\ a < x \le в; \\ 0, & якщо\ x > в. \end{cases}$$

Нормальним називають розподіл ймовірностей неперервної випадкової величини, якщо її щільність розподілу має вигляд

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\delta^2}}, \qquad (9.6)$$

де a - математичне сподівання X; σ – середнє квадратичне відхилення X.

Графік щільності нормального розподілу (крива Гауса) зображений на рис. 9.3

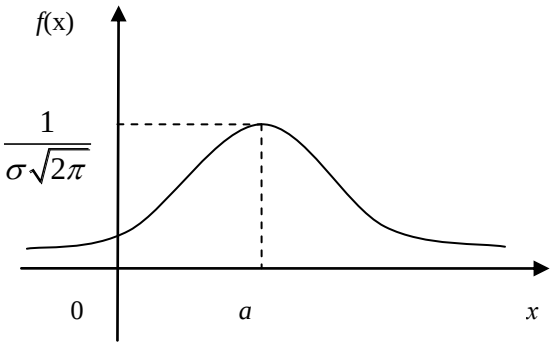


Рисунок 9.3

Теорема про імовірність потрапляння до заданого інтервалу нормально розподіленої випадкової величини.

Імовірність того, що X набуде значення, яке належить до інтервалу (α; β), обчислюється за формулою:

$$Đ(\alpha < X < \beta)=\Phi\left(\frac{\beta-\alpha}{\sigma}\right)-\Phi\left(\frac{\alpha-\grave{a}}{\sigma}\right), \qquad (9.7)$$

де $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$ - функція Лапласа.

Теорема про імовірність заданого відхилення.

Імовірність того, що абсолютна величина відхилення менша за додатне число δ,

$$P(|X-a|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right). \qquad (9.8)$$

Зокрема, при $a = 0$ вірна рівність

$$P(|X|<б)=2\Phi\left(\frac{\delta}{\sigma}\right). \qquad (9.9)$$

Приклад. Знайти математичне сподівання, дисперсію і середнє квадратичне відхилення випадкової величини *X*, яка розподілена рівномірно в інтервалі (*а;в*).

Розв'язання. Враховуючи, що в інтервалі (*a*;*b*) функція $f(x)=\frac{1}{b-a}$, і, застосовуючи послідовно формули (9.2) , (9.5), знаходимо:

$$M(X)=\int_a^b xf(x)dx=\int_a^b x\cdot\frac{1}{b-a}dx=\frac{1}{b-a}\int_a^b xdx=\frac{1}{b-a}\cdot\frac{x^2}{2}\Bigg|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{a+b}{2};$$

$$D(X)=\int_a^b x^2 f(x)dx-M^2(X)=\int_a^b x^2\frac{1}{b-a}dx-\frac{(a+b)^2}{4}=\frac{1}{b-a}\cdot\frac{x^3}{3}\Bigg|_a^b-\frac{(a+b)^2}{4}=$$

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

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{(b-a)^2}{12}}=\frac{b-a}{2\sqrt{3}}.$$

Тут враховано, що $b > a$.

Приклад. Математичне сподівання і середнє квадратичне відхилення нормально розподіленої випадкової величини *X* відповідно дорівнюють 20 і 5. Знайти ймовірність того, що в результаті випробовування X набуде значення, яке належить інтервалу (15; 25).

Розв'язання. Враховуючи, що випадкова величина X розподілена нормально, а також беручи до уваги, що $a=20,\ \sigma=5,\ \alpha=15,\ \beta=25$, за формулою (9.7) знаходимо

$$P(15<X<25)=\Phi\left(\frac{25-20}{5}\right)-\Phi\left(\frac{15-20}{5}\right)=\Phi(1)-\Phi(-1)=\Phi(1)+\Phi(1)=$$

$$=2\Phi(1)=2\cdot 0{,}3413=0{,}6826.$$

Значення функції Лапласа $\Phi(1)=0{,}3413$ знайдено за додатком Б. Враховано також, що $\Phi(-x)=-\Phi(x)$.

Приклад. Автомат виготовляє шарики. Шарик вважається придатним, якщо відхилення *X* діаметра шарика від проектного розміру за абсолютною величиною менше за 0,7 мм. Вважаючи, що випадкова величина *X* розподілена нормально з середнім квадратичним відхиленням $\sigma=0{,}4мм$, знайти, скільки буде придатних шариків серед ста виготовлених.

Розв'язання. Оскільки *X* – відхилення діаметра шарика від проектного розміру, то $M(X)=a=0$.

За формулою (9.9) $P(|X|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right)$.

Підставляючи до цієї формули значення $\sigma=0{,}4$ і $\delta=0{,}7$, отримаємо

$$P(|x|<0{,}7)=2\Phi\left(\frac{0{,}7}{0{,}4}\right)=2\Phi(1{,}75)=2\cdot 0{,}4599=0{,}9198\approx 0{,}92.$$

Таким чином, імовірність відхилення, меншого за 0,7 мм, дорівнює 0,92. Звідси випливає, що приблизно 92 шарика із 100 будуть придатними.

Питання для самоперевірки

1 Дайте означення математичного сподівання, дисперсії та середнього квадратичного відхилення.

2 Що називають рівномірним законом розподілу ймовірностей неперервної випадкової величини?

3 Що називають нормальним законом розподілу ймовірностей неперервної випадкової величини?

ЛЕКЦІЯ 10

10.1 Закон великих чисел. Теореми Бернуллі та Чебишева

10.1 Закон великих чисел. Теореми Бернуллі та Чебишева

Закон великих чисел є загальною назвою так званих граничних теорем теорії ймовірностей, які встановлюють відповідність між теоретичними та дослідними характеристиками вигадкових величин або випадкових подій при великій кількості випробувань. Вище були описані деякі граничні теореми у схемі Бернуллі. ,

Закон великих чисел відіграє важливу роль у різних процесах, пов'язаних з масовим виробництвом. Граничні теореми, які встановлюють граничні закони розподілу випадкових величин, об'єднують загальною назвою - **центральна гранична теорема**. Прикладом такої теореми є теорема Ляпунова.

Теорема (лема Чебишева).

Нехай випадкова величина X приймає лише невід'ємні значення. Тоді ймовірність того, що вона прийме значення більше додатного числа A не більше дробу $\frac{M(X)}{A}$, тобто, $P(X > A) \le \frac{M(X)}{A}$.

Доведення. Розглянемо випадок дискретної випадкової величини $\boldsymbol{X}$, яка приймає значення $x_1, x_2, \ldots, x_n$, причому $\boldsymbol{x_i}>0$.Нехай $x_1, x_2, \ldots, x_k < A$, $x_{k+1}, x_{k+2}, \ldots, x_n > A$. Математичне сподівання ДВВ X обчислюється за формулою

$$M(X) = (x_1 p_1 + \ldots + x_k p_k) + (x_{k+1} p_{k+1} + \ldots + x_n p_n),$$

звідки

$$M(X) \ge x_{k+1} p_{k+1} + \ldots + x_n p_n .$$

Враховуючи, що $x_{к+1}, x_{к+2} \ldots x_n > A$, отримаємо

$$M(X) \ge A(x_{k+1} p_{k+1} + \ldots + x_n p_n),$$

де $p_{k+1} = P(X = x_{k+1}), \ldots\ldots\ldots, p_n = P(X = x_n)$.

Таким чином,

$$M(X) \ge A \cdot P(X > A), \Rightarrow P(X > A) \le \frac{M(X)}{A}.$$

Теорема. (Нерівність Чебишева).

Для довільної випадкової величини X і довільного дійсного числа $\varepsilon > 0$ має місце нерівність

$$P\left(\left|X - M(X)\right| \geq \varepsilon\right) \leq \frac{D(X)}{\varepsilon^2}. \qquad (10.1)$$

Доведення. Подія $\left|X - M(X)\right| \geq \varepsilon$ рівносильна події $\frac{(X - M(X))^2}{\varepsilon^2} \geq 1$, тому $P\left(\left|X - M(X)\right| \geq \varepsilon\right) = P\left(\frac{(X - M(X))^2}{\varepsilon^2} \geq 1\right)$. Випадкова величина $\frac{(X - M(X))^2}{\varepsilon^2} \geq 1$ невід'ємна, тому за попередньою теоремою, властивостями математичного сподівання і означенням дисперсії будемо мати

$$P\left(\frac{(X - M(X))^2}{\varepsilon^2} \geq 1\right) \leq M\left(\frac{(X - M(X))^2}{\varepsilon^2}\right) = \frac{1}{\varepsilon^2} M\left((X - M(X))^2\right) = \frac{D(X)}{\varepsilon^2}.$$

Враховуючи, що $P\left(\left|X - M(X)\right| \geq \varepsilon\right) = P\left(\frac{(X - M(X))^2}{\varepsilon^2} \geq 1\right)$, отримаємо

$$P\left(\left|X - M(X)\right| \geq \varepsilon\right) \leq \frac{D(X)}{\varepsilon^2}.$$

Теорему доведено.

Нерівність Чебишева (10.1) допускає іншу форму. Оскільки подія $\left|X - M(X)\right| < \varepsilon$ протилежна події $\left|X - M(X)\right| \geq \varepsilon$, то має місце рівність

$$P\left(\left|X - M(X)\right| < \varepsilon\right) + P\left(\left|X - M(X)\right| \geq \varepsilon\right) = 1,$$

звідси та з нерівності (10.1) отримуємо нерівність

$$P\left(\left|X - M(X)\right| < \varepsilon\right) \geq 1 - \frac{D(X)}{\varepsilon^2}. \qquad (10.2)$$

Теорема. (Теорема Чебишева, закон великих чисел). Нехай дисперсії випадкових величин $X_1, X_2, \ldots, X_n$ обмежені однією сталою $D(X_i) \leq C$ для $i = 1, \ldots, n$, $\varepsilon > 0$ - довільне дійсне число, і $\bar{X} = \frac{1}{n}\left(X_1 + \ldots + X_n\right)$. Тоді ймовірність нерівності $\left|\bar{X} - M(\bar{X})\right| < \varepsilon$ буде як завгодно близька до 1, якщо число n досить велике, тобто

$$\lim_{n \to \infty} P\left(\left|\bar{X} - M(\bar{X})\right| < \varepsilon\right) = 1. \qquad (10.3)$$

Доведення.

Застосуємо нерівність Чебишева у формі (10.2) до величини $\bar{X}$:

$$P\left(\left|\bar{X} - M(\bar{X})\right| < \varepsilon\right) \geq 1 - \frac{D(\bar{X})}{\varepsilon^2}.$$

За властивостями дисперсії і умовою теореми

$$D(\bar{X}) = \frac{1}{n^2}\left(D(X_1) + \ldots + D(X_n)\right) \le \frac{nC}{n^2} = \frac{C}{n}.$$

Звідси, враховуючи, що ймовірність будь-якої події не перевищує 1, отримаємо

$$1 \ge P\left(\left|\bar{X} - M(\bar{X})\right| < \varepsilon\right) \ge 1 - \frac{C}{n\varepsilon^2}.$$

Переходячи в отриманих нерівностях до границі при $n \to \infty$, будемо мати рівність

$$\lim_{n\to\infty} P\left(\left|\bar{X} - M(\bar{X})\right| < \varepsilon\right) = 1.$$

Теорему доведено.

Відзначимо один її частинний випадок: якщо всі X_i мають рівні математичні сподівання $M(X_i) = a$ для $i = 1,\ldots,n$, і $D(X_i) \le C$ для $i = 1,\ldots,n$, тоді

$$\lim_{n\to\infty} P\left(\left|\bar{X} - a\right| < \varepsilon\right) = 1 \qquad (10.4)$$

Суть теореми Чебишева така. Незважаючи на те, що кожна з незалежних випадкових величин X_i може приймати значення, далекі від математичного сподівання $M(X_i)$**,** середнє арифметичне $\bar{X}$ достатньо великого числа випадкових величин з великою ймовірністю близьке до середнього арифметичного їх математичних сподівань.

Тому теорема Чебишева має важливе практичне значення. Нехай, наприклад, вимірюється деяка величина. Звичайно, за її значення приймають середнє арифметичне результатів кількох вимірювань. Але чи так можна робити? Теорема Чебишева дає позитивну відповідь на це питання.

На теоремі Чебишева ґрунтується широко використовуваний у статистиці вибірковий метод, згідно з яким за порівняно невеликою вибіркою роблять висновок про всю сукупність досліджуваних об’єктів.

Із частинного випадку (10.4) теореми Чебишева випливає теорема Бернуллі, яка є найпростішою формою закону великих чисел.

Теорема (Теорема Бернуллі).

Нехай m - число появ події A в n незалежних випробуваннях, p - ймовірність появи події A в кожному випробуванні. Тоді для довільного дійсного числа $\varepsilon > 0$ має місце рівність

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1. \qquad (10.5)$$

Доведення. Позначимо через X_i випадкову величину, що дорівнює числу появ події A в i - му випробуванні, $i = 1,\ldots,n$. Тоді, очевидно

$$m = X_1 + \ldots + X_n, \quad M(X_i) = p, \quad D(X_i) = pq \le 1,$$

тобто виконані всі умови частинного випадку теореми Чебишева. Тоді рівність (10.4) переходить у рівність (10.5). Теорему Бернуллі доведено.

Практичний зміст теореми Бернуллі такий: при сталості ймовірності випадкової події A в цих випробуваннях можна з імовірністю, як завгодно

близькою до 1, стверджувати, що спостережувана відносна частота події *A* буде як завгодно мало відхилятися від її ймовірності.

Нагадаємо, що **нормальний закон розподілу** має важливе значення в теорії ймовірностей. Цим законом характеризуються ймовірності стрільби по цілі, ймовірності у вимірюваннях, у страховій справі тощо.

Російський математик О.М. Ляпунов (1857-1918) довів так звану центральну граничну теорему, яка дає відповідь на питання, чому нормально розподілені випадкові величини часто зустрічаються на практиці.

Сформулюємо її наступним чином:

Якщо випадкова величина *X є* сумою великого числа незалежних випадкових величин, кожної з яких на всю суму дуже малий, то випадкова величина X має розподіл близький до нормального.

Із нерівності Чебишева випливає, що

$$P(|m-np|<\varepsilon)\geq 1-\frac{npq}{\varepsilon^2},$$

або

$$P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)\geq 1-\frac{pq}{n\varepsilon^2}, \qquad (10.6)$$

а також формула, яка виражає теорему Чебишева:

$$P(|\overline{X}-M(\overline{X})|<\varepsilon)\geq 1-\frac{C}{n\varepsilon^2}. \qquad (10.7)$$

Приклад. Дискретна випадкова величина *X* задана законом розподілу:

X	0,1	0,4	0,6
P	0,2	0,3	0,5

Використовуючи нерівність Чебишева, оцінити ймовірність того, що $|X-M(X)|<\sqrt{0{,}4}$.

Розв'язання. За даним законом розподілу *X* знайдемо $M(X)$ і $D(X)$:

$M(X) = 0{,}1\cdot 0{,}2 + 0{,}4\cdot 0{,}3 + 0{,}6\cdot 0{,}5 = 0{,}02 + 0{,}12 + 0{,}3 = 0{,}44$;

$D(X) = (0{,}1)^2\cdot 0{,}2 + (0{,}4)^2\cdot 0{,}3 + (0{,}6)^2\cdot 0{,}5 - (0{,}44)^2 =$

$= 0{,}002 + 0{,}048 + 0{,}18 - 0{,}1936 = 0{,}0364.$

Підставляючи $M(X) = 0{,}44$, $D(X) = 0{,}0364$, $\varepsilon=\sqrt{0{,}4}$ до нерівності (5.8), остаточно отримаємо:

$$P\left(|X-0{,}44|<\sqrt{0{,}4}\right)\geq 1-\frac{0{,}0364}{0{,}4}=0{,}909.$$

Питання для самоперевірки

1 Що можна оцінити за допомогою нерівності Чебишева? Яким чином?

2 Сформулюйте теорему Чебишева та теорему Бернуллі.

3 Як оцінюються $P\left(\left|\frac{m}{n}-\delta\right|<\varepsilon\right)$ та $P\left(\left|\overline{O}-M(\overline{O})\right|<\varepsilon\right)$?

4 Суть теореми Чебишева?

5 Який практичний зміст теореми Бернулі?

ЛЕКЦІЯ 11

11.1 Генеральна та вибіркова сукупності. Статистичний розподіл вибірки
11.2 Емпірична функція розподілу та її властивості

11.1 Генеральна та вибіркова сукупності. Статистичний розподіл вибірки

Нехай маємо деяку сукупність однорідних об'єктів. Наприклад: а) партія деталей; б) партія ламп.

Як вивчити кількісні та якісні ознаки, які характеризують дану сукупність?

В а) і б) за якісну ознаку можна прийняти стандартність деталі або лампи. За кількісну ознаку у випадку а) можна прийняти контролюючий розмір деталі, у випадку б) – час роботи лампи. Іноді для вивчення кількісної або якісної ознаки проводять так зване суцільне обстеження, тобто обстежують кожний об'єкт, що на практиці не завжди можливо. Тому замість перевірки всіх об'єктів з даної партії перевіряють тільки невелику її частину, вибрану випадково.

Означення 1 Вся сукупність однорідних об'єктів, яка підлягає вивченню, називається генеральною сукупністю.

Означення 2 Частина випадково відібраних об'єктів з генеральної сукупності називається вибірковою сукупністю або вибіркою.

Означення 3 Число об'єктів у генеральній сукупності або вибірки називається об'ємом генеральної сукупності або вибірки і позначається відповідно через N і n.

Означення 4 Вибірка є представницькою (репрезентативною), якщо вона утворена випадково і достатньо добре відтворює генеральну сукупність.

Із означення 4 випливає, що вибірка може бути репрезентативною, якщо правильно організувати спосіб відбору з генеральної сукупності.

Способи відбору:

1 Випадкова вибірка з поверненням (повторний відбір).
2 Випадкова відбірка без повернення (безповторний відбір).
3 Механічний відбір.
4 Типовий відбір.
5 Серійний відбір.

Механічний відбір – вся генеральна сукупність механічно розділяється на стільки частин, скільки елементів повинно бути у вибірки. Потім з кожної частини випадково по одному елементу складається вибірка.

Типовий відбір використовується, коли у виготовлені елементів вибірки приймають участь різні станки, цеха. Вся продукція розбивається на частини і далі як у механічному відбору формують випадкову вибірку.

Серійний відбір – об'єкти вибираються з генеральної сукупності не по одному, а «серіями».

Нехай з генеральної сукупності кількісної ознаки X об'ємом N елементів виконана вибірка об'ємом n елементів, причому ознака X прийняла значення x_1,

$x_2, \ldots, x_к$, при цьому значення x_1 повторилось n_1 разів, x_2 – n_2 разів,…, $x_к$ – n_k разів, де $n_1 + n_2 + \ldots + n_k = \sum_{i=1}^{k} n_i = n$.

Означення. Якщо ознака X приймає тільки ізольовані одне від одного значення, то вона називається дискретною, а якщо вона може приймати значення, які скільки завгодно мало відрізняються одне від одного, то вона називається неперервною.

Означення. Значення $x_1, x_2, \ldots, x_к$, які приймає ознака X, називаються її варіантами.

Означення. Послідовність варіант, виписаних у порядку зростання, називається варіаційним рядом.

Означення. Кількість спостережень $n_1, n_2, \ldots, n_k$ разів появи відповідно варіант $x_1, x_2, \ldots, x_к$ називаються частотами варіант, а відношення $\frac{n_i}{n} = W_i$ $(i = 1, 2, \ldots, k)$ – відносними частотами (частостями).

Означення. Таблиця, яка встановлює відповідність між:

а) варіантами та їх частотами (табл.11.1);

б) варіантами та їх відносними частотами (табл.11.2)

називається статистичним розподілом вибірки, тобто

Таблиця 11.1

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Таблиця 11.2

x_i	x_1	x_2	…	x_k
W_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_k}{n}$

$$\sum_{i=1}^{k} W_i = 1.$$

Статистичний розподіл вибірки є аналогом (реалізацією) розподілу ознаки (випадкової величини) X.

Зауваження. У випадку, коли ознака X є неперервною, будувати варіаційний ряд не дуже зручно за великою кількістю вихідних даних. Тому число інтервалів k за формулою Стерджеса рекомендовано вибирати

$k = 1 + 3{,}322 \lg n$, а величина інтервалу $h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n}$, де $x_{\max}$ – найбільше значення ознаки, $x_{\min}$ – найменше значення ознаки, n – число вихідних даних (об’єм вибірки).

За початок першого інтервалу вибирають величину $x_{поч.} = x_{\min} - \frac{h}{2}$, а кінець останнього інтервалу ($x_{кінц.}$) повинен задовольняти умові

$$x_{кінц.} - h \le x_{\max} < x_{кінц.}.$$

Приклад. Під час дослідження кількісної ознаки X із генеральної сукупності було отримано вибірку

4, 3, 6, 4, 7, 2, 5, 1, 2, 5, 4, 4, 3, 5, 6, 3, 4, 1, 3, 4 . (11.1)

Знайти об'єм вибірки та її статистичний розподіл частоти та відносних частот.

Розв'язання. Оскільки вибірка складається з 20 значень, то об'єм вибірки $n = 20$. Вважаючи означення дискретної ознаки, вона задовольняє її умовам.

Побудуємо варіаційний ряд вибірки, тобто розташуємо варіанти в порядку їх зростання.

1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7**.** (11.2)

У даній вибірці всього сім різних значень, тобто варіант 1, 2, 3, 4, 5, 6, 7. Знайдемо їх частоти $n_1=2;\ n_2=2;\ n_3=4;\ n_4=6;\ n_5=3;\ n_6=2;\ n_7=1$ та їх відносні частоти за формулою $W_i=\frac{n_i}{n}$; $W_1=\frac{2}{20}=0{,}1;\ W_2=\frac{2}{20}=0{,}1;\ W_3=\frac{4}{20}=0{,}2;$

$$W_4=\frac{6}{20}=0{,}3;\ W_5=\frac{3}{20}=0{,}15;\ W_6=\frac{2}{20}=0{,}1;\ W_7=\frac{1}{20}=0{,}05.$$

Запишемо шукані статистичні розподіли (табл. 11.3, табл. 11.4):

Таблиця 11.3

x_i	1	2	3	4	5	6	7
n_i	2	2	4	6	3	2	1

$$N=\sum n_i=20$$

Таблиця 11.4

x_i	1	2	3	4	5	6	7
W_i	0,1	0,1	0,2	0,3	0,15	0,1	0,05

Контроль: $\sum W_i = 1$.

Приклад. При вимірюванні діаметрів валиків після шліфування одержані наступні результати: 6,75; 6,77; 6,77; 6,73; 6,76; 6,74; 6,70; 6,75; 6,71; 6,72; 6,72; 6,79; 6,71; 6,78; 6,73; 6,70; 6,73; 6,77; 6,75; 6,74. Знайти об'єм вибірки та побудувати її статистичні розподіли.

Розв'язання. Об'єм вибірки дорівнює $n = 20$. Так як варіанти мало відрізняються одне від одного, то маємо інтервальний ряд. Для його побудови треба знайти кількість інтервалів $k=1+3{,}322\lg n$, то довжину кожного інтервалу знайдемо за формулою

$$h=\frac{x_{\max}-x_{\min}}{k},\ \text{де}\ k=1+3{,}322\cdot\lg 20=1+3{,}322\cdot 1{,}3=1+4{,}32=5{,}32;$$

$$x_{max}=6{,}79;\ x_{min}=6{,}70;\ h=\frac{6{,}79-6{,}70}{5{,}32}\approx 0{,}02.$$

За початок першого інтервалу вибираємо величину $x_{поч.}=x_{min}-\frac{h}{2}=6{,}70-0{,}01=6{,}69.$

Будуємо статистичні розподіли інтервального варіаційного ряду (табл. 11.5, табл. 11.6), для якого кінцеві значення інтервалів не входять в кількість значень варіантів, які попали в даний інтервал.

Таблиця 11.5

$(x_i;x_{i+1})$	6,69-6,71	6,71-6,73	6,73-6,75	6,75-6,77	6,77-6,79	6,79-6,81
n_i	2	4	5	4	4	1

Контроль: $n = 2+4+5+4+4+1=20$.

Таблиця 11.6

$(x_i;x_{i+1})$	6,69-6,71	6,71-6,73	6,73-6,75	6,75-6,77	6,77-6,79	6,79-6,81
W_i	0,1	0,2	0,25	0,2	0,2	0,05

Контроль: $W=0,1+0,2+0,25+0,2+0,2+0,05=1$.

11.2 Емпірична функція розподілу та її властивості

Під час вивчення варіаційних рядів поряд з поняттям частоти використовується поняття накопиченої частоти, яка позначається $n_i^{нак.}$. Накопичена частота показує, скільки спостерігалось варіант із значенням ознаки, меншим *X*. Відношення накопиченої частоти $n_i^{нак.}$ до загального числа спостережень *n* називається накопиченою відносною частотою $W_i^{нак.}$. Накопичені частоти (частості) для кожного інтервалу знаходяться послідовним додаванням частот (частостей) всіх попередніх інтервалів, включно даний.

Для графічного зображення варіаційних рядів найбільш часто використовуються емпірична функція розподілу (комулята), полігон та гістограма.

Означення. Емпіричною функцією розподілу $F^*(x)$ називається відносна частота (частость) того, що ознака (випадкова величина) *X* прийме значення, менше заданого аргументу *x*, тобто $F^*(x) = W_x^{нак.} = \dfrac{n_x^{нак.}}{n}$ (11.3)

Властивості $F^*(x)$:

1) $0 \le F^*(x) \le 1$;

2) $F^*(x_{\min}) = 0$, де $x_{\min}$ є найменшою варіантою варіаційного ряду;

3) $F^*(x)\big|_{x>x_{\max}} = 1$, де $x_{\max}$ є найбільшою варіантою варіаційного ряду;

4) $F^*(x)$ є неспадною функцією аргументу *x*, тобто $F^*(x_2) \ge F^*(x_1)$ при $x_2 \ge x_1$.

Питання для самоперевірки

1 Дати визначення генеральної та вибіркової сукупності.
2 Яка вибірка є представницькою (репрезентативною)?
3 Способи відбору вибірки, дати означення кожного з них.
4 Дати визначення дискретної та неперервної ознаки *X*.

5 Що називається варіантою, варіаційним рядом?

6 Що таке частота, відносна частота варіант?

7 Дати визначення дискретного статистичного розподілу вибірки.

8 Аналогом чого є статистичний розподіл вибірки?

9 Як будується статистичний розподіл вибірки для неперервної ознаки X?

10 Дати означення емпіричної функції розподілу $F*(x)$.

11 Сформулюйте властивості $F*(x)$.

ЛЕКЦІЯ 12

12.1 Полігон та гістограма

12.2 Генеральна і вибіркова середні. Дисперсія

12.1 Полігон та гістограма

Дискретний статистичний розподіл вибірки можна зобразити графічно у вигляді ламаної лінії, відрізки якої з'єднують точки $(x_1;n_1),(x_2;n_2),\ldots(x_k;n_k)$, або $(x_1;w_1),(x_2;w_2),\ldots,(x_k;w_k)$.

Перша ламана лінія називається полігоном частот, друга – полігоном відносних частот.

Полігон, як правило, застосовується для зображення дискретного варіаційного ряду.

Приклад. Вибірку задано дискретним статистичним розподілом (табл. 12.1).

Таблиця 12.1

x_i	-6	-4	-2	2	4	6
n_i	5	10	15	20	40	10
w_i	0,05	0,1	0,25	0,2	0,4	0,1

Необхідно:

1 Побудувати $F^*(x)$ і графічно її зобразити.

2 Графічно зобразити полігони частот і відносних частот.

Розв'язання.

1 Згідно з означенням та властивостями $F^*(x)$ має такий вигляд:

$$F^*(x)=\frac{n_x^{нак.}}{n}=\begin{cases}0, & x\le -6,\\ 0,05, & -6<x\le -4,\\ 0,15, & -4<x\le -2,\\ 0,3, & -2<x\le 2,\\ 0,5, & 2<x\le 4,\\ 0,9, & 4<x\le 6,\\ 1,0, & x>6.\end{cases}$$

Зобразимо функцію $F^*(x)$ графічно (рис.12.1).

2 Графічно зобразимо полігони частот і відносних частот (рис. 12.2, рис.12.3).

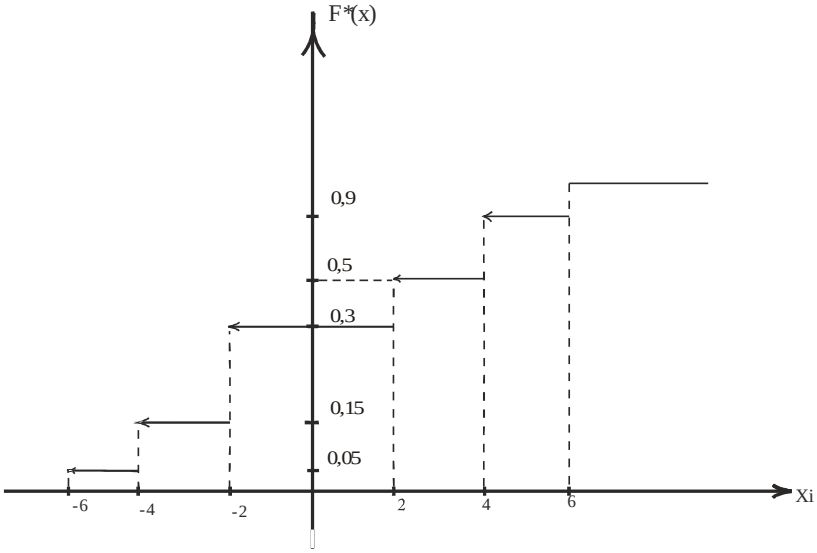


Рис.12.1

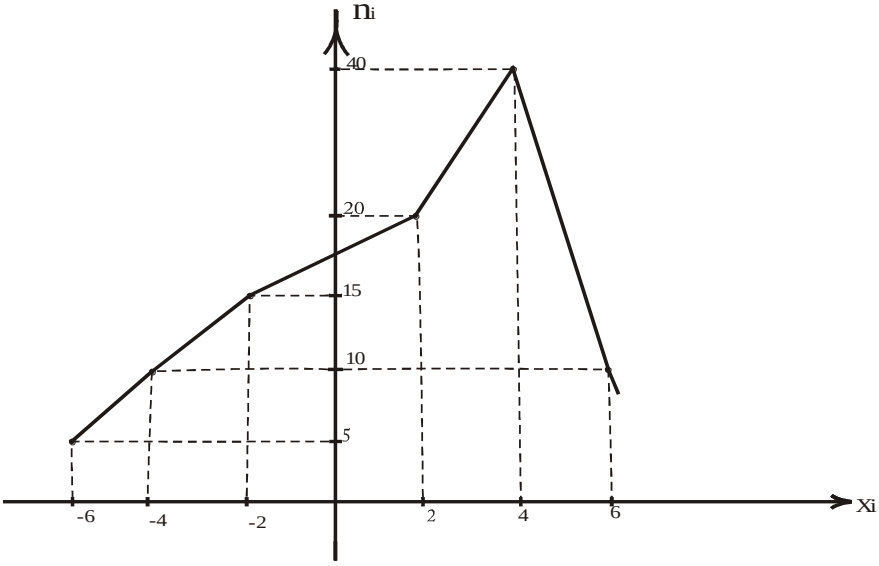


Рис.12.2

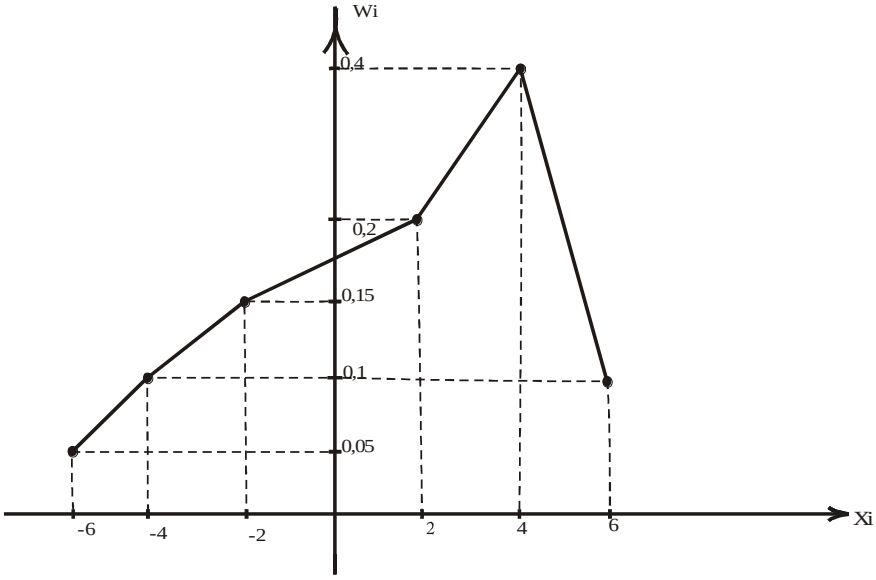


Рис.12.3

Зауваження. У процесі побудови полігона інтервального варіаційного ряду за значення варіанти x_i вибирається середина частинних інтервалів, а в якості частот n_i та відносних частот w_i вибирається їх значення для відповідного інтервалу.

Приклад. Вибірка задана інтервальним розподілом частот (табл. 12.2):

Таблиця 12.2

$(x_i\,;x_{i+1})$	(1;2)	(2;3)	(3;4)	(4;5)	(5;6)	(6;7)	(7;8)
n_i	19	9	12	14	7	22	17

$$\sum n_i = 100\,.$$

Побудувати полігон відносних частот.

Розв'язання. Запишемо інтервальний розподіл відносних частот (табл. 12.3).

Таблиця 12.3

$(x_i\,;x_{i+1})$	(1;2)	(2;3)	(3;4)	(4;5)	(5;6)	(6;7)	(7;8)
w_i	0,19	0,09	0,12	0,14	0,07	0,22	0,17

$\sum w_i = 1$

Знайдемо середини частинних інтервалів

$$x_1^{сер} = \frac{x_1 + x_2}{2} = \frac{1+2}{2} = 1{,}5; \qquad x_2^{сер} = \frac{x_2 + x_3}{2} = \frac{2+3}{2} = 2{,}5;$$

$$x_3^{сер} = \frac{x_3 + x_4}{2} = \frac{3+4}{2} = 3{,}5; \qquad x_4^{сер} = \frac{x_4 + x_5}{2} = \frac{4+5}{2} = 4{,}5;$$

$$x_5^{сер} = \frac{x_5 + x_6}{2} = \frac{5+6}{2} = 5{,}5; \qquad x_6^{сер} = \frac{x_6 + x_7}{2} = \frac{6+7}{2} = 6{,}5;$$

$$x_7^{сер} = \frac{x_7 + x_8}{2} = \frac{7+8}{2} = 7{,}5.$$

Відкладемо на осі абсцис одержані середини частинних інтервалів $x_i^{сер}$, а на осі ординат – значення відповідних їм відносних частот w_i. Послідовно з'єднуючи між собою точки $(x_i^{сер}\,;w_i)$ – відрізками, отримуємо полігон відносних частот (рис.12.4).

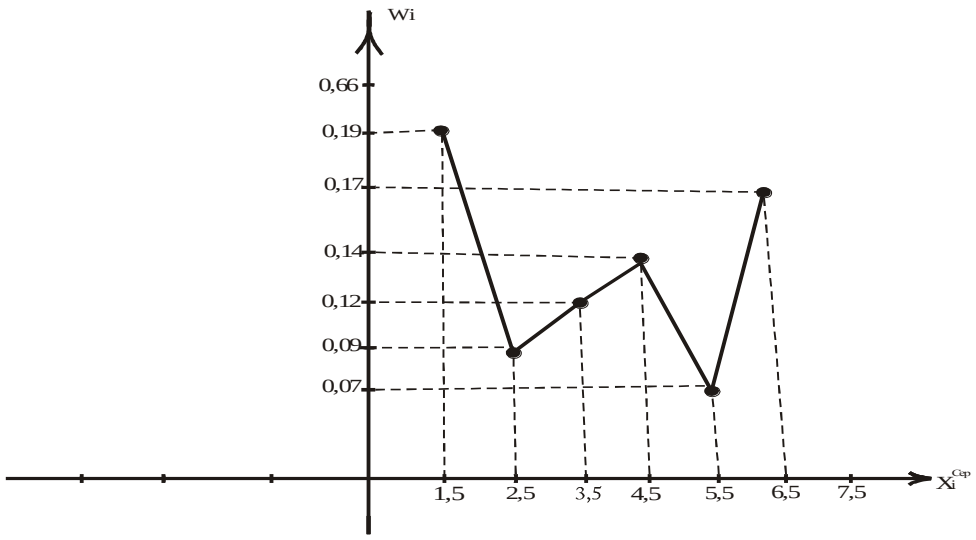


Рис.12.4

Означення. Гістограмою частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є частинні інтервали довжиною h, а висоти дорівнюють відношенню $\frac{n_i}{h}$ (щільність частоти). Площа *i*-го частинного прямокутника $h\frac{n_i}{h} = n_i$, тобто сумі частот варіант, що потрапили в *i*-й частинний інтервал. Площа гістограми частот дорівнює сумі всіх частот, тобто обсягу вибірки, тобто площа гістограми частот дорівнює $\sum n_i = n$.

Аналогічно, гістограмою відносних частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є частинні інтервали довжиною h, а висоти дорівнюють відношення $\frac{w_i}{n}$ (щільність відносної частоти). Площа *i*-го частинного прямокутника дорівнює $h\frac{w_i}{h} = w_i$, тобто $S = \sum_{i=1}^{k} w_i = 1$, площа гістограми відносних частот дорівнює одиниці.

Зауваження 1 Гістограму можна побудувати тільки для інтервального статистичного розподілу.

Зауваження 2 Очевидно, що при збільшенні *n* можна вибрати все більш малі інтервали (*h*), при цьому гістограма буде наближатися до деякої кривої, яка обмежує площу близьку до 1. Ця крива є графіком щільності розподілу випадкової величини *X*.

Зауваження 3 Комулятивна крива (комулята) – крива накопичених частот або відносних частот. Для дискретного статистичного розподілу комулятою є ламана лінія, яка з'єднує точки $(x_i; n_i^{нак.})$ або $(x_i; w_i^{нак.})$ ($i=1,2,\ldots,k$). Для інтервального статистичного розподілу є ламана лінія, яка починається з точки, абсциса яка дорівнює початку першого інтервалу, а ордината – накопиченій частоті, рівній нулю.

Зауваження 4 Полігон і гістограма аналогічні криві розподілу ознаки *X*, а емпірична функція розподілу $F^*(x)$ – функція розподілу випадкової величини *X*.

Розглянемо приклад побудови гістограми, полігона частот для інтервального статистичного розподілу.

Приклад. Вибірку задано інтервальним розподілом частот (табл. 12.4).

Таблиця 12.4

(x_i, x_{i+1})	(1;2)	(2;3)	(3;4)	(4;5)	(5;6)
n_i	13	9	5	16	7

Побудувати гістограму, полігон частот.

Розв'язання.

1 Побудуємо гістограму частот. Для цього знайдемо об'єм вибірки

$$n=13+9+5+16+7=50$$

і довжину інтервалів $h = x_{i+1} - x_i = 1$, а також висоти прямокутників $\frac{n_i}{h}$.

$$\frac{n_1}{h}=\frac{13}{1}=13;\ \frac{n_2}{h}=\frac{9}{1}=9;\ \frac{n_3}{h}=\frac{5}{1}=5;\ \frac{n_4}{h}=\frac{16}{1}=16;\ \frac{n_5}{h}=\frac{7}{1}=7.$$

Відповідно до заданих даних побудуємо прямокутники з основою $h=1$ і висотами $\frac{n_i}{h}$ (рис.12.5).

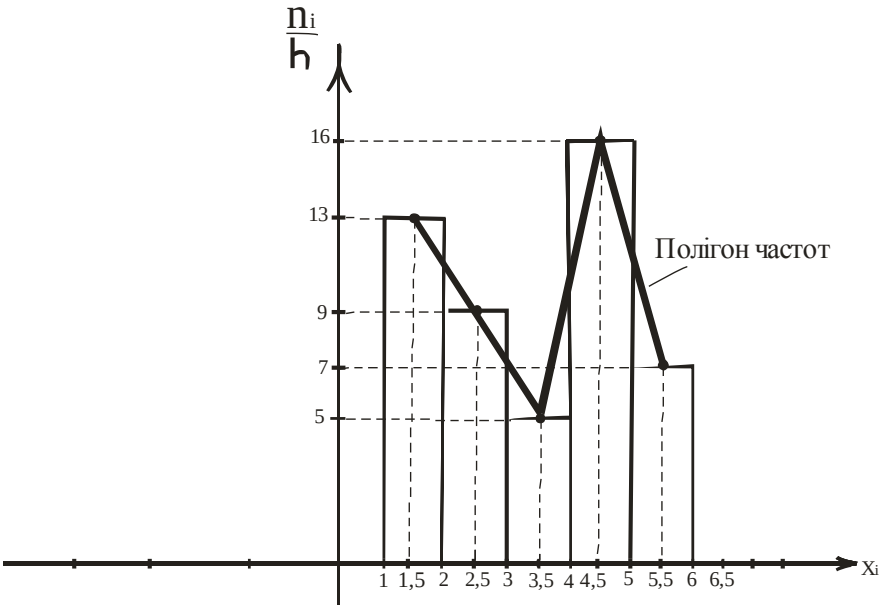


Рис.12.5

Прямокутники утворюють гістограму частот.

2 Для побудови полігона частот треба з'єднати середини верхньої основи кожного прямокутника ламаною лінією.

12.2 Генеральна і вибіркова середні. Дисперсія

Побудувавши статистичний розподіл вибірки і зобразивши його графічно, можна одержати первісне уявлення про закономірності, які мають місце у ряду спостережень. Але на практиці часто цього не достатньо. Така ситуація виникає, коли необхідно уточнити ті або інші відомості про ряд розподілу, або коли виникає необхідність порівняти два ряди і більше. При цьому необхідно порівняти однотипні варіаційні ряди, тобто ряди, які одержані при обробці зрівнювальних статистичних даних. Зрівнювальні розподіли можуть суттєво відрізнятися одне від одного. Тому для подальшого вивчення зміни значень випадкової величини використовують числові характеристики варіаційних рядів, які зображають деякі сталі величини, які подають варіаційний ряд в цілому і відображають властивості сукупності закономірностей, що вивчаються. До таких числових характеристик відносяться середня величина ряду розподілу, величини, які відображають варіацію змін – розмах, дисперсія, середнє квадратичне відхилення та інші.

1 Вибіркова середня величина $\overline{x}_{\hat{a}}$ вибірки. Дана величина визначається формулами:

$$\overline{x}_{\hat{a}} = \frac{\sum_{з=1}^{k} x_i n_i}{n} \text{ (для упорядкованої вибірки) або } \overline{x}_{\hat{a}} = \sum_{i=1}^{n} x_i w_i , \qquad (12.1)$$

де x_i – значення *i*-ої варіанти, n_i – частота *i*-ої варіанти, n – об'єм вибірки, k – кількість варіант у вибірках, w_i – відносна частота *i*-ї варіанти. Для інтервального варіаційного ряду x_i є серединою інтервалів варіаційного ряду.

$$\overline{x}_{в} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ (для невпорядкованої вибірки).} \qquad (12.2)$$

Статистичний зміст середньої величини заключається у тому, що навколо неї концентруються інші результати спостережень.

2 Відхил варіант визначають формулою

$$(x_i - \overline{x_в}) n_i ,\ i = 1, \overline{k} . \qquad (12.3)$$

При цьому очевидно, що $\sum_{i=1}^{k} (x_i - \overline{x_в}) n_i = 0$.

3 Модою *Mo** вибірки варіант називається варіанта, якій відповідає найбільша частота. Якщо варіаційний ряд має одну таку частоту, то ряд називається одномодальним, дві – двомодальним і т.д. Для визначення моди інтервального статистичного розподілу необхідно знайти такий частинний інтервал, що має найбільшу частоту появи.

Мода обчислюється за формулою

$$Mo^* = x_{i-1} + \frac{n_{Mo} - n_{Mo-1}}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}} h, \qquad (12.4)$$

де x_{i-1} – початок модального інтервалу; h – довжина часткового інтервалу; n_{Mo} – частота модального інтервалу; n_{Mo-1} – частота домодального інтервалу; n_{Mo+1} – частота післямодального інтервалу.

4 Медіана *Me**. Медіаною дискретного статистичного розподілу називають варіанту, яка поділяє варіаційний ряд на дві рівні частини за кількістю варіант. Для визначення медіани інтервального статистичного розподілу вибірки необхідно визначити медіанний частковий інтервал. Якщо, наприклад, на *i*-му інтервалі $[x_{i-1}, x_i]$ $F^*(x_{i-1}) < 0{,}5$ і $F^*(x_i) > 0{,}5$, то беручи до уваги те, що досліджувана ознака *X* є неперервною і при цьому $F^*(x)$ є не спадною функцією, всередині інтервалу $[x_{i-1}, x_i]$неодмінно існує таке значення $x = Me^*$, де $F^*(Me^*) = 0{,}5$. Медіана обчислюється за формулою

$$Me^* = x_{i-1} + \frac{0{,}5 - F^*(x_{i-1})}{F^*(x_i) - F^*(x_{i-1})} h, \qquad (12.5)$$

де $h = x_i - x_{i-1}$ – довжина інтервалу.

Приклад. За даними інтервального статистичного розподілу вибірки (табл. 12.5)

Таблиця 12.5

h=4	0–4	4–8	8–12	12–16	16–20	20–24
n_i	6	14	20	25	30	5

побудувати гістограму частот і $F^*(x)$. Визначити Мо*, Ме*.

Розв'язання. Гістограму частот зобразимо на рис. 12.6.

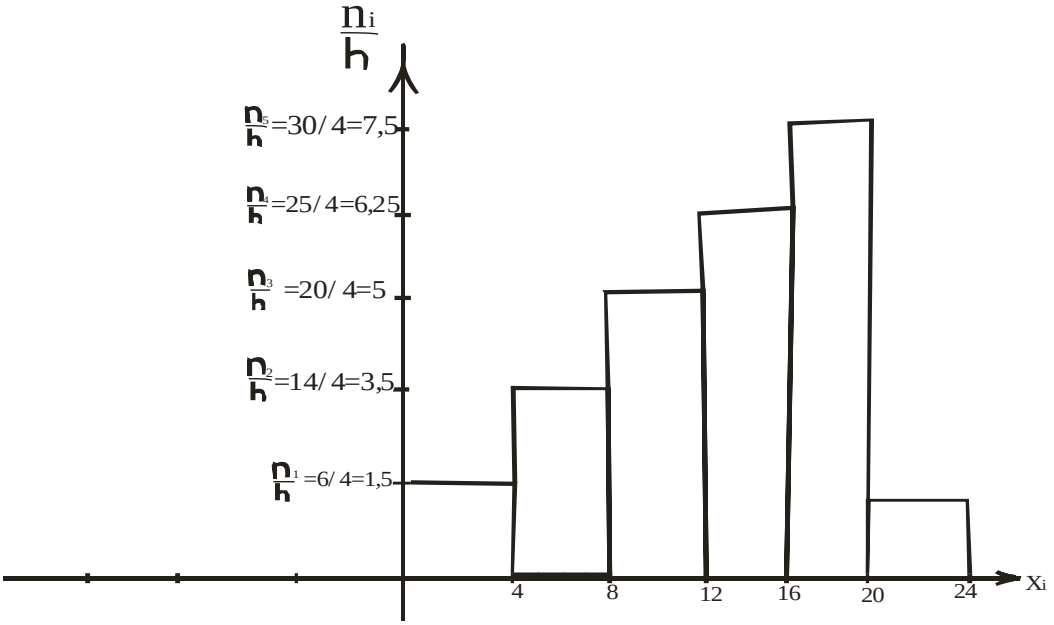


Рис.12.6

Побудуємо графік $F^*(x)$, для цього спочатку розрахуємо її значення на кожному з інтервалів, скориставшись формулою $F^*(x) = \frac{n_x^{нак}}{n}$.

$$F^*(x) = \begin{cases} 0, & x \le 0, \\ \frac{6}{100} = 0,06, & 0 < x \le 4, \\ \frac{20}{100} = 0,2, & 4 < x \le 8, \\ \frac{40}{100} = 0,4, & 8 < x \le 12, \\ \frac{65}{100} = 0,65, & 12 < x \le 16, \\ \frac{95}{100} = 0,95, & 16 < x \le 20, \\ \frac{100}{100} = 1, & 20 < x \le 24. \end{cases}$$

Графік $F^*(x)$ зображено на рис.12.7.

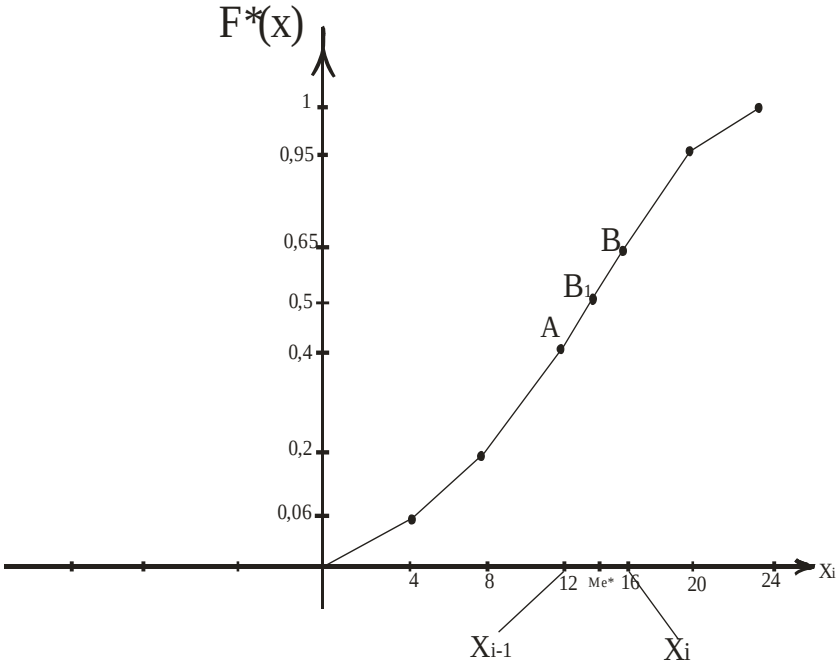


Рис.12.7

Використовуючи рис.12.7 і формулу (12.5), обчислимо значення Me^*:

$$x_{i-1} = 12,\ x_i = 16,\ h = x_i - x_{i-1} = 4,\ F^*(x_{i-1}) = 0,4,\ F^*(x_i) = 0,65,$$

$$Me^* = 12 + \frac{0,5 - 0,4}{0,65 - 0,4} 4 = 13.$$

Відповідно до таблиці 12.5 найбільшу частоту має інтервал (16 – 20), значення якої дорівнює 30, тому за формулою (12.4) обчислимо моду Mo^* інтервального розподілу, використовуючи значення $x_{i-1} = 16$, $h = x_i - x_{i-1} = 4$,

$$h_{Mo} = 30,\ h_{Mo-1} = 25,\ h_{Mo+1} = 5.$$

$$Mo^* = x_{i-1} + \frac{h_{Mo} - h_{Mo-1}}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}} h = 16 + \frac{30 - 25}{2 \cdot 30 - 25 - 5} \cdot 4 = 16,7.$$

Таким чином, для вихідного інтервального статистичного розподілу, заданого таблицею 12.5 медіана *Me*=13,6; Mo*=16,7.*

5 Розмах *R*.

Для грубого розходження варіант відносно середнього значення $\overline{x_в}$ застосовується величина, яка дорівнює різниці між найбільшою і найменшою варіантами варіаційного ряду $R = x_{\max} - x_{\min}$ (12.6)

6 Дисперсія вибірки.

Для більш точного оцінювання розходження варіант вибірки відносно $\overline{x_в}$ вибирається дисперсія.

Вибірковою дисперсією $D_в$ називається середній квадратичний відхил варіант від їх середньої арифметичної, яку можна обчислити за формулами.

$$D_в = \frac{\sum_{i=1}^{k}(x_i - \overline{x_в})^2}{n} \cdot n_i \text{ (для упорядкованої вибірки).} \qquad (12.7)$$

$$D_в^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x_в})^2}{n} \text{(для невпорядкованої вибірки).} \qquad (12.8)$$

Дисперсію $D_в$, як правило, обчислюють за формулою

$$D_в = \frac{\sum_{i=1}^{k} x_i^2 n_i}{n} - (\overline{x_в})^2 . \qquad (12.9)$$

7 Середнє квадратичне відхилення $D_в$ вибірки.

У якості міри розходження (варіації) бажано мати характеристику, яка вимірюється в тих же одиницях, що і значення ознаки. Такою характеристикою є вибіркове середнє квадратичне відхилення $\sigma_в$, яке обчислюється за формулою

$$\sigma_в = \sqrt{D_в} = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \overline{x_в})^2 \cdot n_i}{n}} . \qquad (12.10)$$

8 Коефіцієнт варіації $\overline{v}$.

Для порівняння оцінок варіацій статистичних рядів з різними значеннями $\overline{x_в} \neq 0$ вводиться поняття коефіцієнта варіації, який дорівнює процентному відношення вибіркового середньоквадратичного відхилу $\sigma_в$ до середньої арифметичної $\overline{x_в}$, тобто

$$\overline{v} = \frac{\sigma_в}{\overline{x_в}} \cdot 100\% \quad (\overline{x_в} \neq 0). \qquad (12.11)$$

Зауваження. Для визначення $\overline{x_в}, D_в, \sigma_в$ для інтервального статистичного розподілу вибірки треба перейти до дискретного статистичного розподілу, варіантами якого є середина часткових інтервалів:

$x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$, який має такий вигляд (табл. 12.6):

Таблиця 12.6

$x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$	x_1^*	x_2^*	x_3^*	…	x_k^*
n_i	n_1	n_2	n_3	…	n_k

Приклад. За даними інтервального статистичного розподілу вибірки, в якому наведено розподіл маси новонароджених x_i (табл. 12.7), обчислити $\overline{x_в}$, $D_в$, $\sigma_в$.

Таблиця 12.7

$x_i - x_{i+1}$	1–1,2	1,2–1,4	1,4–1,6	1,6–1,8	1,8–2,0	2,0–2,2	2,2–2,4	2,4–2,6	2,6–2,8	2,8-3,0	3,0-3,2
n_i	5	12	18	22	36	24	19	15	11	9	2

Розв'язання. Побудуємо дискретний статистичний розподіл (табл. 12.8) за заданим інтервальним з урахуванням того, що $h = 0{,}2$.

Таблиця 12.8

$x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$	1,1	1,3	1,5	1,7	1,9	2,1	2,3	2,5	2,7	2,9	3,1
n_i	5	12	18	22	36	24	19	15	11	9	2

Підрахуємо число варіант $n = 5+12+18+22+36+24+19+15+11+9+2=173$

За формулами (25.4), (25.12) і (25.13) обчислимо $\overline{x_в}$, $D_в$, $\sigma_в$.

$$\overline{x_в} = \frac{\sum_{i=1}^{k} x_i^* n_i}{n} = \frac{1{,}1\cdot 5 + 1{,}3\cdot 12 + 1{,}5\cdot 18 + 1{,}7\cdot 22 + 1{,}9\cdot 36 + 2{,}1\cdot 24 + 2{,}3\cdot 19 + 2{,}5\cdot 15 + 2{,}7\cdot 11 + 2{,}8\cdot 9 + 3{,}1\cdot 2}{173} =$$

$$= \frac{347{,}5}{173} = 2{,}008671\text{кг}.$$

Отже, $\overline{x_в} \approx 2{,}009$ кг.

$$D_в = \frac{\sum_{i=1}^{k} x_i^2 n_i}{n} - (\overline{x_в})^2 = \frac{(1{,}1)^2 \cdot 5 + (1{,}5)^2 \cdot 1{,}8 + (1{,}7)^2 \cdot 22 + (1{,}9)^2 \cdot 36}{173} +$$

$$+\frac{(2,1)^2\cdot 24+(2,3)^3\cdot 19+(2,5)^2\cdot 15+(2,7)^2\cdot 11+(2,9)^2\cdot 9+(3,1)^2\cdot 2}{173}-(2,009)^2\approx 0,217.$$

$$D_в\approx 0{,}217,\ \sigma_в=\sqrt{0{,}217}\approx 0{,}466.$$

Питання для самоперевірки

1 Що таке полігон частот та відносних частот і для чого він використовується?

2 Як будується полігон частот та відносних частот для інтервального варіаційного ряду?

3 Дати означення гістограми частот та відносних частот.

4 Дати означення кумулятивної кривої (помуляти) для дискретного і інтервального розподілу ознаки *X*.

5 Записати формули для обчислення $\overline{x_в}$, $D_в$, $\sigma_в$ для дискретного статистичного розподілу вибірки.

6 Записати формули для обчислення $\overline{x_в}$, $D_в$, $\sigma_в$ для інтервального статистичного розподілу вибірки.

7 Як визначається *Me** і *Mo** для інтервального статистичного розподілу?

8 Що називається розмахом, коефіцієнтом варіації?

ЛЕКЦІЯ 13

13.1 Статистичні оцінки параметрів розподілу, їх властивості

13.1 Статистичні оцінки параметрів розподілу, їх властивості

Сформулюємо задачу оцінки параметрів генеральної сукупності у загальному вигляді. Нехай розподіл ознаки *X* генеральної сукупності задається функцією ймовірності $\varphi(x_i,\theta)=P(X=x_i)$ (для дискретної випадкової величини) або щільністю ймовірності $\varphi(x,\theta)$ (для неперервної випадкової величини), яка містить невідомий параметр θ.

Для обчислення параметра θ неможливо дослідити усі елементи генеральної сукупності. Тому параметр θ оцінюють за допомогою вибірки, яка складається із значень $x_1,x_2,...,x_n$. Ці значення можна розглядати як часткові значення (реалізації) *n* незалежних випадкових величин $X_1,X_2,...,X_n$, кожна з яких має той же самий закон розподілу, що і випадкова величина *X*.

Означення. Оцінкою θ^* параметра θ називають будь-яку функцію результатів спостережень над випадковою величиною *X* (інакше – статистикою), за допомогою якої судять про значення параметра θ:

$$\theta^*=\theta_n^*(X_1,X_2,...,X_n).$$

Оскільки $X_1,X_2,...,X_n$ – випадкові величини, то й оцінка θ^* також є випадковою величиною, яка залежить від закону розподілу випадкової величини *X* і числа *n*.

Параметри генеральної сукупності $M(X)=\bar{X}_{\Gamma}, D_{\Gamma}, \sigma_{\Gamma}$, Mo, Me, r_{xy} є величинами сталими, але їх числове значення невідоме. Ці параметри оцінюються параметрами вибірки: $\bar{x}_{\text{В}}, D_{\text{В}}, \sigma_{\text{В}}, \text{Mo}^*$, $\text{Me}^*, r_{\text{В}}$, які дістають у процесі обробці вибірки. Вони є величинами непередбачуваними, тобто випадковими.

Точкові статистичні оцінки параметрів генеральної сукупності

Статистична оцінка θ^* параметра θ, яка визначається одним числом (точкою), називається *точковою*.

Завжди існує множина функцій від результатів спостережень $X_1, X_2, ..., X_n$, які можна розглядати у якості оцінки параметра θ. Тому виникає питання: якими властивостями має володіти оцінка, щоб вважати її якісною у деякому сенсі? Серед найбільш важливих властивостей розглядають наступні: незміщеність, ґрунтовність, ефективність.

Означення. Точкова статистична оцінка називається незміщеною, коли математичне сподівання цієї оцінки точно дорівнює оцінювальному параметру θ, а саме:

$$M\left(\theta^*\right)=\theta. \tag{13.1}$$

Якщо

$$M\left(\theta^*\right)\neq\theta, \tag{13.2}$$

точкова статистична оцінка θ^* називається зміщеною відносно параметра генеральної сукупності θ.

Різниця

$$\theta^*-\theta=\delta \tag{13.3}$$

називається *зміщенням статистичної оцінки* θ^*.

Оцінювальний параметр може мати кілька точкових незміщених статистичних оцінок. Тому порівнюють їх дисперсії як міру розсіювання і виявляють ту, яка має меншу дисперсію.

Означення. Точкова статистична оцінка називається ефективною, коли при заданому обсязі вибірки вона має мінімальну дисперсію.

Означення. Точкова статистична оцінка називається ґрунтовною, якщо у разі необмеженого збільшення обсягу вибірки θ^* наближається до оцінювального параметра θ, а саме:

$$\lim_{n\to\infty} P\left(\left|\theta^*-\theta\right|<\delta\right)=1. \tag{13.4}$$

Методи визначення точкових статистичних оцінок

Існують три методи визначення точкових статистичних оцінок для параметрів генеральної сукупності.

Метод аналогій. Цей метод базується на тому, що для параметрів генеральної сукупності вибирають такі самі параметри вибірки, тобто для оцінки $\bar{X}_{\Gamma}=M(X)$, D_{Γ} вибирають аналогічні статистики — $\bar{x}_{\text{В}}$, $D_{\text{В}}$.

Метод найменших квадратів. Згідно з цим методом статистичні оцінки визначаються з умови мінімізації суми квадратів відхилень варіант вибірки від статистичної оцінки θ^*.

Отож, використовуючи метод найменших квадратів, можна, наприклад, визначити статистичну оцінку для $\overline{X}_{\Gamma} = M(X)$. Для цього скористаємося функцією $u = \sum_{i=0}^{n} (x_i - \theta^*)^2 n_i$. Використовуючи необхідну умову існування екстремуму, дістанемо

$$\frac{\partial u}{\partial \theta^*} = -2\sum_{i=0}^{n} \left(x_i - \theta^* \right) n_i = 0 \ ,$$

звідки

$$\sum_{i=1}^{n} x_i n_i - \sum_{i=1}^{n} n_i \theta^* = 0 \, ,$$

$$\theta^* = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \overline{x}_{\text{В}}.$$

Звідси для $\theta = \overline{X}_{\Gamma}$ точковою статистичною оцінкою буде $\theta^* = \overline{x}_{\text{В}}$ — вибіркова середня. Для генеральної середньої $\overline{X}_{\Gamma} = M(X)$ точковою статистичною оцінкою є $\overline{x}_{\text{В}}$, для генеральної дисперсії D_{Γ} — вибіркова дисперсія $D_{\text{В}}$.

Можна показати, що точковою незміщеною статистичною оцінкою для генеральної середньої $\overline{X}_{\Gamma} = M(X)$ є вибіркова середня $\overline{x}_{\text{В}}$. Дійсно,

$$M(\overline{x}_{\text{В}}) = M\left(\frac{\sum_{i=1}^{n} x_i}{n} \right) = \frac{\sum M(x_i)}{n} = \left| \begin{array}{l} \text{Ураховуючите, що} \\ M(x_i) = \overline{X}_{\Gamma} = a \end{array} \right| = \frac{\sum_{i=1}^{n} a}{n} = \frac{na}{n} = a.$$

Отже, $M(\overline{x}_{\text{В}}) = \overline{X}_{\Gamma}$.

Вибіркова дисперсія $D_{\text{В}}$ є точковою зміщеною статистичною оцінкою для генеральної дисперсії D_{Γ}, де $\frac{n-1}{n}$ — коефіцієнт зміщення, який зменшується зі збільшенням обсягу вибірки n.

Якщо розглянути величину $\frac{n}{n-1} D_{\text{В}}$, то її математичне сподівання

$$M\left(\frac{n}{n-1} D_{\text{В}} \right) = \frac{n}{n-1} M(D_{\text{В}}) = \frac{n}{n-1} \cdot \frac{n-1}{n} D_{\Gamma} = D_{\Gamma}.$$

Величина $\frac{n}{n-1} D_{\text{В}}$ є точковою незміщеною статистичною оцінкою для генеральної дисперсії D_{Γ}, називається виправленою дисперсією і позначається через S^2:

$$S^2 = \frac{n}{n-1} D_{\text{В}}$$

або

$$S^2 = \frac{n}{n-1} \cdot \frac{\sum_{i=1}^{n} (x_i - \bar{x}_В)^2}{n} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_В)^2}{n-1}. \qquad (13.5)$$

Величину

$$S = \sqrt{\frac{n}{n-1} D_В} \qquad (13.6)$$

називають виправленим середнім квадратичним відхиленням. Виправлене середнє квадратичне відхилення є зміщеною точковою статистичною оцінкою для $\sigma_Г$,

Приклад. 200 однотипних деталей були піддані шліфуванню. Результати вимірювання наведені як дискретний статистичний розподіл, поданий у табличній формі:

x_i, мм	3,7	3,8	3,9	4,0	4,1	4,2	4,3	4,4
n_i	1	22	40	79	27	26	4	1

Знайти точкові незміщені статистичні оцінки для $\bar{X}_Г = M(x)$, $D_Г$.

Розв'язання. Оскільки точковою незміщеною оцінкою для $\bar{X}_Г$ є $\bar{x}_В$, то обчислимо

$$\bar{x}_В = \frac{\sum x_i n_i}{n} =$$

$$= \frac{3{,}7 \cdot 1 + 3{,}8 \cdot 22 + 3{,}9 \cdot 40 + 4{,}0 \cdot 79 + 4{,}1 \cdot 27 + 4{,}2 \cdot 26 + 4{,}3 \cdot 4 + 4{,}4 \cdot 1}{200} =$$

$$= \frac{3{,}7 + 83{,}6 + 156 + 316 + 110{,}7 + 109{,}2 + 17{,}2 + 4{,}4}{200} = \frac{808{,}8}{200} = 4{,}004 \text{ мм.}$$

Для визначення точкової незміщеної статистичної оцінки для $D_Г$ обчислимо $D_В$:

$$\frac{\sum x_i^2 n_i}{n} = \frac{(3{,}7)^2 \cdot 1 + (3{,}8)^2 \cdot 22 + (3{,}9)^2 \cdot 40 + (4{,}0)^2 \cdot 79 +}{200} =$$

$$\frac{+(4{,}1)^2 \cdot 27 + (4{,}2)^2 \cdot 26 + (4{,}3)^2 \cdot 4 + (4{,}4)^2 \cdot 1}{200} =$$

$$= \frac{13{,}69 + 317{,}68 + 608{,}4 + 1264 + 453{,}87 + 458{,}64 + 73{,}96 + 19{,}36}{200} =$$

$$= \frac{3209{,}6}{200} = 16{,}048.$$

$$D_В = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_В)^2 = 16{,}048 - (4{,}004)^2 =$$

$$= 16{,}048 - 16{,}032016 = 0{,}015984.$$

Тоді точкова незміщена статистична оцінка для $D_Г$ становитиме:

$$S^2 = \frac{n}{n-1} D_В = \frac{200}{200-1} \cdot 0{,}015984 = \frac{200}{199} \cdot 0{,}015984 = 0{,}01606 \text{ мм}^2.$$

Приклад. Граничне навантаження на сталевий болт x_i, що вимірювалось в лабораторних умовах, задано як інтервальний статистичний розподіл:

x_i, км/мм^2	4,5—5,5	5,5—6,5	6,5—7,5	7,5—8,5	8,5—9,5	9,5—10,5	10,5—11,5	11,5—12,5	12,5—13,5	13,5—14,5
n_i	40	32	28	24	20	18	16	12	8	4

Визначити точкові незміщені статистичні оцінки для $\bar{X}_{\text{Г}} = M(x),\ D_{\text{Г}}$.

Розв'язання. Для визначення точкових незміщених статистичних оцінок $\bar{x}_{\text{В}}$, S^2 перейдемо від інтервального статистичного розподілу до дискретного, який набирає такого вигляду:

$x_i^* = x_{i-1} + \frac{h}{2}$	5	6	7	8	9	10	11	12	13	14
n_i	40	32	28	24	20	18	16	12	8	4

Обчислимо $\bar{x}_{\text{В}}$: $n = \sum n_i = 202$,

$$\bar{x}_{\text{В}} = \frac{\sum x_i^* n_i}{n} = \frac{5\cdot 40 + 6\cdot 32 + 7\cdot 28 + 8\cdot 24 + 9\cdot 20 + 10\cdot 18 +}{202}$$

$$\frac{+11\cdot 16 + 12\cdot 12 + 13\cdot 8 + 14\cdot 4}{202} = \frac{1620}{202} = 8{,}02 \text{ кг/мм}^2.$$

Отже, точкова незміщена статистична оцінка для $\bar{X}_{\text{Г}} = M(x)$, $\bar{x}_{\text{В}} = 8{,}02$ кг/мм^2.

Для визначення S^2 обчислимо $D_{\text{В}}$:

$$\frac{\sum (x_i^*)^2 n_i}{n} = \frac{(5)^2\cdot 40 + (6)^2\cdot 32 + (7)^2\cdot 28 + (8)^2\cdot 24 + (9)^2\cdot 20 +}{202}$$

$$\frac{+(10)^2\cdot 18 + (11)^2\cdot 16 + (12)^2\cdot 12 + (13)^2\cdot 8 + (14)^2\cdot 4}{202} = \frac{14280}{202} \approx 70{,}69.$$

$$D_{\text{В}} = \frac{\sum (x_i^*)^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 70{,}69 - (8{,}02)^2 = 70{,}69 - 64{,}32 \approx 6{,}37 \text{ кг/мм}^2.$$

$$S^2 = \frac{n}{n-1} D_{\text{В}} = \frac{202}{202-1}\cdot 6{,}37 = \frac{202}{201}\cdot 6{,}37 \approx 6{,}4.$$

Звідси точкова незміщена статистична оцінка для $D_{\text{Г}}$ є $S^2 = 6{,}4$ кг/мм^2.

Питання для самоперевірки

1 Що називається точковою статистичною оцінкою?
2 Що таке незміщена точкова статистична оцінка?
3 Що таке зміщена точкова статистична оцінка?
4 Що називають ефективною точковою статистичною оцінкою?
5 Що називають ґрунтовною точковою статистичною оцінкою?
6 У чому сутність методу аналогій?
7 У чому сутність методу найменших квадратів?
8 У чому сутність методу максимальної правдоподібності?

ЛЕКЦІЯ 14

14.1 Інтервальні оцінки. Надійна ймовірність та надійний інтервал

14.2 Надійні інтервали для оцінювання математичного сподівання та середнього квадратичного відхилення при нормальному розподілі

14.1 Інтервальні оцінки. Надійна ймовірність та надійний інтервал

Точкові статистичні оцінки θ^* є випадковими величинами, а тому наближена заміна θ на θ^* часто призводить до істотних похибок, особливо коли обсяг вибірки малий. У цьому разі застосовують інтервальні статистичні оцінки.

Статистична оцінка, що визначається двома числами, кінцями інтервалів, називається інтервальною.

Різниця між статистичною оцінкою θ^* та її оцінювальним параметром θ, взята за абсолютним значенням, називається точністю оцінки, а саме:

$$\left|\theta^* - \theta\right| < \delta, \tag{14.1}$$

де δ є точністю оцінки.

Оскільки θ^* є випадковою величиною, то і δ буде випадковою, тому нерівність (14.1) справджуватиметься з певною ймовірністю.

Імовірність, з якою береться нерівність (26.7), тобто

$$P\left(\left|\theta^* - \theta\right| < \delta\right) = \gamma, \tag{14.2}$$

називають надійністю.

Рівність (14.1) можна записати таким чином:

$$P\left(\theta^* - \delta < \theta < \theta^2 + \delta\right) = \gamma. \tag{14.3}$$

Інтервал $\left[\theta^* - \delta;\ \theta^* + \delta\right]$, що покриває оцінюваний параметр θ генеральної сукупності з заданою надійністю γ, називають надійним.

14.2 Надійні інтервали для оцінювання математичного сподівання та середнього квадратичного відхилення при нормальному розподілі

14.2.1 Побудова надійного інтервалу для $\bar{x}_Г$ при відомому значенні $\sigma_Г$ із заданою надійністю γ

Нехай ознака *X* генеральної сукупності має нормальний закон розподілу. Побудуємо довірчий інтервал для $\bar{X}_Г$, знаючи числове значення середнього квадратичного відхилення генеральної сукупності $\sigma_Г$, із заданою надійністю γ. Оскільки $\bar{x}_В$ як точкова незміщена статистична оцінка для $\bar{X}_Г = M(x)$ має нормальний закон розподілу з числовими характеристиками $M(\bar{x}_В) = \bar{X}_Г = a$, $\sigma(\bar{x}_В) = \frac{\sigma_Г}{\sqrt{n}}$, то, скориставшись (14.2), дістанемо

$$P\left(\left|\bar{x}_В - a\right| < \delta\right) = \gamma. \tag{14.4}$$

Випадкова величина $\bar{x}_В - a$ має нормальний закон розподілу з числовими характеристиками

$$M(\bar{x}_{\text{В}} - a) = M(\bar{x}_{\text{В}}) - a = a - a = 0;$$

$$D(\bar{x}_{\text{В}} - a) = D(\bar{x}_{\text{В}}) = \frac{D_{\text{Г}}}{n};$$

$$\sigma(\bar{x}_{\text{В}}) = \frac{\sigma_{\text{Г}}}{\sqrt{n}}.$$

Тому $\dfrac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}$ матиме нормований нормальний закон розподілу $N(0; 1)$.

Звідси рівність (26.9) можна записати, позначивши $\dfrac{\delta}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}} = x,$ таким чином:

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = \gamma \qquad (14.5)$$

або

$$P\left(\bar{x}_{\text{В}} - \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}} < a < \bar{x}_{\text{В}} + \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}}\right) = \gamma.$$

Згідно з формулою нормованого нормального закону

$$P(|X - a| < \delta) = 2\Phi(\delta)$$

для (14.5) вона набирає такого вигляду:

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = 2\Phi(x) = \gamma. \qquad (14.6)$$

Із рівності (14.6) знаходимо аргументи x, а саме:

$$2\Phi(x) = \gamma \to \Phi(x) = 0{,}5\gamma.$$

Аргумент x знаходимо за значенням функції Лапласа, яка дорівнює 0,5 γ за таблицею (додаток Б).

Отже, надійний інтервал для $\bar{x}_{\text{Г}}$ при відомому значенні $\sigma_{\text{Г}}$ має вигляд

$$\bar{x}_{\text{В}} - \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}} < a < \bar{x}_{\text{В}} + \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}}, \qquad (14.7)$$

що можна зобразити умовно на рис. 14.1.

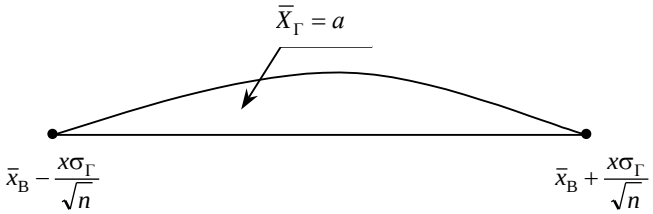


Рис. 14.1

Величина $\dfrac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}}$ називається точністю оцінки, або похибкою вибірки.

Приклад. Вимірявши 40 випадково відібраних після виготовлення деталей, знайшли вибіркову середню, що дорівнює 15 см. Із надійністю $\gamma = 0{,}99$

побудувати надійний інтервал для середньої величини всієї партії деталей, якщо генеральна дисперсія дорівнює $0{,}09\ \text{см}^2$.

Розв'язання. Для побудови довірчого інтервалу необхідно знати:

$$\bar{x}_{\text{В}},\ \sigma_{\text{Г}},\ n,\ x.$$

Із умови задачі маємо: $\bar{x}_{\text{В}} = 15$ см, $\sigma_{\text{Г}} = \sqrt{D_{\text{Г}}} = \sqrt{0{,}09\ \text{см}^2} = 0{,}3\ \text{см}$, $n = 40$, тобто $\sqrt{n} = \sqrt{40} = 6{,}32$. Величина x обчислюється з рівняння

$$\Phi(x) = 0{,}5\gamma = 0{,}5 \cdot 0{,}99 = 0{,}495.$$

$$\Phi(x) = 0{,}495,\quad x = 2{,}58\ [\text{за таблицею значень функції Лапласа}].$$

Знайдемо числові значення кінців довірчого інтервалу:

$$\bar{x}_{\text{В}} - \frac{\sigma_{\text{Г}} \cdot x}{\sqrt{n}} = 15 - \frac{0{,}3 \cdot 2{,}58}{6{,}32} = 15 - 0{,}12 = 14{,}88\ \text{см}.$$

$$\bar{x}_{\text{В}} + \frac{\sigma_{\text{Г}} \cdot x}{\sqrt{n}} = 15 + \frac{0{,}3 \cdot 2{,}58}{6{,}32} = 15 + 0{,}12 = 15{,}12\ \text{см}.$$

Таким чином, маємо:

$$14{,}88 < \bar{X}_{\text{Г}} < 15{,}12.$$

Отже, з надійністю 0,99 (99% гарантії) оцінюваний параметр $\bar{X}_{\text{Г}}$ перебуває усередині інтервалу [14,87; 15,13].

Приклад. Якого значення має набувати надійність оцінки γ, щоб за обсягу вибірки n = 100 похибка її не перевищувала 0,01 при $\sigma_{\text{Г}} = 5$.

Розв'язання. Позначимо похибку вибірки

$$\frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}} = \varepsilon,\ x = \frac{\varepsilon\sqrt{n}}{\sigma_{\text{Г}}} = \frac{0{,}01\sqrt{100}}{5} = \frac{0{,}01 \cdot 10}{5} = 0{,}02.$$

Далі маємо

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\frac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = 2\Phi(x) = 2\Phi(0{,}02) = 2 \cdot 0{,}008 = 0{,}016.$$

Як бачимо, надійність мала.

Приклад. Визначити обсяг вибірки n, за якого похибка $\varepsilon = 0{,}01$ гарантується з імовірністю 0,999, якщо $\sigma_{\text{Г}} = 5$.

Розв'язання. За умовою задачі $P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\frac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = \gamma = 0{,}999$. Оскільки $\frac{x\sigma_{\text{Г}}}{\sqrt{n}} = \varepsilon$, то дістанемо: $n = \frac{x^2\sigma_{\text{Г}}^2}{\varepsilon^2}$. Величину x знаходимо з рівності

$$\Phi(x) = 0{,}5\gamma = 0{,}5 \cdot 0{,}999 = 0{,}4995,\quad x \approx 3{,}4.$$

Тоді $n = \frac{(3{,}4)^2 \cdot 5^2}{(0{,}01)^2} = 2\ 890\ 000.$

14.2.2 Побудова надійного інтервалу для $\bar{x}_{\text{Г}}$ при невідомому значенні $\sigma_{\text{Г}}$ із заданою надійністю γ

Для малих вибірок, з якими стикаємося, досліджуючи різні ознаки в техніці чи економіці, для оцінювання $\bar{X}_{\text{Г}} = a$ при невідомому значенні $\sigma_{\text{Г}}$ неможливо скористатися нормальним законом розподілу. Тому для побудови довірчого інтервалу застосовується випадкова величина

$$t = \frac{\bar{x}_{\text{В}} - a}{\frac{S}{\sqrt{n}}}, \tag{14.8}$$

що має розподіл Стьюдента з $k = n - 1$ ступенями свободи.

Тоді (14.8) набирає такого вигляду:

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\frac{S}{\sqrt{n}}}\right| < t_\gamma\right) = P\left(\bar{x}_{\text{В}} - \frac{t_\gamma \cdot S}{\sqrt{n}} < a < \bar{x}_{\text{В}} + \frac{t_\gamma S}{\sqrt{n}}\right) = 2\int_0^{t_\gamma} f(t) = \gamma,$$

оскільки $f(t)$ для розподілу Стьюдента є функцією парною.

Обчисливши за даним статистичним розподілом $\bar{x}_{\text{В}}$, S і визначивши за таблицею розподілу Стьюдента значення t_γ, будуємо надійний інтервал

$$\bar{x}_{\text{В}} - \frac{t_\gamma \cdot S}{\sqrt{n}} < a < \bar{x}_{\text{В}} + \frac{t_\gamma S}{\sqrt{n}}. \tag{14.9}$$

Тут $t_\gamma(\gamma, k = n - 1)$ обчислюємо за заданою надійністю γ і числом ступенів свободи $k = n - 1$ за таблицею (додаток Г).

Приклад. Випадково вибрана партія з двадцяти приладів була випробувана щодо терміну безвідмовної роботи кожного з них t_i. Результати випробувань наведено у вигляді дискретного статистичного розподілу:

t_i	100	170	240	310	380
n_i	2	5	10	2	1

Із надійністю $\gamma = 0{,}99$ побудувати надійний інтервал для «*a*» (середнього часу безвідмовної роботи приладу).

Розв'язання. Для побудови довірчого інтервалу необхідно знайти середнє вибіркове і виправлене середнє квадратичне відхилення.

Обчислимо $\bar{x}_{\text{В}}$:

$$\bar{x}_{\text{В}} = \frac{\sum t_i n_i}{n} = \frac{100 \cdot 2 + 170 \cdot 5 + 240 \cdot 10 + 310 \cdot 2 + 380 \cdot 1}{20} = \frac{4450}{20} = 222{,}5.$$

Отже, дістали $\bar{x}_{\text{В}} = 222{,}5$ год.

Визначимо $D_{\text{В}}$:

$$\frac{\sum t_i^2 n_i}{n} = \frac{100^2 \cdot 2 + 170^2 \cdot 5 + 240^2 \cdot 10 + 310^2 \cdot 2 + 380^2 \cdot 1}{20} = \frac{1\ 077\ 100}{20} = 53\,855.$$

$$D_{\text{В}} = \frac{\sum t_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 53\,855 - (222{,}5)^2 = 53\,855 - 49\,506{,}25 = 4348{,}75.$$

Отже, $D_{\text{В}} = 4348{,}75$.

Виправлене середнє квадратичне відхилення дорівнюватиме:

$$S = \sqrt{\frac{n}{n-1} D_{\text{В}}} = \sqrt{\frac{20}{20-1} \cdot 4348{,}75} \approx 67{,}66 \text{ год.}$$

За таблицею значень $\int_0^t f(x)dt = \gamma = 0{,}99$ (додаток В) розподілу Стьюдента за заданою надійністю $\gamma = 0{,}99$ і числом ступенів свободи $k = n - 1$ = 20 – 1 = 19 знаходимо значення $t(\gamma = 0{,}99, k = 19) =$ 2,861.

Обчислимо кінці надійного інтервалу:

$$\bar{x}_{\text{В}} - \frac{t_\gamma S}{\sqrt{n}} = 222{,}5 - \frac{2{,}861 \cdot 67{,}66}{\sqrt{20}} = 222{,}5 - \frac{2{,}861 \cdot 67{,}66}{4{,}472} = 179{,}2 \text{ год.}$$

$$\bar{x}_{\text{В}} + \frac{t_\gamma S}{\sqrt{n}} = 222{,}5 + \frac{2{,}861 \cdot 67{,}89}{\sqrt{20}} = 222{,}5 + \frac{2{,}861 \cdot 67{,}66}{4{,}472} = 265{,}8 \text{ год.}$$

Отже, з надійністю $\gamma = 0{,}99$ можна стверджувати, що $\bar{X}_{\text{Г}} = a$ буде міститися в інтервалі

$$179{,}2 < a < 265{,}8.$$

Зауваження. При великих обсягах вибірки, а саме: $n > 30$, на підставі центральної граничної теореми теорії ймовірностей (теореми Ляпунова) розподіл Стьюдента наближається до нормального закону. У цьому разі t_γ знаходиться за таблицею значень функції Лапласа.

Приклад. У таблиці наведено відхилення діаметрів валиків, оброблених на верстаті, від номінального розміру:

h = 5 мк	0 — 5	5 — 10	10 — 15	15—20	20 — 25
n_i	15	75	100	50	10

Із надійністю $\gamma = 0{,}99$ побудувати надійний інтервал для $\bar{X}_{\text{Г}} = a$.

Розв'язання. Для побудови довірчого інтервалу необхідно знайти $\bar{x}_{\text{В}}$, S. Для цього від інтервального статистичного розподілу, наведеного в умові задачі, необхідно перейти до дискретного, а саме:

x_i^*	2,5	7,5	12,5	17,5	22,5
n_i	15	75	100	50	10

Обчислимо $\bar{x}_{\text{В}}$:

$$\bar{x}_{\text{В}} = \frac{\sum x_i^* n_i}{n} = |\text{Оскільки } n = \sum n_i = 250| =$$

$$= \frac{2{,}5 \cdot 15 + 7{,}5 \cdot 75 + 12{,}5 \cdot 100 + 17{,}5 \cdot 50 + 22{,}5 \cdot 10}{250} =$$

$$= \frac{37{,}5 + 562{,}5 + 1250 + 875 + 225}{250} = \frac{2950}{250} = 11{,}8.$$

Отже, $\bar{x}_{\text{В}} = 11{,}8$ мк.

Визначимо $D_{\text{В}}$:

$$\frac{\sum\left(x_i^*\right)^2 n_i}{n}=\frac{(2{,}5)^2\cdot 15+(7{,}5)^2\cdot 75+(12{,}5)^2\cdot 100+(17{,}5)^2\cdot 50+(22{,}5)^2\cdot 10}{250}=$$

$$=\frac{93{,}75+4218{,}75+15625+15312{,}5+5062{,}5}{250}=\frac{40312{,}5}{250}=161{,}25.$$

$$D_{\text{В}}=\frac{\sum\left(x_i^*\right)^2 n_i}{n}-\left(\bar{x}_{\text{В}}\right)^2=161{,}25-(11{,}8)^2=161{,}25-139{,}24=22{,}01.$$

Обчислимо виправлене середнє квадратичне відхилення S:

$$S=\sqrt{\frac{n}{n-1}D_{\text{В}}}=\sqrt{\frac{250}{250-1}\cdot 22{,}01}\approx 4{,}7 \text{ мк.}$$

Із огляду на великий (n = 250) обсяг вибірки можна вважати, що розподіл Стьюдента близький до нормального закону. Тоді за таблицею значення функції Лапласа

$$\Phi(t_\gamma)=0{,}495\rightarrow t_\gamma=2{,}58.$$

Обчислимо кінці інтервалів:

$$\bar{x}_{\text{В}}-\frac{t_\gamma S}{\sqrt{n}}=11{,}8-\frac{2{,}58\cdot 4{,}7}{\sqrt{250}}=11{,}8-\frac{2{,}58\cdot 4{,}7}{15{,}8}=11{,}8-0{,}77=11{,}03 \text{ мк.}$$

$$\bar{x}_{\text{В}}+\frac{t_\gamma S}{\sqrt{n}}=11{,}8+\frac{2{,}58\cdot 4{,}7}{\sqrt{250}}=11{,}8+\frac{2{,}58\cdot 4{,}7}{15{,}8}=11{,}8+0{,}77=12{,}57 \text{ мк.}$$

Отже, надійний інтервал для середнього значення відхилень буде таким:

$$11{,}03<a<12{,}57.$$

Звідси з надійністю $\gamma=0{,}99$ (99%) можна стверджувати, що

$$a\in[11{,}03 \text{ мк; } 12{,}57 \text{ мк}].$$

14.2.3 Побудова надійних інтервалів із заданою надійністю γ для $D_{\text{Г}}$, $\sigma_{\text{Г}}$

У разі, коли ознака X має нормальний закон розподілу, для побудови надійного інтервалу із заданою надійністю γ для $D_{\text{г}}$, $\sigma_{\text{г}}$ застосовуємо випадкову величину

$$\chi^2=\frac{n-1}{\sigma_{\Gamma}^2}S^2, \tag{14.10}$$

що має розподіл χ^2 із $k=n-1$ ступенями свободи.

Оскільки випадкові події

$$A\left(\chi_1^2<\chi^2<\chi_2^2\right) \text{ і } B\left(\frac{1}{\chi_2^2}<\frac{1}{\chi^2}<\frac{1}{\chi_1^2}\right)$$

є рівноймовірними, тобто їх імовірності рівні $(P(A)=P(B))$, маємо:

$$P\left(\chi_1^2<\chi^2<\chi_2^2\right)=P\left(\frac{1}{\chi_2^2}<\frac{1}{\chi^2}<\frac{1}{\chi_1^2}\right). \tag{14.11}$$

Підставляючи в (14.11) $\chi^2=\frac{n-1}{\sigma^2}S^2$, дістанемо

$$P\left(\frac{1}{\chi_2^2}<\frac{1}{\chi^2}<\frac{1}{\chi_1^2}\right)=P\left(\frac{1}{\chi_2^2}<\frac{1}{\frac{n-1}{\sigma_{\Gamma}^2}S^2}<\frac{1}{\chi_1^2}\right)=$$

$$=P\left(\frac{1}{\chi_2^2}<\frac{\sigma_{\Gamma}^2}{(n-1)S^2}<\frac{1}{\chi_1^2}\right)=P\left(\frac{(n-1)S^2}{\chi_2^2}<\sigma_{\Gamma}^2<\frac{(n-1)S^2}{\chi_1^2}\right)=\gamma.$$

Отже, довірчий інтервал для $\sigma_{\Gamma}^2 = D_{\Gamma}$ матиме вигляд

$$\frac{(n-1)S^2}{\chi_2^2} < D_{\Gamma} < \frac{(n-1)S^2}{\chi_1^2}. \qquad (14.12)$$

Тоді надійний інтервал для σ_{Γ} випливає із (14.12) і буде таким:

$$\frac{S\sqrt{n-1}}{\chi_2} < \sigma_{\Gamma} < \frac{S\sqrt{n-1}}{\chi_1}. \qquad (14.13)$$

Значення χ_1^2, χ_2^2 знаходимо за таблицею значень χ_1^2, χ_2^2 згідно з рівностями:

$$P(\chi^2 > \chi_1^2) = 1 - \frac{\alpha}{2}; \qquad (14.14)$$

$$P(\chi^2 > \chi_2^2) = \frac{\alpha}{2}, \qquad (14.15)$$

де $\alpha = 1 - \gamma$.

Питання для самоперевірки

1 Що є точковою незміщеною статистичною оцінкою для $\overline{X}_{\Gamma}$?

2 Що означає точкова незміщена статистична оцінка для D_{Γ}?

3 Що називається виправленою дисперсією, виправленим середнім квадратичним відхиленням?

4 Як обчислюється $M(\overline{x}_{\text{В}})$, $M(D_{\text{В}})$, $M(S_{\text{В}}^2)$?

5 Визначення інтервальної статистичної оцінки для параметрів генеральної сукупності.

6 Що називають точністю і надійністю оцінки?

7 Що називають надійним інтервалом?

8 Як побудувати надійний інтервал із заданою надійністю γ при відомому значенні σ_{Γ}?

8 Як побудувати довірчий надійний інтервал для $\overline{X}_{\Gamma}$ із заданою надійністю γ при невідомому значенні σ_{Γ}?

ЛЕКЦІЯ 15

15.1 Елементи теорії кореляції. Поняття регресії. Коефіцієнт кореляції

15.1 Елементи теорії кореляції. Поняття регресії. Коефіцієнт кореляції

Можливі два варіанти взаємозв’язків між двома змінними X і Y. У першому випадку обидві змінні розглядаються як рівноцінні в тому сенсі, що вони не поділяються на залежну і незалежну змінні. Головним у цьому разі є питання про присутність і силу взаємозв’зку між цими змінними. Наприклад, між ціною товару й об’ємом попиту на нього, між урожаєм картоплі й урожаєм зерна, між інтенсивністю руху транспорта і числом аварій. Під час дослідження сили лінійної залежності між такими змінними звертаються до кореляційного аналізу, основною мірою якого є коефіцієнт кореляції.

У другому випадку взаємозв’язків виділяють одну з величин як незалежну, а іншу як залежну. У цьому разі зміна першої з них може бути

причиною для зміни другої. Наприклад, зростання доходу веде до збільшення споживання; зростання ціни – до зниження попиту; зниження відсоткової ставки збільшує інвестиції. Але така залежність не є однозначною в тому сенсі, що кожному конкретному значенню незалежній змінній може відповідати не одна, а множина значень з деякої області. Іншими словами, кожному конкретному значенню незалежної змінної відповідає деякий ймовірносний розподіл залежної змінної (яку розглядають як випадкову величину). Залежність такого типу виражається співвідношенням $M(Y|x) = f(x)$ і називається функцією регресії Y на X. $M(Y|x)$ – умовне математичне сподівання. Інакше кажучи, регресія – це функціональна залежність між незалежною змінною й умовним математичним сподіванням залежної змінної, яка будується з метою передбачення цього середнього значення при фіксованому значенні першої.

Для відображення того факту, що реальне значення залежної змінної не завжди співпадає з її умовним математичним сподіванням і можуть бути різними при одному й тому ж значенні незалежної змінної, фактична залежність повинна бути доповнена деяким доданком ε, який по суті є випадковою величиною й вказує на стохастичну сутність залежності. Отже, зв'язок між залежною і незалежною змінними виражається співвідношенням $Y = M(Y|x) + \varepsilon,$ яке називається регресійною моделлю.

Найбільш істотними причинами включення випадковою фактора в регресійну модель є:

1 Невключення в модель усіх пояснювальних змінних, оскільки будь-яка регресійна модель є спрощенням реальної ситуації.

2 Неправильний вибір функціональної форми моделі.

3 Помилки вимірювань.

4 Обмеженність статистичних даних.

5 Непередбаченість людського фактора.

Розв'язання задачі побудови якісного рівняння регресії, яке відповідає емпіричним даним можно розбити на три кроки:

1) вибір формули рівняння регресії;
2) оцінка параметрів вибраного рівняння;
3) аналіз якості рівняння і перевірка адекватності рівняння емпіричним даним.

Вибір формули зв'язку змінних називається специфікацією рівняння регресії. У випадку парної регресії вибір формули зазвичай здійснюється за графічним зображенням реальних статистичних даних у вигляді точок в декартовій системі координат, яке називається кореляційним полем (рис. 15.1).

На рисунку 27.1*а* взаємозв'язок між X і Y близький до лінійного. У даному випадку в якості залежності між X і Y можна взяти лінійну функцію $Y = b_0 + b_1 X$.

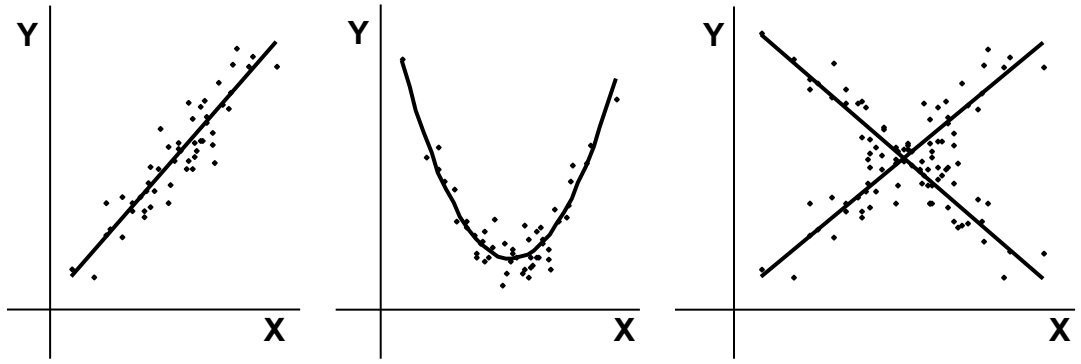
На графіку 27.1*б* реальний зв'язок між X і Y скоріше всього описується квадратичною функцією $Y = aX^2 + bX + c.$

На графіку 27.1*в* явний зв'язок між X і Y відсутній.

Якщо функція регресії лінійна, то кажуть про лінійну регресію. Отже, лінійна регресія являє собою лінійну функцію між умовним математичним сподіванням $M(Y|X = x_i)$ залежної змінної Y і однієї пояснювальної змінної X (x_i – значення незалежної змінної в i – му спостереженні, $i = 1,2,...,n$).

$$M(Y|X = x_i) = b_0 + b_1 x_i.$$

Зауважимо, що принциповою в даному випадку є лінійність за параметрами b_0 і b_1.



а *б* *в*

Рисунок 15.1

Для відображення того факту, що кожне індивідуальне значення y_i відхиляється від відповідного умовного математичного сподівання, необхідно ввести у співвідношення $M(Y|X = x_i) = b_0 + b_1 x_i$ випадковий доданок ε_i:

$$y_i = M(Y|X = x_i) + \varepsilon_i = b_0 + b_1 x_i + \varepsilon_i.$$

Таке співвідношення називається теоретичною лінійною регресійною моделлю; b_0 і b_1 − теоретичними параметрами (коефіцієнтами); ε_i − випадковим відхиленням.

У загальному вигляді теоретичну лінійну регремійну модель подають у вигляді

$$Y = b_0 + b_1 X + \varepsilon.$$

Для визначення значень теоретичних коефіцієнтів регресії необхідно знати всі значення змінних X і Y генеральної сукупності, що практично неможливо.

Отже, задача лінійного регресійного аналізу полягає в тому, щоб за статистичними даними (x_i, y_i), $i = 1, 2, ..., n$ для змінних X і Y:

1) одержати найкращі оцінки невідомих параметрів b_0 і b_1;

2) перевірити статистичні гіпотези про параметри моделі;

3) перевірити, чи достатньо гарно модель погоджується зі статистичними даними.

Таким чином, за вибіркою обмеженого об'єму можна побудувати так зване емпіричне рівняння регресії:

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i,$$

де $\hat{y}_i$ – оцінка (наближене значення) умовного математичного сподівання $M(Y|X = x_i)$; $\hat{b}_0$ і $\hat{b}_1$ – оцінки невідомих параметрів b_0 і b_1, які називають емпіричними коефіцієнтами регресії. Отже, в конкретному випадку

$$y_i = \hat{b}_0 + \hat{b}_1 x_i + e_i,$$

де відхилення e_i – оцінка теоретичного випадкового відхилення ε_i.

Найбільш поширеним і теоретично обгрунтованим методом знаходження коефіцієнтів є метод найменших квадратів (МНК), згідно з яким коефіцієнти $\hat{b}_0$ і $\hat{b}_1$ знаходяться за умови

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{b}_0 - \hat{b}_1 x_i)^2 \to \min.$$

Згідно з МНК оцінки параметрів $\hat{b}_0$ і $\hat{b}_1$ визначаються так:

$$\hat{b}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}, \ \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}.$$

Тут $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{xy} = \frac{1}{n}$.

Коефіцієнт $\hat{b}_1$ можна також обчислити й за формулою $\hat{b}_1 = r_{xy} \cdot \frac{S_y}{S_x}$,

де $r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2} \cdot \sqrt{\overline{y^2} - \bar{y}^2}}$ – вибірковий коефіцієнт кореляції;

$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, $S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$ – стандартні відхилення.

Властивості оцінок коефіцієнтів регресії, а отже, і якості побудованої регресії істотно залежать від властивостей випадкової складової ε. Доведено, що для одержання за МНК найкращих результатів (оцінки обгрунтовані, незміщені й ефективні) необхідно, щоб виконувався ряд передбачень відносно випадкового відхилення:

1) математичне сподівання випадкового відхилення ε_i дорівнює нулю, а саме $M(\varepsilon_i) = 0, i = 1,2,...n$;

2) дисперсія випадкових відхилень ε_i стала $D(\varepsilon_i) = \sigma^2, i = 1,2,...n$;

3) випадкові відхилення ε_i і ε_j є незалежними одне від одного, $i \neq j, \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$;

4) випадкове відхилення повинно бути незалежно від пояснюючих змінних $\mathrm{cov}(\varepsilon_i, x_i) = 0$.

Припускають також, що випадкова величина ε розподілена за нормальним законом.

Питання для самоперевірки

1 Яка залежність називається статистичною?
2 Що називається умовною середньою?
3 Яка залежність називається кореляційною?
4 Рівняння регресії, функція регресії, лінія регресії.

ЛЕКЦІЯ 16

16.1 Знаходження параметрів вибіркового рівняння прямої лінії регресії

16.1 Знаходження параметрів вибіркового рівняння прямої лінії регресії

Емпіричне рівняння регресії визначається на основі скінченного числа статистичних даних. Отже, коефіцієнти емпіричного рівняння регресії є випадковими величинами. Під час проведення статистичного аналізу виникає необхідність порівняння емпіричних коефіцієнтів регресії $\hat{b}_0$ і $\hat{b}_1$ з деякими теоретично очікуваними значеннями b_0 і b_1 цих коефіцієнтів. Даний аналіз проводиться за схемою статистичної перевірки гіпотез.

Для перевірки гіпотези

$$H_0 : \hat{b}_1 = b_1,$$
$$H_1 : \hat{b}_1 \neq b_1$$

використовують статистику $t = \dfrac{\hat{b}_1 - b_1}{S_{\hat{b}_1}}$, яка у разі справедливості H_0 має розподіл Стьюдента з числом степеней свободи $k = n - 2$, $n -$об'єм вибірки. Отже, $H_0 : \hat{b}_1 = b_1$ відхиляється на основі даного критерію, якщо

$$\left| t_{\hat{o}\,\grave{a}\hat{e}\grave{o}} \right| = \left| \frac{\hat{b}_1 - b_1}{S_{\hat{b}_1}} \right| \geq t_{\alpha/2, n-2},$$

де $\alpha -$рівень значущості. У разі невиконання цієї нерівності немає підстав для відхилення H_0.

$S_{\hat{b}_1} -$стандартна помилка оцінки $\hat{b}_1$, яка обчислюється так:

$$S_{\hat{b}_1} = \sqrt{\frac{S^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}},$$

де

$$S^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}.$$

$t_{\alpha/2, n-2}$ – критичне (табличне) значення статистики Стьюдента, яке відповідає рівню значущості α і числу степеней свободи $k = n - 2$.

Найбільш важливою на початковому етапі статистичного аналізу побудованої моделі є задача виявлення присутності лінійного зв'язку між Y і X. Ця проблема може бути вирішена за такою схемою:

$$H_0 : \hat{b}_1 = 0,$$

$$H_1 : \hat{b}_1 \neq 0.$$

Таку гіпотезу називають гіпотезою про статистичну значущість коефіцієнтів регресії. При цьому, якщо H_0 сприймається, то є підстава стверджувати, що величина Y не залежить від X. У цьому випадку кажуть, що коефіцієнт $\hat{b}_1$ статистично незначущий (він близький до нуля). У разі відхилення H_0 припускають, що коефіцієнт $\hat{b}_1$ значущий, тобто є лінійний зв'язок між Y і X.

За аналогічною схемою на основі t – статистики превіряють гіпотезу про статистичну значущість коефіцієнта $\hat{b}_0$, при цьому стандартна помилка $S_{\hat{b}_0}$ обчислюється так:

$$S_{\hat{b}_0} = \sqrt{\frac{S^2 \sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}} = \sqrt{\overline{x^2} \cdot S_{\hat{b}_1}^2}.$$

Слід зазначити, що для парної регресії більш важливим є аналіз статистичної значущості коефіцієнта $\hat{b}_1$, оскільки саме в ньому завуальований вплив змінної X на результат Y.

Надійні інтервали, які з надійністю $\gamma = 1 - \alpha$ покривають параметри b_0 і b_1 мають такий вигляд:

$$(\hat{b}_0 - t_{\alpha/2, n-2} \cdot S_{\hat{b}_0}; \hat{b}_0 + t_{\alpha/2, n-2} \cdot S_{\hat{b}_0}),$$

$$(\hat{b}_1 - t_{\alpha/2, n-2} \cdot S_{\hat{b}_1}; \hat{b}_1 + t_{\alpha/2, n-2} \cdot S_{\hat{b}_1}).$$

Однією з центральних задач регресійного моделювання є передбачення (прогнозування) значень залежної змінної при конкретних значеннях пояснювальної змінної. Можливі два підходи: або передбачити умовне математичне сподівання залежної змінної (прогнозування середнього значення), або прогнозувати деяке конкретне значення залежної змінної (передбачення індивідуального значення).

Нехай побудовано рівняння регресії $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$, на основі якого необхідно передбачити умовне математичне сподівання $M(Y|X = x_p)$ змінної Y при $X = x_p$. У даному випадку значення $\hat{y}_p = \hat{b}_0 + \hat{b}_1 x_p$ є оцінкою $M(Y|X = x_p)$. Тоді природньо виникає питання, як сильно може відхилитися модельне середнє значення $\hat{y}_p$, розраховане за емпірічним рівнянням, від відповідного умовного математичного сподівання? Відповідь на це питання дається на основі інтервальних оцінок, побудованих із заданою надійністю $\gamma = 1 - \alpha$ при будь-якому конкретному значення x_p пояснювальної змінної.

Надійний інтервал інтервал для математичного сподівання $M(Y|X = x_p)$ має такий вигляд:

$$(\hat{b}_0 + \hat{b}_1 x_p - t_{\alpha/2, n-2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}; \hat{b}_0 + \hat{b}_1 x_p + t_{\alpha/2, n-2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}).$$

На практиці іноді більш важливо знайти дисперсію Y, ніж її середнє значення або надійні інтервали для умовних математичних сподівань. Це дозволяє визначити допустимі межі для конкретного значення Y.

Надійний інтервал для конкретного значення змінної Y, яке відповідає значенню $X = x_p$ визначається таким чином:

$$(\hat{b}_0 + \hat{b}_1 x_p - t_{\alpha/2, n-2} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}; \hat{b}_0 + \hat{b}_1 x_p + t_{\alpha/2, n-2} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}).$$

Цей інтервал визначає межі, поза якими може знаходитися не більше $100 \cdot \alpha\%$ точок спостережень при $X = x_p$. Зауважимо, що даний інтервал ширший довірчого інтервалу для умовного математичного сподівання (на рисунку 16.1 межі цього інтервалу показано пунктирною лінією).

Аналізуючи побудовані інтервали, слід зазначити, що найбільш вузькими вони будуть при $x_p = \bar{x}$. У міру віддалення x_p від серенього значення довірчі інтервали розширяються (рисунок 16.1). Отже, необхідно достатньо обережно екстраполювати одержані результати на прогнозові області. З іншого боку, по мірі зростання числа спостережень n ці інтервали звужуються до лінії регресії.

Приклад. Для аналізу залежності об'єма споживання Y (у.о.) домогосподарств від доходу X (у.о.) відібрана вибірка об'єму $n = 12$ (помісячно протягом року), результати якої наведені в таблиці 16.1.

Необхідно:

1) визначити тип залежності;

2) оцінити параметри рівняння регресії Y на X;

3) оцінити силу лінійної залежності між X і Y;

4) оцінити статистичну значущість коефіцієнтів регресії при рівні значущості $\alpha = 0{,}05$;

5) розрахувати 95%-ті довірчі інтервали для теоретичних коефіцієнтів регресії;

6) спрогнозувати споживання при доході $X = 160$; визначити 95%-ті довірчі інтервали для цього прогнозу.

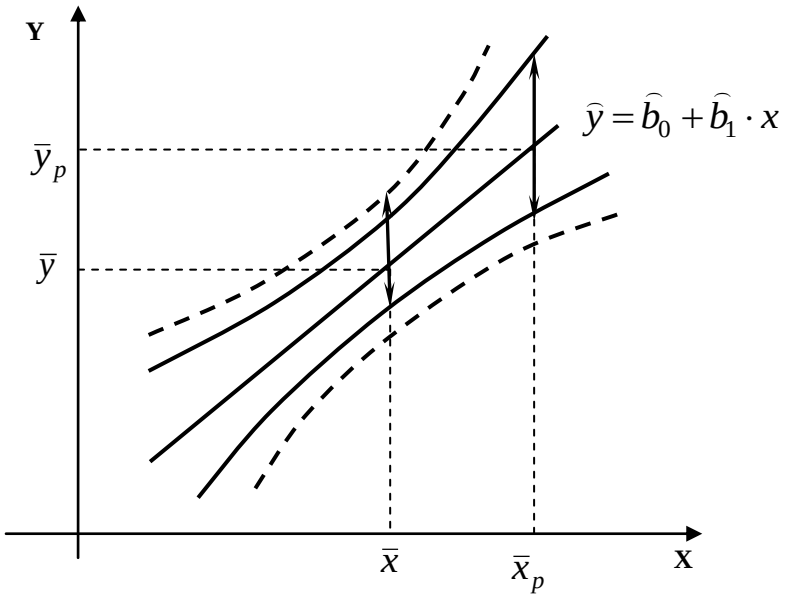


Рис.16.1

Таблиця 16.1

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	107	109	110	113	120	122	123	128	136	140	145	150
y_i	102	105	108	110	115	117	119	125	132	130	141	144

Розв`язання

1 Для визначення типу залежності побудуємо кореляційне поле (рис.16.3).

При розташуванні точок на кореляційн6ому полі припускаємо, що залежність між X і Y лінійна: $\hat{Y} = \hat{b}_0 + \hat{b}_1 X$.

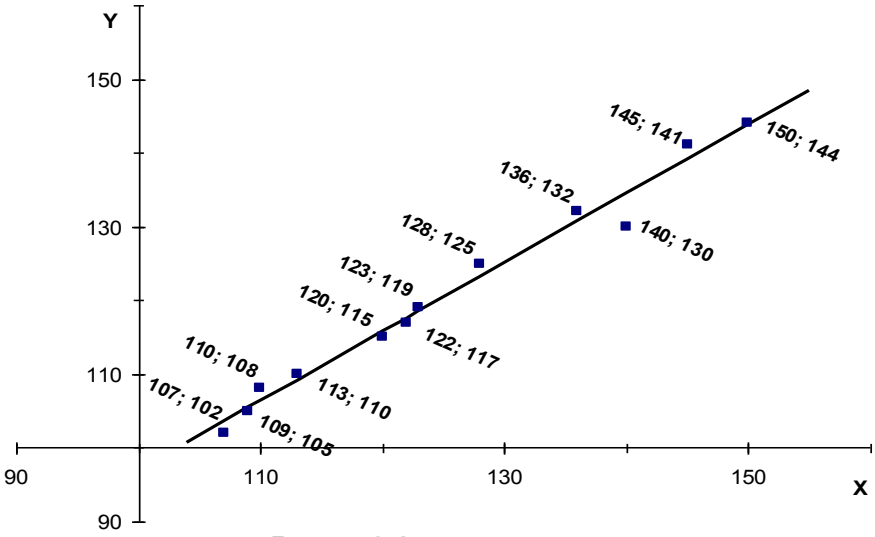


Рис. 16.3

3 Для наочності обчислень за МНК побудуємо таблицю 16.2.

Таблиця 16.2

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$\hat{y}_i$	e_i	e_i^2
1	107	102	11449	10914	10404	103,63	-1,63	2,66
2	109	105	11881	11445	11025	105,49	-0,49	0,24
3	110	108	12100	11880	11664	106,43	1,57	2,46
4	113	110	12769	12430	12100	109,23	0,77	0,59
5	120	115	14400	13800	13225	115,77	-0,77	0,59
6	122	117	14884	14274	13689	117,63	-0,63	0,40
7	123	119	15129	14637	14161	118,57	0,43	0,18
8	128	125	16384	16000	15625	123,24	1,76	3,10
9	136	132	18496	17952	17424	130,71	1,29	1,66
10	140	130	19600	18200	16900	134,45	-4,45	19,8
11	145	141	21025	20445	19881	139,11	1,89	3,57
12	150	144	22500	21600	20736	143,78	0,22	0,05
Сума	1503	1448	190617	183577	176834	-	0	35,3
Середнє	125,25	120,67	15884,75	15298,08	14736,17	-	-	-

У таблиці значення округляються до сотих; ураховуються похибки округлень.

Згідно з МНК, маємо

$$\hat{b}_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{15298,08 - 125,25 \cdot 120,67}{15884,75 - (125,25)^2} = \frac{184,1625}{197,1875} = 0,9339;$$

$$\hat{b}_0 = \overline{y} - \hat{b}_1 \overline{x} = 120,67 - 0,9339 \cdot 125,25 = 3,699.$$

Отже, рівняння парної лінійної регресії має такий вигляд: $\hat{Y} = 3,699 + 0,9339X$. Зобразимо дану пряму регресії на кореляційному полі. За цим рівнянням розрахуємо $\hat{y}_i$, а також $e_i = y_i - \hat{y}_i$.

3 Для аналізу сили лінійної залежності обчислимо коефіцієнт кореляції:

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2} \cdot \sqrt{\overline{y^2} - \overline{y}^2}} = \frac{184,1625}{14,04 \cdot 13,23} = 0,9914.$$

Знайдене значення коефіцієнта кореляції дозволяє зробити висновок про сильну лінійну залежність між змінними X і Y.

4 Оцінимо статистичну значущість коефіцієнтів $\hat{b}_0$ і $\hat{b}_1$.

Маємо

$$S^2_{\hat{b}_1} = \frac{S^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{S^2}{n(\overline{x^2} - \overline{x}^2)} = \frac{\sum_{i=1}^{n} e_i^2}{n(n-2)(\overline{x^2} - \overline{x}^2)} = \frac{\sum_{i=1}^{n}(y_i - \hat{b}_0 - \hat{b}_1 x_i)^2}{n(n-2)(\overline{x^2} - \overline{x}^2)} =$$

$$= 0,0023;$$

$$S_{\hat{b}_1} = \sqrt{0,0023} = 0,0485; \; t_{\hat{b}_1} = \frac{\hat{b}_1}{S_{\hat{b}_1}} = \frac{0,9339}{0,0485} = 19,2557.$$

Критичне значення для рівня значущості $\alpha = 0,05$ дорівнює $t_{крит.} = t_{\alpha/2, n-2} = t_{0,025;10} = 2,23$.

Порівняємо модуль спостережуваного значення $\left|t_{\hat{b}_1}\right| = 19,2557$ з критичним значенням $t_{0,025;10}$. Оскільки $t_{\hat{b}_1} = 19,2557 > 2,23 = t_{êðèò}$, то нульова гіпотеза $H_0 : \hat{b}_1 = 0$ повинна бути відхилена на користь альтернативної гіпотези $H_1 : \hat{b}_1 \neq 0$ при вибраному рівні значущості. Це підтверджує статистичну значущість коефіцієнта регресії $\hat{b}_1$.

Аналогічно перевіряється статистична значущість коефіцієнта $\hat{b}_0$:

$$S^2_{\hat{b}_0} = \overline{x^2} \cdot S^2_{\hat{b}_1} = 0,0023 \cdot 15884,75 = 36,5349;$$

$$S_{\hat{b}_0} = \sqrt{36,5349} = 6,044;$$

$$t_{\hat{b}_0} = \frac{\hat{b}_0}{S_{\hat{b}_0}} = \frac{3,699}{6,044} = 0,612.$$

Оскільки $\left|t_{\hat{b}_0}\right| = 0,612 < 2,23 = t_{êðèò.}$, то гіпотеза $H_0 : \hat{b}_0 = 0$ про статистичну незначущість коефіцієнта $\hat{b}_0$ не відхиляється. Це означає, що у даному випадку вільним членом рівняння регресії можна знехтувати, тобто можна розглядати регресію $Y = \hat{b}_1 \cdot X$.

5 Згідно з формулами

$$(\hat{b}_0 - t_{\alpha/2,n-2} \cdot S_{\hat{b}_0}; \hat{b}_0 + t_{\alpha/2,n-2} \cdot S_{\hat{b}_0}),$$

$$(\hat{b}_1 - t_{\alpha/2,n-2} \cdot S_{\hat{b}_1}; \hat{b}_1 + t_{\alpha/2,n-2} \cdot S_{\hat{b}_1})$$

довірчі інтервали будуть такими:

$(3{,}699 - 2{,}23 \cdot 6{,}044; 3{,}699 + 2{,}23 \cdot 6{,}044) = (-9{,}779; 17{,}177)$ – для коефіцієнта $\hat{b}_0$;

$(0{,}9339 - 2{,}23 \cdot 0{,}0485; 0{,}9339 + 2{,}23 \cdot 0{,}0485) = (0{,}826; 1{,}042)$ – для коефіцієнта $\hat{b}_1$.

Фактично надійні інтервали визначають значення теоретичних коефіцієнтів регресії b_0 і b_1, які будуть прийнятні з надійністю 0,95 при знайдених оцінках $\hat{b}_0 = 3{,}699$ і $\hat{b}_1 = 0{,}9339$.

6 Прогнозове споживання при доході $x = 160$ складе

$$\hat{y}(160) = 3{,}699 + 0{,}9339 \cdot 160 = 153{,}12.$$

Розрахуємо 95%-ті надійний інтервал для умовного математичного сподівання $M(Y|X = x_p)$ при $X = x_p = 160$. Скориставшись формулою

$$(\hat{b}_0 + \hat{b}_1 x_p - t_{\alpha/2,n-2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}; \hat{b}_0 + \hat{b}_1 x_p + t_{\alpha/2,n-2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}),$$

знайдемо межі інтервалу:

$$3{,}699 + 0{,}9339 \cdot 160 \pm 2{,}23 \cdot 1{,}8788 \cdot \sqrt{\frac{1}{12} + \frac{(125{,}25 - 160)^2}{2102{,}1875}}.$$

Таким чином, надійний інтервал для середнього значення Y при $X = 160$ має вигляд (149,728;156,5193). Іншими словами, середнє споживання при доході 160 з імовірністю 0,95 буде знаходитися в інтервалі (149,728;156,5193).

Розрахуємо межі інтервалу, в якому буде зосереджено не менше 95% можливих об’ємів споживання при необмежено великому числі спостережень і рівні доходу $X = 160$. Для цього скористаємося формулою

$$(\hat{b}_0 + \hat{b}_1 x_p - t_{\alpha/2,n-2} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}; \hat{b}_0 + \hat{b}_1 x_p + t_{\alpha/2,n-2} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}).$$

$$3{,}699 + 0{,}9339 \cdot 160 \pm 2{,}23 \cdot 1{,}8788 \cdot \sqrt{1 + \frac{1}{12} + \frac{(125{,}25 - 160)^2}{2102{,}1875}}.$$

Тоді інтервал, у якому буде знаходитися, в крайньому разі, 95% індивідуальних об’ємів споживання при доході $X = 160$, має вигляд

(147,4898;158,7082).

Слід звернути увагу на те, шо він включає в себе надійний інтервал для умовного середнього споживання.

Питання для самоперевірки

1 Сформулюйте дві задачі теорії кореляції.

2 Знаходження параметрів вибіркового рівняння прямої лінії регресії за не згрупованими даними.

3 Вибірковий коефіцієнт кореляції, його властивості.

ДОДАТОК А

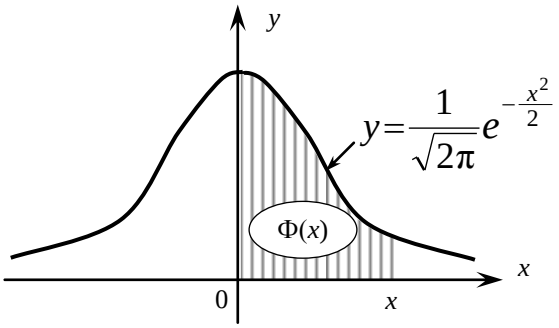
Таблиця значень функції Гауса $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

x	0	1	2	3	4	5	6	7	8	9
0,0	0,399	399	399	399	399	398	398	398	398	397
0,1	397	397	396	396	395	395	394	393	393	392
0,2	391	390	389	389	388	387	386	385	384	383
0,3	381	380	379	378	377	375	374	373	371	370
0,4	368	367	365	364	362	361	359	357	356	354
0,5	352	350	349	347	345	343	341	339	337	335
0,6	333	331	329	327	325	323	321	319	317	314
0,7	312	310	308	306	303	301	299	297	294	292
0,8	290	287	285	283	280	278	276	273	271	269
0,9	266	264	261	259	257	254	252	249	247	244
1,0	242	240	237	235	232	230	228	225	223	220
1,1	218	216	213	211	208	206	204	201	199	197
1,2	194	192	190	187	185	183	180	178	176	174
1,4	150	148	146	144	142	140	137	135	133	132
1,6	111	109	107	106	104	102	101	099	097	096
1,8	079	078	076	075	073	072	071	069	068	067
1,9	066	064	063	062	061	059	058	057	056	055
2,0	054	053	052	051	050	049	048	047	046	045
2,1	044	043	042	041	040	040	039	038	037	036
2,2	036	035	034	033	033	032	031	030	030	029
2,4	022	022	021	021	020	020	019	019	018	018
2,6	014	013	013	013	012	012	012	011	011	011
2,8	008	008	008	007	007	007	007	007	006	006
2,9	006	006	006	006	005	005	005	005	005	005
3,0	004	004	004	004	004	004	004	004	004	003
3,1	003	003	003	003	003	003	003	003	003	003
3,2	002	002	002	002	002	002	002	002	002	002
3,3	002	002	002	002	002	002	001	001	001	001
3,4	001	001	001	001	001	001	001	001	001	001
3,5	001	001	001	001	001	001	001	001	001	001
3,6	001	001	001	001	001	001	001	001	001	000

ДОДАТОК Б

Значення функції Лапласа

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}\,dt\,.$$



x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.0	0.00000				
0.05	0.01994	1.05	0.35314	2.05	0.47982
0.10	0.03983	1.10	0.36433	2.10	0.48214
0.15	0.05962	1.15	0.37493	2.15	0.48422
0.20	0.07926	1.20	0.38493	2.20	0.48610
0.25	0.09871	1.25	0.39435	2.25	0.48778
0.30	0.11791	1.30	0.40320	2.30	0.48928
0.35	0.13683	1.35	0.41149	2.35	0.49061
0.40	0.15542	1.40	0.41924	2.40	0.49180
0.45	0.17364	1.45	0.42647	2.45	0.49286
0.50	0.19146	1.50	0.43319	2.50	0.49379
0.55	0.20884	1.55	0.43943	2.55	0.49461
0.60	0.22575	1.60	0.44520	2.60	0.49534
0.65	0.24215	1.65	0.45053	2.65	0.49598
0.70	0.25804	1.70	0.45543	2.70	0.49653
0.75	0.27337	1.75	0.45994	2.75	0.49702
0.80	0.28814	1.80	0.46407	2.80	0.49744
0.85	0.30234	1.85	0.46784	2.85	0.49781
0.90	0.31594	1.90	0.47128	2.90	0.49813
0.95	0.32894	1.95	0.47441	2.95	0.49841
1.00	0.34134	2.00	0.47725	3.00	0.49865
3.1	0.49903	3.2	0.49931	3.3	0.49952
3.4	0.49966	3.5	0.49977	3.6	0.49984
3.7	0.49989	3.8	0.49993	3.9	0.49995
4.0	0.499968	4.5	0.499997	5.0	0.49999997

ДОДАТОК В

Критичні точки розподілення Стьюдента (t – розподіл)

Число ступенів свободи, k	Рівень значущості, α						
	0,20	0,10	0,05	0,02	0,01	0,002	0,001
1	3,08	6,31	12,7	31,82	63,66	127,32	636,62
2	1,89	2,92	4,30	6,97	9,93	14,09	31,60
3	1,64	2,35	3,18	4,54	5,84	7,45	12,94
4	1,53	2,13	2,78	3,75	4,60	5,60	8,61
5	1,48	2,02	2,57	3,37	4,03	4,77	6,86
6	1,44	1,94	2,45	3,14	3,71	4,32	5,96
7	1,42	1,90	2,36	3,00	3,50	4,03	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	5,04
9	1,38	1,83	2,26	2,82	3,25	3,69	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,59
11	1,36	1,80	2,20	2,72	3,11	3,50	4,44
12	1,36	1,78	2,18	2,68	3,05	3,43	4,32
13	1,35	1,77	2,16	2,65	3,01	3,37	4,22
14	1,34	1,76	2,14	2,62	2,98	3,33	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	4,02
17	1,33	1,74	2,11	2,57	2,90	3,22	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,92
19	1,33	1,73	2,09	2,54	2,86	3,17	3,88
20	1,33	1,73	2,09	2,53	2,85	3,15	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,82
22	1,32	1,72	2,07	2,51	2,82	3,12	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,77
24	1,32	1,71	2,06	2,49	2,80	3,09	3,75
25	1,32	1,71	2,96	2,48	2,79	3,08	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,69
28	1,31	1,70	2,05	2,47	2,76	3,05	3,67
29	1,31	1,70	2,04	2,46	2,76	3,04	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,65
40	1,30	1,68	2,02	2,42	2,70	2,97	3,55
60	1,30	1,67	2,00	2,39	2,66	2,91	3,46
120	1,29	1,66	1,98	2,36	2,62	2,86	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,29

ДОДАТОК Г

Таблиця значень $t_\gamma(\gamma, k)$.

k	γ												
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,158	0,326	0,510	0,727	1,00	1,376	1,963	3,078	6,314	12,706	31,82	63,65	63,66
2	0,142	0,289	0,445	0,617	0,816	1,061	1,336	2,886	2,920	4,303	6,965	9,925	31,60
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	2,638	2,353	3,182	4,541	5,841	12,94
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,694	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,363	4,032	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,401	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,086	1,363	1,796	2,201	2,718	3,106	4,487
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,728	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,103	2,552	2,872	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883

20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,859	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,857	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0.855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646

СПИСОК ДЖЕРЕЛ ІНФОРМАЦІЇ

1 Кудрявцев В.А., Демидович Б.П. Краткий курс высшей математики – М.: Наука, 1975.

2 Берман Г.Н. Сборник задач по курсу математического анализа. – М: Наука, 1985.

3 Щипачёв В.С. Курс высшей математики – Изд. МГУ, 1981.

4 Овчинников П.Ф., Яремчук Ф.П., Михайленко В.М. Высшая математика. Под ред. П.Ф. Овчинникова – К.: Высш. шк., 2001.

5 Пискунов Н.С. Дифференциальное и интегральное исчисление для втузов : В. 2т.- М. : Наука, 1985.

6 Вирченко Н.А., Ляшко И.И., Швецов К.И. Графики функций: Справочник. – К.: Наук. думка, 1979.

7 Гмурман В. Е. Теория вероятностей и математическая статистика – М.: Высш. шк., 1977.

8 Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике – М.: Высш. шк., 1977.

9 Маркович Э.С. Курс высшей математики с элементами теории вероятностей и математической статистики. – М.: Высш.шк., 1976.

10 Данко П.Е., Попов А.Г. Высшая математика в упражнениях и задачах – Т.3. – М.: Высш. шк., 1978.

11 Кремер Н.Ш. Теория вероятностей и математическая статистика – М.: ЮНИТИ, 2010.

12 Натансон И.П. Краткий курс высшей математики.- М.: Физматгиз, 1963.

13 Ефимов Н.В. Краткий курс аналитической геометрии. – М.: Наука, 1972.

14 Клетеник Д.В. Сборник задач по аналитической геометрии. – М.: Наука, 1986.

15 Борисенко О.А. Диференціальна геометрія і топологія. – Х.:ДСВ Основа, 1995.

16 Станішевский С.О. Вища математика. – Х.: ХДАМГ, 2002.

17 Эльсгольц Л.Э. Дифференциальные уравнения и вариационное исчисление. – М.: Наука, 1965.

18 Письменный Д. Конспект лекций по высшей математике: в 2 ч. – М.: Айрис- пресс, 2004.

19 Маркушевич А.И. Замечательные кривые. – М.:Наука, 1978.

ЗМІСТ

Лекція 1 ……... 3
Лекція 2 ……... 7
Лекція 3 ……... 11
Лекція 4 ……... 17
Лекція 5 ……... 19
Лекція 6 ……... 21
Лекція 7 ……... 27
Лекція 8 ……... 30
Лекція 9 ……... 37
Лекція 10 ……..... 42
Лекція 11 ……. 46
Лекція 12 ……. 50
Лекція 13 ……..... 60
Лекція 14 ……. 65
Лекція 15 ……. 71
Лекція 16 ……..... 75
ДОДАТОК А …….. 83
ДОДАТОК Б …….. 84
ДОДАТОК В …….. 85
ДОДАТОК Г …….. 86
Список джерел інформації……. 88

Навчальне видання

Стасенко Олександр Миколайович

МАТЕМАТИКА ДЛЯ ЕКОНОМІСТІВ
(теорія ймовірності і математична статистика**)**

Тексти лекцій

Роботу до видання рекомендував Поклонський Є.В.

За редакцією автора

План 2017р., поз.58,17 Формат 60х84 1/16. Папір друк. № 2
Підп. до друку 25.11.2016 Обл. -вид. арк. Умов. друк. арк. 5
Тираж 50 прим. Зам. № 37287.

ХНУБА, Україна, 61002, Харків, вул. Сумська, 40

Підготовлено та віддруковано РВВ Харківського національного університету будівництва та архітектури