МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Харківський державний університет харчування та торгівлі

Ж. А. Крутовий, М. С. Софронова

ЕЛЕМЕНТИ ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ

Навчальний посібник

Харків
ХДУХТ
2016

УДК 519.22/.25
ББК 22.172
К 84

Рецензенти:
д-р техн. наук, проф. М. С. Синєкоп,
д-р техн. наук, проф. П. П. Пивоваров

Рекомендований до видання вченою радою ХДУХТ
протокол № 7 від 25.02.2016 р.

Крутовий Ж. А.
Елементи теорії ймовірностей та математичної статистики.
К 84 [Електронний ресурс] : навч. посібник / Ж. А. Крутовий,
М. С. Софронова. – Електрон. дані. – Х. : ХДУХТ, 2016. – 1 електрон.
опт. диск (CD-ROM); 12 см. – Назва з тит. екрана.

Навчальний посібник містить основні теоретичні відомості з теорії ймовірностей та математичної статистики. Наведено багато прикладів, що ілюструють викладені теоретичні поняття, у тому числі з галузі.

Призначений для студентів вищих навчальних закладів та осіб, які використовують імовірнісні та статистичні методи під час розв’язання практичних задач. Посібник буде корисним також студентам прискореної та заочної форм навчання.

УДК 519.22/.25
ББК 22.172

ПЕРЕДМОВА

Це навчальне видання написане на основі лекцій, прочитаних авторами протягом багатьох років студентам різних факультетів ХДУХТ.

Доцільність підготовки посібника зумовлена суттєвим зменшенням кількості аудиторних годин, передбачених програмою навчальної дисципліни, що впливає як на обсяг матеріалу, так і на методику викладання.

Мета посібника – забезпечити засвоєння понять курсу теорії ймовірностей та математичної статистики, сприяти формуванню навичок у застосуванні математичних методів, допомогти студентам при самостійному розв'язуванні задач.

Перша частина посібника – теорія ймовірностей – містить 4 лекції з таких тем: події, безпосереднє обчислення ймовірностей; теореми додавання та множення ймовірностей; формула повної ймовірності, повторні випробування, випадкові величини; щільність розподілу ймовірностей, основні числові характеристики та закони розподілу випадкових величин.

Друга частина посібника – математична статистика, зокрема теорія кореляції –містить лекції з таких тем: функціональна та статистична залежності, кореляційна таблиця; кореляційна залежність, емпірична та теоретична лінії регресії; спрощення розрахунків коефіцієнта кореляції; оцінка сили (тісноти) кореляційної залежності у випадку прямолінійної регресії, коефіцієнт кореляції; статистична перевірка гипотез; перевірка гіпотези про нормальний закон розподілу ймовірностей випадкової величини – теплові втрати маси в процесі варіння м'яса.

ЕЛЕМЕНТИ ТЕОРІЇ ЙМОВІРНОСТЕЙ

ЛЕКЦІЯ № 1

ПОДІЇ. БЕЗПОСЕРЕДНЄ ОБЧИСЛЕННЯ ЇХ ЙМОВІРНОСТЕЙ

План

1. Предмет теорії ймовірностей.
2. Приклади закономірностей в випадкових масових явищах.
3. Початкові поняття теорії ймовірностей: подія, достовірні, неможливі, несумісні, протилежні події.
4. Класичне означення ймовірності події.
5. Статистичне поняття ймовірності.
6. Безпосереднє обчислення ймовірностей подій з використанням понять комбінаторики:
   а) розміщення, приклад;
   б) перестановки, переставлення, приклад;
   в) сполучення, приклад.

Предмет теорії ймовірностей

Предметом теорії ймовірностей є дослідження закономірностей в випадкових масових явищах незалежно від їх конкретної природи.

Приклад 1. В посудині міститься газ. Кожна молекула рухається хаотично. Тиск на стінки посудини – це результат удару великої кількості молекул. Сила удару різних молекул різна. Але результат сумарної дії великої кількості молекул в різні моменти часу відрізняється дуже мало, що високоточні прилади цього не відчувають, тобто він являється стійким.

Приклад 2. Розмір взуття, необхідного будь-якому довільному покупцеві, є випадкова величина. Але, якщо розглянути велику кількість покупців, наприклад, 10^6, то доля (частка) покупців, яким необхідне взуття, наприклад, 42 розміру, складає, наприклад, 20% і являється майже сталою величиною. Тобто вона слабо змінюється при заміні одного мільйона покупців іншим мільйоном.

Названа закономірність має практичний інтерес.

Початкові поняття теорії ймовірностей

Означення. Подією в теорії ймовірностей (Т.Й.) називається результат експерименту, явища, спостереження тощо. Позначаються події літерами *A, B, C, ...* .

Приклад. В ящику лежать браковані і небраковані (стандартні) деталі. Подія *A* полягає в тому, що взята навмання деталь буде стандартною.

Означення. Подія називається достовірною (вірогідною), якщо при виконанні певних умов вона не може не відбутися.

Приклад. В ящику лежить 100 стандартних банок. Подія A полягає в тому, що взята навмання банка являється стандартною. Ця подія достовірна.

Означення. Подія називається неможливою, якщо при виконанні певних умов вона не може відбутися.

Приклад. На складі знаходяться стандартні прилади. Подія B полягає в тому, що взятий навмання прилад буде нестандартним. Подія B – неможлива.

Означення. Події A та B називаються несумісними, якщо поява однієї з них виключає можливість появи іншої події.

Приклад. У студента один лотерейний білет. Подія A полягає в тому, що на цей білет припаде виграш – велосипед; подія B – виграш – одеколон. Ці події – несумісні.

Означення. Події A, B, C, ... , M називаються єдино можливими, якщо при виконанні певних умов не може відбутися подія, яка не належить даній сукупності.

Означення. Події A, B, C, ... , M утворюють повну групу подій, якщо вони являються єдино можливими і несумісними.

Приклад. Підкидається гральний кубик. A_1 – подія полягає в тому, що на верхній грані буде одиниця, ... , A_6 – на верхній грані буде шість точок. Події A_1, A_2, ... , A_6 утворюють повну групу подій.

Означення. Дві єдино можливі і несумісні події називаються протилежними.

Приклад. Підкидається монета. Подія A – герб на верхній стороні, $\overline{A}$ – не герб. Події A та $\overline{A}$ – протилежні.

Ймовірність подій

Класичне визначення. Ймовірністю події A називається відношення числа випадків, які сприяють появі цієї події, до загального числа єдино можливих несумісних рівноможливих випадків:

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

Приклад. В ящику 100 деталей, серед них 5 бракованих. Ймовірність того, що взята навмання деталь буде бракованою:

$$P(A)=\frac{5}{100}.$$

Статистичне поняття ймовірностей

Припустимо, що проводиться деякий експеримент n разів. При цьому подія A відбулася μ разів. Тоді відношення $P^*(A) = \dfrac{\mu}{n}$ називається відносною частотою події A.

Якщо проводиться декілька серій з n експериментів (n досить велике), і відносна частота $P^*(A)$ коливається навколо числа p, то це число називається ймовірністю події A. Це визначення ймовірності називається статистичним.

<u>Приклад</u>. Гральний кубик підкинули 10 разів. Число «1» випало 3 рази. Відносна частота події A (вона полягає в тому, що на верхній грані буде число «1») складає

$$P^*(A) = \frac{3}{10},$$

а ймовірність

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

Безпосереднє обчислення ймовірності

<u>Приклад 1</u>. Одночасно підкидають два гральні кубики. Необхідно знайти ймовірність того, що сума очок на верхній грані буде 9.

Подію, ймовірність якої треба знайти, позначимо через A.

Всі можливі випадки зображені в табл. 1.1.

Таблиця 1.1

	1	2	3	4	5	6
1	1;1	1;2	1;3	1;4	1;5	1;6
2	2;1	2;2	2;3	2;4	2;5	2;6
3	3;1	3;2	3;3	3;4	3;5	3;6
4	4;1	4;2	4;3	4;4	4;5	4;6
5	5;1	5;2	5;3	5;4	5;5	5;6
6	6;1	6;2	6;3	6;4	6;5	6;6

Загальне їх число дорівнює n=36. Число випадків, в яких подія A може відбутися, складає m=4. Це випадки: (3;6), (4;5), (5;4), (6;3). Ймовірність події A дорівнює $P(A) = \dfrac{4}{36} = \dfrac{1}{9}$.

Безпосереднє обчислення ймовірностей подій з використанням понять комбінаторики

Задача 1. Таємний замок має 5 дисків. На кожному дискові 9 чисел: 1, 2, 3, ... , 9. Знайти ймовірність того, що замок відриється, якщо набрати навмання п'ятизначне число.

За формулою класичної ймовірності

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

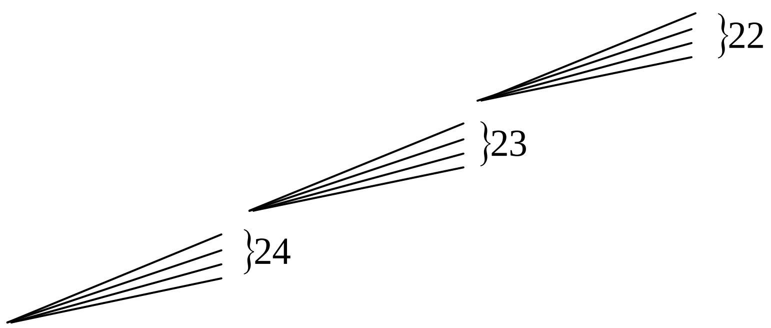
маємо: $m = 1$, $n = 9 \cdot 9 \cdot 9 \cdot 9 \cdot 9 = 9^5$.

Задача 2. В групі ПР-14 25 студентів. Викладач має намір навмання викликати до дошки чотирьох студентів. Знайти ймовірність того, що першим буде названий студент Зуб, другим – Шаповал, третім – Алексєєва. четвертим – Гусєва.

Означення. Розміщеннями із n елементів по m називаються комбінації по m елементів, які відрізняються складом елементів або їх розташуванням (порядком):

$$A_n^m = n(n-1)(n-2)\ldots(n-(m-1)).$$

Тоді $P = \dfrac{1}{A_{25}^4}$ (див. рис. 1.1).



1
2
…
25

Рис. 1.1

$$P = \frac{1}{25 \cdot 24 \cdot 23 \cdot 22}.$$

$$A_n^m = \frac{n!}{(n-m)!}.$$

Задача 3. На однакових картках написані букви Т, С, 0, П, Р. Картки перегорнуті. Їх навмання виймають по одній та складають. Знайти ймовірність того, що буде складене слово «спорт».

Означення. Перестановками (переставленнями) з n елементів називаються комбінації з n елементів, які відрізняються тільки розташуванням елементів.

$$P_n = n(n-1)(n-2)\ldots 1.$$

Число переставлень з n елементів складає

$$P_n = n!$$

Тоді $P(A) = \dfrac{1}{P_5}$.

Задача 4. В групі ПР-14 25 студентів. Викладач має намір навмання викликати до дошки чотирьох студентів. Знайти ймовірність того, що викликаними до дошки будуть студенти: Зуб, Шаповал, Алексєєв, Гусєва.

В цій задачі порядок виходу до дошки значення не має.

Означення. Сполученнями з n елементів по m називаються такі комбінації по m елементів, які відрізняються тільки складом елементів.

Число комбінацій з n елементів по m позначається C_n^m.

$$A_n^m = C_n^m P_m = C_n^m m!,$$

$$C_n^m = \frac{A_n^m}{m!} = \frac{n!}{m!(n-m)!}.$$

Ймовірність шуканої події дорівнює $P = \dfrac{1}{C_{25}^4} = \dfrac{4! \cdot 21!}{25!}$.

ЛЕКЦІЯ № 2

ТЕОРЕМИ ДОДАВАННЯ ТА МНОЖЕННЯ ЙМОВІРНОСТЕЙ

План

1. Ймовірність суми несумісних подій.
2. Залежні та незалежні події. Умовна ймовірність.
3. Ймовірність добутку залежних та незалежних подій.
4. Сума сумісних подій.
5. Ймовірність суми *n* сумісних подій.
6. Ймовірність суми певного числа сумісних подій.

Теорема додавання ймовірностей несумісних подій

Розглянемо задачу.

Задача. В ящику знаходяться яблука першого, другого та третього сортів. Першого сорту 40%, другого – 35%, третього – 25%. Навмання дістають із ящика одне яблуко. Знайти ймовірність того, що це яблуко буде або першого, або другого сорту.

Введемо події:

A – полягає в тому, що взяте навмання яблуко буде першого сорту;

B – полягає в тому, що взяте навмання яблуко буде другого сорту;

C – подія полягає в тому, що взяте навмання яблуко буде або першого, або другого сорту.

Означення. Сумою двох несумісних подій називається подія, яка полягає в тому, що відбудеться будь-яка з них.

$C = A + B$.

$P(C) = P(A + B) = P(A) + P(B) = 0{,}4 + 0{,}35 = 0{,}75$.

Теорема. Ймовірність суми певного числа несумісних подій $A_1, A_2, \ldots, A_n$ дорівнює сумі ймовірностей цих подій:

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

Висновки:

1. Сума ймовірностей подій, які утворюють повну групу подій, дорівнює одиниці.

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

2. Якщо A та $\overline{A}$ – протилежні події, то має місце рівність

$$P(A + \overline{A}) = P(A) + P(\overline{A}) = 1,$$

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

Теорема множення ймовірностей подій

Означення 1. Події *A* та *B* називаються незалежними, якщо ймовірність появи однієї з них не залежить від того, відбудеться чи ні друга подія. В противному випадку події називаються залежними.

Означення 2. Події *A*, *B*, *C*, ... , *M* називаються незалежними у сукупності, якщо будь-яка з них і будь-яка комбінація решти подій (яка містить всю решту або декілька з них) являються незалежними.

Означення 3. Ймовірність події *B*, яка знайдена в припущенні, що подія *A* відбулася, називається умовною ймовірністю події *B*. Вона позначається так: $P_A(B)$.

Приклад. В ящику 10 куль: 7 білих і 3 чорних. Подія *A* полягає в тому, що перша, взята навмання куля – біла; подія *B* полягає в тому, що друга куля, взята навмання, являється білою.

Розглянемо схему без повернення кулі. Тоді

$$P(A)=\frac{7}{10},\quad P_A(B)=\frac{6}{9}.$$

При схемі з поверненням кулі

$$P(A)=\frac{7}{10},\quad P(B)=\frac{7}{10}$$

– це безумовна ймовірність.

Означення 4. Добутком подій *A*, *B*, *C*, ..., *M* називають подію, яка полягає в сумісній появі всіх цих подій.

Теорема. Ймовірність сумісної появи певного числа залежних подій $A_1, A_2, \ldots, A_n$ визначається за формулою:

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

Ймовірність сумісної появи певного числа незалежних подій дорівнює

$$P(A_1\cdot A_2\cdot\ldots\cdot A_n)=P(A_1)\cdot P(A_2)\cdot\ldots\cdot P(A_n).$$

Приклад. В ящику 30 пар взуття. Серед них 5 пар бракованих. Навмання дістають по одній парі три пари. Знайти ймовірність того, що і перша, і друга, і третя пари виявляться бракованими.

Позначимо через A_i – подію, яка полягає в тому, що *i*-та пара взуття бракована, *i*=1,2,3.

$$P(A_1\cdot A_2\cdot A_3)=P(A_1)\cdot P_{A_1}(A_2)\cdot P_{A_1A_2}(A_3).$$

Події A_i, *i*=1,2,3 – залежні. Ймовірності складають:

$$P(A_1)=\frac{5}{30},\ P_{A_1}(A_2)=\frac{4}{29},\ P_{A_1A_2}(A_3)=\frac{3}{28}.$$

Тоді $P(A_1\cdot A_2\cdot A_3)=\frac{5}{30}\cdot\frac{4}{29}\cdot\frac{3}{28}=\frac{2}{203}$.

Ймовірність суми сумісних подій

Означення. Сумою сумісних подій називається подія, яке полягає в тому, що відбудеться, принаймні, одна з цих подій.

Теорема. Ймовірність суми двох сумісних подій визначається за формулою:

$$P(A+B)=P(A)+P(B)-P(AB).$$

Приклад. Ймовірність влучення в ціль при першому пострілі 0,7; при другому – 0,8. Для знищення цілі досить одного влучення. Визначити ймовірність того, що ціль буде знищена, якщо зроблено два постріли.

A_i, i=1,2 – подія, яка полягає в тому, що в ціль влучено при i-му пострілі.

$$C=A_1+A_2,\ P(C)=P(A_1)+P(A_2)-P(A_1A_2),$$

$$P(C)=P(A_1)+P(A_2)-P(A_1)P(A_2)=0{,}7+0{,}8-0{,}7\cdot 0{,}8=0{,}94.$$

Ймовірність суми *n* сумісних подій

$\sum\limits_{i=1}^{n}A_i$ або ймовірність того, що відбудеться принаймні одна з цих подій, визначається за формулою:

$$P\left(\sum_{i=1}^{n}A_i\right)=1-P(\overline{A}_1\cdot\overline{A}_2\cdot\ldots\overline{A_n}).$$

Якщо події $A_1,A_2,\ldots,A_n$ сумісні, але незалежні, то

$$P\left(\sum_{i=1}^{n}A_i\right)=1-P(\overline{A}_1)P(\overline{A}_2)\cdot\ldots\cdot P(\overline{A_n}).$$

Припустимо, що проводиться n незалежних випробувань. Ймовірність того, що подія A відбудеться в будь-якому з них, дорівнює $P(A)=p$. Тоді ймовірність того, що подія A відбудеться принаймні один раз в n незалежних випробуваннях, складає

$$P(C)=1-(1-p)^n \quad \text{або} \quad P(C)=1-q^n,\ \text{де } q=1-p.$$

Приклад. Ймовірність затримки водія в нетверезому стані на будь-якому з пунктів ДАІ складає p=0,5. Знайти ймовірність того, що водія буде затримано, якщо на маршруті 6 пунктів ДАІ.

$$P(C)=1-q^6=1-(1-p)^6=1-(1-0{,}5)^6=1-(0{,}5)^6\approx 1-0{,}016=0{,}984.$$

ЛЕКЦІЯ № 3

ФОРМУЛА ПОВНОЇ ЙМОВІРНОСТІ. ПОВТОРНІ ВИПРОБУВАННЯ. ВИПАДКОВІ ВЕЛИЧИНИ

План

1. Формула повної ймовірності.
2. Формула Бейєса.
3. Повторення незалежних випробувань. Формула Бернуллі.
4. Випадкові величини (ВВ). Дискретні та неперервні.
5. Закони розподілу випадкових величин: ряд розподілу, функція розподілу.
6. Ймовірність попадання випадкової величини на інтервал $[\alpha, \beta]$.

Формула повної ймовірності

Розглянемо задачу.

Задача. В збиральний цех поступають деталі однакові з трьох верстатів. З першого – 40% загальної кількості, з другого – 35%, з третього – 25%. Ймовірність браку складає: 0,05 для першого верстату, 0,1 – для другого, 0,04 – для третього. Знайти ймовірність того, що взята навмання деталь в збиральному цеху буде бракованою.

Позначимо через A подію, ймовірність якої знаходиться. Тоді $A = A_1 + A_2 + A_3$, де A_i – подія, яка полягає в тому, що деталь виготовлена на i-му верстаті і є бракованою. Введемо гіпотези (несумісні події, які складають повну групу подій):

$H_i, i = 1,2,3$ – взята навмання деталь, виготовлена на i-му верстаті.

Тоді $A_i = H_i A, i = 1,2,3.$

$$P(A) = P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3) =$$
$$= P(H_1 A) + P(H_2 A) + P(H_3 A) = P(H_1) P_{H_1}(A) + P(H_2) P_{H_2}(A) +$$
$$+ P(H_3) P_{H_3}(A) = 0{,}4 \cdot 0{,}05 + 0{,}35 \cdot 0{,}1 + 0{,}25 \cdot 0{,}04 = 0{,}65.$$

Теорема. Якщо подія A може відбутися лише при одній з несумісних подій (гіпотез) $H_1, H_2, ..., H_n$, що утворюють повну групу подій, то ймовірність події A дорівнює сумі добутків ймовірностей гіпотез на ймовірності події при відповідних гіпотезах

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

Формула Бейєса

Позначимо: H_i – гіпотези, A – подія, яка відбувається при одній з гіпотез. Тоді

$$P(H_i A) = P(H_i)\cdot P_{H_i}(A) = P(A)\cdot P_A(H_i),\ i=1,2,\ldots,n.$$

Звідси маємо:

$$P_A(H_i) = \frac{P(H_i)\cdot P_{H_i}(A)}{P(A)} = \frac{P(H_i)\cdot P_{H_i}(A)}{\sum\limits_{i=1}^{n} P(H_i)\cdot P_{H_i}(A)}.$$

Це формула Бейєса. За її допомогою можна переоцінити ймовірність гіпотези після того, як подія A відбулася.

<u>Приклад</u>. У магазин надходять електричні лампи з трьох заводів, відповідно 30%, 50%, 20% загальної кількості. Ймовірність браку складає відповідно 0,1; 0,04; 0,05 для різних заводів.

Лампочка, взята в магазині навмання, виявилася бракованою. Чому дорівнює ймовірність того, що лампочка виготовлена на другому заводі?

Нехай A – поява бракованої лампочки, H_i – лампочка надійшла з i-го заводу, $i=1,2,3$. Тоді

$$P_A(H_2) = \frac{P(H_2)\cdot P_{H_2}(A)}{P(A)} = \frac{0{,}5\cdot 0{,}04}{0{,}3\cdot 0{,}1 + 0{,}5\cdot 0{,}04 + 0{,}2\cdot 0{,}05} = \frac{0{,}02}{0{,}06} = \frac{1}{3}.$$

Повторення незалежних випробувань. Формула Бернуллі

Розглянемо задачу.

<u>Задача</u>. В ящику 10 куль, серед них 7 білих, 3 чорних. Розглянемо схему з поверненням, тобто кулю, взяту навмання, роздивляються і повертають в ящик. Необхідно знайти ймовірність того, що в результаті 5 спроб куля білого кольору буде взята 3 рази.

Позначимо: A – подія, яка полягає в тому, що в результаті одного випробування взята біла куля; B – подія, ймовірність якої треба знайти.

$$B_1 = A\cdot A\cdot A\cdot \overline{A}\cdot \overline{A},\ \ B_1 = A\cdot A\cdot \overline{A}\cdot \overline{A}\cdot A,\ \ldots$$

$$B = \sum_i B_i,$$

$$P(B_i) = P^3(A)\cdot P^2(\overline{A}) = p^3 q^2, \text{ де } q = 1-p,\ \ p = P(A).$$

Всього подій B_i буде C_5^3. Тоді шукана ймовірність складає

$$P(B) = C_5^3 p^3 q^2.$$

Теорема Бернуллі. Нехай p – ймовірність деякої події в одному випробуванні. Тоді ймовірність того, що подія, що розглядається, відбудеться m разів у n незалежних випробуваннях, складає

$$P_{m,n} = C_n^m p^m q^{n-m}, \text{ де } q = 1 - p.$$

З використанням цієї формули можна розглядати задачі 4-х типів:

1. Знаходити ймовірність того, що подія відбудеться m разів у n незалежних випробуваннях.
2. Знаходити ймовірність того, що розглянута подія відбудеться не більше m_1 разів у n незалежних випробуваннях.

$$P_n(m \le m_1) = P_{0,n} + P_{1,n} + \ldots + P_{m_1,n}.$$

3. Знаходити ймовірність того, що подія відбудеться не менше, ніж m_2 разів в n незалежних випробуваннях.

$$P_n(m \ge m_2) = P_{m_2,n} + P_{m_2+1,n} + \ldots + P_{n,n}.$$

4. Знаходити ймовірність того, що розглянута подія відбудеться не менше, ніж m_1 разів, але не більше, ніж m_2 разів у n незалежних випробуваннях.

$$P_n(m_1 \le m \le m_2) = P_{m_1,n} + P_{m_1+1,n} + \ldots + P_{m_2,n}.$$

Приклад. Ймовірність народження хлопчика – 0,51. Знайти ймовірність того, що з п'яти народжених буде не менше трьох хлопчиків.

$$P_5(m \ge 3) = P_{3,5} + P_{4,5} + .P_{5,5} =$$
$$= C_5^3(0{,}51)^3(0{,}49)^2 + C_5^4(0{,}51)^4 0{,}49 + C_5^5(0{,}51)^5.$$

Випадкові величини

Означення. Змінна величина, яка приймає різні значення випадково, називається випадковою величиною.

Приклад. Кількість студентів, які запізняться на наступну лекцію; кількість зіпсованих касових апаратів; тривалість життя людини; інтервал часу між двома суднами, які прибувають у порт тощо.

Випадкові величини (ВВ) бувають дискретні та неперервні. Якщо всі значення, які приймають ВВ, можна перерахувати, то величина називається дискретною.

Закон розподілу випадкових величин

Означення. Законом розподілу ВВ називається залежність між значеннями, які приймає ВВ X та їх ймовірностями.

Одним із законів розподілу є ряд розподілу. Це таблиця, в якій задані значення дискретної ВВ (в першому рядку) та їх ймовірності (в другому рядку).

Приклад.

x_i	2	5	7	8
p_i	0,1	0,3	0,2	0,4

Графічно ряд розподілу зображають многокутником розподілу (рис. 3.1)

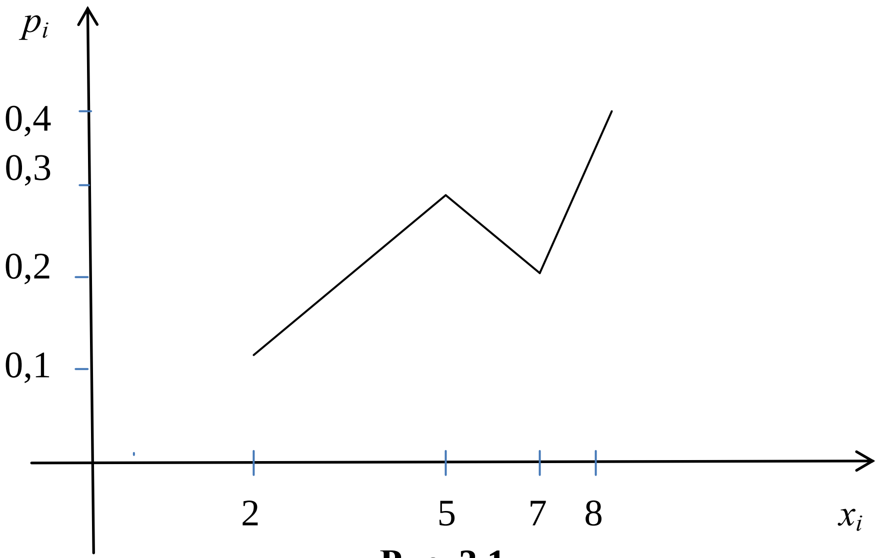


Рис. 3.1

Означення. Функцією розподілу ВВ X називається функція $F(x)$, яка виражає для кожного конкретного значення x ймовірність того, що X приймає будь-яке значення менш, ніж x.

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

Приклад. Записати функцію розподілу для ВВ X, заданої рядом розподілу в попередньому прикладі

$$F(x) = \begin{cases} 0 \quad при \quad x \le 2, \\ 0{,}1 \;\; при \;\; 2 < x \le 5, \\ 0{,}4 \;\; при \;\; 5 < x \le 7, \\ 0{,}6 \;\; при \;\; 7 < x \le 8, \\ 1 \quad при \;\; x > 8. \end{cases}$$

Графік цієї функції зображено на рис. 3.2.

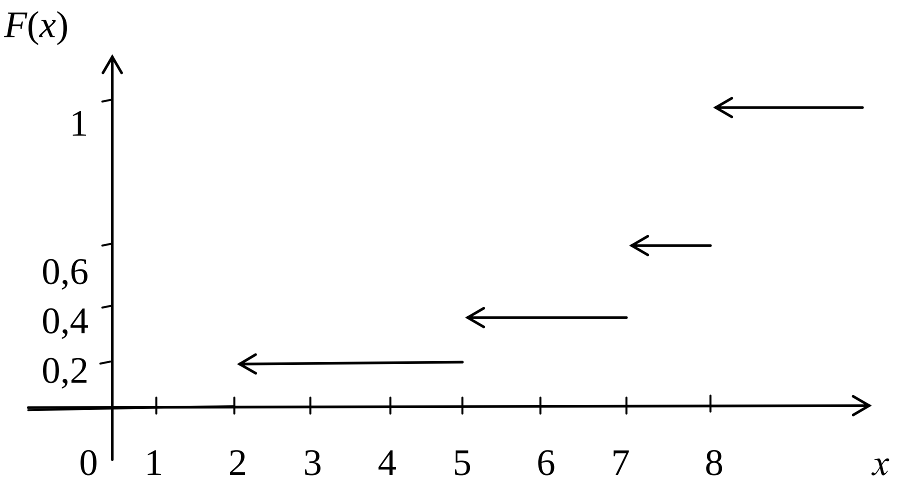


Рис. 3.2

Властивості $F(x)$:

1. $0 \le F(x) \le 1$, $F(-\infty) = 0$, $F(+\infty) = 1$.
2. $F(x_1) \le F(x_2)$, якщо $x_1 < x_2$, тобто $F(x)$ – неспадна.
3. Ймовірність того, що ВВ *X* приймає значення на інтервалі $[\alpha, \beta)$ дорівнює

$$P(\alpha \le X < \beta) = F(\beta) - F(\alpha).$$

Зауваження. Якщо ВВ неперервна, то $P(X = \alpha) = 0$. Тоді

$$P(\alpha \le X < \beta) = P(\alpha < X < \beta).$$

Отже, для неперервних ВВ останню властивість можна переписати і так:

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha).$$

ЛЕКЦІЯ № 4

ЩІЛЬНІСТЬ РОЗПОДІЛУ ЙМОВІРНОСТЕЙ. ОСНОВНІ ЧИСЛОВІ ХАРАКТЕРИСТИКИ ТА ЗАКОНИ РОЗПОДІЛУ ВИПАДКОВИХ ВЕЛИЧИН

План

1. Щільність розподілу ймовірностей ВВ. Властивості.
2. Характеристики ВВ. Математичне сподівання ВВ. Властивості.
3. Дисперсія ВВ. Властивості. Середнє квадратичне відхилення.
4. Нормальний закон розподілу ВВ
5. Показниковий закон розподілу ВВ.
6. Закони розподілу дискретних ВВ: а) біноміальний; б) закон Пуассона. Характеристики ВВ, які мають ці закони розподілу.

Щільність розподілу ймовірностей

Нехай $F(x)$ – неперервна. Розглянемо $F(x) = P(X < x)$,

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

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

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

<u>Означення</u>. Похідна від функції розподілу називається щільністю розподілу ймовірностей і позначається

$$F'(x) = f(x).$$

Ймовірність того, що ВВ *X* приймає значення на інтервалі дорівнює

$$P(x \le X < x + \Delta x) = f(x)\Delta x.$$

Тоді ймовірність того, що ВВ *X* приймає значення на інтервалі $[\alpha, \beta]$ дорівнює

$$P(\alpha \le X \le \beta) = \int_{\alpha}^{\beta} f(x)dx.$$

Якщо задана $f(x)$, то $F(x)$ дорівнює

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

Очевидні властивості щільності розподілу ймовірностей:

1. $f(x) \ge 0$.
2. $\int_{-\infty}^{+\infty} f(x)dx = P(-\infty < X < +\infty) = 1.$

Розглянемо випадок рівномірної щільності $f(x) = const$ на інтервалі $[\alpha, \beta]$.

Тоді

$$\int_{\alpha}^{\beta} C dx = 1, \qquad C(\beta-\alpha)=1, \qquad C=\frac{1}{\beta-\alpha},$$

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha}, x\in[\alpha,\beta],\\ 0, x\notin[\alpha,\beta].\end{cases}$$

Приклад. Випадкова величина X має щільність розподілу

$$f(x)=\begin{cases}\dfrac{1}{2}\cos x, -\dfrac{\pi}{2}\le x\le\dfrac{\pi}{2},\\ 0, |x|\ge\dfrac{\pi}{2},\end{cases}$$

а) знайти ймовірність того, що ВВ приймає значення на інтервалі $\left[0,\dfrac{\pi}{3}\right]$.

$$P(0\le X\le\frac{\pi}{3})=\int_0^{\frac{\pi}{3}}\frac{1}{2}\cos x dx=\frac{1}{2}\sin x\Big|_0^{\frac{\pi}{3}}=\frac{1}{2}\sin\frac{\pi}{3}=\frac{1}{2}\frac{\sqrt{3}}{2}=\frac{\sqrt{3}}{4};$$

б) знайти функцію розподілу заданої ВВ:

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

При $-\infty\le x\le-\dfrac{\pi}{2}$ $f(x)=0$

$$F(x)=\int_{-\infty}^{x} 0dx=0.$$

При $-\dfrac{\pi}{2}<x\le\dfrac{\pi}{2}$

$$F(x)=\int_{-\infty}^{-\frac{\pi}{2}} 0dx+\int_{-\frac{\pi}{2}}^{x}\frac{1}{2}\cos x dx=\frac{1}{2}\sin x\Big|_{-\frac{\pi}{2}}^{x}=\frac{1}{2}\left[\sin x-\sin(-\frac{\pi}{2})\right]=\frac{1}{2}(\sin x+1).$$

При $x>\dfrac{\pi}{2}$

$$F(x)=\int_{-\infty}^{-\frac{\pi}{2}} 0dx+\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{1}{2}\cos x dx+\int_{\frac{\pi}{2}}^{x} 0dx=0+\frac{1}{2}\sin x\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}+0=1.$$

Таким чином:

$$F(x)=\begin{cases} 0 \quad \textit{при} \quad x \le -\dfrac{\pi}{2}, \\ \dfrac{1}{2}(\sin x+1) \quad \textit{при} \quad -\dfrac{\pi}{2} < x \le \dfrac{\pi}{2}, \\ 1 \quad \textit{при} \quad x > \dfrac{\pi}{2}. \end{cases}$$

Числові характеристики випадкових величин

Головними числовими характеристиками ВВ являються: математичне сподівання, дисперсія, середнє квадратичне відхилення тощо.

Математичне сподівання

Означення. Математичним сподіванням ВВ *X* називається середнє очікуване значення ВВ. Математичне сподівання визначається за такими формулами:

– для дискретних ВВ

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

– для неперервних ВВ:

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

де x_i, p_i – різні значення і їх ймовірності для дискретних ВВ, i=1,2,…,n,

$f(x)$ – щільність розподілу ймовірностей неперервної ВВ.

Приклад. Дискретна ВВ задана таким рядом розподілу

x_i	2	3	4	5
p_i	0,2	0,6	0,1	0,1

Знайти її математичне сподівання

$$M(X)=2\cdot 0{,}2+3\cdot 0{,}6+4\cdot 0{,}1+5\cdot 0{,}1=3{,}1.$$

Приклад. Знайти математичне сподівання ВВ, заданої такою щільністю розподілу ймовірностей

$$f(x)=\begin{cases} \dfrac{1}{2}\cos x, \quad x\in\left[-\dfrac{\pi}{2},\dfrac{\pi}{2}\right], \\ 0 \quad \textit{при} \quad |x|>\dfrac{\pi}{2}. \end{cases}$$

$$M(X)=\int\limits_{-\infty}^{+\infty}xf(x)dx=\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}}x\frac{1}{2}\cos x dx=\begin{bmatrix} u=\frac{x}{2}, & du=\frac{dx}{2} \\ dv=\cos x dx, & v=\sin x \end{bmatrix}=$$

$$=\frac{x}{2}\sin x\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}-\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\sin x dx=\frac{1}{2}\left[\frac{\pi}{2}\cdot 1-\left(-\frac{\pi}{2}\right)\cdot(-1)\right]-\frac{1}{2}(-\cos x)\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}=0.$$

Властивості математичного сподівання

1. $M(C)=C\cdot 1$.

Можна вважати, що константа являється ВВ, яка приймає лише одне значення з ймовірністю p=1.

2. $M(kX)=kM(X)$.

3. Математичне сподівання суми певного числа ВВ дорівнює сумі математичних сподівань цих ВВ.

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

4. Математичне сподівання добутку незалежних ВВ дорівнює добутку математичних сподівань цих ВВ.

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

Означення. ВВ X та $У$ називаються незалежними, якщо події, які полягають у тому, що $X=x_i$ і $Y=y_j$ являються незалежними.

5. $M(X+C)=M(X)+C$.

Дисперсія та середнє квадратичне відхилення

Означення. Дисперсією ВВ X називається математичне сподівання квадрата відхилення ВВ від її математичного сподівання

$$D(X)=M[X-M(X)]^2.$$

Дисперсія ВВ характеризує розсіювання значень ВВ відносно центра групування.

Дисперсія ВВ X визначається за формулами:

– для дискретних ВВ

$$D(X)=\sum_i (x_i-\overline{X})^2 p_i ,$$

де $\overline{X}=M(X)$;

– для неперервних ВВ:

$$D(X)=\int\limits_{-\infty}^{\infty}(x-\overline{X})^2 f(x)dx.$$

Приклади.

1. Обчислити дисперсію ВВ, задану рядом розподілу:

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

$M(X)=\overline{X}=(-3)\cdot 0{,}1+(-2)\cdot 0{,}2+0\cdot 0{,}4+2\cdot 0{,}2+3\cdot 0{,}1=0$

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

$+(3-0)^2\cdot 0{,}1=0{,}9+0{,}8+0{,}8+0{,}9=3{,}4$

2. Обчислити дисперсію ВВ, яка характеризується рівномірною щільністю на інтервалі $[a,b]$. Щільність розподілу ймовірності цієї ВВ дорівнює $f(x)=\dfrac{1}{b-a}$, математичне сподівання

$$\overline{X}=\int\limits_a^b\frac{1}{b-a}xdx=\frac{1}{b-a}\frac{x^2}{2}\bigg|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{a+b}{2}.$$

$$D(X)=\int\limits_a^b\left(x-\frac{a+b}{2}\right)^2\frac{1}{b-a}dx=\frac{1}{b-a}\frac{\left(x-\dfrac{a+b}{2}\right)^3}{3}\Bigg|_a^b=$$

$$=\frac{1}{3(b-a)}\left[\left(\frac{b-a}{2}\right)^3-\left(\frac{a-b}{2}\right)^3\right]=\frac{(b-a)^2}{12}.$$

Властивості дисперсії:

1. $D(C)=(C-C)^2\cdot 1=0.$
2. $D(kX)=k^2D(X).$
3. $D(X+Y)=D(X)+D(Y)$, якщо X, Y – незалежні ВВ.
4. $D(X)=M(X^2)-M^2(X).$
5. $D(X-Y)=D(X)+D(-Y)=D(X)+(-1)^2D(Y)=D(X)+D(Y)$,
   де X, $У$ – незалежні ВВ.
6. $D(X+C)=D(X)+D(C)=D(X).$

Середнім квадратичним відхиленням називається величина $\sigma(X)=\sqrt{D(X)}$.

Закони розподілу дискретних ВВ

Біноміальний закон

Нехай p – ймовірність появи деякої події A в будь-яких з n незалежних випробувань. Тоді число появ події A в n випробуваннях є випадкова величина X, яка приймає цілі значення 0,1,2, ... ,n з ймовірностями

$$P(X=m)=C_n^m p^m q^{n-m},$$

де $q=1-p$, $m=0,I,2,...,n$.

ВВ X, що розглядається, має біноміальний закон розподілу з характеристиками

$$M(X)=np, \;\; D(X)=npq, \; \sigma(X)=\sqrt{npq}.$$

Закон Пуассона

Якщо ВВ X приймає тільки цілі значення $m=0,1,2,...,n$, і ймовірність її визначається за формулою

$$P(X=m)=\frac{\lambda^m}{m!}e^{-\lambda},$$

де λ – параметр, то кажуть, що ця ВВ характеризується пуассонівським законом розподілу ймовірностей.

Можна довести, що для ВВ, що розглядається, числові характеристики відповідно дорівнюють:

$$M(X)=\lambda, \;\; D(X)=\lambda.$$

Приклад. Середня кількість звертань до АТС 120/год. Знайти ймовірність того, що протягом хвилини буде 4 звернення (виклики).

$$\lambda=\overline{X}=\frac{120}{60}=2 \;\; \text{за хв.}$$

$$P(X=4)=\frac{2^4}{4!}e^{-2}.$$

Середнім квадратичним відхиленням ВВ називається величина

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

Величина $\sigma(X)$ має розмірність ВВ X.

Існує багато ВВ, для яких досить знати розглянуті числові характеристики, щоб розв'язувати задачі теорії ймовірностей.

Нормальний закон розподілу неперервних ВВ

Означення. Неперервна ВВ X має нормальний закон розподілу, якщо щільність розподілу ймовірностей дорівнює

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

де a, σ – параметри.

Функція $f(x)$ при різних значеннях параметрів a та σ зображені на рис. 4.1 та рис 4.2.

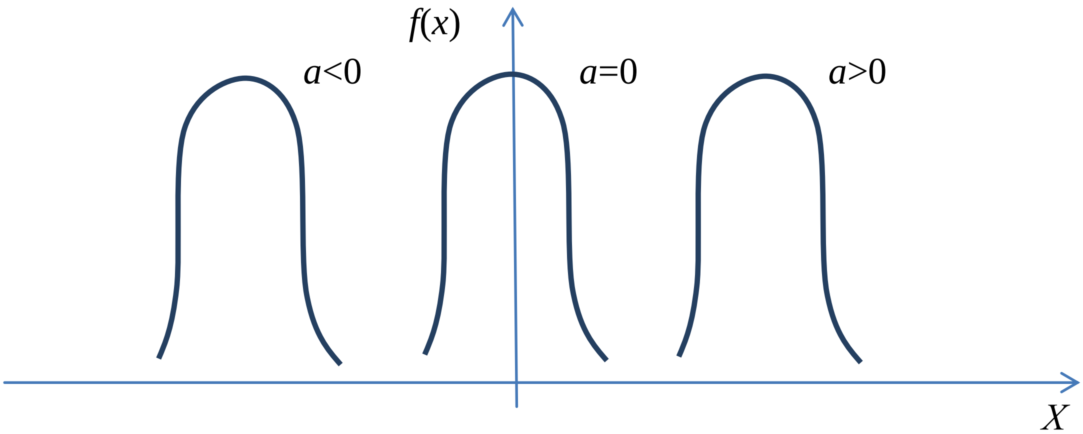


Рис. 4.1

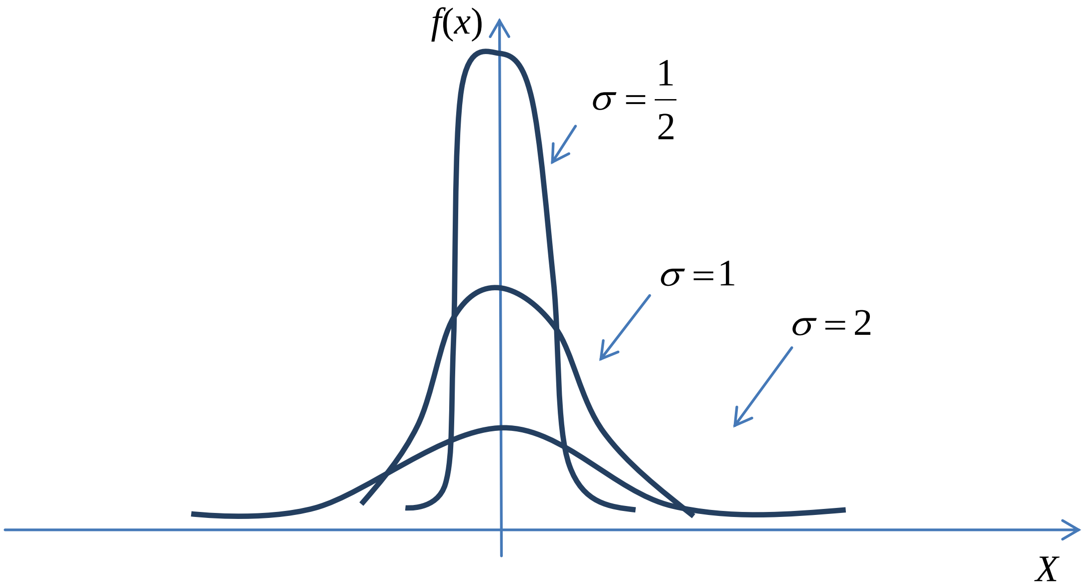


Рис. 4.2

Можна довести, що математичне сподівання та дисперсія ВВ, яка має нормальний закон розподілу, відповідно дорівнюють параметрам a та σ^2 цього закону. Тобто

$$M(X) = \int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = a,$$

$$D(X) = \int_{-\infty}^{+\infty} (x-a)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \sigma^2.$$

Визначимо ймовірність того, що ВВ X, яка має нормальний закон розподілу, приймає значення на інтервалі $[x_1, x_2]$.

$$P(x_1 \le X \le x_2) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left[\begin{array}{ll} \frac{x-a}{\sigma} = t, & t_1 = \frac{x_1 - a}{\sigma} \\ dt = \frac{dx}{\sigma}, & t_2 = \frac{x_2 - a}{\sigma} \end{array} \right] =$$

$$= \int_{t_1}^{t_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{0} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{0}^{t_2} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{0}^{t_2} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_{0}^{t_1} e^{-\frac{t^2}{2}} dt =$$

$$= \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right),$$

де $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt$ – функція Лапласа, значення якої обчислюються за допомогою таблиці. Ця функція має вигляд, зображений на рис. 4.3

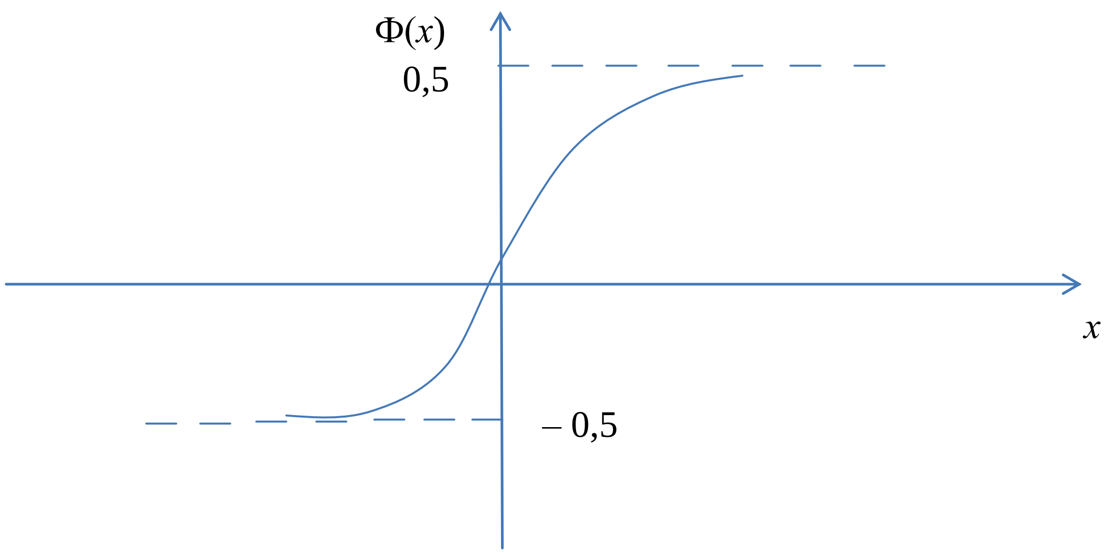


Рис. 4.3

Зауважимо, що Ф(x) – непарна функція, тобто Ф($-x$)= $-$Ф(x). Розглянемо частинний випадок:

$$P([X-a]\le\varepsilon)=P(a-\varepsilon\le X\le a+\varepsilon)=\Phi\left(\frac{a+\varepsilon-a}{\sigma}\right)-\Phi\left(\frac{a-\varepsilon-a}{\sigma}\right)=2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Зокрема, при $\varepsilon=3\sigma$.

$$P([X-a]<3\sigma)=2\Phi\left(\frac{3\sigma}{\sigma}\right).=0{,}9973.$$

Це, так зване, правило «трьох сигм»: ймовірність того, що ВВ приймає значення, які відрізняються від її математичного сподівання по модулю менше, ніж на 3σ, дорівнює 0,9973.

<u>Приклад</u>. Маса виходу кулінарії (котлет) – випадкова величина, яка має нормальний закон розподілу з такими параметрами: a=1000 г. (середня маса 10 котлет), $\sigma=50$ г. Знайти ймовірність того, що маса виходу наступної партії котлет буде в діапазоні від 900 до 1150 г.

$$P(900<X<1150)=\left[\Phi\left(\frac{1150-1000}{50}\right)-\Phi\left(\frac{900-1000}{50}\right)\right]=$$
$$=[\Phi(3)+\Phi(2)]=0{,}499+0{,}477=0{,}976.$$

Показниковий закон розподілу неперервних ВВ

<u>Означення</u>. Неперервна ВВ має показниковий закон розподілу, якщо її щільність розподілу має вигляд

$$f(x)=\mu\cdot e^{-\mu x},$$

де μ – параметр.

Можна довести, що для таких ВВ

$$M(X)=\sigma(X)=\frac{1}{\mu}.$$

<u>Приклад</u>. Тривалість обслуговування в універсамі – ВВ, яка має показниковий закон розподілу.

Елементи теорії кореляції

ЛЕКЦІЯ № 5

ФУНКЦІОНАЛЬНА ТА СТАТИСТИЧНА ЗАЛЕЖНОСТІ. КОРЕЛЯЦІЙНА ТАБЛИЦЯ

Функціональна залежність полягає в тому, що кожному значенню однієї змінної x відповідає деяке значення іншої змінної y. Такі залежності зустрічаються часто, але не завжди.

В економіці, біології і багатьох інших науках зустрічаються інші залежності, які полягають в тому, що кожному значенню однієї змінної x відповідає багато значень іншої змінної.

Означення. Якщо одному значенню змінної x відповідає множина значень змінної y, причому вказана множина значень не залишається незмінною, то кажуть, що між змінними x і y існує *статистична залежність*.

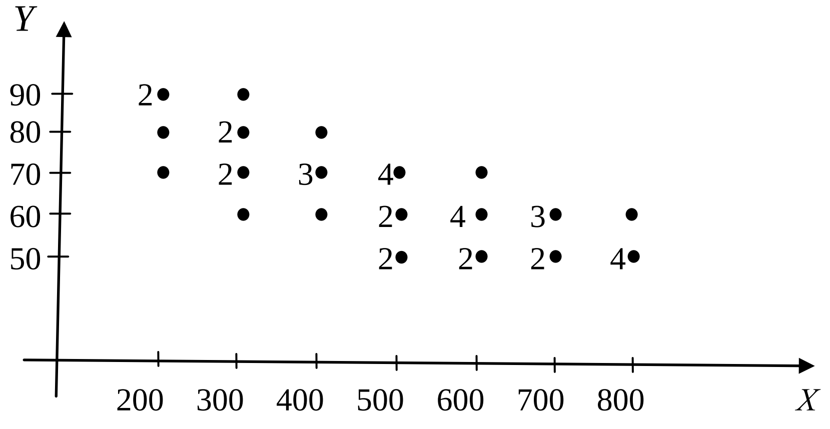
Інакше. Якщо кожному значенню змінної X відповідає розподіл змінної Y, то кажуть, що між x і y існує статистична залежність.

Приклад. Між об'ємом продукції X та собівартістю Y існує статистична залежність. Результат перевірки 40 однотипних підприємств зведено в таблицю.

Y \ X	200	300	400	500	600	700	800	кількість підприємств
50				2	2	2	4	10
60		1	1	2	4	3	1	12
70	1	2	3	4	1			11
80	1	2	1					4
90	2	1						3
кількість підприємств	4	6	5	8	7	5	5	40

X – об'єм продукції, Y – собівартість.

Одержана таблиця називається *кореляційною*. Для наочності дані таблиці зобразимо на графіку, який називається *полем кореляції*.



Аналіз таблиці показує, що зі збільшенням об’єму продукції собівартість в середньому зменшується.

У загальному випадку кореляційна таблиця має вид

$Y \backslash X$	x_1	x_2	…	…	x_i	…	…	x_s	n_k
y_1	m_{11}	m_{12}	…	…	m_{1i}	…	…	m_{1s}	n_1
y_2	m_{21}	m_{22}	…	…	m_{2i}	…	…	m_{2s}	n_2
…	…	…	…	…	…	…	…	…	…
y_k	m_{k1}	m_{k2}	…	…	m_{ki}	…	…	m_{ks}	n_k
…	…	…	…	…	…	…	…	…	…
y_l	m_{l1}	m_{l2}	…	…	m_{li}	…	…	m_{ls}	n_l
m_i	m_1	m_2	…	…	m_i	…	…	m_s	n

$$m_{k1} + m_{k2} + \ldots + m_{ks} = \sum_{i=1}^{s} m_{ki} = n_k, \; k = 1,2,\ldots,l,$$

$$m_{1i} + m_{2i} + \ldots + m_{li} = \sum_{k=1}^{l} m_{ki} = m_i, \; i = 1,2,\ldots,s.$$

Умовні розподілення. Характеристики умовних розподілень

Умовним розподіленням ознаки X, що відповідає значенню ознаки Y рівному $Y = y_k$, називається розподілення

X	x_1	x_2	x_3	…	…	x_s
m	m_{k1}	m_{k2}	m_{k3}	…	…	m_{ks}

Безумовним розподіленням ознаки X називається розподілення

X	x_1	x_2	x_3	…	…	x_s
m_i	m_1	m_2	m_3	…	…	m_s

Аналогічно умовним розподіленням ознаки Y, що відповідає значенню ознаки X=x_i, називається розподілення

Y	y_1	y_2	y_3	…	…	y_l
m	m_{1i}	m_{2i}	m_{3i}	…	…	m_{li}

і безумовним розподіленням ознаки Y називається розподілення

Y	y_1	y_2	y_3	…	…	y_l
n_k	n_1	n_2	n_3	…	…	n_l

Для кожного умовного розподілення ознаки X або *Y* можуть бути визначені числові характеристики: умовна середня та умовна дисперсія.

$$\overline{X}_k = \frac{1}{n_k}\sum_{i=1}^{s} m_{ki} x_i \tag{1}$$

– умовна середня, що відповідає $Y = y_k$,

$$D_k(X) = \frac{1}{n_k}\sum_{i=1}^{s} m_{ki}(x_i - \overline{X}_k)^2, \tag{2}$$

– умовна дисперсія.

Аналогічно для ознаки *Y*

$$\overline{Y_i} = \frac{1}{m_i}\sum_{k=1}^{l} m_{ki} y_k \tag{3}$$

– умовна середня,

$$D_i(Y) = \frac{1}{m_i}\sum_{k=1}^{l} m_{ki}(y_k - \overline{Y_i})^2, \tag{4}$$

– умовна дисперсія.

Часто зручніше обчислювати умовні середні та умовні дисперсії за формулами, які одержують із (1)-(4) введенням змінних

$$x_i' = \frac{x_i - C_1}{h_i} \text{ та } y_k' = \frac{y_k - C_2}{h_2}. \tag{5}$$

В результаті одержимо

$$\overline{X}_k = \frac{h_1}{n_k}\sum_{i=1}^{s} m_{ki} x_i' + \frac{1}{n_k} C_1 \sum_{i=1}^{s} m_{ki} = \frac{h_1}{n_k}\sum_{i=1}^{s} m_{ki} x_i' + C_1, \tag{6}$$

$$\overline{Y_i} = \frac{h_2}{m_i}\sum_{k=1}^{l} m_{ki} y_k' + C_2, \tag{7}$$

$$D_k(X) = \frac{1}{n_k}\sum_{i=1}^{s} m_{ki}(x_i - \overline{X}_k)^2 = \frac{1}{n_k}\sum_{i=1}^{s} m_{ki}(h_i x_i' + C_1 - \overline{X}_k)^2 =$$

$$=\underbrace{\frac{h_1^2}{n_k}\sum_{i=1}^{s} m_{ki}(x_i')^2+(C_1-\overline{X}_k)^2}_{A}+\underbrace{\frac{1}{n_k}\sum_{i=1}^{s} m_{ki}\,2h_i x_i'(C_1-\overline{X}_k)^2}_{B},$$

$$B=-2(\overline{X}_k-C_1)\frac{h_1}{n_k}\sum_{i=1}^{s} m_{ki}x_i'.$$

Оскільки з (6) $\sum_{i=1}^{s} m_{ki}x_i'=\frac{n_k}{h_1}\sum_{i=1}^{s} m_{ki}(\overline{X}-C_1)$, то $B=-2(\overline{X}_k-C_1)^2$

$$D_k(X)=\frac{h_1^2}{n_k}\sum_{i=1}^{s} m_{ki}(x_i')^2-(\overline{X}_k-C_1)^2. \qquad (8)$$

Аналогічно

$$D_i(Y)=\frac{h_2^2}{m_i}\sum_{k=1}^{l} m_{ki}(y_k')^2-(\overline{Y}_i-C_2)^2. \qquad (9)$$

Зауважимо, що величини C_1 та C_2 частіше за все обирають рівними варіантам відповідних ознак X та Y з найбільшими частотами, а величини h_1 та h_2 – величинам кроків відповідно ознак X та Y.

Використання формул (5)-(9) дозволяє спростити обчислення за рахунок здійснення арифметичних операцій з меншими числами.

Визначимо характеристики безумовних розподілень

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{s} m_i x_i, \qquad \overline{Y}=\frac{1}{n}\sum_{k=1}^{l} n_k y_k,$$

$$D(X)=\frac{1}{n}\sum_{i=1}^{s} m_i(x_i-\overline{X})^2, \qquad D(Y)=\frac{1}{n}\sum_{k=1}^{l} n_k(y_k-\overline{Y})^2.$$

Точка з координатами $\overline{X}$, $\overline{Y}$ називається *центром розподілення*.

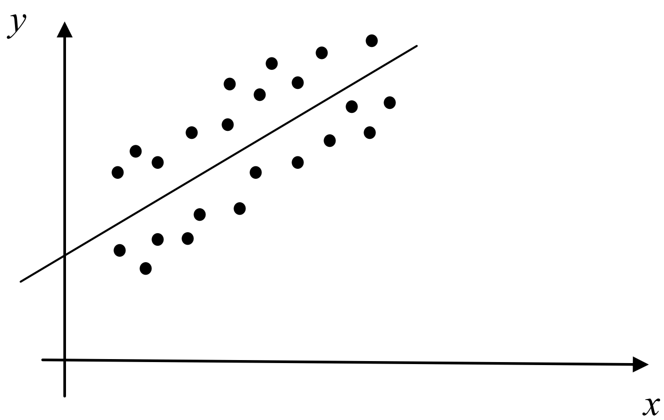
ЛЕКЦІЯ № 6

КОРЕЛЯЦІЙНА ЗАЛЕЖНІСТЬ.
ЕМПІРИЧНА ТА ТЕОРЕТИЧНА ЛІНІЇ РЕГРЕСІЇ

Із сказаного вище випливає, що для кожного значення ознаки $Y = y_k$ можна визначити умовну середню $\overline{X}_k$. Дослідимо зміну умовної середньої $\overline{X}_k$ при зміні Y (або умовної середньої $\overline{Y}_i$ при зміні X).

<u>Означення</u>. Статистична залежність, при якій умовна середня однієї ознаки змінюється при зміні другої ознаки, тобто не залишається сталою, називається *кореляційною залежністю*. Кажуть: змінна Y корелює з X і навпаки.

У загальному випадку при зміні ознаки X змінюється умовна середня $\overline{Y}_x$. Записують $\overline{Y}_x = f(x)$ – *кореляційне рівняння* або теоретична регресія Y на X, а графік називається *теоретичною лінією регресії*.



У цьому випадку ознаку X називають *факторною*, ознаку Y – *результативною*.

Емпірична лінія регресії задається таблицею

X	x_1	x_2	x_3	…	…	x_s
$\overline{Y}_i$	$\overline{Y}_1$	$\overline{Y}_2$	$\overline{Y}_3$	…	…	$\overline{Y}_s$
m_i	m_1	m_2	m_3	…	…	m_s

Теоретична лінія регресії невідома. Вона може бути одержана із емпіричної при необмеженому збільшенні кількості дослідів (спостережень, вимірювань).

Зауважимо, що оскільки ознаки X та Y рівноправні, то можна розглядати регресію X на Y

$$\overline{X}_y = \varphi(y).$$

Емпірична регресія задається таблицею

Y	y_1	y_2	y_3	…	…	y_l
$\overline{X}_k$	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	…	…	$\overline{X}_l$
n_k	n_1	n_2	n_3	…	…	n_l

Задачі математичної статистики полягають:

1) у встановленні форми зв'язку, тобто у визначенні рівняння регресії;
2) у встановленні тісноти і сили зв'язку.

Приступимо до розгляду першої задачі.

Найчастіше зустрічається лінійний кореляційний зв'язок. Це пояснюється наступним: якщо ознаки X та Y є суми великої кількості незалежних випадкових величин, то між ними існує лінійний кореляційний зв'язок.

Крім того, будь-який нелінійний кореляційний зв'язок принаймні на невеликому інтервалі можна представити як лінійний.

Як здійснюється вибір рівняння регресії? Із аналізу кореляційного зв'язку задаються деякою формою зв'язку, тобто припускають, що $\overline{Y}_x = ax + b$ або $\overline{Y}_x = ax^2 + bx + c$, або $\overline{Y}_x = ax^3 + bx^2 + cx + d$, або $\overline{Y}_x = \dfrac{a}{x} + b$ тощо.

Потім параметри того чи іншого з цих рівнянь визначають із умови: щоб сума квадратів відхилень $\varepsilon_i = \overline{Y}_x(x_i) - \overline{Y}_i$ була мінімальною, де $\overline{Y}_x = f(x_i)$. Вказаний спосіб називається *способом найменших квадратів*.

Отже, в загальному випадку складається функція

$$U = \sum_{i=1}^{s} m_i \left(f(x_i) - \overline{Y}_i \right)^2 .$$

Коефіцієнти функції $f(x)$ визначаються з умови мінімума функції U.

Детально розглянемо випадок прямої (прямолінійної) регресії, тобто коли $f(x) = ax + b$.

Пряма регресія

$f(x) = ax + b$, a-?, b-?,

$$\overline{Y}_x = ax + b \tag{1}$$

$\varepsilon_i = \overline{Y}_x(x_i) - \overline{Y}_i$, $\overline{Y}_i = \dfrac{1}{m_i} \sum_{k=1}^{l} m_{ki} y_k$,

$$U = \sum_{i=1}^{s} m_i \varepsilon_i^2 = \sum_{i=1}^{s} m_i \left(a x_i + b - \frac{1}{m_i} \sum_{k=1}^{l} m_{ki} y_k \right)^2 .$$

Запишемо необхідні умови існування екстремума диференційованої функції двох змінних.

$$\begin{cases} \dfrac{\partial U}{\partial a} = 0, & 2\sum\limits_{i=1}^{s} m_i (ax_i + b - \overline{\mathrm{Y}}_i) x_i = 0, \\ \dfrac{\partial U}{\partial b} = 0, & 2\sum\limits_{i=1}^{s} m_i (ax_i + b - \overline{\mathrm{Y}}_i) = 0 \end{cases}$$

або

$$\begin{cases} a\sum\limits_{i=1}^{s} m_i x_i + b\sum\limits_{i=1}^{s} m_i = \sum\limits_{i=1}^{s} m_i \overline{\mathrm{Y}}_i, \\ a\sum\limits_{i=1}^{s} m_i x_i^2 + b\sum\limits_{i=1}^{s} m_i x_i = \sum\limits_{i=1}^{s} m_i \overline{\mathrm{Y}}_{\mathrm{i}} x_i, \end{cases}$$

або

$$\begin{cases} a\overline{\mathrm{X}} + b = \overline{\mathrm{Y}}, \\ a\sum\limits_{i=1}^{s} m_i x_i^2 + bn\overline{\mathrm{X}} = \sum\limits_{i=1}^{s} m_i \overline{\mathrm{Y}}_i x_i. \end{cases}$$

Перше рівняння цієї системи означає, що центр розподілу лежить на прямій лінії регресії.

$$b = \overline{\mathrm{Y}} - a\overline{\mathrm{X}}, \tag{2}$$

$$a\sum_{i=1}^{s} m_i x_i^2 + n\overline{\mathrm{X}}(\overline{\mathrm{Y}} - a\overline{\mathrm{X}}) = \sum_{i=1}^{s} m_i \overline{\mathrm{Y}}_i x_i,$$

$$a = \rho_{y/x} = \frac{\sum\limits_{i=1}^{s} m_i \overline{\mathrm{Y}}_i x_i - n\overline{\mathrm{X}}\overline{\mathrm{Y}}}{\sum\limits_{i=1}^{s} m_i x_i^2 - n(\overline{\mathrm{X}})^2}, \tag{3}$$

$a = \rho_{y/x}$ – *коефіцієнт регресії у* на x.

Із врахуванням (2) та (1) одержимо

$$\overline{\mathrm{Y}}_x = \rho_{y/x} x + \overline{\mathrm{Y}} - \rho_{y/x}\overline{\mathrm{X}}$$

або

$$\overline{\mathrm{Y}}_x - \overline{\mathrm{Y}} = \rho_{y/x}(x - \overline{\mathrm{X}}). \tag{4}$$

Коефіцієнт регресії – є кутовий коефіцієнт прямої лінії регресії У на Х.

Аналогічно запишемо рівняння прямої лінії регресії Х на У.

$$\overline{\mathrm{X}}_y - \overline{\mathrm{X}} = \rho_{x/y}(y - \overline{\mathrm{У}}). \tag{5}$$

При аналізі кореляційної залежності розглядають обидві регресії У на Х та Х на У.

Часто замість розмірного коефіцієнта $\rho_{y/x}$ вводять безрозмірний (безвимірний) коефіцієнт кореляції

$$r = \rho_{y/x} \frac{\sigma(X)}{\sigma(Y)}. \qquad (6)$$

У цьому випадку рівняння лінії регресії прийме вид

$$\overline{\mathrm{Y}}_x - \overline{\mathrm{Y}} = r \frac{\sigma(\mathrm{У})}{\sigma(\mathrm{X})}(x - \overline{\mathrm{X}}). \qquad (7)$$

В чому сенс коефіцієнта r? Про це буде сказано нижче.

Отже, параметри прямолінійної регресії, що визначаються методом найменших квадратів, можуть бути обчислені за формулами (2) та (3).

Перепишемо (3) у більш компактній формі, враховуючи, що $\overline{\mathrm{Y}}_i = \frac{1}{m_i}\sum_{k=1}^{l} m_{ki} y_k$,

$$\sum_{i=1}^{s} m_i \overline{\mathrm{Y}}_i x_i = \sum_{i=1}^{s}(\sum_{k=1}^{l} m_{ki} y_k) x_i = \sum_{i=1}^{s}\sum_{k=1}^{l} m_{ki} x_i y_k \ ,$$

$$\rho_{y/x} = \frac{\sum_{i=1}^{s}\sum_{k=1}^{l} m_{ki} x_i y_k - n\overline{\mathrm{X}}\overline{\mathrm{Y}}}{\sum_{i=1}^{s} m_i x_i^2 - n(\overline{\mathrm{X}})^2}.$$

Введемо

$$\overline{\mathrm{XY}} = \frac{1}{n}\sum_{i=1}^{s}\sum_{k=1}^{l} m_{ki} x_i y_k , \qquad (8)$$

тобто середнє арифметичне добутків змінних x та y.

Одержимо

$$\rho_{y/x} = \frac{n\overline{\mathrm{XY}} - n\overline{\mathrm{X}}\overline{\mathrm{Y}}}{\sum_{i=1}^{s} m_i x^2{}_i - n(\overline{\mathrm{X}})^2} = \frac{\overline{\mathrm{XY}} - \overline{\mathrm{X}}\overline{\mathrm{Y}}}{\sigma^2(\mathrm{X})}. \qquad (9)$$

Коефіцієнт кореляції

$$r = \rho_{y/x} \frac{\sigma(X)}{\sigma(Y)}. = \frac{\overline{\mathrm{XY}} - \overline{\mathrm{X}}\overline{\mathrm{Y}}}{\sigma(\mathrm{X})\sigma(\mathrm{У})}. \qquad (10)$$

Регресія У на X або рівняння кореляції має вид

$$\overline{\mathrm{Y}}_x - \overline{\mathrm{Y}} = r \frac{\sigma(\mathrm{У})}{\sigma(\mathrm{X})}(x - \overline{\mathrm{X}}), \qquad (11)$$

де r визначається за формулою (10).

ЛЕКЦІЯ 7

СПРОЩЕННЯ РОЗРАХУНКІВ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ

З метою спрощення розрахунків коефіцієнта кореляції будемо обчислювати не за формулою (10) (у лекції 2), а за аналогічною, яку одержимо після переходу до нових змінних

$$\mathrm{U}=\frac{x-C_1}{h_1},\ \mathrm{V}=\frac{y-C_2}{h_2}, \tag{12}$$

де величини C_1 та C_2 – варіанти ознак відповідно X та У з найбільшою частотою; h_1 та h_2 – величини кроків, з якими задані відповідно ознаки X та У.

Вказаний перехід, повторюємо, здійснимо з метою спрощення розрахунків.

Отже, з (12) маємо

$$x=C_1+h_1\mathrm{U},\ y=C_2+h_2\mathrm{V},$$

$$\overline{\mathrm{X}}=\frac{1}{n}\sum_{i=1}^{s}m_i(C_1+h_1u_i)=\frac{\cancel{n}C_1}{\cancel{n}}+h_1\overline{\mathrm{U}}=C_1+h_1\overline{\mathrm{U}}, \tag{13}$$

$$\overline{\mathrm{Y}}=\frac{1}{n}\sum_{k=1}^{l}n_k(C_2+h_2v_k)=C_2+h_2\overline{\mathrm{V}}, \tag{14}$$

$$\sigma^2(\mathrm{X})=\frac{1}{n}\sum_{i=1}^{s}m_i(x_i-\overline{\mathrm{X}})=\frac{1}{n}\sum_{i=1}^{s}m_i\left(\cancel{C}_1+h_1u_i-(\cancel{C}_1+h_1\overline{\mathrm{U}})\right)^2=$$

$$=h_1^2\frac{1}{n}\sum_{i=1}^{s}m_i\left(u_i-\overline{\mathrm{U}}\right)^2=h_1^2\sigma^2(\mathrm{U}),$$

$$\overline{\mathrm{U}}=\frac{1}{n}\sum_{i=1}^{s}m_iu_i,\ \overline{\mathrm{V}}=\frac{1}{n}\sum_{k=1}^{l}n_kv_k,$$

$$\sigma(\mathrm{X})=h_1\,\sigma(\mathrm{U}), \tag{15}$$

$$\sigma(\mathrm{Y})=h_2\sigma(\mathrm{V}), \tag{16}$$

$$\sigma^2(\mathrm{U})=\frac{1}{n}\sum_{i=1}^{s}m_iu_i^2-(\overline{\mathrm{U}})^2,\ \sigma^2(\mathrm{V})=\frac{1}{n}\sum_{k=1}^{l}n_kv_k^2-(\overline{\mathrm{V}})^2,$$

$$\overline{\mathrm{XY}}=\frac{1}{n}\sum_{i=1}^{s}\sum_{k=1}^{l}m_{ki}x_iy_k=\frac{1}{n}\sum_{i=1}^{s}\sum_{k=1}^{l}m_{ki}(C_1+h_1u_i)(C_2+h_2v_k)=$$

$$=C_1C_2+C_1h_2\frac{1}{n}\sum_{i=1}^{s}\sum_{k=1}^{l}m_{ki}v_k+C_2h_1\frac{1}{n}\sum_{i=1}^{s}\sum_{k=1}^{l}m_{ki}u_i+h_1h_2\frac{1}{n}\sum_{i=1}^{s}\sum_{k=1}^{l}m_{ki}u_iv_k,$$

$$\overline{\mathrm{V}}=\frac{1}{n}\sum_{k=1}^{l}n_kv_k=\frac{1}{n}\sum_{k=1}^{l}n_k\left(\sum_{i=1}^{s}m_{ki}\right)v_k, \tag{17}$$

$$\overline{\mathrm{XY}} = C_1C_2 + C_1h_2\overline{\mathrm{V}} + C_2h_1\overline{\mathrm{U}} + h_1h_2\overline{\mathrm{UV}}, \tag{18}$$

Підставимо (13)-(18) в (10). Одержимо

$$r = \frac{C_1C_2 + C_1h_2\overline{\mathrm{V}} + C_2h_1\overline{\mathrm{U}} + h_1h_2\overline{\mathrm{UV}} - (C_1 + h_1\overline{\mathrm{U}})(C_2 + h_2\overline{\mathrm{V}}))}{h_1\sigma(\mathrm{U})h_2\sigma(\mathrm{V})},$$

$$r = \frac{\overline{\mathrm{UV}} - \overline{\mathrm{U}}\cdot\overline{\mathrm{V}}}{\sigma(\mathrm{U})\sigma(\mathrm{V})} \tag{19}$$

або

$$r = \frac{\sum\limits_{i=1}^{s}\sum\limits_{k=1}^{l} m_{ki}u_iv_k - n\overline{\mathrm{UV}}}{n\sigma(\mathrm{U})\sigma(\mathrm{V})}. \tag{20}$$

Приклад обчислення коефіцієнта кореляції здійснимо з використанням наступної таблиці.

	U	–2	–1	0	1	2			
V	X / Y	25	35	45	55	65	n_k	n_kv_k	$n_kv_k^2$
–4	0	2					2	–8	32
–3	1	1	5				6	–18	54
–2	2		3	4			7	–14	28
–1	3			2			2	–2	2
0	4			**4**	2		6	0	0
1	5				3		3	3	3
2	6				1	6	7	14	28
3	7				5	2	7	21	63
	m_i	3	8	10	11	8	40	**–4**	**210**
	m_iu_i	–6	–8	0	11	16	$\sum\limits_{i=1}^{5} m_iu_i = \mathbf{13}$		
	$m_iu_i^2$	12	8	0	11	32	$\sum\limits_{i=1}^{5} m_iu_i^2 = \mathbf{63}$		
	$\sum\limits_{k=1}^{l} m_{ki}v_k$	–11	–21	–10	20	18			
	$u_i\sum\limits_{k=1}^{l} m_{ki}v_k$	22	21	0	20	36	$\sum\limits_{i=1}^{5}\sum\limits_{k=1}^{8} m_{ki}u_iv_k = \mathbf{99}$		

$$\mathrm{U}=\frac{x-45}{10},\ \mathrm{V}=\frac{y-4}{1},\ \overline{\mathrm{U}}=\frac{1}{n}\sum_{i=1}^{5}m_i u_i=\frac{13}{40}=0{,}325,\ \overline{\mathrm{V}}=\frac{1}{n}\sum_{k=1}^{8}n_k v_k=-\frac{4}{40}=-0{,}1,$$

$$\sigma^2(\mathrm{U})=\frac{1}{n}\sum_{i=1}^{5}m_i u_i^2-(\overline{\mathrm{U}})^2=\frac{63}{40}-(0{,}325)^2=1{,}46,\qquad \sigma(\mathrm{U})\approx 1{,}2,$$

$$\sigma^2(\mathrm{V})=\frac{1}{n}\sum_{k=1}^{8}n_k v_k^2-(\overline{\mathrm{V}})^2=\frac{210}{40}-(-0{,}1)^2=5{,}24,\qquad \sigma(\mathrm{V})\approx 2{,}3.$$

За формулою (20)

$$r=\frac{99-40\cdot 0{,}325\cdot(-0{,}1)}{40\cdot 1{,}2\cdot 2{,}3}\approx 0{,}9.$$

За формулами (15) та (16) визначимо

$$\sigma(\mathrm{X})=h_1\,\sigma(\mathrm{U})=10\cdot 1{,}2=12,$$

$$\sigma(\mathrm{Y})=h_2\sigma(\mathrm{V})=1\cdot 2{,}3=2{,}3.$$

За формулами (13) та (14) одержимо

$$\overline{\mathrm{X}}=C_1+h_1\overline{\mathrm{U}}=45+10\cdot 0{,}325=48{,}25,$$

$$\overline{\mathrm{Y}}=C_2+h_2\overline{\mathrm{V}}=4+1\cdot(-0{,}1)=3{,}9.$$

За формулою (11) маємо наступне рівняння лінійної кореляції

$$\overline{\mathrm{Y}}_x=3{,}9+0{,}9\cdot\frac{2{,}3}{12}(x-48{,}25).$$

ЛЕКЦІЯ 8

ОЦІНКА СИЛИ (ТІСНОТИ) КОРЕЛЯЦІЙНОЇ ЗАЛЕЖНОСТІ У ВИПАДКУ ПРЯМОЛІНІЙНОЇ РЕГРЕСІЇ. КОЕФІЦІЄНТ КОРЕЛЯЦІЇ

Розглянемо рівняння лінійної регресії У на Х

$$\overline{\mathrm{Y}}_x - \overline{\mathrm{Y}} = r\frac{\sigma(\mathrm{Y})}{\sigma(\mathrm{X})}(x - \overline{\mathrm{X}}). \qquad (1)$$

Коефіцієнт кореляції r в рівнянні (1) характеризує силу (тісноту) лінійної кореляційної залежності у випадку прямолінійної регресії.

Із (1) видно, що при r=0 маємо

$$\overline{\mathrm{Y}}_x = \overline{\mathrm{Y}} = const, \qquad (2)$$

тобто ознака У лінійно не корельована з ознакою Х.

Лінійна кореляційна залежність відсутня. Строго кажучи, тільки лінійна, оскільки могла ж бути

$$\overline{\mathrm{Y}}_x = ax + bx^2 + \ldots . \qquad (3)$$

При a=0 немає лінійної кореляційної залежності.

Можна довести, що коефіцієнт кореляції змінюється від –1 до +1.

Якщо r=0, то лінійної кореляційної залежності немає (рис. 8.1).

При r=1 (рис. 8.2)

$$\overline{\mathrm{Y}}_x - \overline{\mathrm{Y}} = \frac{\sigma(\mathrm{Y})}{\sigma(\mathrm{X})}(x - \overline{\mathrm{X}}), \qquad (4)$$

тобто між змінними x та $\overline{\mathrm{Y}}_{\mathrm{x}}$ існує функціональна лінійна залежність.

В тих випадках, коли $r \neq 0$ і $|r| \neq 1$ лінійна кореляційна залежність тим сильніша, чим більший коефіцієнт кореляції.

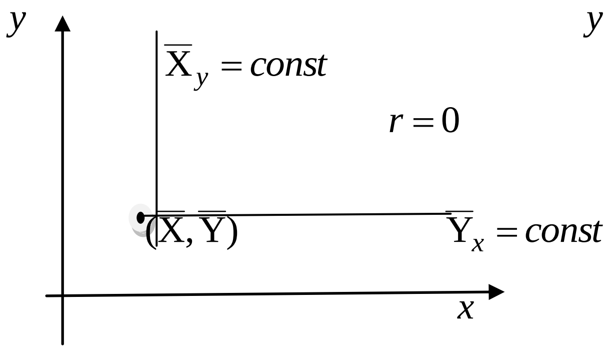


Рис. 8.1

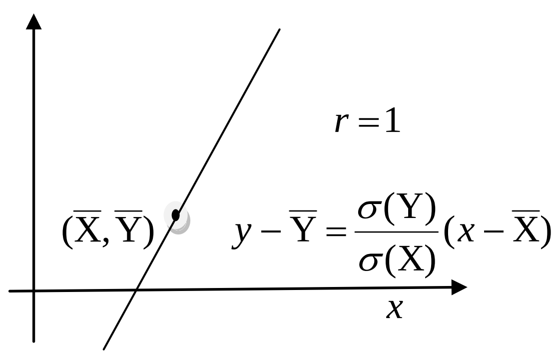


Рис. 8.2

За допомогою коефіцієнта кореляції можна характеризувати тісноту кореляційної залежності у тому випадку, коли емпірична лінія регресії майже пряма.

Точки (y_j, x_i) тим тісніше лежать навколо прямої регресії $\overline{\mathrm{Y}}_x$, чим ближче коефіцієнт кореляції до ± 1.

<u>Висновок</u>. Коефіцієнт r є мірою тісноти лінійного кореляційного зв’язку між X та У. Коли $r=0$, то X та Y не можуть знаходитись у лінійній кореляційній залежності. Ступінь їх кореляційної зв’язаності зростає при наближенні r до ± 1.

Причому лінійний зв'язок буде функціональним, коли коефіцієнт кореляції дорівнює ± 1.

Наведемо приклад достатньо сильної (тісної) лінійної кореляційної залежності між ознаками Y та X.

	U	–3	–2	–1	0	1	2	3			
V	X / Y	5	15	25	35	45	55	65	n_k	$n_k v_k$	$n_k v_k^2$
–2	4	2		2					4	–8	16
–1	8		1	4					5	–5	5
0	12		4	3	**10**				17	0	0
1	16		2		2	3	6		13	13	13
2	20					5	4		9	18	36
3	24						1	1	2	6	18
	m_i	2	7	9	12	8	11	1	50	**24**	**88**
	$m_i u_i$	–6	–14	–9	0	8	22	3	$\sum\limits_{i=1}^{7} m_i u_i = \mathbf{4}$		
	$m_i u_i^2$	18	28	9	0	8	44	9	$\sum\limits_{i=1}^{7} m_i u_i^2 = \mathbf{116}$		
	$\sum\limits_{k=1}^{l} m_{ki} v_k$	–4	1	–8	2	13	17	3			
	$u_i \sum\limits_{k=1}^{l} m_{ki} v_k$	12	–2	8	0	13	34	9	$\sum\limits_{i=1}^{7}\sum\limits_{k=1}^{6} m_{ki} u_i v_k = \mathbf{74}$		

$$\mathrm{U} = \frac{x-35}{10},\ \mathrm{V} = \frac{y-12}{4},\ \overline{\mathrm{U}} = \frac{1}{n}\sum_{i=1}^{7} m_i u_i = \frac{4}{50} = 0{,}08,\ \overline{\mathrm{V}} = \frac{1}{n}\sum_{k=1}^{6} n_k v_k = \frac{24}{50} = 0{,}48,$$

$$\sigma^2(\mathrm{U}) = \frac{1}{n}\sum_{i=1}^{7} m_i u_i^2 - (\overline{\mathrm{U}})^2 = \frac{116}{50} - (0{,}08)^2 = 2{,}14 - 0{,}0064 \approx 2{,}13, \quad \sigma(\mathrm{U}) \approx 1{,}52,$$

$$\sigma^2(\mathrm{V}) = \frac{1}{n}\sum_{k=1}^{6} n_k v_k^2 - (\overline{\mathrm{V}})^2 = \frac{88}{50} - (0{,}48)^2 = 1{,}76 - 0{,}25 \approx 1{,}51, \qquad \sigma(\mathrm{V}) \approx 1{,}24,$$

$$r = \frac{\overline{\mathrm{UV}} - n\overline{\mathrm{U}}\cdot\overline{\mathrm{V}}}{n\sigma(\mathrm{U})\sigma(\mathrm{V})} = \frac{74 - 50\cdot 0{,}08\cdot 0{,}48}{50\cdot 15{,}2\cdot 1{,}24} \approx 0{,}76.$$

$$\overline{\mathrm{X}} = C_1 + h_1\overline{\mathrm{U}} = 35 + 10\cdot 0{,}08 = 35{,}8,\ \overline{\mathrm{Y}} = C_2 + h_2\overline{\mathrm{V}} = 12 + 4\cdot 0{,}48 = 13{,}92,$$

$$\sigma(\mathrm{X}) = h_1\sigma(\mathrm{U}) = 10 \cdot 1{,}52 = 15{,}2,\ \ \sigma(\mathrm{Y}) = h_2\sigma(\mathrm{V}) = 4 \cdot 1{,}24 = 4{,}96\,,$$

$$\overline{\mathrm{Y}}_x = \overline{\mathrm{Y}} - r\frac{\sigma(\mathrm{Y})}{\sigma(\mathrm{X})}(x - \overline{\mathrm{X}}),$$

$$\overline{\mathrm{Y}}_x = 13{,}92 - 0{,}76 \cdot \frac{4{,}96}{15{,}2}(x - 35{,}8),$$

$$\overline{\mathrm{Y}}_x = 0{,}25x + 4{,}97.$$

ЛЕКЦІЯ 9

СТАТИСТИЧНА ПЕРЕВІРКА ГІПОТЕЗ

План

1. Статистична гіпотеза. Перевірка гіпотези про генеральну частку (ймовірність).
2. Ознаки висунення гіпотез про закон розподілу ВВ (випадкових величин).
3. Критерії перевірки гіпотез (узгодження). Рівень значущості, область допустимих значень, критична область, критична точка.
4. Критерій χ^2 Пірсона.
5. Перевірка потоку вимог (закону розподілу числа покупців) в універсамі.
6. Зауваження:
   а) про помилки першого і другого роду при виборі рівня значущості;
   б) про перевірку закону розподілу тривалості обслуговування покупців в універсамі.

Другим після статистичної оцінки параметрів розподілу, з точки зору важливості, є статистична перевірка гіпотез (припущень).

Статистична гіпотеза.
Перевірка гіпотези про генеральну частку (ймовірність)

Нехай проведено деякий експеримент. Деяка подія A відбулась m=151 раз з n=280 раз (дослідах, експериментах).

Позначимо X – випадкову величину – кількість появ події A в n незалежних випробуваннях. Ймовірність появи події A в одному випробуванні P(A)=p.

<u>Сформулюємо гіпотезу</u>: ймовірність появи події в одному випробуванні дорівнює $P(A)=p=\frac{1}{2}$.

<u>Питання.</u> Як перевірити: результати спостереження узгоджуються з гіпотезою про ймовірність $p=\frac{1}{2}$ чи ні?

Якщо не узгоджуються, то гіпотезу необхідно відкинути.

Можна довести, що характеристики ВВ X відповідно дорівнюють

$$\overline{X}=np,\quad \sigma^2(X)=npq,\quad \sigma(X)=\sqrt{npq}.$$

У нашому випадку

$$\overline{X}=280\cdot\frac{1}{2}=140,\ \sigma(X)=\sqrt{280\cdot\frac{1}{2}(1-\frac{1}{2})}=8{,}37.$$

Питання. Чи можна вважати частоту m=151 достатньо близькою до теоретичної 140, що відповідає гіпотезі?

Яке відхилення можна вважати допустимим?

Як знайти границю (межу) допустимих відхилень?

Як знайти критичне відхилення, перевищення якого малоймовірне настільки, що його можна вважати практично неможливим?

І якщо воно фактично спостерігається, то з цього випливає, що наша гіпотеза несумісна з результатами спостережень (генеральна частка, що мала місце в експерименті значимо, суттєво відрізняється від ймовірності p).

Якщо ж фактично відхилення менше критичного, то кажуть, що результати спостережень не протирічать гіпотезі (відхилення, що спостерігались пояснюються випадковістю випробувань).

Найчастіше використовуються наступні *рівні значущості*: 0,05, 0,01, 0,001.

В різних дослідах, сферах діяльності вони вибираються різними.

Можна довести, що відхилення $m–np$ характеризується законом розподілу, близьким до нормального.

$$P(|m-np|<\varepsilon)\approx 2\Phi\left(\frac{\varepsilon}{\sigma}\right),$$

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

Тоді при $\varepsilon = 1{,}96\sigma$ маємо

$$P(|m-np|<1{,}96\sigma)\approx 2\Phi\left(\frac{1{,}96\sigma}{\sigma}\right)=0{,}95,$$

$$P(|m-np|>1{,}96\sigma)\approx 2\Phi\left(\frac{1{,}96\sigma}{\sigma}\right)=0{,}05$$

Отже, якщо задамось 5 % рівнем значущості, то критичні точки (границі, відхилення) будуть складати

$np \pm 1{,}96\sigma$ або 140–$1{,}96\cdot 8{,}37$, 140+$1{,}96\cdot 8{,}37$

або 140–16,4; 140+16,4

або [123,6; 156,4].

Отже, значення 151, що спостерігалось, не протирічить гіпотезі про ймовірність $p=\frac{1}{2}$ або гіпотеза про ймовірність $p=\frac{1}{2}$ не протирічить результатам спостережень.

Означення. Статистична гіпотеза – це припущення про тип або властивості розподілу ВВ, які спостерігаються під час експерименту.

Гіпотези бувають різні. Одним з найважливіших класів гіпотез є гіпотези про закон розподілу (ВВ).

Ознаки висунення гіпотез про закон розподілу ВВ

Як висунути ту чи іншу гіпотезу? Як перевірити, що результати спостережень узгоджуються з гіпотезою чи навпаки?

Приведемо ознаки, які дозволяють формулювати гіпотези про закони розподілу ВВ; про ті закони, які найчастіше зустрічаються при дослідженні процесів торгівлі та харчування.

1. Якщо ВВ є сумою великого числа випадкових величин, причому дія (вплив) кожного доданку невелика в порівнянні з сумарною дією, то згідно граничних теорем теорії ймовірностей, сумарна ВВ має закон розподілу нормальний або близький до нього.

2. Якщо ВВ X приймає цілі значення, причому M(X) наближено дорівнює дисперсії D(X), то можна сформулювати гіпотезу про Пуассонівський закон розподілу цієї ВВ:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \lambda = \overline{\mathrm{X}}.$$

3. Якщо ВВ X є неперервною, причому M(X)$\approx \sigma(X)$, то можна сформулювати гіпотезу про показниковий закон розподілу

$$f(x) = \mu \cdot e^{-\mu x},\ \frac{1}{\mu} = \overline{\mathrm{X}} = \sigma(X).$$

Крім сформульованих ознак, перед висуненням гіпотези про закон розподілу неперервної ВВ доцільно проаналізувати гістограму відносної частоти. В торгівлі та харчуванні найчастіше зустрічаються ВВ, які мають саме ці закони розподілу: нормальний, Пуассонівський та показниковий.

Критерії перевірки статистичних гіпотез

Означення. Критерієм перевірки гіпотез (критерієм узгодження) називається правило, у відповідності з яким робиться висновок про те, що результати спостережень не протирічать, узгоджуються із сформульованою гіпотезою і гіпотеза залишається, або навпаки – результати спостережень протирічать гіпотезі і гіпотеза відхиляється.

Позначимо:

K – деякий критерій узгодження;

K_γ – значення критерію, яке відповідає рівню значущості γ (K_γ – критична точка).

Кількісне значення критерію K характеризує міру (є мірою) відхилення вибіркового розподілу від гіпотетичного (рис. 9.1).

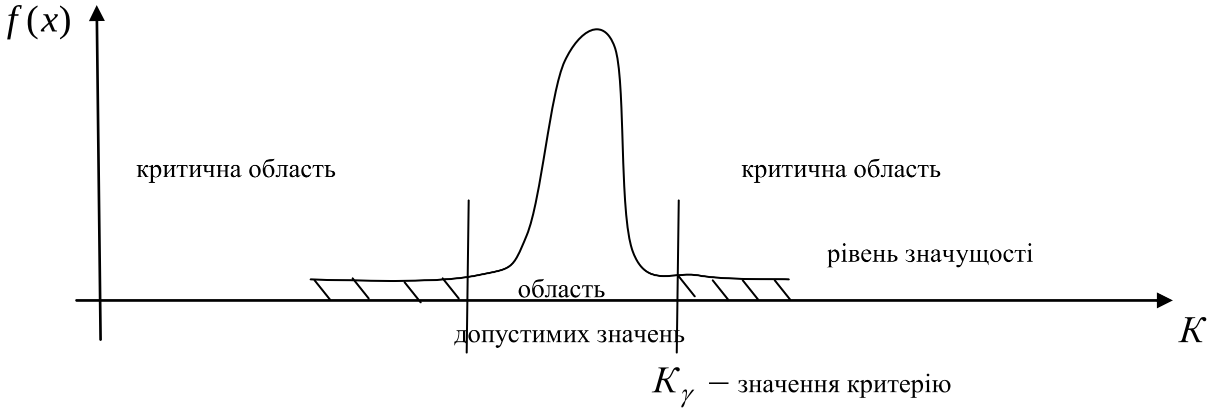


Рис. 9.1

Існують різні формули для обчислення такого відхилення (критерій Пірсона, Колмогорова тощо).

Тоді $P(K > K_\gamma) = \gamma$, $\gamma = 0{,}05$ або $\gamma = 0{,}01$ тощо.

Подія $K > K_\gamma$ – практично неможлива подія (див рис. 9.1).

Критерій χ^2 Пірсона

Це критерій, яким частіше за все користуються для перевірки статистичних гіпотез.

Порядок роботи:

1. В результаті аналізу даних спостережень формулюється гіпотеза про закон розподілу конкретної ВВ. Обчислюються параметри гіпотетичного закону розподілу.

2. Виходячи з гіпотетичного закону розподілу знаходяться теоретичні значення частот n_k^T.

3. Визначається розрахункове значення критерію Пірсона за формулою

$$\chi_p^2 = \sum_{k=1}^{l} \frac{(n_k^T - n_k^l)^2}{n_k^T},$$

де n_k^l – експериментальні частоти;

l – число варіант (кількість груп емпіричного розподілу).

4. Знаходиться число ступенів вільності за формулою $\nu = l - r$, де r – дорівнює сумі параметрів гіпотетичного закону, збільшеній на одиницю.

Зауважимо, що для нормального розподілу r=1+2 (два параметри: $\overline{X}, \sigma(X)$); для Пуассонівського закону r=1+1 (один параметр $\lambda = \overline{X}$); для показникового закону r=1+1 (один параметр $\mu = \dfrac{1}{\overline{X}}$).

5. Для рівня значущості γ та ν ступенів вільності знаходиться (за допомогою таблиць) критична точка χ^2_γ

$$P(\chi^2 > \chi^2_\gamma) = \gamma .$$

6. Робиться висновок у відповідності з критерієм і сформульована гіпотеза або приймається, або відхиляється.

Якщо $\chi^2_p < \chi^2_\gamma$, то робиться висновок: з ймовірністю $1-\gamma$ результати спостережень узгоджуються із сформульованою гіпотезою; гіпотеза приймається.

Якщо $\chi^2_p > \chi^2_\gamma$, то з ймовірністю $1-\gamma$ результати спостережень протирічать сформульованій гіпотезі; гіпотеза відхиляється.

Перевірка гіпотези про Пуассонівський закон розподілу кількості покупців, що потребують обслуговування в універсамі в одиницю часу

В таблиці наведено результати спостережень потока покупців в одному з магазинів. У першому стовпці надано кількість вимог (k), що поступили протягом хвилинних інтервалів, у другому стовпці – кількість хвилинних інтервалів, що містили k вимог.

Кількість вимог, що поступили протягом хвилинного інтервалу k	Кількість інтервалів з k вимогами n_k^l	Значення ймовірностей знайдені за формулою Пуассона P_k	Теоретичні значення частот $n_k^T = n \cdot P_k$	$\dfrac{(n_k^T - n_k^l)^2}{n_k^T}$
0	70	0,135335	68	0,059
1	130	0,270671	135	0,185
2	140	0,270671	135	0,185
3	90	0,180447	90	0
4	40	0,090224	45	0,556
5	22	0,036089	18	0,222
6	6	0,012030	6	0
7	2	0,003437	2	0
8	0	0,000859	0	0
	n=500			$\chi^2_p = 1{,}607$

Сумарна кількість інтервалів спостережень складає n=500. За час спостереження поступило 1000 покупців.

$$\overline{\chi}_{в} = \frac{\sum_{k=0}^{8} k \cdot n_k^l}{n} = \frac{1000}{500} = 2\frac{пок.}{за\ \ інт.},$$

$$S_{в}^2 = \frac{1}{n-1}\sum_{k=0}^{8}\left[k - \overline{\mathrm{X}}\right]^2 \cdot n_k^l = 1{,}91.$$

Бачимо, що оцінки дисперсії і математичного сподівання мають близькі значення. Висуваємо гіпотезу про Пуассонівський закон розподілу числа покупців, тобто

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \text{ де } \lambda = \overline{\mathrm{X}},$$

$$P_k = \frac{2^k}{k!}e^{-2}.$$

Величини $P_k = P(X = k)$ надано в третьому стовпчику, в четвертому – теоретичні частоти, в п'ятому – квадрати різниці частот, розділені на n_k^T .

Розрахункова величина критерію Пірсона дорівнює χ_p^2=1,607. Число степенів волі складає $\nu = 9 - (1+1) = 7$.

Табличне значення критерію, що відповідає рівню значущості $\gamma = 0{,}95$ і $\nu = 7$ дорівнює ступеням вільності дорівнює

$$\chi_\gamma^2 = 14{,}1.$$

Оскільки $\chi_p^2 < \chi_\gamma^2$ (1,607<14,1), робимо висновок: з ймовірністю 0,95 стверджуємо, що результати спостережень не протирічать гіпотезі про Пуассонівський закон розподілу числа покупців, які зробили покупки в універсамі протягом певного інтервалу часу.

Зауваження.

1. Чим більше рівень значущості, тим вужча область допустимих значень, тим більша можливість не припустити не тільки хибну, але й правильну гіпотезу.

2. Чим менше рівень значущості, тим ширша область допустимих значень критерія, тим більша можливість прийняти не тільки правильну гіпотезу, але й неправильну.

3. Зроблена перевірка (див. книгу [Крутового Ж.А «Економіко-математичне моделювання», ч.1]) гіпотези про показниковий закон розподілу тривалості обслуговування касирами універсаму.

ЛЕКЦІЯ 10

ПЕРЕВІРКА ГІПОТЕЗИ ПРО НОРМАЛЬНИЙ ЗАКОН РОЗПОДІЛУ ЙМОВІРНОСТЕЙ ВИПАДКОВОЇ ВЕЛИЧИНИ – ТЕПЛОВІ ВТРАТИ МАСИ В ПРОЦЕСІ ВАРІННЯ М'ЯСА

У таблиці 10.1 наведено результати спостережень (50 експериментів) теплових втрат маси в процесі варіння м'яса.

Перевірку статистичної гіпотези будемо здійснювати, користуючись критерієм χ^2 Пірсона.

Прийняті позначення:

X – ВВ відносного зменшення (у відсотках) початкової ваги м'яса;

$\overline{X}$, $D(X)$ – відповідно, математичне сподівання та дисперсія ВВ, яка є характеристикою розсіювання значень ВВ;

n – загальна кількість здійснених експериментів (спостережень);

n_i – експериментальні частоти;

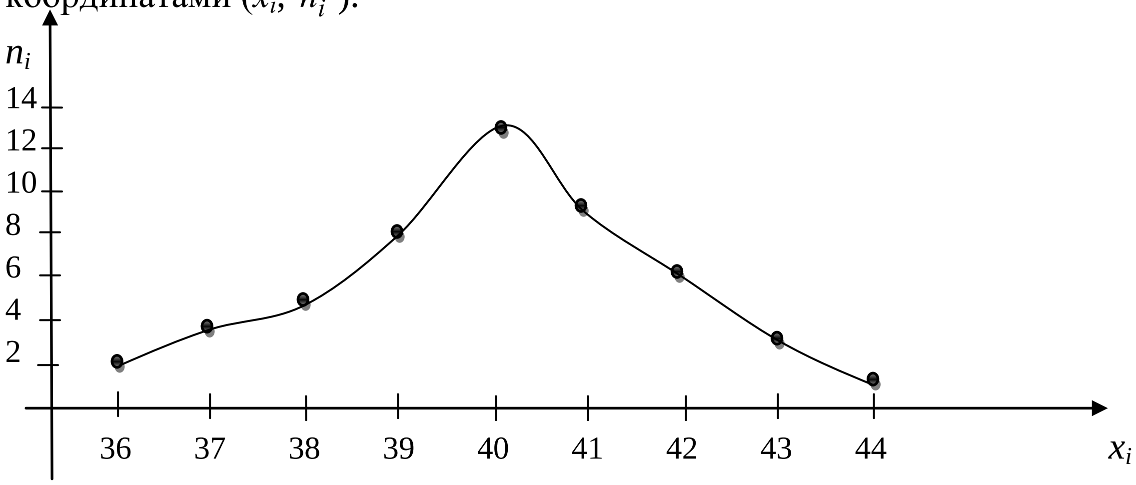
x_i – теплові втрати маси м'яса, одержані в процесі i-ого спостереження;

$S_в$ – виправлене середнє квадратичне відхилення ВВ.

Таблиця 10.1 – **Результати спостережень теплових втрат маси в процесі варіння м'яса**

Інтервал втрати маси м'яса, %	35,5–36,5	36,5–37,5	37,5–38,5	38,5–39,5	39,5–40,5	40,5–41,5	41,5–42,5	42,5–43,5	43,5–44,5
Середина i-го інтервалу x_i	36	37	38	39	40	41	42	43	44
Експериментальна частота n_i	1	4	5	8	13	9	6	3	1

Побудуємо полігон частот – ламану лінію, відрізки якої з'єднують точки з координатами (x_i, n_i).



Зауважимо, що величина зменшення ваги м’яса в процесі варіння зумовлена дією великого числа випадкових факторів.

Попередні дослідження фахівців показують, що результат дії кожного з факторів є невеликим порівняно з сумарним результатом. Згідно зі сказаним, користуючись граничними теоремами теорії ймовірностей, а також виходячи з аналізу полігона частот, можна висунути гіпотезу про нормальний закон розподілу ймовірностей ознаки, що аналізується.

Спроможні, незміщені та ефективні оцінки $\overline{X}_в$, $S_в^2$ величин $\overline{X}$ та $D(X)$ розраховувалися, використовуючи результати експерименту за формулами:

$$\overline{X}_в = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad S_в^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X}_в)^2 .$$

Якщо ВВ характеризується нормальним законом розподілу, то щільність розподілу ймовірностей цієї ВВ визначається, як відомо, двома параметрами $\overline{X}_в$, $S_в^2$, які обчислюються з використанням результатів спостережень за вказаними формулами.

Усі розрахунки, пов’язані з перевіркою сформульованої статистичної гіпотези, надано в табл. 10.2.

Таблиця 10.2 – **Результати перевірки статистичної гіпотези**

x_i	n_i	x_i^2	$x_i n_i$	$n_i x_i^2$	$x_i - \overline{X}_в$	$U_i = \frac{x_i - \overline{X}_в}{S_в}$	$\varphi(U_i)$	n_i^T
36	1	1296	36	1296	–4	–2,265	0,0310	1
37	4	1369	148	5476	–3	–1,698	0,0957	2
38	5	1444	190	7220	–2	–1,132	0,2131	6
39	8	1521	312	12168	–1	–0,566	0,3391	10
40	13	1600	520	20800	0	0	0,3989	11
41	9	1681	369	15129	1	0,566	0,3391	10
42	6	1764	252	10284	2	1,132	0,2131	6
43	3	1849	129	5547	3	1,698	0,0957	2
44	1	1936	44	1936	4	2,265	0,0310	1
	50		2000	80156				49

Обчислення:

$$\overline{X}_в = \frac{\sum_{i=1}^{9} n_i x_i}{\sum_{i=1}^{9} n_i} = \frac{2000}{50} = 40,$$

$$S_в = \left[\frac{\sum_{i=1}^{9} n_i \cdot x_i^2}{\sum_{i=1}^{9} n_i} - \left(\frac{\sum_{i=1}^{9} n_i \cdot x_i}{\sum_{i=1}^{9} n_i} \right)^2 \right]^{\frac{1}{2}} = \left[\frac{80156}{50} - (40)^2 \right]^{\frac{1}{2}} \approx 1{,}766,$$

$$U_i = \frac{x_i - \overline{X}_в}{S_в}, \quad n_i^T = \frac{nh}{S_в} \varphi(U_i) = 28{,}31 \varphi(U_i),$$

де h – різниця між двома сусідніми варіантами (крок).

$$\varphi(U_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{U_i^2}{2}}, \qquad \overline{\chi}_p = \sum_{i=1}^{9} \frac{(n_i - n_i^T)^2}{n_i^T} = 3{,}25,$$

де $\overline{\chi}_p$ – розрахункове значення критерію Пірсона.

Значення функції $\varphi(U_i)$ визначається з використанням відповідних таблиць математичної статистики.

Кількість варіантів l (кількість груп емпіричного розподілу) дорівнює 9. Число ступенів вільності становить

$$\nu = l - r = 9 - 3 = 6,$$

де r дорівнює сумі параметрів гіпотетичного закону розподілу, збільшеній на одиницю.

Використовуючи таблиці розподілу Пірсона для різних рівнів значущості: γ=0,01; 0,02; 0,05 та 6 ступенів вільності, визначимо критичні точки:

$$\chi_{0,01}^2 = 16{,}8; \ \chi_{0,02}^2 = 15{,}0; \ \chi_{0,05}^2 = 12{,}6.$$

Оскільки $\chi_p^2 < \chi_\gamma^2$ (3,25<12,6, тим більше що 3,25<16,8 і 3,25<15,0), можна дійти висновку: з ймовірністю 0,99 (0,98; 0,95) результати спостережень узгоджуються зі сформульованою гіпотезою про нормальний закон розподілу величини відносного зменшення ваги м’яса в процесі варіння. Отже, приймається гіпотеза про нормальний закон розподілу ВВ з такою щільністю ймовірностей

$$f(x) = \frac{1}{S_в \sqrt{2\pi}} e^{-\frac{(x-\bar{x}_в)^2}{2S_в^2}} = \frac{1}{1{,}766\sqrt{2\pi}} e^{-\frac{(x-40)^2}{2\cdot(1{,}766)^2}} \text{ або } f(x) \approx 0{,}2e^{-\frac{(x-40)^2}{6{,}2}}.$$

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учебное пособие. – 11-е изд., перераб. и доп. – Москва : Юрайт, 2011. – 405 с.
2. Гмурман В. Е. Теория вероятностей и математическая статистика : [учебное пособие для студентов вузов]. – 12-е изд., перераб. – М. : Высшее образование, 2008. – 478 с.
3. Полевич В. В. Елементи теорії ймовірностей і математичної статистики – Х. : ХДУХТ, 2007. – 218 с.
4. Барковський В. В. Теорія ймовірностей та математична статистика / В. В. Барковський, Н. В. Барковська, О. К. Лопатін – К. : ВАТ «Книжкова друкарня наукової книги», 2006.
5. Кулініч Г. Л. Вища математика: підручник : у 2 кн. – 2-ге вид., перероб. і доп. – К. : Либідь, 2003. – Кн. 2. Спеціальні розділи / Г. Л. Кулініч, Є. Ю. Таран, В. М. Бурим та ін.; За ред. Г. Л. Кулініча. – 368с.
6. Лихолетов И. И. Высшая математика, теория вероятностей и математическая статистика. – М., 1976. – 718 с.

ЗМІСТ

Навчальне електронне видання
комбінованого використання
Можна використовувати в локальному та мережному режимах

КРУТОВИЙ Жорж Андрійович
СОФРОНОВА Марина Сергіївна

**ЕЛЕМЕНТИ ТЕОРІЇ ЙМОВІРНОСТЕЙ
ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ**

Навчальний посібник

Відповідальний за випуск зав. кафедри фізико-математичних та інженерно-технічних дисциплін д-р техн. наук, проф. М. І. Погожих

Технічний редактор О. В. Щегельська

План 2016 р., поз. 33

Підп. до друку 16.05.2016 . Один електронний оптичний диск (CD-ROM); супровідна документація. Об’єм даних 1,68 Мб. Тираж 100 прим.

Видавець і виготівник
Харківський державний університет харчування та торгівлі.
вул. Клочківська, 333, Харків, 61051.
Свідоцтво суб’єкта видавничої справи ДК № 4417 від 10.10.12 р.