

МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ЛЕКЦІЙНИХ ЗАНЯТЬ

дисципліна Теорія ймовірностей
та математична статистика
(шифр і назва навчальної дисципліни)

спеціальність 123 Комп'ютерна інженерія
(шифр і назва спеціальності)

освітня програма Обслуговування комп'ютерних систем і мереж
(назва освітньої програми)

відділення Комп'ютерних технологій та систем
(назва відділення)

Розробив: Буряк Г.І.

Розглянуто та схвалено
на засіданні циклової комісії

математики та інформатики

Протокол № _____ від _____ 2023 р

Голова комісії _____ Малик О.М.

Лекція № 1

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 1.1	Основні поняття теорії ймовірностей та комбінаторики Предмет теорії ймовірностей. Коротка історична довідка. Випробування та події. Види та алгебра випадкових подій.
Мета заняття	Надати уявлення щодо предмету вивчення дисципліни, ознайомити з термінами та позначеннями.

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН
конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Предмет теорії ймовірностей. 2 Коротка історична довідка. 3 Випробування та події. 4 Види та алгебра випадкових подій.
4. Заключна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е. Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.
3. Буряк, Г.І. Методика застосування засобів Microsoft Excel при розв'язанні задач теорії ймовірностей [Текст] / Г.І. Буряк, – КРКМ ДНУ ім. О. Гончара, 2017. – 67 с.

Навчальні матеріали лекції

ВСТУП

Теорія ймовірностей та математична статистика — самостійні математичні науки, які становлять обов'язкову частину математичного циклу дисциплін оскільки містять теоретичну основу викладання для багатьох спеціальностей.

Теорія ймовірностей відіграє важливу роль у виявленні кількісних закономірностей та якісних тверджень у природничо-наукових, інженерно-технічних і гуманітарних дослідженнях. Необхідність її вивчення диктується сучасним рівнем розвитку науки, коли математика стає універсальною мовою науки та, як наслідок — елементом загальної культури.

Засвоєння курсу дозволить майбутнім фахівцям досліджувати математичні моделі, вирішувати завдання в умовах ризику та приймати необхідні управлінські рішення.

Аксіоматична побудова теорії ймовірностей

Предметом дослідження теорії ймовірності є особливі залежності, притаманні результатам масових однорідних випробувань.

Результатом випробування є **подія** — всякий факт, який у результаті випробування може відбуватися або не відбуватися

Події, як правило, позначають великими латинськими літерами: $A, B, C, \dots$ або в разі їх значної кількості — великою латинською літерою з індексом: $A_1, A_2, \dots, A_n$, а її зміст подають у фігурних дужках.

Наприклад, подія $A = \{\text{Рейс до Києва виконано за розкладом}\}$.

Достовірною називається подія, яка неодмінно відбувається при проведенні випробування.

Неможливою називається подія, яка при проведенні випробування не відбудеться ні за яких умов.

Наприклад

При виборі навмання одного виробу із комплекту, який містить лише стандартні вироби, подія

$A = \{\text{вибрано стандартний виріб}\}$ — достовірна

$B = \{\text{вибрано нестандартний виріб}\}$ — неможлива

Випадковою називається подія, яка в результаті проведеного випробування може відбутися або не відбутися залежно від випадкових факторів.

Наприклад

При виборі навмання одного виробу з комплекту, який складається зі стандартних та нестандартних виробів, подія

$A = \{\text{вибрано стандартний виріб}\}$ — випадкова

Несумісними називають події, якщо поява однієї з них виключає появу інших подій в одному і тому ж випробуванні.

Наприклад

Під час розіграшу м'яча

$A = \{\text{очко зараховано гравцеві зліва}\}$

$B = \{\text{очко зараховано гравцю праворуч}\}$

Сумісними називають події, якщо поява однієї з них не виключає можливості появи інших.

Наприклад

Два стрільці стріляють в мішень

$A_1 = \{\text{перший стрілок влучив у мішень}\}$

$A_2 = \{\text{другий стрілок влучив у мішень}\}$

Рівноможливими називають події, якщо немає причин стверджувати, що будь-яка з них можлива за інші.

Наприклад

При умові, що центр ваги шестигранного кубика не зміщений, рівно можливими є такі події:

$A_1 = \{\text{випала грань 1}\}$

$A_2 = \{\text{випала грань 2}\}$

$A_3 = \{\text{випала грань 3}\}$

$A_4 = \{\text{випала грань 4}\}$

$A_5 = \{\text{випала грань 5}\}$

$A_6 = \{\text{випала грань 6}\}$

Випадкові події $A_1, A_2, A_3, \dots, A_n$, утворюють **повну групу подій**, якщо в наслідок випробування хоча б одна з них з'явиться обов'язково

Наприклад

При однократному кидку грального кубика події:

$A_1 = \{\text{випала грань 1}\}, A_2 = \{\text{випала грань 2}\}, A_3 = \{\text{випала грань 3}\},$

$A_4 = \{\text{випала грань 4}\}, A_5 = \{\text{випала грань 5}\}, A_6 = \{\text{випала грань 6}\}$

утворюють повну групу.

Дві несумісні події, які утворюють повну групу, називають **протилежними**
 $\bar{A}$ — подія протилежна події A .

Наприклад

Якщо позначити через A подію, що при одноразовому киданні монети з'явиться герб, то подія з'явиться надпис буде протилежною

$A = \{\text{поява герба}\}$

$\bar{A} = \{\text{поява надпису}\}$

Елементарними наслідками називають такі події, які не можливо розділити на більш прості.

Наприклад

При киданні монети можуть бути два можливих наслідки:

- герб;
- надпис

При киданні грального кубика можуть бути шість можливих наслідків.

При випробуванні «постріл у мішень» можна розглядати такі наслідки:

- влучення у мішень;
- кількість вибитих очок;
- координати точки влучення

Отже, що приймати за наслідок випробування залежить від умови задачі.

Множину усіх можливих елементарних наслідків називають **простором елементарних наслідків**

Простір елементарних наслідків може містити таку множину елементів:

- скінчену
- злічену
- незлічену

Наприклад

При двократному киданні монети простір елементарних наслідків містить складає 4 точки: (Г, Г), (Г, Н), (Н, Г), (Н, Н).

При киданні двох гральних кубиків простір елементарних наслідків містить 36 точок.

При стрільбі одиночними пострілами до першого влучення можливі такі елементарні події:

$\omega_1 = \{\text{влучення при першому пострілі}\}$

$\omega_2 = \{\text{влучення при другому пострілі}\}$

$\omega_3 = \{\text{влучення при третьому пострілі}\}$

...

$\omega_n = \{\text{влучення при } n\text{-му пострілі}\}$

При виробництві рідкокристалічних дисплеїв неоднакові умови технологічного процесу, тому час роботи пристроїв відрізняється від його номінального значення, тобто буде випадковою подією. Простір елементарних наслідків у цьому випадку буде нескінченною незліченою множиною, елементи якої неможливо пронумерувати

Питання та завдання до контролю знань студентів

1. Які події називають достовірними, неможливими, випадковими?
2. Що є предметом теорії ймовірностей?
3. Як визначають та позначають протилежну подію,
4. Які події називають сумісними, несумісними, рівноможливими?
5. Які події називають елементарними наслідками?
6. При яких умовах можна стверджувати, що події утворюють повну групу?
7. Що називають простором елементарних наслідків?
8. Яку множену елементів може утворювати простір елементарних наслідків?

Лекція № 2

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 1.1	Основні поняття теорії ймовірностей та комбінаторики Класичне та статистичне означення ймовірностей. Властивості ймовірності та частоти. Геометричні ймовірності. Елементи комбінаторики. Елементарні комбінаторні задачі. Біном Ньютона.
Мета заняття	Ознайомити з філософським змістом ймовірності, а також сформулювати означення ймовірності та методи її обчислення, дослідження комбінаторних конфігурацій, алгоритмів побудови, їх оптимізація, застосування при розв'язанні задач.

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН
конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Класичне означення ймовірності. 2 Геометрична ймовірність. 3 Властивості ймовірності та частоти. 4 Перестановки, розміщення, сполучення. 5 Біном Ньютона.
4. Заключна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е. Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.

ВСТУП

Імовірність події — є численна міра степеня об'єктивної можливості цієї події.

Це означення імовірності визначає філософську суть ймовірності, але не вказує закону знаходження ймовірності будь-якої події

Комбінаторика — розділ математики, присвячений розв'язанню задач вибору та розташуванню елементів деякої, зазвичай, скінченної множини відповідно до заданих правил. Кожне таке правило визначає спосіб побудови деякої конструкції із елементів вихідної множини, що зветься комбінаторною конфігурацією.

Тому на меті комбінаторного аналізу стоїть дослідження комбінаторних конфігурацій, алгоритмів їх побудови, оптимізація таких алгоритмів, а також розв'язання задач переліку.

Термін «комбінаторика» ввів Лейбніц, який у 1666 році опублікував свою працю «Міркування про комбінаторне мистецтво».

Найпростішими прикладами комбінаторних конфігурацій є перестановки, розміщення, комбінація та розбиття.

1. Класичне означення ймовірності події

Імовірність події A дорівнює відношенню числа елементарних наслідків, які сприяють події A , до загального числа усіх єдиноможливих та рівноможливих елементарних наслідків.

При класичному визначенні ймовірність події визначається рівністю

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

m — число елементарних наслідків, що сприяють події A

n — число усіх єдиноможливих та рівноможливих наслідків

Відносною частотою (частістю) події A називають відношення числа випробувань, у яких подія A з'явиться, до числа фактично виконаних випробувань.

Ймовірність $P(A)$ події A обчислюється до випробування, а частість $W(A)$ обчислюється після випробування.

Відносна частість події визначається рівністю

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

m — кількість випробувань, у яких з'явиться подія A

n — кількість усіх випробувань

Частість має властивість стійкості:

$$P(A) = \lim_{n \rightarrow \infty} W(A)$$

Приклад

Кинуті два гральні кубики. Знайти ймовірність того, що

- а) сума очок, що випала гранях — парна, причому на грані хоча б одного з кубиків з'явиться шістка (подія A),
- б) сума очок, що випала дорівнює 4 (подія B),
- в) добуток очок буде дорівнює семи (подія C);
- г) виявиться не менше 20 (подія D);
- д) сума очок, що випала дорівнює п'яти, а добуток — чотирьом (подія E).

Розв'язання

Число усіх єдиноможливих та рівноможливих наслідків — 36, оскільки простір елементарних наслідків складається з комбінацій, що наведено в таблиці

- а) Число елементарних наслідків, що сприяють події A — 5, оскільки

1	2	3	4	5
				

Отже ймовірність події: $P(A) = \frac{5}{36}$

б) Число елементарних наслідків, що сприяють події B — 3, оскільки

Отже ймовірність події:

$$P(B) = \frac{3}{36} = \frac{1}{12}$$

1	2	3
 	 	 

в) Число елементарних наслідків, що сприяють події C — 0, оскільки для даної події не існує сприятливих результатів. Отже ймовірність події:

$$P(C) = \frac{0}{36} = 0, \text{ тобто подія } C \text{ — неможлива.}$$

г) Число елементарних наслідків, що сприяють події D — 8, оскільки

$$P(D) = \frac{8}{36} = \frac{2}{9}$$

1	3	4	5
 	 	 	 
2	3	4	5
 	 	 	 

д) Число елементарних наслідків, що сприяють події E — 2, оскільки

Отже ймовірність події:

$$4 + 1 = 5$$

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

$$4 \cdot 1 = 4$$

1	2
 	 

Відповідь. $P(A) = 5/36$, $P(B) = 1/12$, $P(C) = 0$, $P(D) = 2/9$, $P(E) = 1/18$.

2. Геометричне означення ймовірності

Класичне означення ймовірності має місце лише тоді, коли m та n скінченні, усі елементарні наслідки рівноможливі.

Якщо множина усіх елементарних наслідків нескінченна і, як наслідок займає деяку область G , а події A сприяє лише частина $g \in G$, то обчислення ймовірності події A виконують згідно геометричного означення ймовірності.

Імовірність випадкової події A дорівнює відношенню

$$P(A) = \frac{\text{mes } g}{\text{mes } G}$$

міри g до міри G .

Приклад

На площині накреслені два концентричні кола, радіуси яких 5 та 10 см відповідно. Знайти ймовірність того, що точка, кинута навмання у велике коло, потрапить в кільце, що утворене побудованими колами. Передбачається, що ймовірність попадання точки у плоску фігуру пропорційна площі цієї фігури та не залежить від її розташування відносно великого кола.

Розв'язання

Позначимо через $A = \{\text{точка потрапить у кільце}\}$.

Площа великого кола ($\text{mes } G$): $\text{mes } G = \pi \cdot R^2 = 100\pi$

Площа кільця ($\text{mes } g$): $\text{mes } g = \pi \cdot R^2 - \pi \cdot r^2 = \pi(R^2 - r^2) = \pi(10^2 - 5^2) = 75\pi$

Шукана ймовірність дорівнює відношенню відповідних площ:

$$P(A) = \frac{\text{mes } g}{\text{mes } G} = \frac{75\pi}{100\pi} = 0,75$$

Відповідь. $P(A) = 0,75$.

Приклад

До сигналізатора надходять сигнали від двох пристроїв, причому надходження кожного із сигналів рівноможливе в будь-який момент проміжку часу тривалістю T . Моменти надходження сигналів незалежні один від іншого. Сигналізатор спрацьовує, якщо різниця між моментами надходження сигналів менша t ($t < T$). Знайдіть ймовірність того, що сигналізатор спрацьовує за час T , якщо кожен з пристроїв пошле по одному сигналу.

Розв'язання

Нехай x та y — моменти надходження сигналів

Можливі значення задовольняють подвійним нерівностям:

$$0 \leq x \leq T; \quad 0 \leq y \leq T$$

Введемо прямокутну систему координат xOy . У цій системі подвійним нерівностям

$0 \leq x \leq T; \quad 0 \leq y \leq T$ задовольняють координати будь-якої точки, що належить

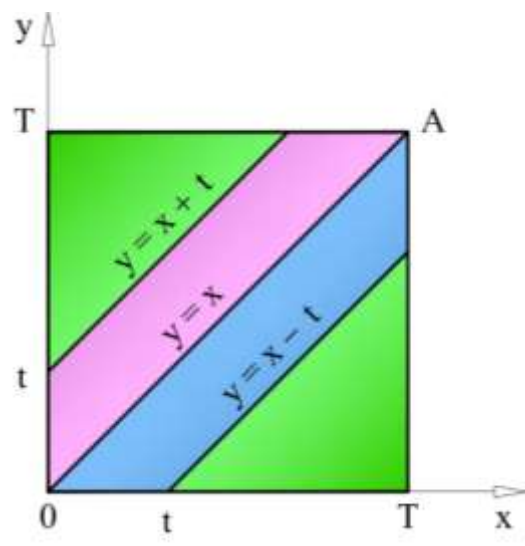
квадрату OTAT (див. рисунок 1 а)), площа якого $\in \text{mes } G$, тобто усі можливі значення моментів надходження сигналів.

Сприятливі значення задовольняють нерівностям: $y - x < t$; $x - y < t$,

а рівносильні їм нерівності

$y < x + t$ — виконуються для координат тих точок, що лежать вище прямої $y = x$ і нижче прямої $y = x + t$

$y > x - t$ — виконуються для координат тих точок, що лежать нижче прямої $y = x$ і вище прямої $y = x - t$.



а)

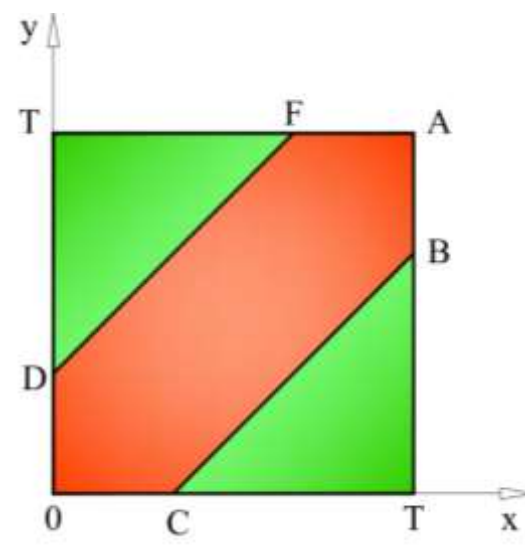
Отже, сприятливими вважатимемо координати будь-якої точки, що належить шестикутнику ABCODF (див. рисунок 1 б)), площа якого $\in \text{mes } g$.

$$\text{mes } G = S_{OTAT} = T^2$$

$$\text{mes } g = S_{ABCODF} = S_{OTAT} - (S_{\triangle DTF} + S_{\triangle BTC}) = T^2 - (T - t)$$

Шукана імовірність дорівнює відношенню відповідних площ:

$$P(A) = \frac{\text{mes } g}{\text{mes } G} = \frac{T^2(T - t)}{T^2}$$



б)

3. Властивості ймовірностей

1 Ймовірність достовірної події дорівнює одиниці

$$P(A) = 1$$

2 Ймовірність неможливої події дорівнює нулю:

$$P(A) = 0$$

3 Ймовірність випадкової події є додатне число, менше одиниці:

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

4. Поняття сполуки. Правило суми і добутку.

Вибрані (або вибрані і розміщені) групи елементів називають сполуками. Якщо всі елементи сполуки різні, то одержуємо сполуки без повторень, а якщо елементи можуть повторюватися, то одержуємо сполуки з повтореннями.

4.1 Правило суми

Якщо елемент А можна вибрати m способами, а елемент В — n способами, то А або В можна вибрати $(m + n)$ способами.

Наприклад

Якщо на тарілці лежить 5 груш і 4 яблука, то вибрати один фрукт (тобто грушу або яблуко) можна 9 способами ($5 + 4 = 9$).

4.2 Правило добутку

Якщо елемент А можна вибрати m способами, а після цього елемент В — n способами, то А і В можна вибрати $(m \cdot n)$ способами.

Наприклад

Якщо в кіоску продають ручки 5-ти видів і зошити 4-х видів, то вибрати набір з ручки і зошита (тобто пару — ручка і зошит) можна $5 \cdot 4 = 20$ способами (оскільки до кожної з 5 ручок можна взяти будь-який із 4 зошитів).

Висновок

- кількості способів вибору кожного елементу додають, якщо доводиться вибирати або перший елемент, або другий, або третій і т. д. елемент;
- кількості способів вибору кожного елементу перемножують, якщо доводиться вибирати набір, у який входить і перший, і другий, і третій, і т. д. елементи.

5. Впорядковані множини

Для впорядкованих множин суттєвим є порядок слідування їх елементів, тобто те, який елемент записано на першому місці, який на другому тощо. Зокрема, якщо одні й ті самі елементи записати в різному порядку, то отримаємо різні упорядковані множини. Щоб відрізнити запис впорядкованої множини від неупорядкованої, елементи впорядкованої множини часто записують у круглих дужках, наприклад $(1; 2; 3) \neq (1; 3; 2)$.

Розглядаючи впорядковані множини, слід враховувати, що впорядкованість не є властивістю самої неупорядкованої множини (з якої ми одержали впорядковану), оскільки одну й ту саму множину можна по-різному впорядкувати.

Наприклад

Множину з трьох чисел $\{-5; 1; 3\}$ можна впорядкувати:

- за зростанням: $(-5; 1; 3)$;
- за спаданням: $(3; 1; -5)$;
- за зростанням абсолютної величини числа: $(1; 3; -5)$

Будемо розуміти, що для того щоб задати скінченну впорядковану множину з n елементів, досить вказати, який елемент знаходиться на першому місці, який на другому, ..., який на n -му.

5.1 Перестановки

Перестановкою з n елементів називається будь-яка впорядкована множина з n елементів.

Інакше кажучи, це така множина, для якої указано, який елемент знаходиться на першому місці, який — на другому, ..., який — на n -му).

Кількість перестановок без повторень з n елементів позначається P_n (P — перша літера французького слова *permutation* — перестановка).

$$(P_n) = n!, \quad (2.1)$$

де $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$ (читається: «Ен факторіал»)

Наприклад

Кількість різних шестизначних чисел, які можна скласти з цифр 1, 2, 3, 4, 5, 6, не повторюючи ці цифри в одному числі, дорівнює

$$P_6 = 6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720$$

5.2 Розміщення

Розміщенням з n елементів по k називається будь-яка впорядкована множина з k елементів, складена з елементів n -елементної множини.

Наприклад

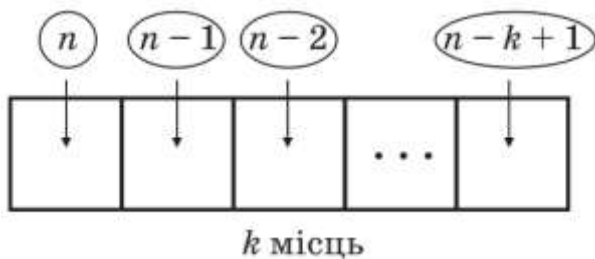
Із множини з трьох цифр $\{1; 5; 7\}$ можна скласти такі розміщення з двох елементів без повторень:

$$(1; 5), (1; 7), (5; 7), (5; 1), (7; 1), (7; 5).$$

Кількість розміщень з n елементів по k позначається A_n^k (читається: «А з n по k », А — перша літера французького слова arrangement — розміщення).

Отже, $A_3^2 = 6$.

Складання кількості розміщень з n елементів по k (без повторень) можна уявити, як послідовне заповнення k місць, які зображені у вигляді клітинок.



На перше місце можна вибрати один з n елементів заданої множини (тобто елемент для першої клітинки можна вибрати n способами).

Якщо елементи не можна повторювати, то на друге місце можна вибрати тільки один елемент із тих, що залишилися, тобто з $(n - 1)$.

Тепер уже два елементи використані і на третє місце можна вибрати тільки один з $(n - 2)$ елементів і т. д. На k -те місце можна вибрати тільки один з $n - (k - 1) = n - k + 1$ елементів.

Оскільки потрібно вибрати елементи і на перше місце, і на друге, ..., і на k -те, то використовується правило добутку. Отже число розміщень з n елементів по k :

$$A_n^k = n(n-1)(n-2)\dots(n-k+1) \quad (2.2)$$

$$A_n^k = \frac{n!}{(n-k)!} \quad (2.3)$$

Наприклад

Кількість різних тризначних чисел, які можна скласти з цифр 1, 2, 3, 4, 5, 6, якщо цифри не можуть повторюватися, дорівнює

$$A_6^3 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{1 \cdot 2 \cdot 3} = 4 \cdot 5 \cdot 6 = 120$$

5.3 Комбінації (сполучення)

Комбінацією без повторень з n елементів по k називається будь-яка k -елементна підмножина n -елементної множини.

Наприклад

З множини $\{a, b, c, d\}$ можна скласти такі комбінації без повторень з трьох елементів: $\{a, b, c\}$, $\{a, b, d\}$, $\{a, c, d\}$, $\{b, c, d\}$.

Кількість комбінацій без повторень з n елементів по k позначається символом C_n^k (читається: «Число комбінацій із n по k » або «це із n по k », C — перша літера французького слова *combinaison* — комбінація).

Отже, $C_4^3 = 4$.

З'ясувати, скільки всього можна скласти комбінацій без повторень з n елементів по k можна, використавши формули числа розміщень і перестановок.

Складання розміщення без повторень з n елементів по k проводять у два етапи. Спочатку слід вибрати k різних елементів із заданої n -елементної множини, не враховуючи порядок вибору цих елементів. Після цього одержану множину з k різних елементів впорядковують. Її можна впорядковувати $P_k = k!$ способами. Одержано розміщення без повторень з n елементів по k . Отже, кількість розміщень без повторень з n елементів по k в $k!$ разів більша за число комбінацій без повторень з n елементів по k .

Тобто

$$A_n^k = C_n^k \cdot k!$$

Звідси

$$C_n^k = \frac{A_n^k}{k!}$$

Враховуючи, що за формулою (2.3)

$$A_n^k = \frac{n!}{(n-k)!}, \text{ одержуємо}$$

$$C_n^k = \frac{n!}{k!(n-k)!} \quad (2.4)$$

Наприклад

Із групи, що складається з 25 студентів, можна виділити 5 студентів для участі в олімпіаді з математики C_{25}^5 способами, тобто

$$C_{25}^5 = \frac{25!}{5!(25-5)!} = \frac{25!}{5! \cdot 20!} = \frac{21 \cdot 22 \cdot 23 \cdot 24 \cdot 25}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 53130$$

Властивості числа комбінацій без повторень

- 1) $C_n^k = C_n^{n-k}$
- 2) $C_n^0 + C_n^1 + C_n^2 + C_n^3 + \dots + C_n^n = 2^n$
- 3) $C_n^k + C_n^{k+1} = C_{n+1}^{k+1}$
- 4) $C_n^n = 1$
- 5) $C_n^1 = n$

Використовуючи формулу (2.4), легко обґрунтувати першу властивість числа комбінацій без повторень.

Оскільки

$$C_n^{n-k} = \frac{n!}{(n-k)! \cdot (n-(n-k))!} = \frac{n!}{(n-k)! \cdot k!} = C_n^k$$

Для того щоб властивість 1 можна було використовувати і при $k = n$, домовилися вважати, що $C_n^0 = 1$. Тоді $C_n^n = C_n^0 = 1$.

Якщо у формулі (2.4) скоротити чисельник і знаменник на $(n-k)!$, то дістанемо формулу, за якою зручно обчислювати C_n^k при малих значеннях k :

$$C_n^k = \frac{\overbrace{n(n-1)(n-2)\dots n(n-k+1)}^{k \text{ множників}}}{k!} = \frac{n(n-1)(n-2)\dots n(n-k+1)}{\underbrace{1 \cdot 2 \cdot \dots \cdot k}_{k \text{ множників}}} \quad (2.5)$$

Наприклад

При обчисленні числа комбінацій C_{25}^2 задіяні два множники

$$C_{25}^2 = \frac{\overbrace{25 \cdot 24}^{2 \text{ множники}}}{1 \cdot 2} = 25 \cdot 12 = 300$$

При обчисленні числа комбінацій C_8^3 задіяні три множники

$$C_8^3 = \frac{\overbrace{8 \cdot 7 \cdot 6}^{3 \text{ множники}}}{1 \cdot 2 \cdot 3} = 8 \cdot 7 = 56$$

При розв'язуванні найпростіших комбінаторних задач важливо правильно вибрати формулу, за якою будуть проводитися обчислення кількості сполук. Для цього досить з'ясувати таке:

- чи враховується порядок слідування елементів у сполуці?
- чи всі задані елементи входять до одержаної сполуки?

Якщо, наприклад, порядок слідування елементів враховується і з n за даних елементів у сполуці використовується тільки k елементів, то за означенням — це розміщення з n елементів по k . Після визначення виду сполуки слід також з'ясувати, чи можуть елементи в сполуці повторюватися, тобто з'ясувати, яку формулу потрібно використати — для кількості сполук без повторень чи з повтореннями.

Чи враховується порядок слідування елементів у сполуці?					
Так			Ні		
Чи усі елементи входять до сполуки?					
Так		Ні			
Перестановки		Розміщення		Комбінації	
без повторень	з повтореннями	без повторень	з повтореннями	без повторень	з повтореннями
$P_n = n!$	$\bar{P}_n = \frac{n!}{k_1! k_2! \dots k_m!},$ де $k_1 + k_2 + \dots + k_m = n$	$A_n^k = \frac{n!}{(n-k)!}$	$\bar{A}_n^k = n^k$	$C_n^k = \frac{n!}{k!(n-k)!}$	$\bar{C}_n^k = C_{n+k-1}^k$

5.4 Обчислення числа комбінацій без повторень за допомогою трикутника Паскаля

Обчислення числа комбінацій без повторень можна організувати послідовним обчисленням відповідних значень, використавши властивість 3. Ця рівність дозволяє послідовно обчислювати значення C_n^k за допомогою спеціальної таблиці, яка називається трикутником Паскаля. Якщо вважати, що $C_0^0 = 1$, то кожний рядок цієї таблиці починається з одиниці і закінчується одиницею ($C_n^0 = C_n^n = 1$).

Якщо якийсь рядок вже записано, наприклад, третій, то в четвертому рядку потрібно записати на першому місці одиницю. На другому місці записуємо число, яке дорівнює сумі двох чисел третього рядка, які стоять над ним ліворуч і праворуч (оскільки за формулою $C_4^1 = C_3^0 + C_3^1 = 1 + 3 = 4$). На третьому місці записуємо число, яке дорівнює сумі двох наступних чисел третього рядка, які стоять над ним ліворуч і праворуч ($C_4^2 = C_3^1 + C_3^2 = 3 + 3 = 6$), і т. д. (а на останньому місці знову записуємо одиницю).

Значення C_n^k									
$k \backslash n$	0	1	2	3	4	5	6	...	
0	1								
1	1	1							
2	1	2	1						
3	1	3	3	1					
4	1	4	6	4	1				
5	1	5	10	10	5	1			
6	1	6	15	20	15	6	1		
...			•	•	•				
n	C_n^0	C_n^1	C_n^2	C_n^3	C_n^4	C_n^5	C_n^6	...	

6. Біном Ньютона

6.1 Розклад степеня бінома

Двочлен виду $a + x$ також називають біномом. З курсу алгебри відомо, що

$$(a + x)^1 = a + x = 1 \cdot a + 1 \cdot x;$$

$$(a + x)^2 = a^2 + 2ax + x^2 = 1 \cdot a^2 + 2 \cdot ax + 1 \cdot x^2;$$

$$(a + x)^3 = a^3 + 3a^2x + 3ax^2 + x^3 = 1 \cdot a^3 + 3 \cdot a^2x + 3 \cdot ax^2 + 1 \cdot x^3.$$

Можна помітити, що коефіцієнти розкладу степеня бінома $(a + x)$ при $n = 1, 2, 3$ співпадають з числами у відповідному рядку трикутника Паскаля. Виявляється, що ця властивість виконується і для довільного натурального n , тобто правдива формула

$$(a + x)^n = C_n^0 a^n + C_n^1 a^{n-1}x + C_n^2 a^{n-2}x^2 + C_n^3 a^{n-3}x^3 + \dots + C_n^k a^{n-k}x^k + C_n^n x^n \quad (3.1)$$

Формулу (3.1) називають біномом Ньютона. Права частина цієї рівності називається розкладом степеня бінома $(a + x)^n$, а числа C_n^k (при $k = 0, 1, 2, \dots, n$) називають біноміальними коефіцієнтами. Загальний член розкладу степеня бінома має вигляд

$$T_{k+1} = C_n^k \cdot a^{n-k} \cdot x^k \quad (\text{де } k = 0, 1, 2, \dots, n) \quad (3.2)$$

Записуючи степінь двочлена за формулою бінома Ньютона для невеликих значень n , біноміальні коефіцієнти можна обчислювати за трикутником Паскаля, який зручно записувати так

Степінь	Коефіцієнти розкладу	Орієнтир
$(a + x)^0$	1	У кожному рядку по краях стоять одиниці, а кожне зрешти чисел дорівнює сумі двох чисел, що знаходяться над ним ліворуч і праворуч
$(a + x)^1$	1 1	
$(a + x)^2$	1 2 1	
$(a + x)^3$	1 3 3 1	
$(a + x)^4$	1 4 6 4 1	
$(a + x)^5$	1 5 10 10 5 1	
$(a + x)^6$	1 6 15 20 15 6 1	
...	...	

Наприклад

$$(a + b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5.$$

Якщо в формулі бінома Ньютона замінити x на $(-x)$, то одержимо формулу піднесення до степеня різниці $a - x$.

$$(a - x)^n = C_n^0 \cdot a^n - C_n^1 \cdot a^{n-1} \cdot x + C_n^2 \cdot a^{n-2} \cdot x^2 - C_n^3 \cdot a^{n-3} \cdot x^3 + \dots + (-1)^n \cdot x^n \quad (3.3)$$

6.2 Властивості біноміальних коефіцієнтів

1) Число біноміальних коефіцієнтів (а отже, і число доданків у розкладі n -го степеня бінома) дорівнює $n + 1$, оскільки розклад містить усі степені x від 0 до n (і інших доданків не містить).

2) Коефіцієнти членів, рівновіддалених від початку і кінця розкладу, рівні між собою, оскільки $C_n^k = C_n^{n-k}$

3) Сума всіх біноміальних коефіцієнтів дорівнює 2^n .

4) Сума біноміальних коефіцієнтів, що стоять на парних місцях, дорівнює сумі біноміальних коефіцієнтів, що стоять на непарних місцях.

Висновки

При розв'язуванні багатьох практичних задач доводиться вибирати з певної сукупності об'єктів елементи, що мають ті або інші властивості, розміщувати ці елементи в певному порядку і т. д. Оскільки в таких задачах йдеться про ті або інші комбінації об'єктів, то такі задачі називають комбінаторними. Отже методи розв'язування комбінаторних задач розглядаються як вибір і розміщення елементів деякої скінченної множини на основі якихось умов. Вибрані (або вибрані і розміщені) групи елементів називають сполуками. Якщо всі елементи сполуки різні, то одержуємо сполуки без повторень, а якщо елементи можуть повторюватися, то одержуємо сполуки з повтореннями.

Питання та завдання до контролю знань студентів

1. Сформулюйте і поясніть на прикладах правило суми і правило добутку для розв'язування комбінаторних задач.
2. Що називається розміщенням з n елементів по k без повторень, навести приклади використання.
3. Обґрунтуйте формулу для обчислення числа розміщень з n елементів по k без повторень.
4. Поясніть, що називається перестановкою з n елементів (без повторень). Наведіть приклади.
5. Обґрунтуйте формулу для обчислення числа перестановок з n елементів без повторень.
6. Поясніть, що називається комбінаціями з n елементів по k без повторень. Наведіть приклади.
7. Запишіть формулу для обчислення числа комбінацій з n елементів по k без повторень. Наведіть приклади її використання.
8. Обґрунтуйте формулу для обчислення числа комбінацій з n елементів по k без повторень.
9. Поясніть, як можна послідовно обчислювати значення за допомогою спеціальної таблиці — трикутника Паскаля.
10. Поясніть на прикладах, як можна вибирати відповідну формулу при розв'язуванні найпростіших комбінаторних задач. Сформулюйте і доведіть властивості біноміальних коефіцієнтів.

Лекція № 3

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 1.2 Основні теореми теорії ймовірностей

Випадкові події. Дії з випадковими подіями. Теореми додавання та множення ймовірностей. Залежні та незалежні події, умовні ймовірності. Повна ймовірність, формули Байєса. Ймовірність гіпотез.

Мета заняття Ознайомити з види випадкових подій, сформулювати основні теореми, розглянути метод обчислення повної ймовірності, навчити виконувати переоцінку гіпотези

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН
конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Теореми додавання ймовірностей несумісних та сумісних подій. Наслідки. 2 Теорема множення ймовірностей. 3 Формула повної ймовірності. 4 Поняття гіпотези. Переоцінка гіпотези за формулою Байєса.
4. Заключна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е.Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.

Навчальні матеріали лекції

Додавання ймовірностей

Теорема 1 (додавання ймовірностей несумісних подій)

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

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

Наслідок

Ймовірність появи однієї з декількох попарно несумісних подій

$A_1, A_2, \dots, A_n$, байдуже якої, дорівнює сумі ймовірностей цих подій:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

Теорема 2 (сума повної групи випадкових подій)

Сума ймовірностей повної групи випадкових подій $A_1, A_2, A_3, \dots, A_n$ дорівнює одиниці

$$P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1$$

Теорема 2 (сума повної групи випадкових подій)

Сума ймовірностей повної групи випадкових подій $A_1, A_2, A_3, \dots, A_n$ дорівнює одиниці

$$P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1$$

Теорема 3 (сума протилежних подій)

Сума ймовірностей протилежних подій A та $\bar{A}$ дорівнює одиниці

$$P(A) + P(\bar{A}) = 1$$

Теорема 4 (додавання ймовірностей сумісних подій)

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

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

Приклад 1

На стелажі в бібліотеці в довільному порядку розставлено 15 підручників, причому п'ять з них у палітурці. Бібліотекар бере навмання три підручника. Знайти ймовірність того, що хоча б один з взятих підручників опиниться в палітурці (подія A).

Розв'язання

I спосіб.

Вимога — $A = \{\text{хоча б один з взятих підручників в палітурці}\}$ — буде здійснене, якщо відбудеться будь-яка з наступних трьох несумісних подій: $B = \{\text{один підручник в палітурці}\}$

$C = \{\text{два підручника в палітурці}\}$

$D = \{\text{три підручника в палітурці}\}$

Отже, подію A можна представити у вигляді суми подій:

$$A = B + C + D$$

Ймовірності подій B , C та D :

$$P(B) = \frac{C_5^1 \cdot C_{10}^2}{C_{15}^3} = \frac{45}{91}; \quad P(C) = \frac{C_5^2 \cdot C_{10}^1}{C_{15}^3} = \frac{20}{91}; \quad P(D) = \frac{C_5^3}{C_{15}^3} = \frac{2}{91}$$

Отже, ймовірність події A обчислюється за теоремою 1:

$$P(A) = P(B) + P(C) + P(D) = \frac{45}{91} + \frac{20}{91} + \frac{2}{91} = \frac{67}{91}$$

Зауваження

Під час розв'язання задач на відшукування ймовірності події A доцільно спочатку обчислити ймовірність події $\bar{A}$, а потім знайти шукану ймовірність за формулою

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

ІІ спосіб.

Якщо подія $A = \{\text{хоча б один з взятих підручників в палітурці}\}$, тоді

$$\bar{A} = \{\text{жоден із взятих підручників в палітурці}\}$$

Ймовірності подій $\bar{A}$:

$$P(\bar{A}) = \frac{C_{10}^3}{C_{15}^3} = \frac{24}{91};$$

Тоді ймовірність події A :

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{24}{91} = \frac{67}{91}$$

Відповідь. $P(A) = 67/91$.

Множення ймовірностей

Теорема 5 (множення ймовірностей)

Ймовірність спільної появи двох подій дорівнює добутку ймовірності однієї з них на умовну ймовірність іншої, обчислену в припущенні, що перша подія вже настала

$$P(A \cdot B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A)$$

Зокрема, для незалежних подій

Ймовірність спільної появи двох незалежних подій дорівнює добутку ймовірностей цих подій

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

Наслідок

Ймовірність спільної появи декількох подій дорівнює добутку ймовірностей однієї з них на умовні ймовірності всіх інших, причому ймовірність кожної наступної події обчислюють у припущенні, що всі попередні події вже настали:

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

Зокрема, для незалежних подій

Ймовірність спільної появи декількох подій, незалежних в сукупності, дорівнює добутку ймовірностей цих подій:

$$P(A_1 A_2 A_3 \dots A_n) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \dots \cdot P(A_n)$$

Приклад 2

Для сигналізації про аварію встановлено два незалежно працюючих сигналізатори. Ймовірність того, що при аварії спрацює перший сигналізатор дорівнює 0,95, а другий — 0,9 відповідно. Знайти ймовірність того, що при аварії спрацює лише один сигналізатор.

Розв'язання

Введемо позначення подій:

$A_1 = \{\text{спрацював перший сигналізатор}\}$, $A_2 = \{\text{спрацював другий сигналізатор}\}$

$B_1 = \{\text{з'явилась лише подія } A_1\}$, $B_2 = \{\text{з'явилась лише подія } A_2\}$

Поява події B_1 рівносильне появі події $A_1 \bar{A}_2$, тобто $B_1 = A_1 \cdot \bar{A}_2$

Поява події B_2 рівносильне появі події $\bar{A}_1 A_2$, тобто $B_2 = \bar{A}_1 \cdot A_2$

Оскільки події A_1 та A_2 — незалежні, то події $\bar{A}_1$ та $\bar{A}_2$ — теж незалежні, отже:

$$P(B_1) = P(A_1 \cdot \bar{A}_2) = P(A_1) \cdot P(\bar{A}_2) = 0,95 \cdot (1 - 0,9) = 0,095$$

$$P(B_2) = P(\bar{A}_1 \cdot A_2) = P(\bar{A}_1) \cdot P(A_2) = (1 - 0,95) \cdot 0,9 = 0,045$$

Відповідь. Отже, шукана ймовірність: $P(A_1 \cdot \bar{A}_2 \cup \bar{A}_1 \cdot A_2) = 0,095 + 0,045 = 0,14$

Приклад 3

У розпорядженні системного адміністратора 3 маршрутизатора (Router) та 7 комутаторів (Switch). Для організації локальної мережі системний адміністратор задіяв спочатку один пристрій, а потім інший.

Знайти ймовірність того, що перший з узятих пристроїв — маршрутизатор, а другий — комутатор.

Розв'язання

Нехай події $A = \{\text{задіяний маршрутизатор}\}$,

$B = \{\text{задіяний комутатор}\}$

Ймовірність того, що перший з задіяних пристроїв маршрутизатор: $P(A) = \frac{2}{10}$

Ймовірність події B — умовна та обчислена з припущенням, що подія A відбулася: $P_A(B) = \frac{7}{9}$

За теоремою 5 $P(A \cdot B) = P(A) \cdot P_A(B) = \frac{3}{10} \cdot \frac{7}{9} = \frac{7}{30}$

Відповідь. $P(A) = 7/30$.

Приклад 4

У деякому людському суспільстві 70% палять, 40% хворіють на рак легенів та 28% палять та мають рак легенів. Знайти імовірність того, що навмання обрана особа з цього суспільства

- а) не палить, але має рак легенів;
- б) палить, але не має раку легенів;
- в) ніколи не палила і не має раку легенів;
- г) палить та має рак легенів;
- д) або палить або має рак легенів.

Розв'язання

Позначимо події:

$A = \{\text{особа палить}\}$

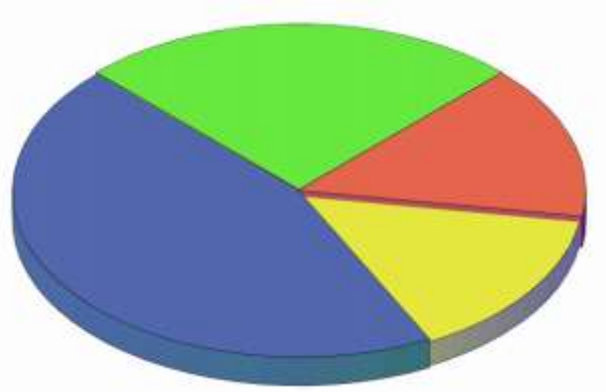
$B = \{\text{особа хвора на рак легенів}\}$

За умовою задачі:

$$P(A) = 0,7; \quad P(B) = 0,4;$$

$$P(\bar{A}) = 0,3; \quad P(\bar{B}) = 0,6;$$

$$P(A \cdot B) = 0,28$$



а) Ймовірність того, що навмання обрана особа не палить, але має рак легенів:

$$P(\bar{A} \cdot B) = 0,3 \cdot 0,4 = 0,12$$

б) Ймовірність того, що навмання обрана особа палить, але не має рак легенів:

$$P(A \cdot \bar{B}) = 0,7 \cdot 0,6 = 0,42$$

в) Ймовірність того, що навмання обрана особа ніколи не палила і не має раку легенів: $P(\bar{A} \cdot \bar{B}) = 0,3 \cdot 0,6 = 0,18$

г) Ймовірність того, що навмання обрана особа палить та має рак легенів:

$$P(A \cup B) = P(A) + P(B) - P(A \cdot B) = 0,7 + 0,4 - 0,28 = 0,82$$

д) Ймовірність того, що навмання обрана особа або палить або має рак легенів: $P(A \cdot \bar{B} \cup \bar{A} \cdot B) = P(A \cdot \bar{B}) + P(\bar{A} \cdot B) = 0,42 + 0,12 = 0,54$

Формули повної імовірності

Теорема 6

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

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

Приклад 5

У першому ящику 20 деталей, з яких 15 стандартних. У другому ящику 10 деталей, з яких 9 стандартних. З другого ящика беруть навмання одну деталь і перекладають її до першого ящика. Знайти ймовірність того, що взята після цього навмання деталь з першого ящика стандартна.

Розв'язання

Позначимо події:

$A = \{\text{з першого ящика взято стандартну деталь}\}$

$B_1 = \{\text{з другого ящика переклали до першого стандартну деталь}\}$

$B_2 = \{\text{з другого ящика переклали до першого нестандартну деталь}\}$

Події B_1, B_2 — несумісні, а подія A може з'явитися лише сумісно з однією з них:

Ймовірності гіпотез:

$$P(B_1) = \frac{9}{10}; \quad P(B_2) = \frac{1}{10}$$

Тоді умовні ймовірності події A :

$$P_{B_1}(A) = \frac{16}{21} \quad P_{B_2}(A) = \frac{15}{21}$$

Отже, ймовірність події обчислена за теоремою 6:

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) = \frac{9}{10} \cdot \frac{16}{21} + \frac{1}{10} \cdot \frac{15}{21} = \frac{144 + 15}{210} = \frac{159}{210} = \frac{53}{70}$$

Відповідь. $P(A) = 53/70$.

Зауваження

Оскільки в умовах теореми 6 невідомо, з якою подією із несумісних подій $B_1, B_2, \dots, B_n$ з'явиться подія A , кожному з цих подій вважають гіпотезою

Якщо випробування проведено та в результаті його подія A з'явилась, то умовна імовірність $P_A(B_k)$ може не дорівнювати $P(B_k)$, а їх порівняння дозволяє переоцінити ймовірність гіпотези при умові, що подія A з'явилась.

$$P_A(B_k) = \frac{P(B_k) \cdot P_{B_k}(A)}{\sum_{k=1}^n P(B_k) \cdot P_{B_k}(A)}, \quad k = 1, 2, \dots, n$$

Приклад 6

Друковані плати поступають для перевірки їх стандартності до одного з двох контролерів. Ймовірність того, що плата попаде до першого контролера, дорівнює 0,6, а до другого — 0,4. Імовірність того, що придатний виріб буде визнаний стандартним першим контролером, дорівнює 0,94, а другим — 0,98. Придатний виріб при перевірці визнаний стандартним. Знайти імовірність того, що виріб перевіряв перший контролер.

Розв'язання

Позначимо події $A = \{\text{придатний виріб признано стандартним}\}$

$B_1 = \{\text{виріб перевірено першим контролером}\}$

$B_2 = \{\text{виріб перевірено другим контролером}\}$

За умовою задачі $P(B_1) = 0,6$; $P(B_2) = 0,4$; $P_{B_1}(A) = 0,94$; $P_{B_2}(A) = 0,98$

$$P_A(B_1) = \frac{P(B_1) \cdot P_{B_1}(A)}{P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A)} = \frac{0,6 \cdot 0,94}{0,6 \cdot 0,94 + 0,4 \cdot 0,98} = 0,59$$

Відповідь. Після появи події A ймовірність перевірки виробу першим контролером — $P_A(B_1) = 0,59$ поменшала.

Приклад 7

Імовірність знищення БПЛА з одного пострілу з першого ПЗРК дорівнює 0,2, а з другого — 0,1. Кожен комплекс робить по одному пострілу, причому було одне влучення в апарат. Яка імовірність того, що влучив перший ПЗРК?

Розв'язання

Позначимо події: $A = \{\text{знищення БПЛА з одного пострілу першим ПЗРК}\}$

$B = \{\text{знищення БПЛА з одного пострілу другим ПЗРК}\}$

$C = \{\text{одне влучення в БПЛА}\}$

Маємо такі гіпотези які утворюють повну групу подій: $H_1 = A \cdot B; \quad H_2 = A \cdot \bar{B};$

Ймовірність цих подій: $H_3 = \bar{A} \cdot B; \quad H_4 = \bar{A} \cdot \bar{B}$

$$P(H_1) = 0,2 \cdot 0,1 = 0,02; \quad P(H_2) = 0,2 \cdot (1 - 0,1) = 0,18;$$

$$P(H_3) = (1 - 0,2) \cdot 0,1 = 0,08; \quad P(H_4) = (1 - 0,2) \cdot (1 - 0,1) = 0,72;$$

Тоді умовні імовірності події C будуть: $P_{H_1}(C) = 0; \quad P_{H_2}(C) = 1;$

$$P_{H_3}(C) = 1; \quad P_{H_4}(C) = 0;$$

Отже, ймовірність того, що влучив перший ПЗРК:

$$\begin{aligned} P_C(H_2) &= \frac{P(H_2) \cdot P_{H_2}(C)}{P(H_1) \cdot P_{H_1}(C) + P(H_2) \cdot P_{H_2}(C) + P(H_3) \cdot P_{H_3}(C) + P(H_4) \cdot P_{H_4}(C)} = \\ &= \frac{0,18 \cdot 1}{0,02 \cdot 0 + 0,18 \cdot 1 + 0,08 \cdot 1 + 0,72 \cdot 0} = 0,69 \end{aligned}$$

Відповідь. Ймовірність того, що влучив перший ПЗРК — 0,69.

Питання та завдання до контролю знань студентів

1. Як формулюють і якими формулами записують теореми додавання ймовірностей сумісних та несумісних подій?
2. Які випадкові події називають незалежними?
3. Як визначають та позначають умовну ймовірність?
4. Як формулюють і якими формулами записують теореми множення ймовірностей залежних та незалежних випадкових подій?
5. За якою формулою можна обчислити ймовірність появи хоча б однієї з n сумісних подій?
6. Яким умовам повинна задовольняти подія, щоб її ймовірність можна було знаходити за формулою повної ймовірності? Який вигляд має ця формула?
7. Коли застосовують формули Байєса та як їх записують?
8. Що називається повною групою подій?
9. Які події називаються сумісними?
10. Як записується формула повної ймовірності?
11. Що таке гіпотеза в формулі повної ймовірності?
12. Які обмеження накладаються на гіпотези у формулі повної ймовірності?
13. При яких умовах застосовується теорема гіпотез?
14. Що дозволяє оцінювати формула Байєса?
15. Чи можна переоцінити ймовірність гіпотези до того, як став відомий результат випробування?

Лекція № 4

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 1.3 Послідовності випробувань

Схема та формула Бернуллі. Граничні теореми у схемі Бернуллі. Локальна та інтегральна теореми Лапласа.

Мета заняття Обумовити доцільність використання формули Бернуллі, надати уявлення щодо використання апроксимаційних формул.

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН

конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Формула Бернуллі. 2 Апроксимаційна формула Пуасона. 3 Локальна функція Лапласа. 4 Інтегральна функція Лапласа.
4. Заклучна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е. Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.
3. Буряк, Г.І. Методика застосування засобів Microsoft Excel при розв'язанні задач теорії ймовірностей [Текст] / Г.І. Буряк, – КРКМ ДНУ ім. О. Гончара, 2017. – 67 с.

ВСТУП

У багатьох задачах теорії ймовірностей, статистики та повсякденної практики треба досліджувати послідовність (серію) n випробувань. Наприклад, випробування «кинуто 1000 однакових монет» можна розглядати як послідовність 1000 більш простих випробувань — «кинута одна монета». При киданні 1000 монет імовірність появи герба або надпису на одній монеті не залежить від того, що з'явиться на інших монетах. Тому можна казати, що у цьому випадку випробування повторюються 1000 разів незалежним чином.

Конкретну відповідь на питання "чи можна передбачити, яка приблизно буде відносна частота появ події?" дає теорема, доведена Якобом Бернуллі (опублікована в 1713 р.), яка отримала назву «закону великих чисел» і поклала початок теорії ймовірностей як науки. Доказ Бернуллі було складним більш просте доказ дано П. Л. Чебишевим в 1846 р.

1. Формула Бернуллі

Якщо усі n випробувань проводити в однакових умовах та ймовірність появи події $p = P(A)$ в усіх випробуваннях однакова і не залежить від появи або не появи $q = P(\bar{A})$ в інших випробуваннях, то таку послідовність незалежних випробувань називають схемою Бернуллі.

Ймовірність того, що в n незалежних випробуваннях, у кожному з яких імовірність появи події дорівнює p ($0 < p < 1$), подія наступить рівно m разів (байдуже, в якій послідовності), дорівнює

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

Зауваження 1.1

Ймовірність появи події A в n випробуваннях схеми Бернуллі **менше m** разів знаходять за формулою:

$$P_n(k < m) = P_n(0) + P_n(1) + P_n(2) + \dots + P_n(m-1)$$

Зауваження 1.2

Ймовірність появи події A в n випробуваннях схеми Бернуллі **не менше m** разів знаходять за формулою:

$$P_n(k \leq m \leq n) = P_n(m) + P_n(m+1) + P_n(m+2) + \dots + P_n(n)$$

Зауваження 1.3

Ймовірність появи події A в n випробуваннях схеми Бернуллі **хоча б один раз** знаходять за формулою:

$$P_n(1 \leq m \leq n) = 1 - q^n$$

Зауваження 2.1

У багатьох випадках треба знаходити **найбільш імовірне значення m_0** числа m появ події A . Це значення m визначається співвідношеннями:

$$np - q \leq m_0 \leq np - q$$

або

$$(n+1)p - 1 \leq m_0 \leq (n+1)p$$

Зауваження 2.2

Число m_0 повинно бути цілим. Якщо $(n+1)p$ — ціле число, тоді найбільше значення імовірності має при двох числах:

$$m_1 = (n+1)p - 1$$

та

$$m_2 = (n+1)p$$

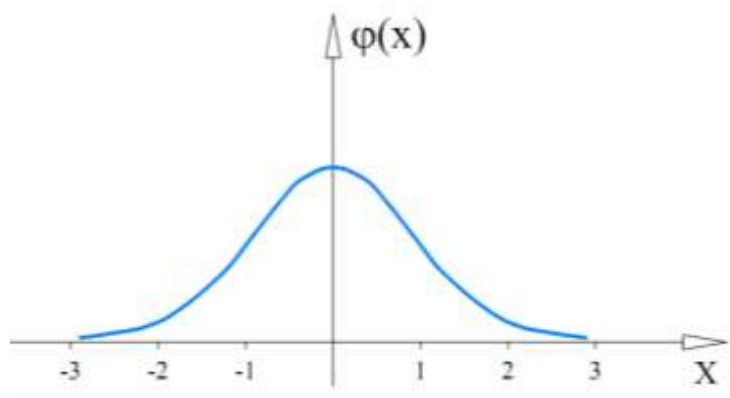
Зауваження 3

Якщо імовірність появи події A в кожному випробуванні дорівнює p , то кількість n випробувань, які необхідно здійснити, щоб з імовірністю P можна було стверджувати, що подія A з'явиться хоча б один раз, знаходять за формулою:

$$n > \frac{\ln(1-P)}{\ln(1-p)}$$

2. Локальна теорема Лапласа

Ймовірність того, що в n незалежних випробуваннях, у кожному з яких імовірність появи події дорівнює p ($0 < p < 1$), подія наступить рівно m разів (байдуже, в якій послідовності), наближено дорівнює (тим точніше, чим більше n)



$$P_n(m) = \frac{1}{\sqrt{npq}} \varphi(x)$$

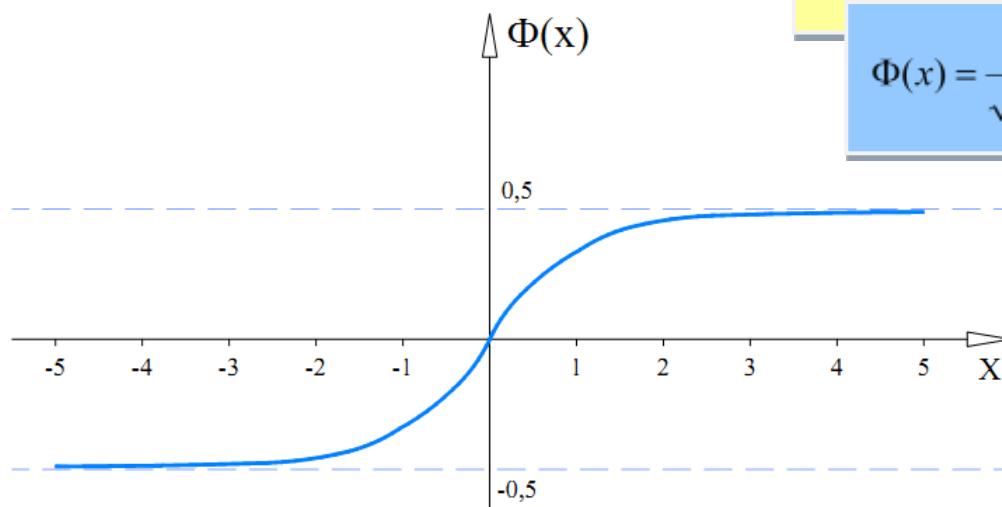
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$x = \frac{m - np}{\sqrt{npq}}$$

Рисунок 1 – Графік локальної функції Лапласа

3. Інтегральна теорема Лапласа

Ймовірність того, що в n незалежних випробуваннях, у кожному з яких імовірність появи події дорівнює p ($0 < p < 1$), подія наступить не менш m_1 разів і не більше m_2 разів, наближено дорівнює



$$P(m_1; m_2) = \Phi(x'') - \Phi(x')$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$$

$$x' = \frac{m_1 - np}{\sqrt{npq}}$$

$$x'' = \frac{m_2 - np}{\sqrt{npq}}$$

Рисунок 2 – Графік інтегральної функції Лапласа

У схемі Бернуллі імовірність появи події A в усіх випробуваннях однакова. Але у практичній діяльності іноді зустрічаються і такі випадки, коли у n незалежних випробуваннях імовірності появи події A різні. У цьому випадку не можна обчислювати за формулою Бернуллі імовірність появи події A m разів у n випробуваннях, а треба використовувати твірну функцію.

4. Послідовність випробувань із різними ймовірностями

Якщо у n незалежних випробуваннях ймовірності появи події A різні $p_1, p_2, \dots, p_n$. Тоді імовірності не появи події A тобто $\bar{A}$ також будуть різними:

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

У цьому випадку обчислювати ймовірність появи події A m разів у n випробуваннях треба з використанням твірної функції.

Правило

Шукана імовірність $P_n(m)$ дорівнює коефіцієнту, що стоїть при z^m

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

Приклад

Ймовірність відмови кожного з 4 ноутбуків у 4 незалежних випробуваннях різні та дорівнюють $p_1 = 0,1$ $p_2 = 0,2$ $p_3 = 0,3$ $p_4 = 0,4$

Знайти ймовірність того, що внаслідок випробувань:

- а) не відмовить жоден ноутбук;
- б) відмовлять один, два, три, чотири ноутбуки;
- в) відмовить хоча б один ноутбук;
- г) відмовлять не менше двох ноутбуків

Розв'язання

Твірна функція має вигляд: $\varphi_4(z) = (q_1 + p_1z)(q_2 + p_2z)(q_3 + p_3z)(q_4 + p_4z)$

$$\varphi_4(z) = (0,9 + 0,1z)(0,8 + 0,2z)(0,7 + 0,3z)(0,6 + 0,4z);$$

$$\varphi_4(z) = 0,3024 + 0,4404z + 0,2144z^2 + 0,0404z^3 + 0,0024z^4$$

а) $P_4(0) = 0,3024$

б) $P_4(1) = 0,4404$; $P_4(2) = 0,2144$; $P_4(3) = 0,0404$; $P_4(4) = 0,0024$

в) $P_4(1 \leq m \leq 4) = 1 - P_4(0) = 1 - 0,3024 = 0,6976$

г) $P_4(m \geq 2) = 1 - (P_4(0) + P_4(1)) = 1 - (0,3024 + 0,4404) = 0,2572$

5. Ймовірність відхилення відносної частоти від постійної ймовірності

Границя імовірності відхилення відносної частоти події **A** від її імовірності на величину, що більше або дорівнює **a**, дорівнює нулеві.

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{m}{n} - p\right| \geq a\right) = 0$$

Згідно означенню границі рівність

$$P\left(\left|\frac{m}{n} - p\right| \geq a\right) = \alpha \quad \text{— нескінченно мала величина}$$

$$\left|\frac{m}{n} - p\right| \geq a \quad \text{— практично неможлива}$$

$$\left|\frac{m}{n} - p\right| < a \quad \text{— практично достовірна}$$

$$\left|\frac{m}{n} - p\right| = 0 \quad \begin{array}{l} \text{може відрізнитись від} \\ \text{практично достовірної} \\ \text{події на нескінченно малу} \\ \text{величину} \end{array} \quad \left|\frac{m}{n} - p\right| < a$$

Це означає, що прямує до ймовірності, тобто відносна частота (частість) події A відрізняється від ймовірності події A на нескінченно малу величину, яку практично можна не враховувати.

Зауваження

З використанням інтегральної функції Лапласа $\Phi(x)$ формула набуває вигляд

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) = 2\Phi\left(\varepsilon \cdot \sqrt{\frac{n}{pq}}\right)$$

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

Питання та завдання до контролю знань студентів

1. Яка послідовність випробувань утворює схему Бернуллі? Яку формулу звуть формулою Бернуллі і що вона дозволяє обчислювати?
2. За якими формулами знаходять ймовірність появи події A менше m або не менше за m разів у n випробуваннях схеми Бернуллі?
3. За якою формулою знаходять ймовірність появи події A хоча б один раз у n випробуваннях?
4. Як можна знайти найбільш ймовірне значення числа появ події A у схемі Бернуллі?
5. Як можна знайти кількість випробувань у схемі Бернуллі, яка дозволяє з ймовірністю P стверджувати, що подія A з'явиться хоча б один раз?
6. У яких випадках доцільно використовувати граничні теореми у схемі Бернуллі?
7. Коли доцільно застосовувати формули Пуассона, локальну або інтегральну формули Муавра-Лапласа? Як визначаються і які мають властивості локальна та інтегральна функції Лапласа?
8. Як формулюється теорема Бернуллі і який вона має наслідок?
9. Який існує зв'язок між твердженням теореми Бернуллі та інтегральною функцією Лапласа? Які задачі дозволяє розв'язувати цей зв'язок?
10. Яка послідовність випробувань не вписується у схему Бернуллі?

Лекція № 5

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 1.4 Випадкові величини

Види випадкових величин та способи їх задання. Закони розподілу та числові характеристики дискретних випадкових величин.

Мета заняття Дати уявлення щодо випадкової величини, класифікувати, показати закони розподілу ймовірностей.

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН
конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Біноміальний розподіл. 2 Розподіл Пуассона, найпростіший потік подій. 3 Геометричний та гіпергеометричний розподіли. 4 Інтегральна та диференціальна функції розподілу.
4. Заклучна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е.Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.

ВСТУП

Для комфорту у будівлі — випадкові величини,
їх розподілу закони вимагають пам'ятати:
як знаходять інтеграли, похідні як треба брати.
Головні характеристики випадкових величин є основою статистики.
Тому треба без лінця вивчити їх до кінця.

При дослідженні багатьох проблем виникають такі випадкові події, наслідком яких є поява деякого числа, заздалегідь невідомого. Тому такі числові значення — випадкові.

Прикладом такої події є: кількість очок, що випадає при киданні грального кубика; кількість студентів, які прийдуть на лекцію; кількість цукрового буряка, який чекають одержати з одного гектара.

1. Випадкові величини

Випадковою величиною називають таку величину, яка в наслідок випробування може прийняти лише одне числове значення, заздалегідь невідоме і обумовлене випадковими причинами

Випадкові величини доцільно позначати великими літерами латинського алфавіту, а їх можливі значення — відповідними малими літерами з індексами:

$$X: x_1, x_2, x_3, \dots x_n$$

$$Y: y_1, y_2, y_3, \dots y_n$$

$$Z: z_1, z_2, z_3, \dots z_n$$

Через p_i позначають ймовірність події — випадкова величина X набула значення x_i . Це можна записати так: $P(X = x_i) = p_i$ (де $i = 1, 2, \dots, 11$).

2. Дискретні випадкові величини

Дискретною випадковою величиною (ДВВ) називають таку величину, яка може приймати відокремлені ізольовані одне від одного числові значення (їх можна пронумерувати) з відповідними ймовірностями.

Між двома сусідніми можливими значеннями дискретної випадкової величини немає можливих значень.

Іншими словами, можливі значення дискретної випадкової величини можна перенумерувати. Число можливих значень дискретної випадкової величини може бути скінченим або нескінченним (в останньому випадку множину всіх можливих значень називають зліченою).

3. Законом розподілу дискретної випадкової величини

Законом розподілу дискретної випадкової величини називають таке співвідношення, яке встановлює зв'язок між можливими значеннями випадкової величини і відповідними їм ймовірностями.

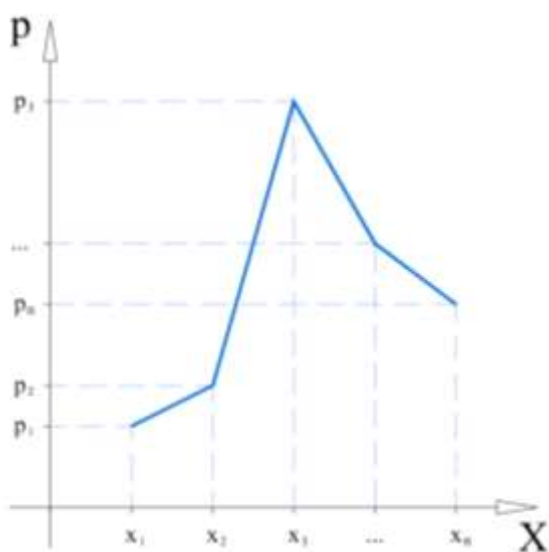


Рисунок 1

При табличному заданні закону розподілу дискретної випадкової величини перший рядок таблиці містить можливі значення, а друга — їх ймовірності.

повна група					
X	x_1	x_2	x_3	$\dots$	x_n
p	p_1	p_2	p_3	$\dots$	p_n

$\Sigma p_k = 1$

4. Аналітичні закони розподілу дискретної випадкової величини

Біноміальний закон розподілу

Нехай проводиться n незалежних випробувань, у кожному з яких подія A може з'явитися або не з'явитися. Ймовірність настання події в усіх випробуваннях постійна і дорівнює p . Отже випробування вписуються в схему Бернуллі.

Біноміальним називають закон розподілу дискретної випадкової величини X — числа появ події в n незалежних випробуваннях, у кожному з яких імовірність появи події дорівнює p ; ймовірність можливого значення $X = m$; (числа m появ події) обчислюють за формулою Бернуллі:

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

Закон розподілу Пуассона

Якщо у схемі незалежних повторних випробувань n досить велике, а p або q прямує до нуля, то біноміальний розподіл апроксимує розподіл Пуассона, параметр якого $\lambda = np$, причому при $p \leq 0,1$ або $p \geq 0,9$ ця апроксимація дає гарні результати незалежно від величини n .

Геометричний розподіл

Геометричний розподіл застосовують у різноманітних задачах теорії надійності та у страхових розрахунках. Ряд ймовірностей цього розподілу буде нескінченно спадною геометричною прогресією із знаменником q , сума якої дорівнює одиниці.

Геометричний розподіл вказує імовірність появи m елементів з певною властивістю серед n елементів, взятих із сукупності N елементів, яка містить m елементів саме такої властивості.

$$P_n(X = m) = pq^{m-1}$$

Геометричний розподіл

Геометричний розподіл використовують у багатьох задачах статистичного контролю якості.

Геометричний розподіл вказує імовірність появи m елементів з певною властивістю серед n елементів, взятих із сукупності N елементів, яка містить k елементів саме такої властивості.

$$P_n(X = m) = \frac{C_k^m \cdot C_{N-k}^{n-m}}{C_N^n}$$

Якщо об'єм вибірки n малий у порівнянні з об'ємом N сукупності, тобто $\frac{n}{N} \leq 0,1$; $\frac{k}{N} \leq 0,1$, то імовірності у гіпергеометричному розподілі будуть близькими до відповідних ймовірностей біноміального розподілу з $p = \frac{k}{N}$.

У статистиці це означає, що розрахунки ймовірностей для без повторної вибірки будуть мало відрізнятись від розрахунків ймовірностей для повторної вибірки.

5. Числові характеристики випадкових величин

В практичній діяльності не завжди вдається одержати закон розподілу, або закон надто складний для практичних розрахунків. Тому з'явилася потреба характеризувати дискретні випадкові величини за допомогою числових характеристик, які характеризують особливості випадкових величин достатньо.

Математичне сподівання

Математичним сподіванням дискретної випадкової величини називають число, яке дорівнює сумі добутків усіх її можливих значень на відповідні їм ймовірності.

$$M(X) = \sum_{k=1}^n x_k \cdot p_k$$

Основні властивості математичного сподівання

1) Математичне сподівання постійної величини дорівнює самій постійній

$$M(C) = C$$

2) Постійний множник можна виносити за знак математичного сподівання

$$M(CX) = C \cdot M(X)$$

3) Математичне сподівання добутку декількох взаємно незалежних дискретних випадкових величин дорівнює добутку їх математичних сподівань, тобто

$$M(X_1 \cdot X_2 \cdot \dots \cdot X_n) = M(X_1) \cdot M(X_2) \cdot \dots \cdot M(X_n)$$

4) Математичне сподівання суми випадкових величин дорівнює сумі їх математичних сподівань, тобто

$$M(X_1 + X_2 + \dots + X_n) = M(X_1) + M(X_2) + \dots + M(X_n)$$

Дисперсією дискретної випадкової величини

Дисперсією дискретної випадкової величини називають число, яке дорівнює математичному сподіванню квадрата відхилення дискретної випадкової величини від її математичного сподівання.

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

Основні властивості дисперсії

Дисперсія сталої дорівнює нулю

$$D(C) = 0$$

Дисперсія будь-якої дискретної випадкової величини — невід'ємна

$$D(X) \geq 0$$

Постійний множник можна виносити за знак дисперсії, при цьому постійний множник треба піднести у квадрат

$$D(CX) = C^2 \cdot D(X)$$

Дисперсія дискретної випадкової величини дорівнює різниці між математичним сподіванням квадрата випадкової величини X та квадрата її математичного сподівання

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

Дисперсія алгебраїчної суми дискретних випадкових величин X та Y дорівнює сумі їх дисперсій

$$D(X \pm Y) = D(X) + D(Y)$$

Середнє квадратичне відхилення

Середнє квадратичне відхилення дорівнює квадратному кореню з дисперсії

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

Початковим моментом порядку k випадкової величини X називають математичне сподівання величини X^k

$$\nu_k = M(X^k), k = 1, 2, \dots, n$$

Центральним моментом порядку k випадкової величини X називають математичне сподівання величини $[X - M(X)]^k$

$$\mu_k = M([X - M(X)]^k)$$

Висновки

Оскільки дискретні випадкові величини широко застосовуються в багатьох галузях, таких як статистика, економіка, інженерія, інформатика та інші, то розуміння дискретних випадкових величин є важливим для аналізу та моделювання різних реальних явищ, де випадковість відіграє роль. Бажано розуміти, що функції розподілу дискретних випадкових величин допомагають обчислити ймовірності подій та характеристики розподілу.

Питання та завдання до контролю знань студентів

- 1 Що таке дискретна випадкова величина і як вона відрізняється від неперервної випадкової величини?
- 2 Які основні характеристики використовуються для опису дискретних випадкових величин?
- 3 Що таке ймовірнісна функція маси (PMF) і як вона пов'язана з дискретною випадковою величиною?
- 4 Які властивості має кумулятивна функція розподілу (CDF) для дискретних випадкових величин?
- 5 Як обчислити очікуване значення (expected value) дискретної випадкової величини?
- 6 Як виміряти розсіювання дискретної випадкової величини і які характеристики використовуються для цього?
- 7 Які основні дискретні випадкові величини ви знаєте і які їх особливості?
- 8 Як визначити ймовірності подій, пов'язаних з дискретними випадковими величинами?
- 9 Які методи можна використовувати для моделювання дискретних випадкових величин?
- 10 Які приклади застосування дискретних випадкових величин у різних галузях, наприклад, в статистиці, економіці або інженерії?

Лекція № 6

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 1.4 Випадкові величини

Закони розподілу та числові характеристики законів розподілу неперервних випадкових величин.

Мета заняття Розглянути систему двох випадкових величин, розподіл функції одного та двох випадкових аргументів.

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН

конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Поняття щодо системи декількох випадкових величин. 2 Функція та числові характеристики двовимірних випадкових величини. 3 Умовні закони розподілу системи. 4 Дискретна та неперервна двовимірні випадкові величини.
4. Заклучна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е.Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.

ВСТУП

Неперервні випадкові величини є важливою складовою статистики та імовірнісного аналізу і використовуються для моделювання та вивчення різноманітних явищ у багатьох галузях науки та практиці. Вони відрізняються від дискретних випадкових величин тим, що можуть приймати значення на нескінченному проміжку або на континуумі чисел.

Неперервні випадкові величини відображають реальні фізичні величини, такі як час, відстань, швидкість, маса, температура тощо. Одним з головних понять, пов'язаних з неперервними випадковими величинами, є їх щільність розподілу.

Щільність розподілу та її властивості

Щільністю розподілу

ймовірностей неперервної випадкової величини X називають функцію $f(x)$ — першу похідну від функції розподілу $F(x)$:

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

Теорема

Ймовірність того, що неперервна випадкова величина X прийме значення, що належить інтервалу $(a; b)$, дорівнює визначеному інтегралу від щільності розподілу, взятому в межах від a до b :

$$P(a < X < b) = \int_a^b f(x) dx$$

Властивість 1

Щільність розподілу — невід'ємна функція

$$f(x) \geq 0$$

Властивість 2

Невласний інтеграл від щільності розподілу в межах від $-\infty$ до ∞ дорівнює одиниці:

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

Приклад 6.1 Дискретна випадкова величина X задана законом розподілу. Знайти центральні моменти першого, другого, третього і четвертого порядків.

$$F(X) = \begin{cases} 0 & \text{при } x \leq 0, \\ \sin 2x & \text{при } 0 < x \leq \frac{\pi}{4}, \\ 1 & \text{при } x > \frac{\pi}{4}. \end{cases}$$

Розв'язання

Оскільки щільність розподілу дорівнює першій похідній від функції розподілу,

то

$$f(x) = F'(X) = \begin{cases} (0)' & \text{при } x \leq 0, \\ (\sin 2x)' & \text{при } 0 < x \leq \frac{\pi}{4}, \\ (1)' & \text{при } x > \frac{\pi}{4}. \end{cases}$$

Похідна складеної функції:

$$(\sin 2x)' = \left\{ \begin{array}{l} y = \sin u, \\ u = 2x \end{array} \right\} = y'_u \cdot u'_x = (\sin u)'_u \cdot (2x)'_x = \cos u \cdot 2 = 2 \cos 2x.$$

Тоді щільність розподілу:

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ 2 \cos 2x & \text{при } 0 < x \leq \frac{\pi}{4}, \\ 1 & \text{при } x > \frac{\pi}{4}. \end{cases}$$

Приклад 6.2 Знайти функцію розподілу по заданій щільності розподілу:

Розв'язання

Використаємо формулу

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

$$F(x) = \int_{-\infty}^x f(x) dx \text{ розглянемо:}$$

Якщо $x \leq a$, то $f(x) = 0$, отже $F(x) = 0$.

Якщо $a < x \leq b$, то $f(x) = \frac{1}{b-a}$, отже

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(x) dx = \int_{-\infty}^a 0 dx + \int_a^x \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^x dx = \frac{x}{b-a} \Big|_a^x = \\ &= \frac{x}{b-a} - \frac{a}{b-a} = \frac{x-a}{b-a}. \end{aligned}$$

Якщо $x > b$, то $f(x) = 0$, отже

$$F(x) = \int_{-\infty}^a 0 dx + \int_a^b \frac{dx}{b-a} + \int_b^x 0 dx = \frac{1}{b-a} \int_a^b dx = \frac{x}{b-a} \Big|_a^b = \frac{b}{b-a} - \frac{a}{b-a} = \frac{b-a}{b-a} = 1.$$

На рисунку 6.1 наведено графік функції розподілу неперервної випадкової величини.

Тоді функція розподілу:

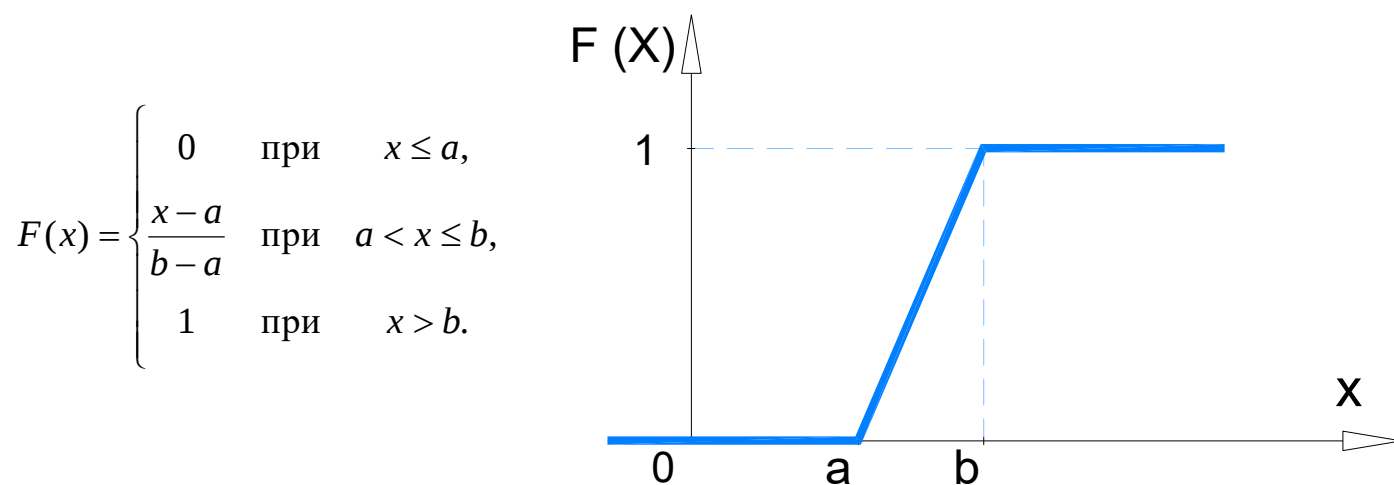


Рисунок 6.1 — Графік функції розподілу

Приклад 6.3 Щільності розподілу випадкової величини X : $f(x) = \frac{a}{e^{-x} + e^x}$

Знайти параметр a

Розв'язання

Оскільки щільність розподілу має задовольняти умові $\int_{-\infty}^{\infty} f(x) dx = 1$, то має

виконуватись рівність:

$$\int_{-\infty}^{\infty} \frac{a}{e^{-x} + e^x} dx = 1$$

$$\text{Звідси дістаємо } a \int_{-\infty}^{\infty} \frac{dx}{e^{-x} + e^x} = 1, \quad \Rightarrow \quad a = \frac{1}{\int_{-\infty}^{\infty} \frac{dx}{e^{-x} + e^x}}.$$

Розглянемо невизначений інтеграл

$$\int \frac{dx}{e^{-x} + e^x} = \int \frac{dx}{e^{-x}(1 + e^{2x})} = \int \frac{e^x dx}{1 + e^{2x}} = \left\{ \begin{array}{l} t = e^x; \\ dt = e^x dx \end{array} \right\} = \int \frac{dt}{1 + t^2} = \arctg t = \arctg e^x$$

Тоді

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{dx}{e^{-x} + e^x} &= \lim_{b \rightarrow -\infty} \int_b^0 \frac{dx}{e^{-x} + e^x} + \lim_{c \rightarrow \infty} \int_0^c \frac{dx}{e^{-x} + e^x} = \\ &= \lim_{b \rightarrow -\infty} (\arctg e^x \Big|_b^0) + \lim_{c \rightarrow \infty} (\arctg e^x \Big|_0^c) = \lim_{b \rightarrow -\infty} (-\arctg e^b) + \lim_{c \rightarrow \infty} (\arctg e^c) = \frac{\pi}{2}. \end{aligned}$$

Отже, параметр а:

$$a = \frac{1}{\pi/2} = \frac{2}{\pi}.$$

Числові характеристики неперервних випадкових величин

Числові характеристики дозволяють краще розуміти розподіл неперервної випадкової величини, її центральні тенденції, розсіювання та форму розподілу. Вони є важливими інструментами для аналізу та порівняння різних неперервних випадкових величин.

Неперервні випадкові величини мають ряд числових характеристик, які дозволяють описати їх розподіл та властивості.

Математичне сподівання неперервної випадкової величини

Математичне сподівання неперервної випадкової величини

X , можливі

значення якої належать всій осі Ox , визначається рівністю

$$M(X) = \int_{-\infty}^{\infty} x f(x) dx$$

де $f(x)$ — щільність розподілу неперервної випадкової величини X .

Теорема 1

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

$$M(X) = \int_a^b x f(x) dx$$

Теорема 2

Якщо $f(x)$ є щільність ймовірності X , неперервна випадкова величина Y є функцією випадкової величини X , тобто $Y = \varphi(X)$, тоді математичне сподівання Y знаходиться за формулою

$$M(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x) f(x) dx$$

Якщо щільність імовірності $f(x)$ — парна функція $f(-x) = f(x)$ то центр розподілу X співпадає з початком $M(X) = 0$. Якщо графік функції $f(x)$ симетричний відносно прямої $x = a$, то $M(X) = a$.

Дисперсія неперервної випадкової величини

Дисперсія неперервної випадкової величини

X , можливі значення якої належать всій осі Ox , визначається рівністю

$$D(X) = \int_{-\infty}^{\infty} [x - M(X)]^2 f(x) dx$$

Рівносильна рівність

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

Зауваження 1

Якщо всі можливі значення X належать лише інтервалу (a, b) , то

$$D(X) = \int_a^b [x - M(X)]^2 f(x) dx$$

Рівносильна рівність

$$D(X) = \int_a^b x^2 f(x) dx - M^2(X)$$

Зауваження 2

Якщо $Y = \varphi(X)$ — функція випадкового аргументу X , причому можливі значення X належать інтервалу (a, b) , то

$$D[\varphi(X)] = \int_a^b [\varphi(x) - M(\varphi(X))]^2 f(x) dx$$

Рівносильна рівність

$$D[\varphi(X)] = \int_a^b \varphi^2(x) f(x) dx - M^2(\varphi(X))$$

Приклад 6.4 Задана функція розподілу випадкової величини X . Знайти математичне сподівання та дисперсію:

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ x & \text{при } 0 < x \leq 1, \\ 1 & \text{при } x > 1. \end{cases}$$

Розв'язання

За означенням щільність розподілу:

$$f(x) = F'(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ 1 & \text{при } 0 < x \leq 1, \\ 0 & \text{при } x > 1. \end{cases}$$

Тоді математичне сподівання неперервної випадкової величини X :

$$M(X) = \int_0^1 x \cdot 1 dx = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2}.$$

Дисперсія неперервної випадкової величини X :

$$D(X) = \int_0^1 x^2 \cdot 1 dx - \left(\frac{1}{2} \right)^2 = \left. \frac{x^3}{3} \right|_0^1 - \frac{1}{4} = \frac{1}{12}.$$

Середнє квадратичне відхилення неперервної випадкової величини X :

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{1}{12}} = \frac{1}{2\sqrt{3}} \approx 0,2887.$$

Приклад 6.5 Випадкова величина X задана щільністю розподілу в інтервалі $(0; \pi)$; поза інтервалом $f(x) = 0$. Знайти математичне сподівання функції $Y = \varphi(X) = X^2$

Розв'язання

Скористуємось формулою для обчислення математичного сподівання функції $\varphi(X)$ від випадкового аргументу X :

$$M(X^2) = \frac{1}{2} \int_0^{\pi} x^2 \sin x dx$$

Обчислимо по частинах невизначений інтеграл $\frac{1}{2} \int x^2 \sin x dx$:

$$\begin{aligned} \frac{1}{2} \int x^2 \sin x dx &= \left\{ \begin{array}{l} u = x^2; \quad du = 2x dx; \\ dv = \sin x dx; \quad v = \int \sin x dx = -\cos x \end{array} \right\} = \frac{1}{2} \left[-x^2 \cos x + 2 \int x \cos x dx \right] = \\ &= \left\{ \begin{array}{l} u = x; \quad du = dx; \\ dv = \cos x dx; \quad v = \int \cos x dx = \sin x \end{array} \right\} = \frac{1}{2} \left[-x^2 \cos x + 2(x \sin x - \int \sin x dx) \right] = \\ &= \frac{-x^2 \cos x + 2x \sin x + 2 \cos x}{2}. \end{aligned}$$

Тоді математичне сподівання:

$$\begin{aligned} M(X^2) &= \frac{1}{2} \int_0^{\pi} x^2 \sin x dx \\ M(X^2) &= \frac{-x^2 \cos x + 2x \sin x + 2 \cos x}{2} \Big|_0^{\pi} = \\ &= \frac{-\pi^2 \cos \pi + 2\pi \sin \pi + 2 \cos \pi}{2} - \frac{2 \cos 0^\circ}{2} = \frac{\pi^2 - 2}{2} - 1 = \frac{\pi^2 - 4}{2}. \end{aligned}$$

Початкові та центральні моменти неперервної випадкової величини

Початковим моментом порядку k

неперервної випадкової величини X називають математичне сподівання величини X^k

$$\nu_k = \int_{-\infty}^{\infty} x^k f(x) dx$$

Центральний моментом порядку k

неперервної випадкової величини X визначається рівністю

$$\mu_k = \int_{-\infty}^{\infty} [x - M(X)]^k f(x) dx$$

Зауваження 1

Якщо всі можливі значення X належать лише інтервалу (a, b) , то

$$\nu_k = \int_a^b x^k f(x) dx$$

$$\mu_k = \int_a^b [x - M(X)]^k f(x) dx$$

Зауваження 2

Для обчислення центральних моментів зручно користуватися формулами, що виражають центральні моменти через початкові

$$\mu_1 = 0$$

$$\mu_2 = \nu_2 - \nu_1^2$$

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

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

Приклад 6.6 Випадкова величина X задана щільністю розподілу $f(x) = 0,5x$ в інтервалі $(0; 2)$; поза інтервалом $f(x) = 0$. Знайти початкові та центральні моменти першого, другого, третього, четвертого порядків.

Розв'язання

Виконаємо розрахунок початкових моментів:

– початковий момент першого порядку

$$v_1 = \int_0^2 x \cdot (0,5x) dx = 0,5 \cdot \left. \frac{x^3}{3} \right|_0^2 = \frac{0,5 \cdot 2^3}{3} - \frac{0,5 \cdot 0^3}{3} = \frac{4}{3};$$

– початковий момент другого порядку

$$v_2 = \int_0^2 x^2 \cdot (0,5x) dx = 0,5 \cdot \left. \frac{x^4}{4} \right|_0^2 = \frac{0,5 \cdot 2^4}{4} - \frac{0,5 \cdot 0^4}{4} = 2;$$

– початковий момент третього порядку

$$v_3 = \int_0^2 x^3 \cdot (0,5x) dx = 0,5 \cdot \left. \frac{x^5}{5} \right|_0^2 = \frac{0,5 \cdot 2^5}{5} - \frac{0,5 \cdot 0^5}{5} = 3,2;$$

– початковий момент четвертого порядку

$$v_4 = \int_0^2 x^4 \cdot (0,5x) dx = 0,5 \cdot \left. \frac{x^6}{6} \right|_0^2 = \frac{0,5 \cdot 2^6}{6} - \frac{0,5 \cdot 0^6}{6} = \frac{16}{3}.$$

Для обчислення центральних моментів використаємо формули, що виражають центральні моменти через початкові:

$$\mu_1 = 0$$

$$\mu_2 = v_2 - v_1^2 = \frac{2}{9}$$

$$\mu_3 = v_3 - 3v_2v_1 + 2v_1^3 = -\frac{8}{135}$$

$$\mu_4 = v_4 - 4v_3v_1 + 6v_2v_1^2 - 3v_1^4 = \frac{16}{135}$$

Розподіли неперервних випадкових величин

Рівномірний розподіл

Рівномірним

називають розподіл ймовірностей неперервної випадкової величини X , якщо на інтервалі $(a; b)$, якому належать всі можливі значення X , щільність зберігає постійне значення, а саме

$$f(x) = \frac{1}{b-a}$$

поза цим інтервалом —

$$f(x) = 0$$

На рисунку 6.2 наведено графік щільності рівномірного розподілу неперервної випадкової величини X .

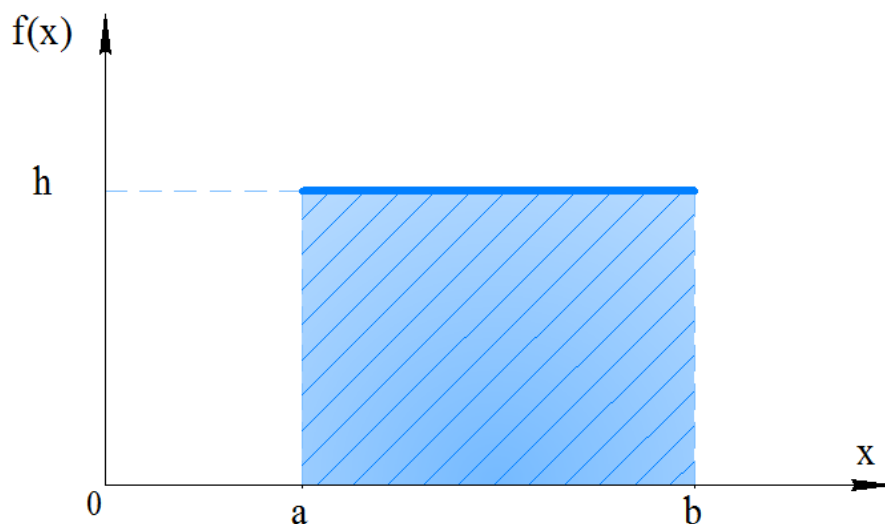


Рисунок 6.2 – Графік щільності розподілу неперервної випадкової величини X

Щільність функції рівномірного розподілу неперервної випадкової величини X :

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

Математичне сподівання неперервної випадкової величини X , у якій щільність зберігає постійне значення знаходиться за формулою:

$$M(X) = \int_a^b x f(x) dx = \frac{a+b}{2}$$

Дисперсія неперервної випадкової величини X , у якій щільність зберігає постійне значення знаходиться за формулою:

$$D(X) = \int_a^b x^2 f(x) dx = \frac{(b-a)^2}{12}$$

Середнє квадратичне відхилення неперервної випадкової величини X , у якій щільність зберігає постійне значення знаходиться за формулою:

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

Приклад 6.7



Ціна поділки шкали амперметра дорівнює 0,1 А. Показання амперметра округлюють до найближчого цілого ділення.

Знайти ймовірність того, що при відліку буде зроблена помилка, що перевищує 0,02 А.

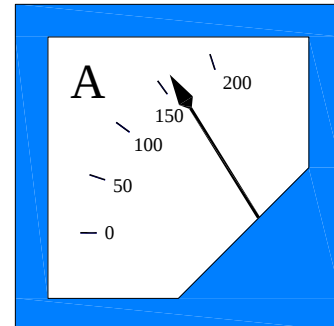
Розв'язання

Розглянемо помилку округлення відліку як випадкову величину X , яка розподілена рівномірно в інтервалі між двома сусідніми цілими поділами.

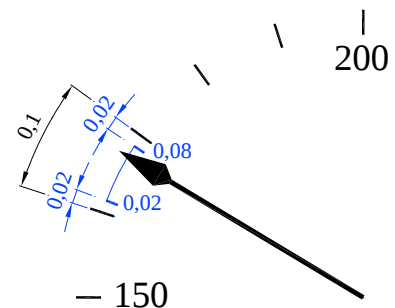
Оскільки довжина інтервалу, в якому замкнені можливі значення X , дорівнює 0,1 тому щільність — $f(x) = \frac{1}{0,1} = 10$.

Помилка відліку перевищить 0,02, якщо вона буде замкнена в інтервалі: (0,02; 0,08)

Отже, шукана ймовірність:



$$P(0,02 < X < 0,08) = \int_{0,02}^{0,08} 10dx = 10x \Big|_{0,02}^{0,08} = 0,6$$



Нормальний розподіл

Нормальним

називають розподіл ймовірностей неперервної випадкової величини X , який описується щільністю

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

де a — математичне сподівання;

σ — середнє квадратичне відхилення X .

Загальним

називають нормальний розподіл з довільними параметрами a та σ ($\sigma > 0$)

Нормованим

називають нормальний розподіл з параметрами $a = 0$ та $\sigma = 1$, а щільність нормованого розподілу:

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

Зокрема

Ймовірність того, що X , набуде значення, яке належить інтервалу $(\alpha; \beta)$

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$$

функція Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$$



П'єр-Сімон де Лаплас

1749 – 1827

Приклад 6.8 Математичне сподівання нормально розподіленої випадкової величини X дорівнює $M(X) = 3$, а дисперсія $D(X) = 16$. Скласти щільність ймовірності X .

Розв'язання

За формулою щільність нормованого розподілу:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} = \left\{ \begin{array}{l} a = M(X) = 3; \\ \sigma = \sqrt{D(X)} = \sqrt{16} = 4 \end{array} \right\} = \frac{1}{4\sqrt{2\pi}} e^{-\frac{(x-3)^2}{32}}.$$

Нормальною кривою

називають графік щільності нормального розподілу

На рисунку 6.3 наведено графік щільності нормального розподілу.

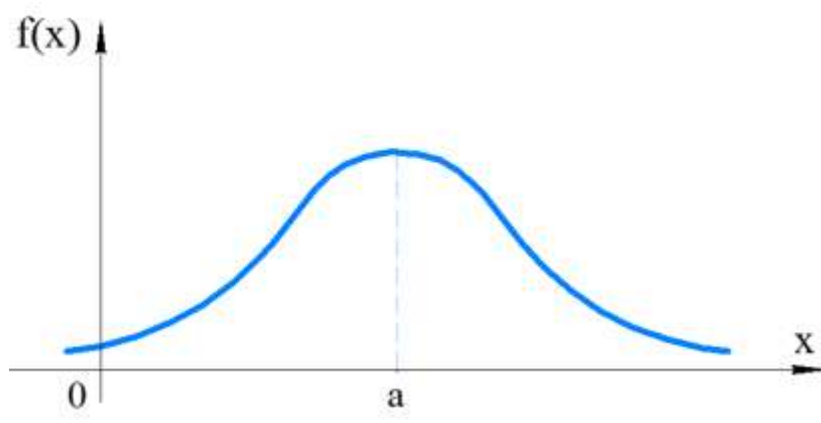


Рисунок 6.3 – Графік щільності нормального розподілу

Властивості функції щільності нормального розподілу

Область визначення функції щільності нормального розподілу $D(f) = (-\infty; \infty)$ — множина дійсних чисел.

Область значень функції щільності нормального розподілу $E(f) = (0; \frac{1}{\sigma\sqrt{2\pi}}]$.

Отже, нормальна крива розташована над Ox .

Границя функції при необмеженому зростанні x (по абсолютній величині) дорівнює нулю $\lim_{|x| \rightarrow \infty} f(x) = 0$

Перша похідна функції щільності нормального розподілу має вигляд:

$$f'(x) = -\frac{x-a}{\sigma^3\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Якщо перша похідна функції щільності нормального розподілу дорівнює нулю, то $f'(x) = 0 \Rightarrow x = a$.

Якщо перша похідна функції щільності нормального розподілу приймає додатне значення, то $f'(x) > 0 \Rightarrow x < a$.

Якщо перша похідна функції щільності нормального розподілу приймає від'ємне значення, то $f'(x) < 0 \Rightarrow x > a$.

Максимальне значення функція щільності приймає при значенні аргумента, що дорівнює математичному сподіванню випадкової величини нормального розподілу

$$f_{\max} = f(a) = \frac{1}{\sigma\sqrt{2\pi}}.$$

Друга похідна функції щільності нормального розподілу має вигляд:

$$f''(x) = -\frac{x-a}{\sigma^3\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \left[1 - \frac{(x-a)^2}{\sigma^2} \right]$$

Точки перегину графіка функції щільності нормального розподілу мають такі координати:

$$\left(a - \sigma; \frac{1}{\sigma\sqrt{2\pi e}} \right), \quad \left(a + \sigma; \frac{1}{\sigma\sqrt{2\pi e}} \right)$$

Вплив параметрів нормального розподілу на форму нормальної кривої

Зміна величини параметра a (математичного сподівання) не змінює форми нормальної кривої, а призводить лише до її зсуву вздовж осі Ox .

На рисунку 6.4 наведено результат розташування нормальної кривої, якщо математичне сподівання дорівнює нулю.

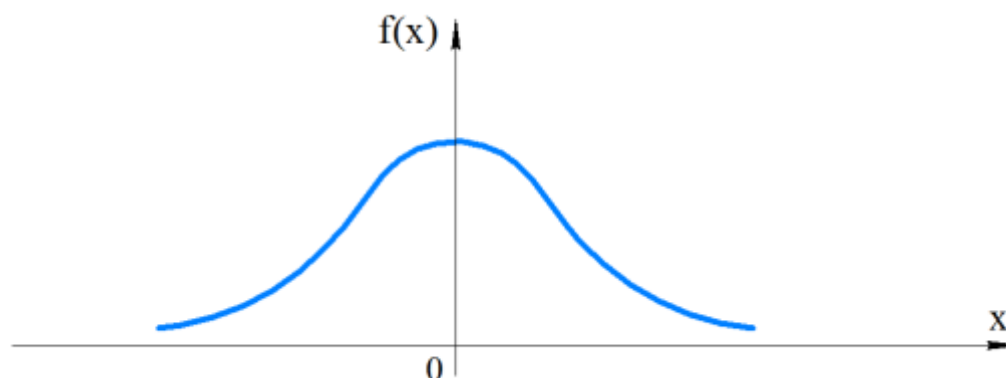


Рисунок 6.4 – Розташування нормальної кривої при нульовому значенні параметра a

На рисунку 6.5 наведено результат розташування нормальної кривої, якщо математичне сподівання зростає.

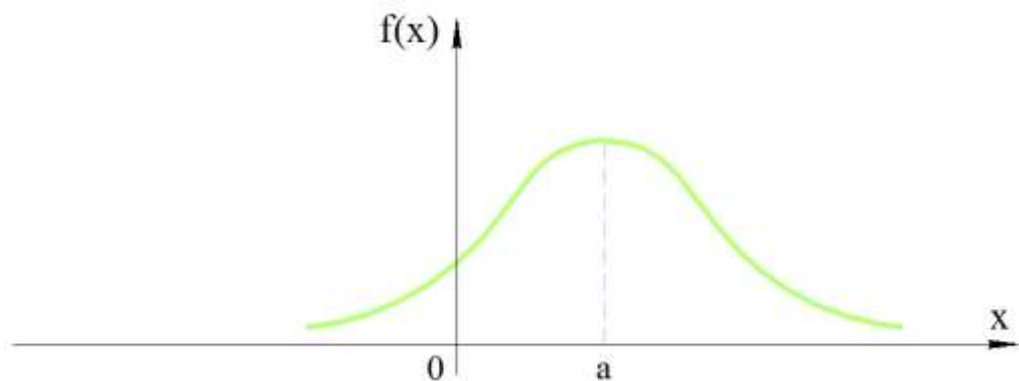


Рисунок 6.5 – Розташування нормальної кривої при зростанні параметра a

На рисунку 6.6 наведено результат розташування нормальної кривої, якщо математичне сподівання спадає.

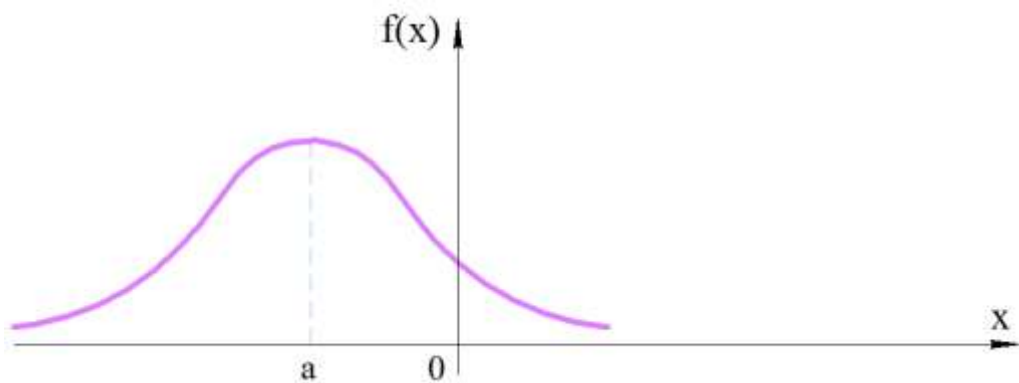


Рисунок 6.6 – Розташування нормальної кривої при спаданні параметра a

Зі зростанням параметра σ максимальна ордината нормальної кривої зменшується, а сама крива стає більш пологою, тобто стискається до осі Ox . Зі спаданням параметра σ нормальна крива стає більш гостровершинною та розтягується вздовж додатного напрямку осі Oy .

На рисунку 6.7 наведено результат впливу параметра σ на форму нормальної кривої.

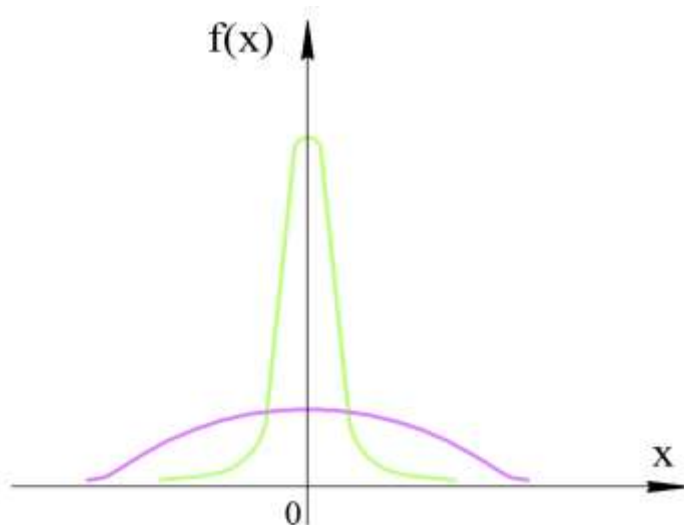


Рисунок 6.7 – Результат впливу на форму нормальної кривої при зміні параметра σ

Якщо дві випадкові величини нормально розподілені і $a = 0$, то ймовірність прийняти значення, що належить інтервалу $(-\delta; \delta)$, більше у той величини, яка має менше значення σ .

На рисунку 6.8 наведено ймовірнісний зміст параметра σ .

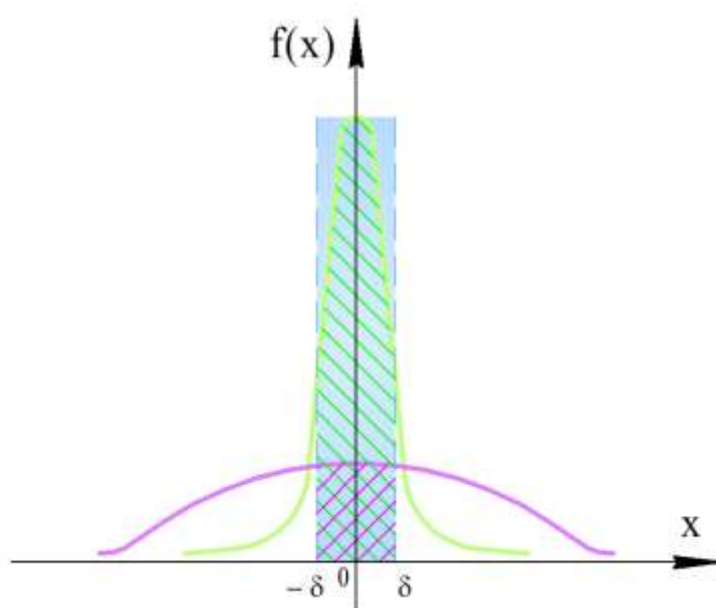


Рисунок 6.8 – Ймовірнісний зміст параметра σ

Обчислення ймовірності заданого відхилення у випадку, коли параметр $a \neq 0$:

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

Обчислення ймовірності заданого відхилення у випадку, коли параметр $a = 0$:

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

Правило трьох σ

Якщо $\delta = \sigma t$, тоді $P(|X - a| < \sigma t) = 2\Phi(t)$

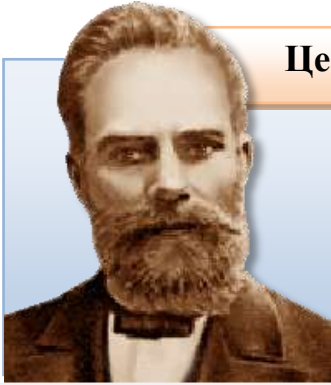
Якщо $t = 3$, тоді $\sigma t = 3\sigma$, тобто

$$P(|X - a| < 3\sigma) = 2\Phi(3) = 2 \cdot 0,49865 = 0,9973$$

Отже, ймовірність того, що відхилення за абсолютною величиною буде менше потроєного середнього квадратичного відхилення дуже мала та лише в 0,27% випадках таке може статися.

Теорема

Якщо випадкова величина розподілена нормально, то абсолютна величина її відхилення від математичного сподівання не перевершує потроєного середнього квадратичного відхилення.



А.М. Ляпунов

Центральна гранична теорема

Якщо випадкова величина X являє собою суму дуже великого числа взаємно незалежних випадкових величин, вплив кожної з яких на всю суму мізерно малий, то X має розподіл, близький до нормального.

До послідовності $X_1, X_2, \dots$ може бути застосована центральна гранична теорема, якщо при будь-якому x функція розподілу нормованої суми при $n \rightarrow \infty$ прямує до нормальної функції розподілу.

Оцінка відхилення теоретичного розподілу від нормального

При вивченні розподілів, відмінних від нормального, виникає необхідність кількісно оцінити цю різницю. З цією метою вводять спеціальні характеристики:



- асиметрією — A_s
- ексцес — E_k

Для нормального розподілу ці характеристики дорівнюють нулю.

Асиметрією теоретичного розподілу

Якщо випадкова величина називають відношення центрального моменту третього порядку до кубу середнього квадратичного відхилення

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

На практиці визначають знак асиметрії по розташуванню кривої розподілу відносно моди:

- якщо «довга частина» кривої розташована правіше моди, то асиметрія додатна;
- якщо лівіше — від'ємна.

На рисунку 6.9 наведено графік відхилення теоретичного розподілу від нормального при додатному значенні асиметрії

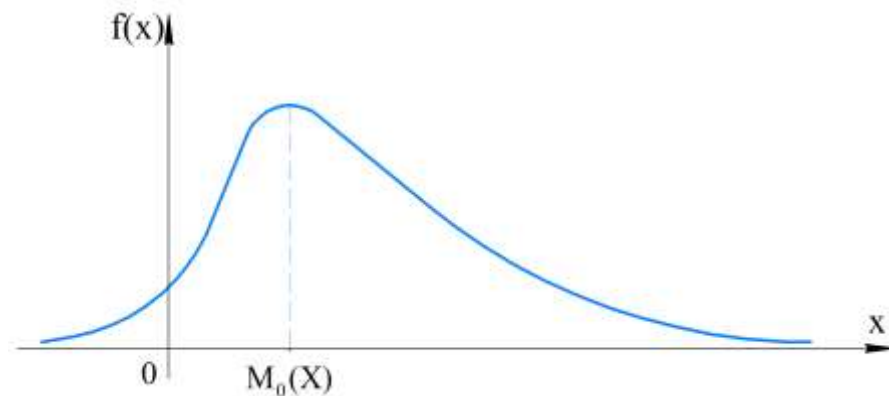


Рисунок 6.9– Відхилення теоретичного розподілу від нормального, якщо $A_S > 0$

На рисунку 6.10 наведено графік відхилення теоретичного розподілу від нормального при додатному значенні асиметрії.

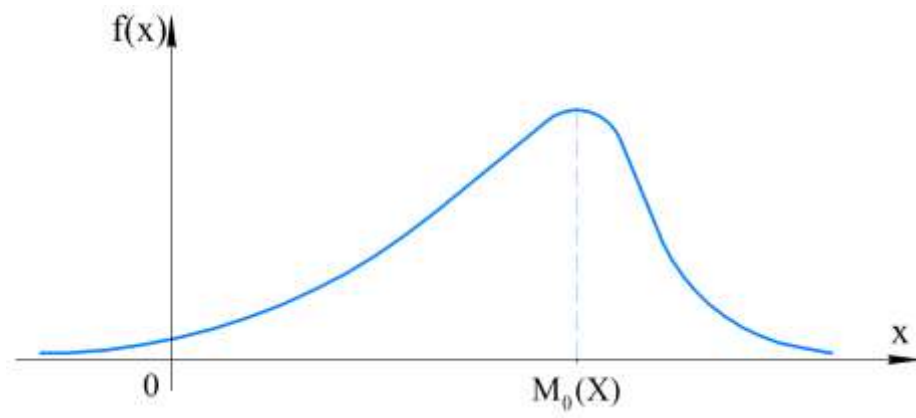


Рисунок 6.10 – Відхилення теоретичного розподілу від нормального, якщо $A_S < 0$

Ексцес теоретичного розподілу

— характеристика, що служить для оцінки більшого або меншого підйому кривої розподілу у порівнянні з нормальною кривою яка визначається рівністю

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

Якщо ексцес додатний, то крива має більш високу та «гостру» вершину, ніж нормальна крива.

На рисунку 6.11 наведено графік відхилення теоретичного розподілу від нормального при додатному значенні ексцесу.

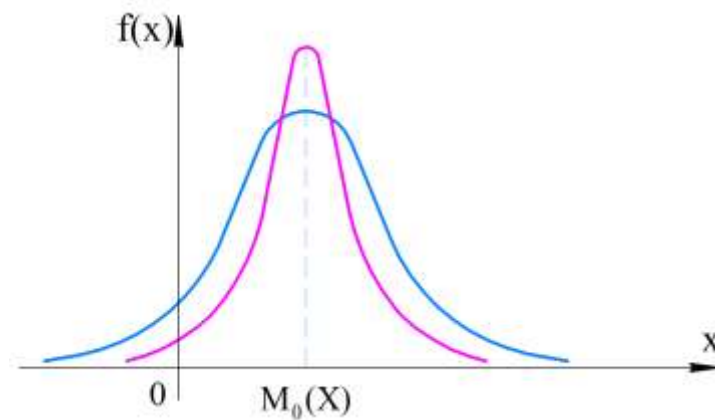


Рисунок 6.11 – Відхилення теоретичного розподілу від нормального, якщо $E_k > 0$

Якщо ексцес від'ємний, то порівнювана крива має нижчу та «плоску» вершину, ніж нормальна крива

На рисунку 6.12 наведено графік відхилення теоретичного розподілу від нормального при від'ємному значенні ексцесу.

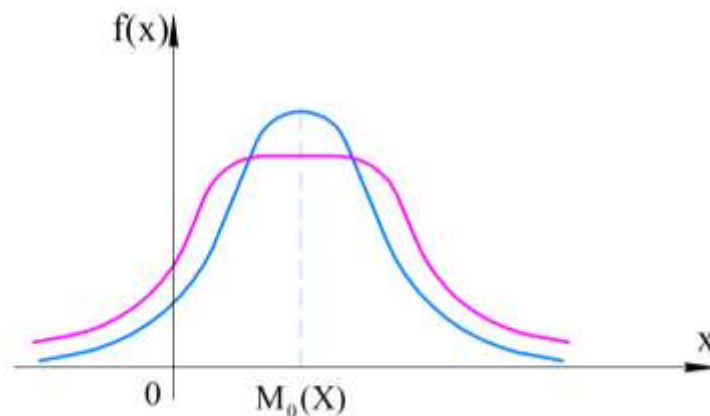


Рисунок 6.12 – Відхилення теоретичного розподілу від нормального, якщо $E_k < 0$

Показниковий розподіл

Неперервна випадкова величина X розподілена за *показниковим* (експоненціальним) законом з параметром λ , якщо її щільність розподілу має вигляд

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Показниковий розподіл задається одним параметром λ . Ця особливість показникового розподілу вказує на його перевагу порівняно з розподілами, які залежать від більшого числа параметрів.

Функцію розподілу $F(x)$ при $x > 0$ знаходимо за формулою

$$F(x) = \int_0^x f(x)dx = \lambda \int_0^x e^{-\lambda x} dx = -e^{-\lambda x} \Big|_0^x = 1 - e^{-\lambda x}.$$

Математичне сподівання $M(X)$ знайдемо застосувавши інтегрування частинами

$$M(X) = \lambda \int_0^{\infty} x e^{-\lambda x} dx = \lambda \lim_{b \rightarrow \infty} \int_0^b x e^{-\lambda x} dx = \frac{1}{\lambda}$$

Дисперсію $D(X)$ знаходимо так

$$D(X) = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2} = \lambda \lim_{b \rightarrow \infty} \int_0^b x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2}.$$

Інтегруючи частинами і застосувавши одержану формулу для $M(X)$, маємо

$$D(X) = \frac{1}{\lambda^2} \text{ і відповідно } \sigma(X) = \frac{1}{\lambda}.$$

Імовірність того, що випадкова величина X , яка має показниковий розподіл, у результаті випробування набуває можливого значення з інтервалу (a, b) , при $a > 0, b > 0$, обчислюється за формулою

$$P\{\alpha < X < \beta\} = \lambda \int_a^b e^{-\lambda x} dx = -e^{-\lambda x} \Big|_a^b = e^{-\lambda a} - e^{-\lambda b}.$$

Висновки

Закони розподілу неперервних випадкових величин є важливими для статистичного аналізу, прогнозування та прийняття рішень в багатьох галузях, включаючи науку, інженерію, фінанси та інші. Найбільш відомим і широко застосовуваним законом розподілу для неперервних випадкових величин є нормальний розподіл. Моделювання подій, що відбуваються з певною стійкою інтенсивністю здійснюється через залучення експоненційного розподілу, а рівномірний розподіл знаходить своє застосування в багатьох областях, де випадковість є рівномірною.

Розуміння та використання законів розподілу неперервних випадкових величин є важливими навичками для ефективного використання статистики та імовірнісного аналізу у дослідженнях та реальних ситуаціях.

Питання та завдання до контролю знань студентів

- 1 Що таке закон розподілу неперервної випадкової величини?
- 2 Які основні закони розподілу неперервних випадкових величин ви знаєте?
- 3 Які властивості характеризують нормальний розподіл?
- 4 В яких випадках застосовується експоненційний розподіл?
- 5 Які характеристики випадкової величини можна визначити за допомогою законів розподілу?
- 6 Які функції використовуються для опису законів розподілу неперервних випадкових величин?
- 7 Які інші закони розподілу неперервних випадкових величин можуть бути корисними для моделювання реальних процесів?
- 8 Які методи можуть використовуватись для оцінювання параметрів законів розподілу неперервних випадкових величин?
- 9 Чому вибір відповідного закону розподілу є важливим аспектом при моделюванні даних?
- 10 Які застосування мають закони розподілу неперервних випадкових величин у реальному житті?

Лекція № 7

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 2.1 Основні поняття та статистичний розподіл

Генеральна та вибіркова сукупності. Джерела даних, способи їх вибору та організації. Емпірична функція розподілу та її властивості. Графічне зображення статистичних розподілів.

Мета заняття Ознайомити із статистичним розподілом вибірки, надати уявлення щодо емпіричної функції розподілу.

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН
конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Емпірична функція розподілу. 2 Правила побудови полігону. 3 Правила побудови гістограми.
4. Заключна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е.Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.
3. Буряк, Г.І. Методика застосування засобів Microsoft Excel при розв'язанні задач теорії ймовірностей [Текст] / Г.І. Буряк, – КРКМ ДНУ ім. О. Гончара, 2017. – 67 с.
4. Дубовик, В.П. Вища математика. Збірник задач. [Текст] / В.П.Дубовик, І.І.Юрик. – К.: «А.С.К.», 2001. – 480 с.

ВСТУП

Перша задача математичної статистики вказати способи збору та угруповання статистичних відомостей, отриманих в результаті спостережень або в результаті спеціально поставлених експериментів. Друга задача математичної статистики розробити методи аналізу статистичних даних в залежності від цілей дослідження. Сюди відносяться:

- оцінка невідомої ймовірності події;
- оцінка невідомої функції розподілу;
- оцінка параметрів розподілу;
- оцінка залежності випадкової величини, тощо;
- перевірка статистичних гіпотез про вид невідомого розподілу або про величину параметрів розподілу.

Предмет математичної статистики полягає в розробці методів збору та обробки статистичних даних для одержання наукових та практичних висновків

Кожен об'єкт, який спостерігають, має декілька ознак. Розглядаючи лише одну ознаку кожного об'єкта, ми припускаємо, що інші ознаки рівноправні, або що множина об'єктів однорідна.

Статистичною сукупністю

називають множини однорідних об'єктів

Перевірку сукупності об'єктів можна провести двома способами:

- провести суцільне обстеження усіх об'єктів сукупності;
- обстежити лише певну частину об'єктів.

Якщо обстеження об'єкта пов'язано з його знищенням, то проводити суцільне обстеження практично не має сенсу

Якщо обстеження об'єкта пов'язано з його знищенням,



Генеральна та вибіркова сукупності

Вибірковою сукупністю (вибіркою)

називають сукупність випадково вибраних об'єктів.

Генеральною сукупністю

називають сукупність об'єктів, з яких зроблено вибірку.

Об'ємом сукупності (вибіркової або генеральної)

називають кількість об'єктів цієї сукупності.

Повторною

називають вибірку, при якій відібраний об'єкт повертається до генеральної сукупності перед відбором іншого об'єкта.

Безповторною

називають вибірку, якщо взятий об'єкт до генеральної сукупності не повертається.

Альтернативою вибірки є перепис — обстеження, що має своєю метою дослідження кожного елементу сукупності, що вивчається.

Зразками перепису можна вважати:

- перепис населення у країні;
- звіт про виробничі показники усіх підприємств аерокосмічної галузі

Вибірку можна ефективно використовувати для вивчення відповідної ознаки генеральної сукупності лише тоді, коли дані вибірки вірно відображають цю ознаку. Коротко ця умова формулюється так: вибірка повинна бути репрезентативною, тобто представницькою.

Згідно із законом великих чисел теорії ймовірностей можна стверджувати, що вибірка буде репрезентативною лише тоді, коли її здійснюють випадково.

Принципово способи відбору можна поділити на два види:

Відбір, який не потребує розчленування генеральної сукупності на частини

- простий випадковий безповторний відбір;
- простий випадковий повторний відбір.

Відбір, при якому генеральну сукупність розбивають на частини.

- типовий відбір;
- механічний відбір;
- серійний відбір.

Способи відбору

Простим випадковим

є такий відбір з генеральної сукупності, при якому кожний об'єкт, що витягається, має однакову ймовірність потрапити до вибірки.

Типовим

називають відбір, при якому об'єкти відбираються не з усієї генеральної сукупності, а з кожної її «типової» частини.

Механічним

називають відбір, при якому генеральну сукупність «механічно» ділять на стільки груп, скільки об'єктів повинно увійти у вибірку, а з кожної групи відбирають один об'єкт.

Серійним

називають відбір, при якому об'єкти відбирають з генеральної сукупності не по одному, а «серіями», які піддаються суцільному обстеженню.

На практиці часто застосовується комбінований відбір, при якому поєднуються вищезазначені способи.



Дані у статистиці надходять до дослідника у вигляді неорганізованої маси, незалежно від того, чи є вони вибірковими даними, чи даними з генеральної сукупності.

У математичній статистиці замість слова «дані» вживається термін «варіанти».

Числову характеристику варіанти при цьому називають ознакою.

Варіаційним рядом

називають послідовність варіант, записаних у зростаючому порядку $x_1, x_2, \dots, x_k$.

Розподіл частот

Нехай із генеральної сукупності взята вибірка із n варіант $x_1, x_2, \dots, x_k$ ознака X прийняла: x_1 прийняте значення n_1 раз,

x_2 прийняте значення n_2 раз,

x_3 прийняте значення n_3 раз,

$\vdots$ $\vdots$

x_k прийняте значення n_k раз,

Частотою

називають додатне число, що вказує, скільки раз та чи інша варіанта зустрічається в таблиці даних, а ряд $n_1, n_2, \dots, n_k$ — рядом частот.

Відносною частотою (частістю)

називають відношення частоти n_i варіанти x_i до об'єму вибірки n .

$$W_i = \frac{n_i}{n}$$

Статистичний розподіл вибірки

називають перелік варіант x_i варіаційного ряду та відповідних їм частот n_i або відносних частот W_i

Об'єм вибірки

— сума усіх частот вибірки

$$n_1 + n_2 + \dots + n_k = \sum_{i=1}^k n_i = n$$

Зауваження

Сума усіх відносних частот повинна дорівнювати одиниці

$$W_1 + W_2 + \dots + W_k = 1; \quad \sum_{i=1}^k W_i = 1$$

Приклад 7.1 Для вивчення потреб у певних розмірах взуття власник магазину спостерігає розміри взуття, проданого на протязі дня:



40, 35, 37, 39, 40, 41, 36, 42, 40, 39,
36, 43, 43, 41, 38, 37, 36, 42, 40, 38.

Розв'язання

В таблиці 7.1 наведено статистичний розподіл заданої вибірки.

Таблиця 7.1 – Розподіл частоти розміру взуття

x_i	35	36	37	38	39	40	41	42	43
n_i	1	2	2	3	2	2	2	2	4

Контроль:

$$1 + 2 + 2 + 3 + 2 + 2 + 2 + 2 + 4 = 20.$$

Приклад 7.2 Задано розподіл частот вибірки об'ємом $n = 20$ (таблиця 7.2).

Написати розподіл відносних частот

Таблиця 7.2 – Розподіл частот

x_i	2	6	12
n_i	3	10	7

Розв'язання

Обчислимо значення відносних частот:

$$W_1 = \frac{n_1}{n} = \frac{3}{20} = 0,15;$$

$$W_2 = \frac{n_2}{n} = \frac{10}{20} = \frac{1}{2} = 0,50;$$

$$W_3 = \frac{n_3}{n} = \frac{7}{20} = 0,35.$$

Таблиця 7.3 – Розподіл відносних частот:

x_i	2	6	12
W_i	0,15	0,50	0,35

Контроль: $\sum_{i=1}^3 W_i = 0,15 + 0,50 + 0,35 = 1.$

Емпірична функція розподілу

Емпіричною функцією розподілу

називають функцію $F^*(x)$, що визначає для кожного значення x відносну частоту події $X < x$.

$$F^*(x) = \frac{n_x}{n}$$

den_x — число варіант, менших за x

Теоретичною функцією розподілу

називають функцію розподілу $F(x)$ генеральної сукупності

$$F^*(x) = \frac{n_x}{n}$$

Різниця між емпіричною та теоретичною функціями у тому, що теоретична функція $F(x)$ визначає ймовірність події $X < x$, а емпірична функція $F^*(x)$ визначає відносну частоту цієї ж події. Отже, емпірична функція розподілу вибірки служить для оцінки теоретичної функції розподілу генеральної сукупності.

Приклад 7.3 Побудувати емпіричну функцію розподілу за наданим розподілом вибірки:

Таблиця 7.4 – Розподіл вибірки:

x_i	2	6	10
n_i	12	18	30

Розв'язання

Знайдемо об'єм вибірки:

$$n = \sum_{i=1}^3 n_i = n_1 + n_2 + n_3 = 12 + 18 + 30 = 60.$$

Найменша варіанта дорівнює 2

$$F^*(x) = 0 \text{ при } x \leq 2$$

Значення $X < 6$ а саме $x_1 = 2$ спостерігалось 12 раз, отже

$$F^*(x) = \frac{n_{x_1}}{n} = \frac{12}{60} = 0,2 \text{ при } 2 < x \leq 6.$$

Значення $X < 10$ а саме $x_1 = 2$ та $x_2 = 6$, спостерігались $12 + 18$ раз, отже

$$F^*(x) = \frac{n_{x_1+x_2}}{n} = \frac{12+18}{60} = \frac{30}{60} = 0,5 \text{ при } 6 < x \leq 10.$$

Оскільки $x = 10$ — найбільша варіанта, то

$$F^*(x) = 1 \text{ при } x > 10.$$

Шукана емпірична функція:

$$F^*(x) = \begin{cases} 0 & \text{при } x \leq 2, \\ 0,2 & \text{при } 2 < x \leq 6, \\ 0,5 & \text{при } 6 < x \leq 10, \\ 1 & \text{при } x > 10. \end{cases}$$

На рисунку 7.1 наведено графік емпіричної функції розподілу.

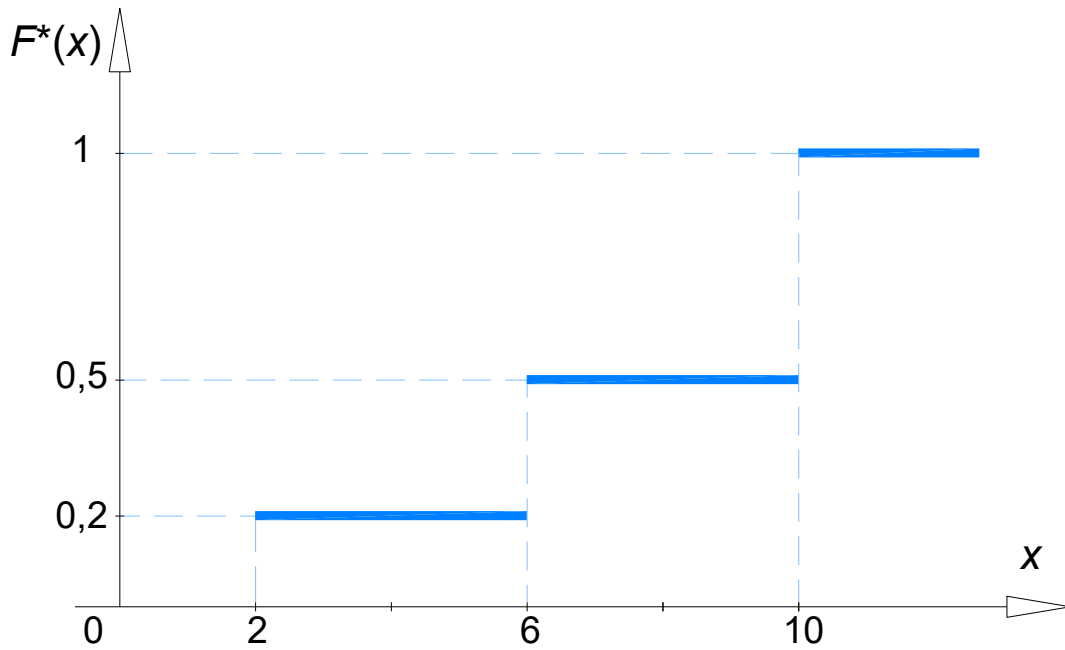


Рисунок 7.1 – Графік емпіричної функції розподілу

Полігон та гістограма



Рисунок 7.2 – Полігон частот

Полігоном відносних частот

називають ламану, відрізки якої з'єднують точки $(x_1; W_1), (x_2; W_2), \dots, (x_k; W_k)$

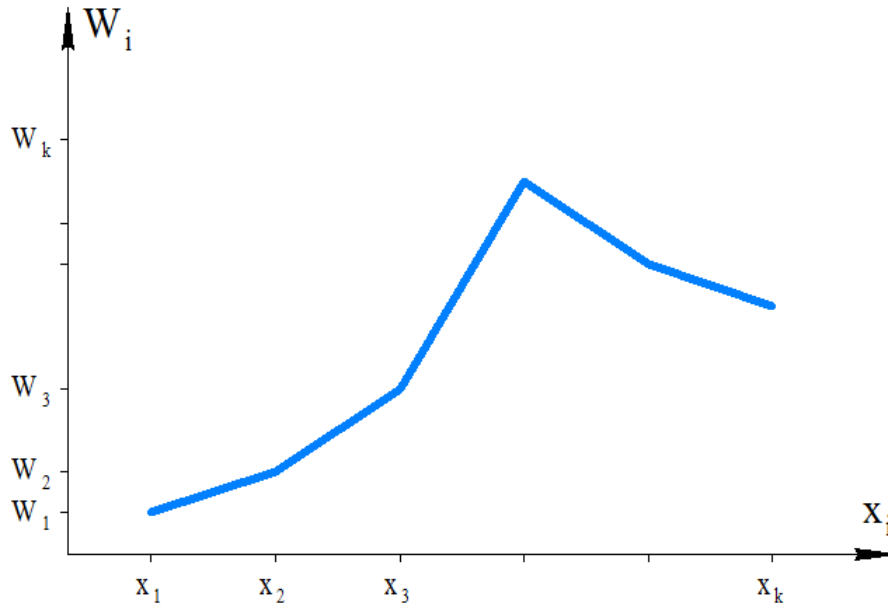
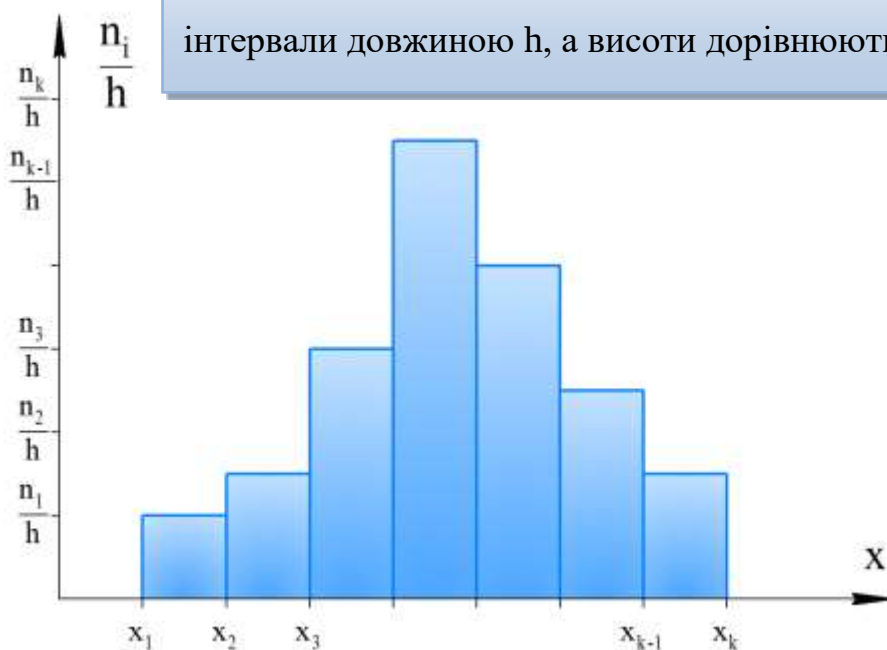


Рисунок 7.3 – Полігон відносних частот

Гістограмою частот

називають ступінчасту фігуру, що складається з прямокутників, основами яких служать часткові інтервали довжиною h , а висоти дорівнюють відношенню.



$$\frac{n_j}{h}$$

Рисунок 7.4 – Гістограма частот

Гістограмою відносних частот

називають ступінчасту фігуру, що складається з прямокутників, основами яких служать часткові інтервали довжиною h , а висоти дорівнюють відношенню.

$$\frac{W_i}{h}$$



Рисунок 7.5 – Гістограма відносних частот

Приклад 7.4 Побудувати полігон відносних частот за наданим розподілом вибірки (таблиця 7.5).

Таблиця 7.5 – Розподіл вибірки:

x_i	2	4	5	7	10
n_i	6	8	4	4	18

Розв'язання

Визначимо об'єм вибірки:
$$n = \sum_{i=1}^5 n_i = 6 + 8 + 4 + 4 + 18 = 40.$$

В таблиці 7.6. наведено розподіл відносних частот вибірки.

Таблиця 7.6 – Розподіл відносних частот

x_i	2	4	5	7	10
W_i	0,15	0,2	0,1	0,1	0,45

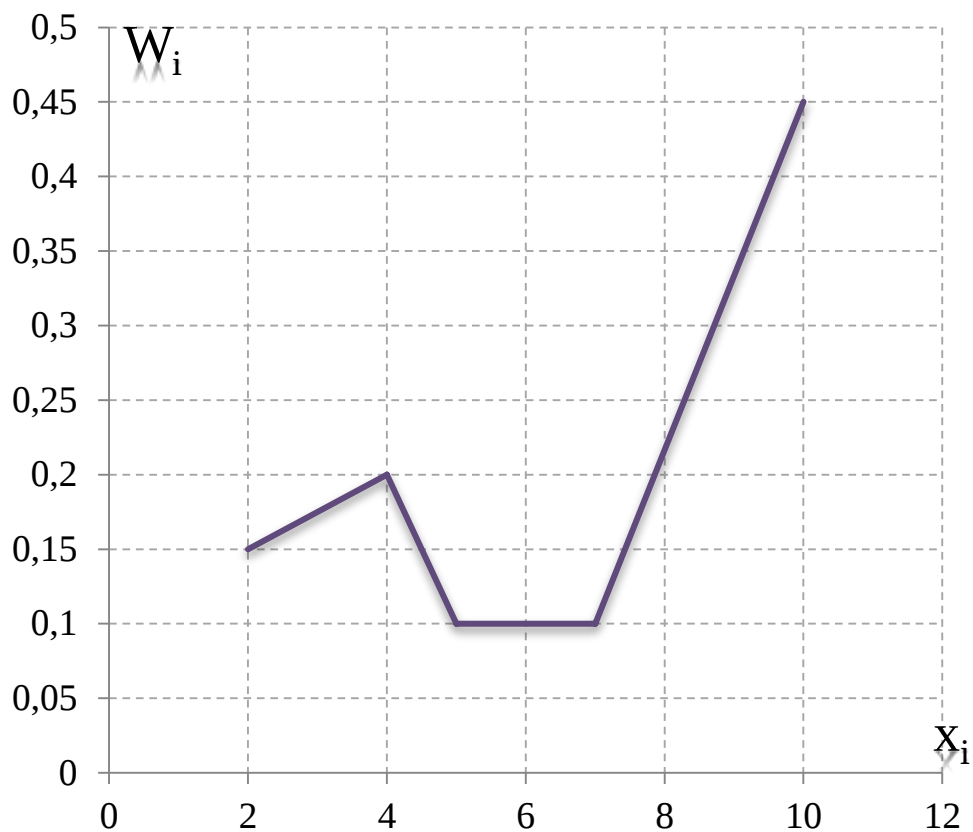


Рисунок 7.6 – Полігон відносних частот

Приклад 7.5 Побудувати гістограму частот за наданим розподілом вибірки об'єму $n = 100$:

Таблиця 8.5.3 – Розподіл вибірки:

Номер інтервалу i	Частинний інтервал h	Сума частот варіант інтервалу n_i	Щільність частоти n_i/h
1	1 – 5	10	2,5
2	5 – 9	20	5
3	9 – 13	50	12,5
4	13 – 17	12	3
5	17 – 21	8	2

Розв'язання

Відкладемо на осі абсцис задані інтервали довжини $h = 4$.

Проведемо над цими інтервалами відрізки, паралельні осі абсцис, що знаходяться від неї на відстані, яка дорівнює відповідним щільностям частоти n_i/h .

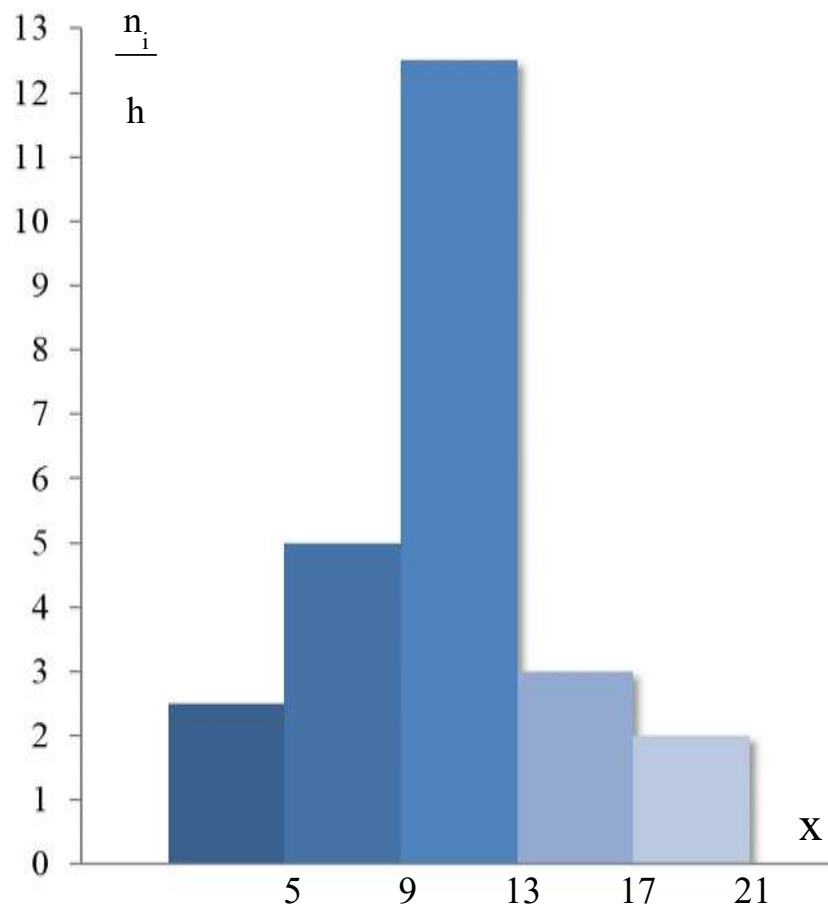


Рисунок 8.5.8 – Гістограма відносних частот

Висновки

Отже, знання основних понять математичної статистики допоможуть розуміти та застосовувати методи статистичного аналізу у власних дослідженнях та роботі з даними. Математична статистика має широке застосування в науці, медицині, економіці, соціальних науках та багатьох інших галузях, де аналіз даних є необхідним для прийняття обґрунтованих рішень. Власне тому важливість врахування випадкової помилки та контролю якості даних не може бути недооцінена в математичній статистиці.

Статистичний розподіл є інструментом, що дозволяє аналізувати та описувати випадкові явища, події або змінні. Для опису форми статистичного розподілу можна використовувати характеристики, такі як симетрія, широта та висота горба, хвостатість і мода.

Для візуалізації статистичних розподілів можна використовувати графіки, такі як гістограми, діаграми розсіювання, графіки ймовірності та функції щільності.

Квантилі допомагають описати статистичний розподіл за його відсотковою точкою, наприклад, медіаною або перцентилями. Для порівняння двох або більше статистичних розподілів використовуються різні статистичні методи, такі як тест Колмогорова-Смірнова, t-тест, аналіз дисперсії тощо.

Питання та завдання до контролю знань студентів

- 1 Що таке статистичний розподіл і яке його призначення в статистиці?
- 2 Які основні поняття пов'язані зі статистичним розподілом?
- 3 Що таке вибіркового розподіл і в чому його відмінність від генерального розподілу?
- 4 Що означає нормалізація статистичного розподілу?
- 5 Які характеристики статистичного розподілу можуть бути використані для опису його форми?
- 6 Які основні типи статистичних розподілів ви знаєте і які їх характеристики?
- 7 Що таке середнє значення і стандартне відхилення в статистичному розподілі?
- 8 Які інструменти використовуються для візуалізації статистичних розподілів?
- 9 Що таке квантиль і як його використовують для опису статистичного розподілу?
- 10 Які статистичні методи використовуються для порівняння двох або більше статистичних розподілів?

Лекція № 8

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 2.2 **Статистичні оцінки параметрів розподілу**
Статистичні оцінки параметрів генеральної сукупності.
Основні вимоги до статистичних оцінок.

Мета заняття Ознайомити із статистичними оцінками параметрів генеральної сукупності та вимогами до оцінок.

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН
конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Основні вимоги до статистичних оцінок. 2 Генеральна та вибіркова середні, генеральна, дисперсії, генеральне, вибіркоче та виправлене відхилення. 3 Параметри генеральної сукупності.
4. Заклучна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е.Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.
3. Дубовик, В.П. Вища математика. Збірник задач. [Текст] / В.П.Дубовик, І.І.Юрик. – К.: «А.С.К.», 2001. – 480 с.

ВСТУП

Статистичні оцінки параметрів розподілу виконують важливу роль у математичній статистиці, дозволяючи нам отримувати числові оцінки невідомих параметрів популяції на основі вибірових даних. Вони дозволяють нам робити висновки про популяцію на основі обмеженої інформації, що маємо з вибірки.

Параметри розподілу, такі як середнє значення, дисперсія, медіана чи коефіцієнти асиметрії та ексцесу, є важливими характеристиками, які дозволяють описувати розподіл випадкової величини. Однак, в реальних ситуаціях, коли ми не маємо повної інформації про популяцію, нам потрібно здійснювати статистичні оцінки цих параметрів на основі вибірки.

Цей процес статистичного оцінювання параметрів розподілу включає в себе вибір відповідного методу оцінювання та обчислення оцінок параметрів на основі доступних даних. Методи оцінювання можуть бути розділені на дві основні категорії: методи точкового оцінювання, які надають одну конкретну оцінку параметра, і інтервальні методи, які надають довірчі інтервали, в межах яких з високою ймовірністю міститься справжнє значення параметра.

При виборі методу оцінювання параметрів розподілу необхідно враховувати властивості даних та припущення про їх розподіл. Точність та ефективність оцінок також є важливими критеріями, які враховуються при виборі методу.

У даній роботі ми розглянемо різні методи статистичного оцінювання параметрів розподілу, вивчимо їх властивості та зробимо практичні висновки щодо їх застосування. Знання процесу статистичного оцінювання допоможе нам робити обґрунтовані висновки на основі даних і робити надійні припущення про популяцію, навіть коли ми маємо лише обмежену вибірку.

При дослідженні ознаки X зібрати та обробити всі її можливі значення здебільшого буває неможливо. Тому з усієї множини N даних про ознаку X , яка називається *генеральною сукупністю* обсягом N , випадковим чином відбирається її частина, що містить n даних і називається *вибірковою сукупністю*, або *вибіркою* обсягом n . Обсяг вибірки n має бути суттєво меншим за обсяг генеральної сукупності N .

Різні можливі значення x_i ознаки X , які потрапили до вибірки, називаються *варіантами*, а система варіант $x_1, x_2, \dots, x_k$, розміщена в зростаючому порядку, називається *варіаційним рядом*.

Якщо варіанти $x_1, x_2, \dots, x_k$ спостерігаються у вибірці відповідно $n_1, n_2, \dots, n_k$ раз, то значення $n_1, n_2, \dots, n_k$ називаються *частотами варіант*, і для них виконується умова

$$\sum_{i=1}^k n_i = n,$$

а відношення

$$\omega_1 = \frac{n_1}{n}, \omega_2 = \frac{n_2}{n}, \dots, \omega_k = \frac{n_k}{n}$$

називаються *відносними частотами варіант*, і для них виконується умова

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

Таблиця, в першому рядку якої наведено впорядковані варіанти x_i (варіаційний ряд), а в другому — відповідні їм частоти n_i (або відносні частоти ω_i), називається *дискретним статистичним розподілом вибірки*.

x_i	x_1	x_2		x_{k-1}	x_k
n_i	n_1	n_2		n_{k-1}	n_k

Якщо кількість варіант у вибірці надто велика або досліджувана знака X є неперервною випадковою величиною, то будується інтервальний статистичний розподіл вибірки, аналогічний дискретному, у першому рядку якого замість дискретного варіаційного ряду записується інтервальний

$$(x_1; x_2), (x_2; x_3), \dots, (x_m; x_{m+1}).$$

Кількість інтервалів m для вибірки обсягом n орієнтовно вважається такою, що дорівнює $\sqrt{n}$, а розмір (крок) h інтервалу обчислюється за формулою

$$h = \frac{x_k - x_1}{m},$$

де x_1 і x_k — відповідно найменша і найбільша варіанти у вибірці.

Частоти n_i^* варіант, які потрапили в кожний частинний інтервал $(x_i ; x_{i+1})$, записуються в другий рядок таблиці.

Якщо межа двох частинних інтервалів збігається зі значенням варіанти, то частота цієї варіанти або ділиться порівну між цими двома інтервалами, або в усіх таких випадках цілком заноситься у правий (або лівий) від варіанти інтервал.

Статистичні оцінки параметрів розподілу

Параметри досліджуваної ознаки X генеральної сукупності, з якої зроблено вибірку $x_1, x_2, \dots, x_n$ є величинами сталими, але їхні числові значення невідомі. Ці параметри оцінюються параметрами вибірки. Нехай θ - невідомий параметр (наприклад, математичне сподівання, дисперсія тощо) досліджуваної ознаки X .

Вибірковою оцінкою параметра θ називається число θ^* , знайдене за даними вибірки або за статистичним розподілом, залежне від $x_1, x_2, \dots, x_n$ і таке, що наближено дорівнює оцінюваному параметру:

$$\theta^* \approx \theta.$$

Ця оцінка, яка визначається у даній вибірці одним числом, точкою, називається *точковою*.

Оскільки θ^* є випадковою величиною, то вона може бути зміненою і незміщеною. Коли математичне сподівання цієї оцінки точно дорівнює оцінюваному параметру θ , тобто:

$$M(\theta^*) = \theta,$$

то оцінка θ називається *незміненою*.

Якщо $M(\theta^*) \neq \theta$, то статистична оцінка θ^* називається *зміщеною* відносно параметра θ генеральної сукупності.

Різниця

$$\theta^* - \theta = \delta$$

називається *зміщенням* статистичної оцінки θ^* .

Точкова статистична оцінка θ^* називається *спроможною*, якщо, у разі необмеженого збільшення обсягу вибірки, θ^* наближається до оцінюваного параметра θ .

$$\lim_{n \rightarrow \infty} P\{|\theta^* - \theta| < \delta\} = 1.$$

Оцінюваний параметр може мати кілька точкових незміщених статистичних оцінок. Точкова статистична оцінка називається *ефективною*, коли при заданому обсязі вибірки вона має мінімальну дисперсію.

Для невідомого математичного сподівання $M(X)$ досліджуваної ознаки X точковою оцінкою є *вибіркова середня*, яка позначається $\bar{x}_B$ і обчислюється за формулою

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

Вона задовольняє всі три перелічені умови. Для невідомої дисперсії $D(X)$ спроможною оцінкою є *вибіркова дисперсія* D_B , обчислювана за формулою

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

Проте ця оцінка є зміщеною. Тому для оцінки невідомої дисперсії застосовують так звану *виправлену дисперсію*, яку позначають s^2 і обчислюють за формулою

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

Оцінкою для середнього квадратичного відхилення $\sigma(X)$ є виправлене *вибіркове середнє квадратичне відхилення*

$$\sigma_B = \sqrt{S^2}.$$

Приклад 1. У результаті вибіркового аналізу добової виручки авіакомпанії дістали вибірку обсягом $n = 40$ (млн. грн):

0,87	0,94	0,99	0,90	0,90	0,87	0,85	0,87
0,90	0,94	0,87	0,87	0,82	0,90	0,94	0,90
0,85	0,85	0,87	0,94	0,81	0,82	0,87	0,97
0,90	0,94	0,85	0,81	0,87	0,85	0,90	0,82
0,99	0,90	0,94	0,82	0,97	0,81	0,85	0,87

Скласти

1. варіаційний ряд
2. дискретний статистичний розподіл вибірки
3. інтервальний розподіл.

Розв'язання. 1. Випишемо різні значення варіант, які потрапили у вибірку:

0,87 0,94 0,99 0,90 0,85 0,82 0,81 0,97.

Розмістивши їх у порядку зростання дістанемо дискретний варіаційний ряд:

0,81 0,82 0,85 0,87 0,90 0,94 0,97 0,99.

2. Обчислимо частоту кожної варіанти із варіаційного ряду і складемо таблицю. Дістанемо дискретний статистичний розподіл вибірки.

x_i	0,81	0,82	0,85	0,87	0,90	0,94	0,97	0,99
p_i	3	4	6	9	8	6	2	2

3. За обсягом вибірки $n = 40$ визначаємо орієнтовну кількість $m = 6$ частинних інтервалів в інтервальному статистичному розподілі, обчислюємо крок інтервалу

$$h = \frac{0,99-0,81}{6} = 0,03.$$

Тоді інтервальний варіаційний ряд запишеться у вигляді

$$(0,81;0,84) \quad (0,84;0,87) \quad (0,87;0,90) \quad (0,90;0,93) \quad (0,93;0,96) \quad (0,96;0,99).$$

Підсумуємо частоти варіант, які потрапили в кожний із частинних інтервалів, при цьому частоти варіант, які збіглися з межами інтервалів (0,87 і 0,90), поділимо порівну між суміжними інтервалами.

Тоді інтервальний статистичний розподіл вибірки матиме вигляд

Інтервали	(0,81;0,84)	(0,84;0,87)	(0,87;0,90)	(0,90;0,93)	(0,93;0,96)	(0,96;0,99)
n_i^*	7	10	9	4	6	4

Приклад 2. Знайти точкові оцінки математичного сподівання, дисперсії і середньоквадратичного відхилення добової виручки авіакомпанії за даними вибірки з прикладу 1.

Розв'язання. За дискретним статистичним розподілом обчислюємо вибірку середню $\bar{x}_B$, виправлену вибірку дисперсію s^2 і виправлене вибіркве середнє квадратичне відхилення σ_B :

$$\bar{x}_B = \frac{1}{40} (0,81 \cdot 3 + 0,82 \cdot 4 + 0,85 \cdot 6 + 0,87 \cdot 9 + 0,9 \cdot 8 + 0,94 \cdot 6 + 0,97 \cdot 2 + 0,99 \cdot 2) = 0,885,$$

$$s^2 = \frac{1}{39} [(0,81 - 0,885)^2 \cdot 3 + (0,82 - 0,885)^2 \cdot 4 + (0,85 - 0,885)^2 \cdot 6 + (0,87 - 0,885)^2 \cdot 9 + (0,9 - 0,885)^2 \cdot 8 + (0,94 - 0,885)^2 \cdot 6 + (0,97 - 0,885)^2 \cdot 2 + (0,99 - 0,885)^2 \cdot 2] = 0,0025,$$

$$\sigma_B = 0,05.$$

Знайдемо оцінки $\bar{x}_B, s^2, \sigma_B$ за інтервальний статистичним розподілом, замінюючи x_i центрами частинних інтервалів x_i^* . Для цього перейдемо спочатку від інтервального статистичного розподілу до дискретного, який набуває такого вигляду:

$x_i^* = x_{i-1} + \frac{h}{2}$	0,825	0,855	0,885	0,915	0,945	0,975
n_i	7	10	9	4	6	4

Тоді

$$\bar{x}_B = \frac{1}{40} (0,825 \cdot 7 + 0,855 \cdot 10 + 0,885 \cdot 9 + 0,915 \cdot 4 + 0,945 \cdot 6 + 0,975 \cdot 4) = 0,888,$$

$$s^2 = \frac{1}{39} [(0,825 - 0,888)^2 \cdot 7 + (0,855 - 0,888)^2 \cdot 10 + (0,885 - 0,888)^2 \cdot 9 + (0,915 - 0,888)^2 \cdot 4 + (0,945 - 0,888)^2 \cdot 6 + (0,975 - 0,888)^2 \cdot 4] = 0,0023,$$

$$\sigma_B = 0,048$$

На практиці часто доводиться на основі результатів випробувань (вибірки) знати закон розподілу генеральної сукупності. Якщо закон розподілу невідомий, але є підстава вважати, що він має певний вигляд (наприклад, назвемо його R), то висувають гіпотезу: генеральна сукупність розподілена за законом R, тобто в цій гіпотезі мова йде про вигляд передбаченого розподілу.

Можливі випадки, коли закон розподілу відомий, але його параметри невідомі. Якщо є підстава припустити, що його невідомий параметр θ дорівнює певному значенню θ_0 , то висувають гіпотезу: $\theta = \theta_0$; у цьому випадку гіпотеза припускає оцінку параметру конкретного розподілу.

Можливі й інші гіпотези: про рівність параметрів двох чи декількох розподілів, про незалежність вибірок, про значимість вибіркового коефіцієнта кореляції тощо.

Статистичною називають гіпотезу про вигляд невідомого розподілу або про параметри невідомих розподілів. Наприклад, статистичними є гіпотези:

- 1) генеральна сукупність розподілена за нормальним законом;
- 2) дисперсії двох нормальних сукупностей рівні між собою.

Перевірку гіпотез на основі вибірових статистичних даних називають *статистичною перевіркою гіпотез*.

Разом із висунутою гіпотезою розглядають і гіпотезу, що перечить їй. Якщо висунута гіпотеза буде знехтувана, то має місце гіпотеза, що перечить їй. З цієї причини гіпотези доцільно розрізняти.

Одну з висунутих гіпотез виділяють в ролі *основної (нульової)* і позначають, як правило H_0 . Поряд з нею висувають альтернативну (*конкуруючу*) гіпотезу, яка суперечить основній і позначають H_1 .

Наприклад, якщо нульова гіпотеза полягає в припущенні, що математичне сподівання a нормального розподілу рівне 5, то альтернативна гіпотеза, зокрема, може полягати в тому, $a \neq 5$. Коротко це записують так:

$$H_0: a = 5; \quad H_1: a \neq 5.$$

Розрізняють також гіпотези за кількістю припущень.

Простою називається гіпотеза, що має лише одне припущення. Наприклад, якщо λ - параметр показникового розподілу, то гіпотеза $H_0: \lambda = 5$ – проста.

Гіпотеза називається *складною*, якщо складається зі скінченного чи нескінченного числа простих гіпотез. Наприклад, складна гіпотеза $H_0: \lambda > 5$, складається з нескінченної множини простих гіпотез: $H_i: \lambda = b_i$, де b_i – довільне число, більше 5.

На основі статистичних даних дуже важко, а іноді і неможливо, зробити безпомилкові висновки щодо гіпотез. У підсумку може бути прийняте неправильне рішення, тобто можуть бути допущені помилки двох родів.

Помилка першого роду полягає в тому, що буде відхилена правильна гіпотеза.

Помилка другого роду полягає в тому, що буде прийнята неправильна гіпотеза.

Правильне рішення може бути прийняте також у двох випадках:

- а) гіпотеза приймається, причому і в дійсності вона правильна;
- б) гіпотеза відхиляється, причому і в дійсності вона неправильна.

Ймовірність здійснити помилку першого роду позначають через α і називають її *рівнем значимості*. Число α задають малим і найчастіше використовують значення α , що дорівнюють 0,05; 0,001 і т.д. Якщо, наприклад, $\alpha = 0,01$, то це означає, що в одному випадку зі ста є ризик допустити помилку першого роду (відхилити правильну гіпотезу).

Для перевірки нульової гіпотези використовують спеціально підібрану випадкову величину, точний чи наближений розподіл якої відомо. Цю величину позначають через Z , якщо вона розподілена нормально, F – по закону Фішера-Снедекора, T – по закону Стьюдента, χ^2 – по закону «хи квадрат» і т.д. Оскільки конкретний вигляд розподілу до уваги не береться, то позначають цю величину через K .

Статистичним критерієм (просто *критерієм*) називають випадкову величину K , що служить для перевірки нульової гіпотези. Для різних гіпотез ці критерії є різними.

Наприклад, коли перевіряють гіпотезу про рівність дисперсії двох нормальних генеральних сукупностей, то в ролі критерію K беруть відношення виправлених вибірових дисперсій:

$$F = \frac{s_1^2}{s_2^2}.$$

Ця величина випадкова, тому в різних випробуваннях дисперсії приймають різні, наперед невідомі значення і розподілена за законом Фішера-Снедекора.

Спостережуваним значенням $K_{\text{спост}}$ називається значення критерію, обчислене по результатах вибірки.

Після вибору певного критерію, множину всіх його можливих значень розбивають на дві множини, що не перетинаються: одна з них містить значення критерію, при яких нульова гіпотеза відхиляється, а друга – при яких приймається.

Критичною областю називають сукупність значень критерію, при яких нульову гіпотезу відхиляють.

Значення статистичного критерію, при яких нульова гіпотеза приймається, називається *областю прийняття гіпотези* (областю допустимих значень).

Основний принцип перевірки статистичних гіпотез формується так: якщо спостережуване значення критерію належить області прийняття гіпотези — гіпотезу приймають, якщо ж спостережуване значення критерію належить критичній області — гіпотезу відхиляють.

Оскільки критерій K — одновимірною випадковою величиною, то всі її можливі значення належать деякому інтервалу. Тому область прийняття гіпотези і критична область також є інтервалами, а, значить, існують точки, які їх розділяють.

Критичними точками (границями) $\kappa_{кр}$ називають точки, які відділяють критичну область від області прийняття рішень.

Розрізняють односторонню (правосторонню чи лівосторонню) і двосторонню критичні області.

Правосторонньою називають критичну область, що визначається нерівністю $K > \kappa_{кр}$, де $\kappa_{кр}$ — додатне число.

Лівосторонньою називають критичну область, що визначається нерівністю $K < \kappa_{кр}$, де $\kappa_{кр}$ — від'ємне число.

Односторонньою називають правосторонню чи лівосторонню критичну область.

Двосторонньою називають критичну область, що визначається нерівностями

$$K < \kappa_1, K > \kappa_2, \kappa_2 > \kappa_1.$$

Перевірка статистичних гіпотез будь-якої природи здійснюється за такою схемою:

1. Формулюється статистична гіпотеза H_0 .
2. Вибирається статистичний критерій відповідно до сформульованої нульової гіпотези H_0 .
3. Залежно від гіпотези H_0 і альтернативної гіпотези H_1 вибирається одностороння або двостороння критична область. Щоб побудувати критичні області, необхідно знайти значення критичних точок.
4. За результатами вибірки обчислюється спостережене значення критерію.
5. Виходячи з вимоги, що при правильності гіпотези H_0 ймовірність того, що Ксп потрапить у критичну область, має дорівнювати прийнятому рівню значимості α , перевіряється статистична гіпотеза.

Якщо К потрапляє у критичну область, а ця подія малоімовірна і вона все-таки здійснилася, то нульова гіпотеза H_0 відхиляється. У протилежному разі – приймається.

Приклад 8.1 Генеральна сукупність задана таблицею розподілу.

Знайти генеральну середню

Таблиця 8.1 – Розподіл генеральної сукупності:

x_i	2	4	5	6
N_i	8	9	10	3

Розв'язання

Визначимо об'єм сукупності:
$$N = \sum_{i=1}^6 N_i = 2 + 3 + 7 + 5 + 2 + 6 = 25.$$

Обчислимо генеральну середню:

$$\bar{x}_r = \frac{1}{N} \sum_{i=1}^6 x_i \cdot N_i = \frac{1}{30} \cdot (3 \cdot 2 + 4 \cdot 3 + 5 \cdot 7 + 6 \cdot 5 + 7 \cdot 2 + 8 \cdot 6) = \frac{145}{25} = 5,8.$$

Приклад 8.2 Генеральна сукупність задана таблицею розподілу (таблиця 8.2).

Таблиця 8.2 – Розподіл генеральної сукупності:

x_i	2	4	5	6
N_i	8	9	10	3

Знайти генеральну середню

Розв'язання

Визначимо об'єм сукупності: $N = \sum_{i=1}^4 N_i = 8 + 9 + 10 + 3 = 30$.

Обчислимо генеральну середню:

$$\bar{x}_r = \frac{1}{N} \sum_{i=1}^4 x_i \cdot N_i = \frac{1}{30} \cdot (8 \cdot 2 + 9 \cdot 4 + 10 \cdot 5 + 3 \cdot 6) = \frac{120}{30} = 4.$$

Отже, генеральна дисперсія: $D_r = \frac{1}{N} \sum_{i=1}^4 N_i (x_i - \bar{x}_r)^2 =$

$$= \frac{1}{30} \cdot (8 \cdot (2 - 4)^2 + 9 \cdot (4 - 4)^2 + 10 \cdot (5 - 4)^2 + 3 \cdot (6 - 4)^2) = \frac{54}{30} = 1,8.$$

Приклад 8.3 З генеральної сукупності вилучена вибірка (таблиця 8.3).

Таблиця 8.3 – Вилучена вибірка:

x_i	2	5	7	10
n_i	6	12	8	14

Знайти незсунуту оцінку
генеральної середньої

Розв'язання

Визначимо об'єм вибірки:

$$n = \sum_{i=1}^4 n_i = 6 + 12 + 8 + 14 = 50.$$

Незсунутою оцінкою генеральної середньої є вибірка середня:

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^4 n_i x_i = \frac{1}{50} (16 \cdot 2 + 12 \cdot 5 + 8 \cdot 7 + 14 \cdot 10) = 5,76.$$

Приклад 8.4 Знайти вибіркoву середню за наданим розподілом вибірки (таблиця 8.4).

Таблиця 8.4 – Розподіл вибірки

x_i	2560	2600	2620	2650	2700
n_i	2	3	10	4	1

Розв'язання

Визначимо об'єм вибірки: $n = \sum_{i=1}^5 n_i = 2 + 3 + 10 + 4 + 1 = 20$.

Початкові варіанти — великі числа, тому перейдемо до умовних варіантів, обравши у якості $C = 2620$.

Тоді умовні варіанти:

$$\begin{aligned} u_1 &= x_1 - C = 2560 - 2620 = -60, \\ u_2 &= x_2 - C = 2600 - 2620 = -20, \\ u_3 &= x_3 - C = 2620 - 2620 = 0, \\ u_4 &= x_4 - C = 2650 - 2620 = 30, \\ u_5 &= x_5 - C = 2700 - 2620 = 80. \end{aligned}$$

В таблиці 8.5 наведено результат розподілу умовних варіант.

Таблиця 8.5 – Розподіл умовних варіант

$u_i = x_i - C$	- 60	- 20	0	30	80
n_i	2	3	10	4	1

Знайдемо вибіркoву середню:

$$\bar{x}_B = C + \frac{1}{n} \sum_{i=1}^5 n_i u_i = 2620 + \frac{1}{20} \cdot (2 \cdot (-60) + 3 \cdot (-20) + 10 \cdot 0 + 4 \cdot 30 + 1 \cdot 80) = 2621.$$

Приклад 8.5 За результати вимірювання росту випадково відібраних 100 студентів (таблиця 8.6) знайти вибірку середню та вибірку дисперсію росту обстежених студентів.

Таблиця 8.6 – Результати вимірювання росту випадково відібраних студентів

Ріст, см	154 - 158	158 - 162	162 - 166	166 - 170	170 - 174	174 - 178	178 - 182
Кількість студентів	10	14	26	28	12	8	2

Розв'язання

Знайдемо середини інтервалів та приймемо їх в якості варіант. В таблиці 8.7 наведено розподіл вибірки.

Таблиця 8.7 – Розподіл вибірки

x_i	156	160	164	168	172	176	180
n_i	10	14	26	28	12	8	2

Перейдемо до умовних варіантів, обравши у якості параметра $C = 168$.

В таблиці 8.8 наведено розподіл умовних варіант.

Таблиця 8.8 – Розподіл умовних варіант

u_i	- 12	- 8	- 4	0	4	8	12
n_i	10	14	26	28	12	8	2



Обчислимо вибірку середню:

$$\bar{x}_B = C + \frac{1}{n} \sum_{i=1}^7 n_i u_i =$$
$$= 168 + \frac{10 \cdot (-12) + 14 \cdot (-8) + 26 \cdot (-4) + 28 \cdot 0 + 12 \cdot 4 + 8 \cdot 8 + 2 \cdot 12}{100} = 166.$$

Обчислимо вибірку дисперсію:

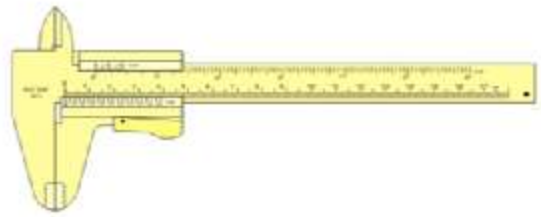
$$D_B = \frac{1}{n} \sum_{i=1}^k n_i u_i^2 - \left(\frac{1}{n} \sum_{i=1}^k n_i u_i \right)^2 =$$
$$= \frac{10 \cdot (-12)^2 + 14 \cdot (-8)^2 + 26 \cdot (-4)^2 + 28 \cdot 0^2 + 12 \cdot 4^2 + 8 \cdot 8^2 + 2 \cdot 12^2}{100} - (-2)^2 = 33,44.$$

Приклад 8.6 В результаті п'яти вимірювань деякої фізичної величини одним приладом отримані наступні результати:

92; 94; 103; 105; 106

Знайти:

- а) вибірку середню результатів вимірювань;
- б) вибірку та виправлену дисперсії помилок приладу.



Розв'язання

Знайдемо середини інтервалів та приймемо їх в якості варіант.

Результат виконаних дій наведено в таблиці 8.9.

Таблиця 8.9 – Розподіл умовних варіант

x_i	92	94	103	105	106
n_i	1	1	1	1	1

Обчислимо вибіркoву середню:

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^5 n_i x_i = \frac{1}{100} \cdot (92 + 94 + 103 + 105 + 106) = 100.$$

Обчислимо вибіркoву дисперсію:

$$\begin{aligned} D_B &= \frac{1}{n} \sum_{i=1}^5 n_i (x_i - \bar{x}_B)^2 = \\ &= \frac{(92 - 100)^2 + (94 - 100)^2 + (103 - 100)^2 + (105 - 100)^2 + (106 - 100)^2}{5} = 34. \end{aligned}$$

Отже, виправлена дисперсія:

$$s^2 = \frac{n}{n-1} D_B = \frac{5}{4} \cdot 34 = 42,5.$$

Висновки

Статистичні оцінки параметрів розподілу є важливим інструментом у математичній статистиці, дозволяючи отримати числові оцінки невідомих параметрів популяції на основі вибіркoвих даних. Правильне використання статистичних оцінок параметрів розподілу є важливим кроком для забезпечення надійності та обґрунтованості статистичних аналізів.

Статистичні оцінки параметрів розподілу є інструментом, що дозволяє отримувати числові оцінки невідомих параметрів популяції на основі вибірових даних.

Оцінки параметрів розподілу є основою для прийняття рішень, роботи зі статистичними моделями та проведення статистичних тестів.

Питання та завдання до контролю знань студентів

- 1 Що таке статистичні оцінки параметрів розподілу і яка їхня роль у математичній статистиці?
- 2 Які методи можна використовувати для отримання статистичних оцінок параметрів розподілу?
- 3 Що таке точкова оцінка параметра розподілу і як її можна обчислити?
- 4 Які інтервальні оцінки параметрів розподілу і як їх можна побудувати?
- 5 Як впливає величина вибірки на точність та ефективність статистичних оцінок параметрів розподілу?
- 6 Які фактори слід враховувати при виборі методу оцінювання параметрів розподілу?
- 7 Як можна перевірити якість та надійність статистичних оцінок параметрів розподілу?
- 8 Які обмеження і припущення супроводжують статистичні оцінки параметрів розподілу?
- 9 Як можна використовувати статистичні оцінки параметрів розподілу для прийняття рішень та проведення статистичних тестів?
- 10 Як використовувати статистичні оцінки параметрів розподілу у практичних дослідженнях та роботі з даними?

Лекція № 9

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 2.2 Статистичні оцінки параметрів розподілу

Точкові та інтервальні оцінки. Довірчий інтервал для оцінки математичного сподівання нормального розподілу.

Мета заняття Ознайомити зі спрощеними методами розрахунку зведених показників вибірки та моментами.

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН

конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Методи добутків та сум. 2 Зведення первинних варіант до рівновіддалених. 3 Оцінка відхилення емпіричного розподілу.
4. Заключна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е. Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.
3. Дубовик, В.П. Вища математика. Збірник задач. [Текст] / В.П. Дубовик, І.І. Юрик. – К.: «А.С.К.», 2001. – 480 с.

ВСТУП

В математичній статистиці точкові та інтервальні оцінки є ключовими концепціями, що дозволяють виробляти оцінки невідомих параметрів на основі обмеженої вибірки даних. Вони надають інструменти для отримання числових оцінок та довірчих інтервалів, що визначають діапазон значень параметрів з певною ймовірністю.

Точкові оцінки представляють собою одне конкретне числове значення, яке припускається близьким до справжнього значення параметра розподілу. Вони можуть бути отримані за допомогою різних методів, таких як метод моментів, метод найбільшої правдоподібності або інші. Точкові оцінки дозволяють зводити невідомі параметри до конкретних чисел, що спрощує аналіз та порівняння різних розподілів.

Інтервальні оцінки, з свого боку, надають діапазон значень, в межах якого з високою ймовірністю міститься справжнє значення параметра. Вони дають більш повну інформацію про невизначеність оцінки та враховують змінність вибірових даних. Інтервальні оцінки зазвичай побудовані з використанням стандартних помилок, квантилей розподілу та різних методів оцінювання.

Дослідження цих оцінок дозволить зрозуміти, як ефективно використовувати отримані результати та забезпечити достовірність статистичних висновків.

Точкові оцінки параметрів розподілу генеральної сукупності

Точкові оцінки параметрів розподілу генеральної сукупності є числовими значеннями, що використовуються для наближеного визначення невідомих параметрів цієї сукупності. Вони є результатом статистичного оцінювання на основі обмеженої вибірки даних з генеральної сукупності.

Точкова оцінка представляє собою конкретне число, що вважається приблизно близьким до справжнього значення параметра розподілу. Наприклад, якщо ми цікавимося середнім значенням (математичним очікуванням) генеральної сукупності, то точкова оцінка цього параметра може бути конкретним числовим значенням, яке ми отримали на основі вибірки.

Точкові оцінки можуть бути отримані за допомогою різних методів, таких як метод моментів, метод найбільшої правдоподібності, метод квазімаксимальної

правдоподібності та інші. Кожен метод має свої особливості і базується на різних припущеннях щодо розподілу даних.

Важливо розуміти, що точкова оцінка є лише наближенням та не завжди точно відображає справжнє значення параметра розподілу. Вона залежить від обсягу вибірки та властивостей самої вибірки. Тому при використанні точкових оцінок важливо розглядати їхній контекст та обґрунтування, а також враховувати інші статистичні міри для оцінки невизначеності та надійності отриманих результатів.

Інтервальна оцінка

Інтервальна оцінка є статистичним методом, який надає діапазон значень, що, з високою ймовірністю, містить справжнє значення параметра розподілу генеральної сукупності. Вона враховує невизначеність та варіативність вибірових даних, дозволяючи оцінити рівень достовірності отриманої інформації.

Інтервальна оцінка представляється у вигляді діапазону значень, який визначається двома числами: нижньою та верхньою межами інтервалу. Ці межі визначаються на основі вибірки даних і деяких статистичних методів. Зазвичай, інтервальні оцінки побудовані з використанням стандартних помилок, стандартних відхилень, довірчих інтервалів та квантилей розподілу. Рівень достовірності, що використовується в інтервальній оцінці, вказує ймовірність того, що справжнє значення параметра знаходиться всередині побудованого інтервалу.

Наприклад, інтервальна оцінка середнього значення генеральної сукупності може мати вигляд "95% довірчий інтервал $[a, b]$ ", де a та b — це нижня та верхня межі інтервалу, а 95% вказує на ймовірність, що справжнє значення середнього лежить всередині цього інтервалу.

Інтервальні оцінки дають нам більш повну інформацію про невизначеність та надійність отриманих результатів, порівняно з точковими оцінками, оскільки вони враховують варіативність вибірових даних. Це дозволяє нам робити більш об'єктивні та достовірні висновки на основі обмеженої вибірки.

Надійність (довірча імовірність) оцінки параметра

Надійність (довірча імовірність) оцінки параметра відображає ймовірність того, що даний інтервал, отриманий за допомогою інтервальної оцінки, містить справжнє значення параметра розподілу генеральної сукупності. Довірча імовірність виражається у відсотках і зазвичай позначається значенням між 0% і 100%.

Наприклад, якщо ми використовуємо 95% довірчий інтервал для оцінки параметра, це означає, що в довірчій імовірності 95% цей інтервал буде містити справжнє значення параметра. Тобто, якщо ми проведемо багато подібних вибірок і побудуємо 95% довірчі інтервали для кожної з них, то близько 95% з них будуть містити справжнє значення параметра.

Важливо розуміти, що надійність оцінки не гарантує, що сама оцінка точна. Довірча імовірність відноситься до вірогідності того, що діапазон, побудований на основі вибірки, містить справжнє значення параметра, а не до точності самої оцінки. Точність оцінки може залежати від розміру вибірки, властивостей розподілу та інших факторів.

Таким чином, довірча імовірність оцінки вказує на рівень достовірності та впевненості в тому, що отриманий інтервал містить справжнє значення параметра. Чим вища довірча імовірність, тим більша надійність оцінки, але це не означає, що вона точна з абсолютною впевненістю.

Довірчий інтервал для оцінки математичного сподівання (середнього значення) нормального розподілу можна побудувати з використанням стандартного помилки та критичного значення нормального розподілу.

Для нормального розподілу з відомим стандартним відхиленням, довірчий інтервал можна обчислити за формулою:

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

Якщо стандартне відхилення генеральної сукупності не відоме, але ми маємо вибіркове стандартне відхилення s , то можна використовувати розподіл Стюдента та відповідне критичне значення t .

Таким чином, довірчий інтервал для оцінки математичного сподівання нормального розподілу буде мати вигляд:

$$\bar{x} \pm t \frac{s}{\sqrt{n}}$$

де значення t вибирається з таблиць розподілу Стюдента в залежності від обраного рівня довірчої імовірності та ступенів свободи ($n - 1$).

Довірчий інтервал дозволяє нам з високою ймовірністю включити справжнє значення математичного сподівання генеральної сукупності в побудований інтервал на основі вибірки даних. Це дає нам інформацію про розмах можливих значень та невизначеність оцінки середнього значення.

Обробка вибірки методом найменших квадратів

Метод найменших квадратів є одним з основних методів обробки вибірки для побудови математичної моделі або знаходження найкращої апроксимації для набору даних. Цей метод використовується для знаходження лінійної залежності між залежною та незалежними змінними.

Основна ідея методу найменших квадратів полягає в тому, щоб знайти параметри моделі, які мінімізують суму квадратів різниць між спостережуваними значеннями залежної змінної та їх прогнозами, отриманими за допомогою моделі. Це означає, що ми шукаємо такі параметри моделі, які найкраще відповідають наявним даним.

Для застосування методу найменших квадратів необхідно виконати наступні кроки:

1. Вибрати підхідну модель, яка апроксимує залежну змінну від незалежних змінних.
2. Зібрати вибірку даних, що містить значення залежної та незалежних змінних.
3. Сформулювати функцію втрати, якою може бути сума квадратів різниць між спостережуваними та прогнозованими значеннями.

4. Знайти такі значення параметрів моделі, які мінімізують функцію втрати. Це може бути досягнуто за допомогою аналітичних методів або чисельних оптимізаційних алгоритмів.

5. Оцінити якість апроксимації, використовуючи метрики, такі як середньоквадратичне відхилення (MSE), коефіцієнт детермінації (R-squared) та інші.

Метод найменших квадратів є потужним інструментом в статистиці та аналізі даних. Він дозволяє робити розрахунки та отримувати числові результати на основі великої кількості даних. Основні переваги методу найменших квадратів включають:

1. Мінімізація помилок: Метод найменших квадратів дозволяє знайти такі параметри моделі, які мінімізують суму квадратів різниць між спостережуваними значеннями та їх прогнозами. Це допомагає зменшити вплив випадкових помилок та забезпечує кращу апроксимацію даних.

2. Використання комплексних моделей: Метод найменших квадратів може бути застосований для моделювання складних залежностей, включаючи лінійні та нелінійні моделі. Це дозволяє аналізувати різноманітні дані та виявляти складні взаємозв'язки між змінними.

3. Широке застосування: Метод найменших квадратів використовується в різних галузях, включаючи економіку, фінанси, фізику, інженерію та багато інших. Він може бути застосований для вирішення різноманітних завдань, від прогнозування та апроксимації до виявлення та аналізу залежностей між змінними.

4. Статистична інтерпретація: Метод найменших квадратів надає статистичну інтерпретацію результатів, таку як оцінки параметрів, стандартні помилки, довірчі інтервали та статистичні тестування. Це дозволяє оцінювати достовірність отриманих результатів та робити статистичні висновки.

Загалом, метод найменших квадратів є потужним інструментом для обробки даних та моделювання залежностей. Він дозволяє отримувати кількісні оцінки та робити висновки на основі наявних даних, що сприяє розвитку наукових досліджень, прийняттю рішень та вирішенню реальних проблем.

Оцінка параметрів лінійної функції

Оцінка параметрів лінійної функції є важливою задачею в статистиці та аналізі даних. Лінійна функція може бути виражена у вигляді:

$$y = \beta_0 + \beta_1 x$$

де y — залежна змінна,

x — незалежна змінна,

β_0 — зсув (інтерсепт),

β_1 — нахил (коефіцієнт нахилу).

Оцінка параметрів лінійної функції полягає в знаходженні оптимальних значень β_0 та β_1 , які найкраще апроксимують дані. Для здійснення оцінки використовують метод найменших квадратів.

Метод найменших квадратів вимагає мінімізувати суму квадратів різниць між спостережуваними значеннями залежної змінної та їх прогнозами за лінійною функцією. Це досягається шляхом обчислення похідних від функції втрати та вирішення системи рівнянь, що дорівнює нулю. Це дає нам оцінки параметрів β_0 та β_1 , які є найбільш оптимальними з точки зору мінімізації функції втрати.

Отримані оцінки параметрів можуть бути використані для прогнозування значень залежної змінної на основі незалежної змінної, виявлення залежностей між змінними та зроблення статистичних висновків.

Важливо врахувати, що оцінки параметрів є статистичними і їх точність може залежати від розміру вибірки та інших факторів. Тому важливо також враховувати стандартні помилки, довірчі інтервали та статистичні тести для оцінки достовірності отриманих результатів.

Після знаходження оцінок параметрів лінійної функції можна провести їх статистичну оцінку, включаючи обчислення стандартних помилок, довірчих інтервалів та статистичних тестів. Отримані оцінки параметрів лінійної функції дозволяють встановити залежність між змінними та робити прогнози для нових значень.

Крім того, вони можуть служити основою для подальшого аналізу, порівняння різних моделей та прийняття рішень.

В цілому, оцінка параметрів лінійної функції є важливим етапом статистичного моделювання та дозволяє отримати кількісні характеристики та висновки на основі наявних даних.

Оцінка параметрів параболічної функціональної залежності

Оцінка параметрів параболічної функціональної залежності є важливою задачею в статистичному аналізі та моделюванні. Параболічна функція може бути представлена у вигляді

$$y = ax^2 + bx + c,$$

де a , b та c — параметри, які визначають форму та положення параболи.

Для оцінки параметрів параболічної функції використовуються різні методи, зокрема метод найменших квадратів. Мета полягає у знаходженні таких значень параметрів, які найкраще підходять до наявних даних.

Метод найменших квадратів використовується для мінімізації суми квадратів різниць між спостережуваними значеннями y та їх прогнозами, отриманими за допомогою параболічної функції. Це досягається шляхом знаходження таких значень параметрів a , b та c , які мінімізують суму квадратів резидуальних (залишкових) величин.

Після знаходження оцінок параметрів параболічної функції можна провести їх статистичну оцінку, включаючи обчислення стандартних помилок, довірчих інтервалів та статистичних тестів.

Отримані оцінки параметрів параболічної функції дозволяють встановити залежність між змінними та робити прогнози для нових значень. Вони також можуть бути використані для порівняння різних моделей та аналізу впливу різних факторів на форму параболи.

В цілому, оцінка параметрів параболічної функції дозволяє отримати кількісні характеристики та висновки на основі наявних даних, що допомагає в розумінні та описі залежностей між змінними та вирішенні реальних задач.

Висновки

Отже, точкові оцінки надають одну конкретну оцінку параметра, тоді як інтервальні оцінки надають діапазон значень, в межах якого з високою ймовірністю міститься справжнє значення параметра. Некоректний вибір методу може призвести до ненадійних або неправильних оцінок. Ефективна оцінка характеризується малим розбіжністю між оцінкою та справжнім значенням параметра, а точна оцінка знаходиться близько до справжнього значення. При оцінюванні параметрів розподілу важливо враховувати можливі джерела помилок, такі як вибіркова випадковість, зміщеність та невідповідність припущень. Коректне розуміння цих факторів допомагає вибрати відповідну методологію оцінювання. Додаткові методи, такі як метод максимальної правдоподібності та метод моментів, можуть бути використані для отримання статистичних оцінок параметрів розподілу. Кожен метод має свої переваги та обмеження, і вибір методу залежить від конкретної задачі та властивостей даних.

Питання та завдання до контролю знань студентів

- 1 Що таке точкова оцінка і як вона використовується в статистиці?
- 2 Які методи можна використовувати для отримання точкових оцінок?
- 3 Як обчислити середнє значення (середнє арифметичне) і його точкову оцінку?
- 4 Які властивості мають точкові оцінки і чому вони можуть бути недостатніми для повного висновку?
- 5 Що таке інтервальна оцінка і як вона відрізняється від точкової оцінки?
- 6 Які методи можна використовувати для отримання інтервальних оцінок?
- 7 Як обчислити довірчий інтервал для середнього значення?
- 8 Як обчислити довірчий інтервал для пропорції чи частоти?
- 9 Як визначити рівень довіри для інтервальної оцінки і чому це важливо?
- 10 Як інтервальні оцінки допомагають уникнути недостовірних висновків та роблять статистичні оцінки більш надійними?

Лекція № 10

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 2.3 Методи розрахунку зведених показників вибірки

Метод добутків для обчислення вибірових середньої та дисперсії. Зведення первинних варіант до рівновіддалених. Емпіричні та теоретичні частоти. Метод сум для обчислення вибірових середньої та дисперсії. Оцінка відхилення емпіричного розподілу від нормального.

Мета заняття Ознайомитися із методами обчислення вибірових середньої та дисперсії

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН
конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 1 Функціональна залежність. 2 Статистична залежність. 3 Кореляційна залежність.
4. Заклучна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е.Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.
3. Буряк, Г.І. Методика застосування засобів Microsoft Excel при розв'язанні задач теорії ймовірностей [Текст] / Г.І. Буряк, – КРКМ ДНУ ім. О. Гончара, 2017. – 67 с.

ВСТУП

У сучасному світі, де величезний обсяг даних стає все більш доступним і важливим, виникає потреба в розробці ефективних методів обробки цих даних. Одним з найпоширеніших і важливих завдань в аналізі даних є обчислення вибірових середніх та дисперсії. Вони є ключовими показниками, що допомагають отримати уявлення про центральні та варіаційні характеристики вибірки. Ці показники допомагають зрозуміти основні властивості досліджуваної групи об'єктів.

Звичайні, початкові та центральні емпіричні моменти

Звичайним емпіричним моментом порядку k

k -х степенів різниць $x_i - C$.

називають середнє значення

$$M'_k = \frac{1}{n} \sum_{i=1}^m n_i (x_i - C)^k$$

Початковим емпіричним моментом порядку k

момент порядку k при $C = 0$.

називають

звичайний

$$M_k = \frac{1}{n} \sum_{i=1}^m n_i x_i^k$$

Зауваження 1

Початковий емпіричний момент першого порядку дорівнює вибірковій середній.

$$M_1 = \frac{1}{n} \sum_{i=1}^m n_i x_i = \bar{x}_v$$

Центральним емпіричним моментом порядку k

момент порядку k при $C = \bar{x}_v$.

називають

звичайний

$$m_k = \frac{1}{n} \sum_{i=1}^m n_i (x_i - \bar{x}_v)^k$$

Зауваження 2

Центральний емпіричний момент другого порядку дорівнює вибірковій дисперсії.

$$m_2 = \frac{1}{n} \sum_{i=1}^m n_i (x_i - \bar{x}_e)^2 = D_v$$

Умовні емпіричні моменти**Умовним емпіричним моментом порядку k**

називають початковий момент порядку k, який обчислено для умовних варіант:

$$M_k^* = \frac{1}{n} \sum_{i=1}^m n_i u_i^k = \frac{1}{n} \sum_{i=1}^m n_i \left(\frac{x_i - C}{h} \right)^k$$

Зауваження 1

Зокрема, умовний емпіричний момент першого порядку може бути розрахований:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m n_i \left(\frac{x_i - C}{h} \right) = \frac{1}{h} \left[\frac{1}{n} \sum_{i=1}^m n_i x_i - C \cdot \frac{1}{n} \sum_{i=1}^m n_i \right]$$

Зауваження 2

Для того щоб знайти вибіркoву середню, достатньо обчислити умовний момент першого порядку, помножити його на h та до результату додати відносний нуль.

$$\bar{x}_e = M_1^* \cdot h + C$$

$$M_1^* = \frac{1}{h} (\bar{x}_e - C)$$

Вираз центральних моментів через звичайні

Центральний момент другого порядку може бути виражений через звичайні в наведений спосіб:

$$m_2 = M_2' - (M_1')^2$$

Центральний момент третього порядку може бути виражений через звичайні в наведений спосіб:

$$m_3 = M'_3 - 3M'_2M'_1 + 2(M'_1)^3$$

Центральний момент третього порядку може бути виражений через звичайні в наведений спосіб:

$$m_4 = M'_4 - 4M'_3M'_1 + 6M'_2(M'_1)^2 - 3(M'_1)^4$$

Вираз звичайних моментів через умовні

Для того щоб знайти звичайний момент порядку k , достатньо умовний момент того ж порядку помножити на h^k , тобто

$$M'_k = M_k^* \cdot h^k$$

$$M_k^* = \frac{1}{h^k} \left[\frac{1}{n} \sum_{i=1}^m n_i (x_i - C)^k \right]$$

Зручні для обчислень формули, що виражають центральні моменти через умовні:

$$m_2 = [M_2^* - (M_1^*)^2] h^2$$

$$m_3 = [M_3^* - 3M_2^*M_1^* + 2(M_1^*)^3] h^3$$

$$m_4 = [M_4^* - 4M_3^*M_1^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4] h^4$$

Метод добутків для обчислення вибірових середньої і дисперсії

Метод добутків знайшов широке застосування в різних галузях, таких як статистика, економіка, фінанси, соціологія та інші. Його перевагою є те, що він дозволяє швидко обчислити вибірові середні та дисперсії навіть для великих вибірок, зменшуючи при цьому обчислювальну складність.

Нехай вибірка задана у вигляді розподілу рівновіддалених варіант та відповідних їм частот (таблиця 10.1):

Таблиця 10.1 – Розподіл рівновіддалених варіант та відповідних їм частот

x_i	x_1	x_2	x_3	x_4	...	x_m
n_i	n_1	n_2	n_3	n_4	...	n_m

У цьому випадку зручно знаходити вибірові середню та дисперсію методом добутків. В процесі виконання розрахунків доцільно користуватися розрахунковою таблицею 10.2.

Таблиця 10.2 – Розрахункова таблиця

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i (u_i + 1)^2$
x_1	n_1	-3	$n_1 \cdot u_1$	$n_1 \cdot u_1^2$	$n_1 (u_1 + 1)^2$
x_2	n_2	-2	$n_2 \cdot u_2$	$n_2 \cdot u_2^2$	$n_2 (u_2 + 1)^2$
x_3	n_3	-1	$n_3 \cdot u_3$	$n_3 \cdot u_3^2$	$n_3 (u_3 + 1)^2$
x_4	n_4	0	A_1		$n_4 (u_4 + 1)^2$
...	...	1	...	...	...
...	...	2	...	...	...
...	...	...	...	...	...
x_m	n_m	u_m	$n_m \cdot u_m$	$n_m \cdot u_m^2$	$n_m (u_m + 1)^2$
			A_2		
	$n = \sum n_i$		$\sum n_i u_i$	$\sum n_i \cdot u_i^2$	$\sum n_i (u_i + 1)^2$

Алгоритм складання розрахункової таблиці при обчислення вибірових середньої та дисперсії методом добутків

В перший стовпець таблиці записують вибірові початкові варіанти, розташовані в порядку зростання.

В другій стовпець записують частоти варіант, а об'єм вибірки поміщають в нижню клітину стовпчика

В третій стовпець записують умовні варіанти, причому в якості відносного нуля C вибирають варіанту, яка розташована приблизно в середині варіаційного ряду. Умовні варіанти обчислюють за формулою

$$u_i = \frac{x_i - C}{h}$$

та вважають, що h дорівнює різниці між будь-якими двома сусідніми варіантами. Але фактично третій стовпець заповнюється так:

- в клітці рядка, що містить відносний нуль, пишуть 0;
- в клітинах над нулем пишуть послідовно $-1, -2, -3$ і т.д.;
- під нулем — $1, 2, 3$ і т.д.

Доцільно складати від'ємні та додатні числа третього стовпця окремо.

В четвертий стовпець записують результат добутку $n_i \cdot u_i$. Суму A_1 та A_2 записують у відповідні клітинки четвертого стовпця.

В п'ятий стовпець записують результат добутку $n_i \cdot u_i^2$. При обчисленні добутків п'ятого стовпця доцільно числа $n_i \cdot u_i$ четвертого стовпця помножити на u_i .

В шостий стовпець записують результат $n_i (u_i + 1)^2$. Фактично шостий стовбець служить для контролю обчислень. Якщо виконується умова

$$\sum n_i (u_i + 1)^2 = \sum n_i \cdot u_i^2 + 2 \sum n_i \cdot u_i + n$$

то обчислення проведені вірно.

Після того як розрахункова таблиця заповнена та перевірена правильність обчислень, обчислюють умовні моменти. Умовний момент першого порядку обчислюється за формулою:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m n_i u_i$$

Умовний момент другого порядку розраховують за формулою:

$$M_2^* = \frac{1}{n} \sum_{i=1}^m n_i u_i^2$$

Обчислення вибіркової середньої здійснюється за формулою:

$$\bar{x}_B = M_1^* h + C$$

Обчислення дисперсії здійснюється за формулою:

$$D_B = [M_2^* - (M_1^*)^2] h^2$$

Безпосереднє обчислення вибірових середньої та дисперсії методом добутоків

Приклад 10.1 Знайти методом добутоків вибірові середню та дисперсію наданого статистичного розподілу (таблиця 10.3).

Таблиця 10.3 – Статистичний розподіл

x_i	10,2	10,4	10,6	10,8	11	11,2	11,4	11,6	11,8	12
n_i	2	3	8	13	25	20	12	10	6	1

Розв'язання

Складемо розрахункову таблицю 10.4:

Таблиця 10.4 – Розрахункова таблиця за даними статистичного розподілу

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i (u_i + 1)^2$
10,2	2	− 4	− 8	32	18
10,4	3	− 3	− 9	27	12
10,6	8	− 2	− 16	32	8
10,8	13	− 1	− 13	13	0
11,0	25	0	$A_1 = -46$		25
11,2	20	1	20	20	80
11,4	12	2	24	48	108
11,6	10	3	30	90	160
11,8	6	4	24	96	150
12,0	1	5	5	25	36
			$A_2 = 103$		
	$n = 100$		$\sum n_i \cdot u_i = 57$	$\sum n_i \cdot u_i^2 = 383$	$\sum n_i (u_i + 1)^2 = 597$

Виконаємо контроль правильності виконання розрахунків розрахункової таблиці:

$$\sum_{i=1}^k n_i x_i^2 + 2 \sum_{i=1}^k n_i x_i + n = 383 + 2 \cdot 57 + 100 = 597$$

Оскільки умова $\sum n_i (u_i + 1)^2 = \sum n_i \cdot u_i^2 + 2 \sum n_i \cdot u_i + n$ виконується, то обчислення проведені вірно.

Умовні моменти першого та другого порядків:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m n_i u_i = \frac{57}{100} = 0,57$$

$$M_2^* = \frac{1}{n} \sum_{i=1}^m n_i u_i^2 = \frac{383}{100} = 3,83$$

Вибіркові середня та дисперсія:

$$\bar{x}_B = M_1^* h + C = 0,57 \cdot (10,4 - 10,2) + 11 = 11,1$$

$$D_B = \left[M_2^* - (M_1^*)^2 \right] h^2 = (3,83 - 0,57^2) \cdot 0,2^2 = 0,14$$

Зведення первинних варіант до рівновіддалених

На практиці, як правило, дані спостережень не є рівновіддаленими числами тому відповідною обробкою значення спостережуваної ознаки зводять до випадку рівновіддалених варіант. Алгоритм зведення первинних варіант до рівновіддалених зводиться до виконання наступних дій:

- інтервал, в якому укладені всі початкові варіанти, поділити на кілька рівних часткових інтервалів;
- знайти середини часткових інтервалів, які і утворюють послідовність рівновіддалених варіант;
- в якості частоти кожної «нової» варіанти прийняти загальне число первинних варіант, що потрапили у відповідний частковий інтервал.

При обчисленні вибіркової дисперсії для зменшення помилки, що пов'язана угрупованням, роблять поправку Шеппарда, а саме віднімають з обчисленої дисперсії $1/12$ квадрата довжини часткового інтервалу.

$$D'_B = D_B + \frac{1}{12} h^2$$

Приклад 10.2 Вибіркова сукупність задана таблицею 10.5.

Скласти варіанти розмірностей, знайти вибіркові середню та дисперсію.

Таблиця 10.5 – Вибіркова сукупність

x_i	n_i	x_i	n_i	x_i	n_i
1,00	1	1,19	2	1,37	6
1,03	3	1,20	4	1,38	2
1,05	6	1,23	4	1,39	1
1,06	4	1,25	8	1,40	2
1,08	2	1,26	4	1,44	3
1,10	4	1,29	4	1,45	3
1,12	3	1,30	6	1,46	2
1,15	6	1,32	4	1,49	4
1,16	5	1,33	5	1,50	2

Розв'язання

Розіб'ємо інтервал 1,00 – 1,50 на наступні 5 часткових інтервалів:

1,00 – 1,10; 1,10 – 1,20; 1,20 – 1,30; 1,30 – 1,40; 1,40 – 1,50.

Середини часткових інтервалів:

$y_1 = 1,05$; $y_2 = 1,15$; $y_3 = 1,25$; $y_4 = 1,35$; $y_5 = 1,45$.

Частота варіанти y_1 :

$$n_1 = 1 + 3 + 6 + 4 + 2 + \frac{4}{2} = 18$$

Оскільки початкова варіанта 1,10 одночасно є кінцем першого часткового інтервалу і початком другого, частота 4 цієї варіанти порівну розподілена між обома частковими інтервалами.

Частота варіанти y_2 :

$$n_2 = \frac{4}{2} + 3 + 6 + 5 + 2 + \frac{4}{2} = 20$$

Частоти інших варіант: $n_3 = 25$; $n_4 = 22$; $n_5 = 15$

Отриманий розподіл рівновіддалених варіант наведено в таблиці 10.6.

Таблиця 10.6 – Розрахункова таблиця за даними статистичного розподілу

y_i	1,05	1,15	1,25	1,35	1,45
n_i	18	20	25	22	15

Складемо розрахункову таблицю 10.7.

Таблиця 10.7 – Розрахункова таблиця за даними статистичного розподілу

y_i	n_i	u_i	$n_i \cdot u_i$	$n_i \cdot u_i^2$	$n_i(u_i + 1)^2$
1,05	18	-2	-36	72	18
1,15	20	-1	-20	20	0
1,25	25	0	$A_1 = -56$	0	25
1,35	22	1	22	22	88
1,45	15	2	30	60	135
			$A_2 = 52$		
	$n = 100$		$\sum n_i \cdot u_i = -4$	$\sum n_i \cdot u_i^2 = 174$	$\sum n_i (u_i + 1)^2 = 266$

Виконаємо контроль правильності виконання розрахунків розрахункової таблиці:

$$\sum_{i=1}^k n_i \cdot u_i^2 + 2 \sum_{i=1}^k n_i \cdot u_i + n = 174 + 2 \cdot (-4) + 100 = 266.$$

Оскільки умова $\sum n_i (u_i + 1)^2 = \sum n_i \cdot u_i^2 + 2 \sum n_i \cdot u_i + n$ виконується, то обчислення проведені вірно.

Умовні моменти першого та другого порядків:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m n_i u_i = -0,04$$

$$M_2^* = \frac{1}{n} \sum_{i=1}^m n_i u_i^2 = 1,74$$

Вибіркові середня та дисперсія:

$$\bar{y}_s = M_1^* h + C = 1,246$$

$$D_s(y) = [M_2^* - (M_1^*)^2] h^2 = 0,017$$

Емпіричні та теоретичні частоти

Емпіричними частотами

називають фактично спостережувані частоти n_i .

Теоретичними частотами

називають частоти n_i' , знайдені за допомогою розрахунків.

У разі безперервного розподілу інтервал можливих значень ділять на k непересічних інтервалів та обчислюють ймовірності P_i попадання X в i -й частковий інтервал, а потім, як і для дискретного розподілу, множать число випробувань на ці ймовірності.

Приклад 10.3 В результаті експерименту, що складається з $n = 520$ випробувань, в кожному з яких реєструвалося число x_i появ деякої події, отримано наступний емпіричний розподіл:

Таблиця 10.8 – Емпіричний розподіл:

x_i	0	1	2	3	4	5	6	7
n_i	120	167	130	69	27	5	1	1

Знайти теоретичні частоти n_i' в припущенні, що випадкова величина X (генеральна сукупність) розподілена згідно із законом Пуассона.

Розв'язання

Оскільки параметр λ дорівнює математичному сподіванню розподілу Пуассона, а в якості оцінки приймають вибірккову середню, то

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^k n_i x_i = \frac{780}{520} = 1,5$$

Отже, приймаємо $\lambda = 1,5$.

Тоді формула Пуассона набуває вигляду $P_{520}(m) = \frac{1,5^m}{m!} e^{-1,5}$

За формулою обчислимо ймовірності $P_{520}(m)$ при $m = 0, 1, 2, 3, 4, 5, 6, 7$.

$$P_{520}(0) = 0,22313; \quad P_{520}(1) = 0,33469; \quad P_{520}(2) = 0,251021; \quad P_{520}(3) = 0,125511;$$

$$P_{520}(4) = 0,047066; \quad P_{520}(5) = 0,014120; \quad P_{520}(6) = 0,003530; \quad P_{520}(7) = 0,000755.$$

Теоретичні частоти:

$$n_1' = n \cdot P_{520}(0) = 520 \cdot 0,22313 = 116, \quad n_5' = n \cdot P_{520}(4) = 520 \cdot 0,047066 = 25,$$

$$n_2' = n \cdot P_{520}(1) = 520 \cdot 0,33469 = 174, \quad n_6' = n \cdot P_{520}(5) = 520 \cdot 0,014120 = 7,$$

$$n_3' = n \cdot P_{520}(2) = 520 \cdot 0,251021 = 131, \quad n_7' = n \cdot P_{520}(6) = 520 \cdot 0,003530 = 2,$$

$$n_4' = n \cdot P_{520}(3) = 520 \cdot 0,125511 = 65, \quad n_8' = n \cdot P_{520}(7) = 520 \cdot 0,000755 = 0.$$

Результати порівняння емпіричних та теоретичних частот приведені в таблиці 10.9.

Таблиця 10.9 – Результати порівняння емпіричних та теоретичних частот

n_i	120	167	130	69	27	5	1	1
n_i'	116	174	131	65	25	7	2	0

Порівняно невелике розходження емпіричних і теоретичних частот підтверджує припущення, що розподіл який розглядається підпорядкований закону Пуассона.

Метод сум обчислення вибірових середньої та дисперсії

Метод сум є широко застосовуваним в багатьох галузях, таких як статистика, економіка, соціологія та інші. Його перевагою є простота розрахунку та низька обчислювальна складність, що робить його популярним інструментом для обробки великих обсягів даних.

Нехай вибірка задана у вигляді розподілу рівновіддалених варіант та відповідних їм частот (таблиця 10.10).

Таблиця 10.10 – Розподіл розподілу рівновіддалених варіант та відповідних їм частот

x_i	x_1	x_2	x_3	x_4	...	x_k
n_i	n_1	n_2	n_3	n_4	...	n_k

При використанні методу сум умовні моменти першого та другого порядків знаходять за формулами:

$$M_1^* = \frac{d_1}{n},$$

$$M_2^* = \frac{s_1 + 2s_2}{n},$$

Алгоритм складання розрахункової таблиці при обчислення вибірових середньої та дисперсії методом сум

В перший стовпець таблиці записують вибірові початкові варіанти, розташовані в порядку зростання.

В другій стовпець записують частоти варіант, а об'єм вибірки поміщають в нижню клітину стовпчика.

Вибирають варіанту, яка розташована приблизно в середині варіаційного ряду в якості відносного нуля C . В клітинах рядка, що містить відносний нуль, записують нулі.

У незаповнені над нулем клітини третього стовпця записують послідовно накопичені частоти, а їх сума записується, як параметр b_1 . У незаповнені під нулем клітини третього стовпця записують послідовно накопичені частоти, а їх сума записується, як параметр a_1 .

В четвертому стовпці над та під вже записаним нулем дописують ще по одному нулю. У незаповнені над нулем клітини четвертого стовпця записують послідовно накопичені частоти, а їх сума записується, як параметр b_2 . У незаповнені під нулем клітини четвертого стовпця записують послідовно накопичені частоти, а їх сума записується, як параметр a_2 .

Таблиця 10.11 – Розрахункова таблиця для обчислення методом сум

x_i	n_i	$b_1 = B_1$	$b_2 = B_2$	
x_1	n_1	n_1		} $\Sigma = B_2$
x_2	n_2	$n_1 + n_2$		
x_3	n_3	$n_1 + n_2 + n_3$	0	
x_4	n_4	0	0	
...	...	...	0	
...	...	$n_k + n_{k-1} + n_{k-2}$		} $\Sigma = A_2$
...	...	$n_k + n_{k-1}$		
x_k	n_k	n_k		
	$n = \Sigma n_i$	$a_1 = A_1$	$a_2 = A_2$	

Безпосереднє обчислення вибірових середньої та дисперсії методом сум

Приклад 10.5 Знайти методом сум вибірову середню та вибірову дисперсію по заданому розподілу вибірки (таблиця 10.12)

Таблиця 10.12 – Розподіл вибірки

x_i	48	52	56	60	64	68	72	76	80	84
n_i	2	4	6	8	12	30	18	8	7	5

Розв'язання

Запишемо початкові варіанти та їх частоти в перший та другий стовпець розрахункової таблиці.

В якості відносного нуля оберемо варіанту, яка має найбільшу частоту.

Запишемо над нулем послідовно накопичені частоти:

$$2; \quad 2 + 4 = 6; \quad 6 + 6 = 12; \quad 12 + 8 = 20; \quad 20 + 12 = 32$$

Сума накопичених частот: $b_1 = 72$.

Запишемо під нулем послідовно накопичені частоти:

$$5; \quad 5 + 7 = 12; \quad 12 + 8 = 20; \quad 20 + 18 = 38$$

Сума накопичених частот: $a_1 = 75$.

Виконаємо аналогічні розрахунки для заповнення четвертого стовпця.

В таблиці 10.6.4 приведені результати заповнення розрахункової таблиці.

Таблиця 10.6.4 – Розрахункова таблиця

x_i	n_i	$b_1 = 72$	$b_2 = 70$
48	2	2	2
52	4	6	8
56	6	12	20
60	8	20	40
64	12	32	0
68	30	0	0
72	18	38	0
76	8	20	37
80	7	12	17
84	5	5	5
	$n = 100$	$a_1 = 75$	$a_2 = 59$

Обчислимо параметри d_1 , s_1 , s_2 :

$$d_1 = a_1 - b_1 = 75 - 72 = 3;$$

$$s_1 = a_1 + b_1 = 75 + 72 = 147;$$

$$s_2 = a_2 + b_2 = 59 + 70 = 129.$$

Умовні моменти першого та другого порядків

$$M_1^* = \frac{d_1}{n} = \frac{3}{100} = 0,03;$$

$$M_2^* = \frac{s_1 + 2s_2}{n} = \frac{147 + 2 \cdot 129}{100} = 4,05$$

Вибіркові середня та дисперсія:

$$\bar{x}_B = M_1^* h + C = 0,03 \cdot (52 - 48) + 68 = 68,12;$$

$$D_B = [M_2^* - (M_1^*)^2] h^2 = (4,05 - 0,03^2) \cdot 4^2 \approx 64,78$$

Оцінка відхилення емпіричного розподілу від нормального

Асиметрією емпіричного розподілу

називають відношення центрального моменту третього порядку до кубу середнього квадратичного відхилення.

$$a_s = \frac{m_3}{\sigma^3}$$

Ексцес емпіричного розподілу

— характеристика, що служить для оцінки більшого або меншого підйому кривої розподілу у порівнянні з нормальною кривою яка визначається рівністю.

$$e_k = \frac{m_4}{\sigma^4} - 3$$

Зауваження 1

Моменти зручно обчислювати методом добутків із використанням формул:

$$m_3 = [M_3^* - 3M_2^*M_1^* + 2(M_1^*)^3] h^3$$

$$m_4 = [M_4^* - 4M_3^*M_1^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4] h^4$$

Приклад 10.6 Знайти методом добутків асиметрію та ексцес за наданим розподілом вибірки (таблиця 10.7.1)

Таблиця 10.7.1 – Розподіл вибірки:

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

Розв'язання

Запишемо початкові варіанти та їх частоти в перший та другий стовпець розрахункової таблиці, а в якості відносного нуля оберемо варіанту 16.

В таблиці 10.7.2 наведено результат виконаних обчислень.

Таблиця 10.7.2 – Розрахункова таблиця

x_i	n_i	u_i	$n_i \cdot u_i$	$n_i \cdot u_i^2$	$n_i \cdot u_i^3$	$n_i \cdot u_i^4$	$n_i(u_i + 1)^4$
12	5	-2	-10	20	-40	80	5
14	15	-1	-15	15	-15	15	—
16	50	0	$A_1 = -25$	—	$A_3 = -55$	—	50
18	16	1	16	16	16	16	256
20	10	2	20	40	80	160	810
22	4	3	12	36	108	324	1024
			$A_2 = 48$		$A_4 = 204$		
$h = 2$	$n = 100$		$\sum n_i \cdot u_i$	$\sum n_i \cdot u_i^2$	$\sum n_i \cdot u_i^3$	$\sum n_i \cdot u_i^4$	$\sum n_i (u_i + 1)^4$
			23	127	149	595	2145

Реалізуємо контроль правильності виконання розрахунків:

$$\sum_{i=1}^k n_i \cdot u_i^4 + 4 \sum_{i=1}^k n_i \cdot u_i^3 + 6 \sum_{i=1}^k n_i \cdot u_i^2 + 4 \sum_{i=1}^k n_i \cdot u_i + n = 2145.$$

Умова виконується, отже обчислення проведені вірно.

Умовні моменти першого, другого, третього та четвертого порядків:

$$M_1^* = \frac{1}{n} \sum_{i=1}^m n_i u_i = 0,23; \quad M_2^* = \frac{1}{n} \sum_{i=1}^m n_i u_i^2 = 1,27;$$

$$M_3^* = \frac{1}{n} \sum_{i=1}^m n_i u_i^3 = 1,49; \quad M_4^* = \frac{1}{n} \sum_{i=1}^m n_i u_i^4 = 5,95.$$

Центральні емпіричні моменти m_3, m_4 :

$$m_3 = [M_3^* - 3M_2^*M_1^* + 2(M_1^*)^3] h^3 = 5,104$$

$$m_4 = [M_4^* - 4M_3^*M_1^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4] h^4 = 79,582$$

Вибіркова дисперсія:

$$D_{\epsilon} = [M_2^* - (M_1^*)^2] h^2 = 4,87.$$

Вибіркове середньоквадратичне відхилення (стандарт):

$$\sigma_{\epsilon} = \sqrt{D_{\epsilon}} = \sqrt{4,87}$$

Асиметрія емпіричного розподілу:

$$a_s = \frac{m_3}{\sigma_{\epsilon}^3} = \frac{5,104}{(\sqrt{4,87})^3} = 0,47.$$

Екссес емпіричного розподілу:

$$e_k = \frac{m_4}{\sigma_{\epsilon}^4} - 3 = \frac{79,582}{(\sqrt{4,87})^4} - 3 = 0,36.$$

Висновки

Розуміння цього методу добутків та методу сум дозволяє зробити більш точні та ефективні обчислення, що, в свою чергу, може привести до нових відкриттів та висновків у різних наукових дисциплінах. Обидва методи мають свої переваги та обмеження, і їх вибір для конкретної задачі залежить від контексту та характеристик вибірки. Метод добутків забезпечує достатню точність при обчисленні вибірових середніх та дисперсій і може бути швидким в обчисленнях. Метод сум особливо підходить для невеликих вибірок або ситуацій, де немає необхідності високої точності.

Обидва методи мають свої обмеження. Метод добутків може бути чутливим до викидів у вибірці, що може призвести до неточностей у розрахунках. З іншого боку, метод сум може бути менш точним для великих вибірок або в ситуаціях, де вибірка має нерівномірний розподіл.

Отже, вибір методу для обчислення вибірових середніх та дисперсій залежить від конкретної задачі, обсягу даних, доступності ресурсів та потреби в точності. В багатьох випадках, комбінування методів або використання альтернативних підходів може бути доцільним для отримання найкращих результатів.

Питання та завдання до контролю знань студентів

1 Які переваги методу добутків в порівнянні з методом сум при обчисленні вибірових середніх та дисперсій?

2 Які обмеження можуть виникнути при використанні методу добутків для обчислення вибірових середніх та дисперсій?

3 Які ситуації або типи даних можуть бути більш підходящими для використання методу сум замість методу добутків при обчисленні вибірових середніх та дисперсій?

4 Які можливі наслідки використання методу добутків або методу сум з помилковими або неправильно введеними даними при обчисленні вибірових середніх та дисперсій?

5 Чи існують альтернативні методи обчислення вибірових середніх та дисперсій, які можуть бути ефективнішими або точнішими за методи добутків та сум, і в яких ситуаціях їх варто розглядати?

Лекція № 11

з навчальної дисципліни Теорія ймовірностей та математична статистика

Тема 2.4 Елементи теорії кореляції

Елементи кореляційно-регресійного аналізу.

Функціональна, статистична та кореляційна залежності.

Умовні середні та вибіркові рівняння регресії.

Мета заняття Ознайомити з елементами Елементи кореляційно-регресійного аналізу та рівняннями кореляції

Матеріально-технічне забезпечення та дидактичні засоби, ТЗН

конспект лекцій, мультимедійна презентація, опорний конспект

Час – 2 академічні години

Час – 80 хвилин

План проведення лекції

Структура лекції	Відведений час	Методичні вказівки
1. Організаційна частина	5	Привітання, визначення присутності студентів на занятті
2. Актуалізація опорних знань, перевірка вивченого матеріалу та мотивація навчальної діяльності студентів	10	Мотивація актуальності теми, формулювання мети лекції, огляд головних питань теми
3. Основна частина (викладення навчальних питань лекції)	55	Викладення основного змісту матеріалу лекції: 4 Функціональна залежність. 5 Статистична залежність. 6 Кореляційна залежність.
4. Заклучна частина Домашнє завдання: (відповідно до робочої програми)	10	Узагальнення, теоретичні висновки, закріплення вивченого на лекції матеріалу фронтальним опитуванням студентів за матеріалами лекції

Література

1. Барковський, В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. Посібник / В.В. Барковський, Н.В. Барковська, О.К. Лопатін. – К: Центр навчальної літератури, 2006. – 424с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е.Гмурман. – М.: Издательство Юрайт, 2012. – 479 с.
3. Дубовик, В.П. Вища математика. Збірник задач. [Текст] / В.П.Дубовик, І.І.Юрик. – К.: «А.С.К.», 2001. – 480 с.

ВСТУП

У сучасному світі, де велика кількість даних стає доступною для аналізу, розуміння зв'язків та взаємозв'язків між змінними стає надзвичайно важливим. Одним з ключових понять для виявлення та вимірювання таких зв'язків є кореляція.

Теорія кореляції є важливою галуззю статистики, яка досліджує ступінь залежності між двома або більшими змінними. Вона надає нам інструменти та метрики для вимірювання сили та напрямку цих зв'язків, дозволяючи нам краще розуміти та пояснювати взаємозв'язки між різними явищами.

Теорія кореляції є невід'ємною складовою багатьох наукових галузей, включаючи статистику, психологію, економіку, соціологію та багато інших. Розуміння основних принципів кореляції допомагає нам виявляти та аналізувати залежності, робити прогнози та приймати обґрунтовані рішення.

Кореляційна залежність

В багатьох задачах потрібно встановити та оцінити залежність величини Y , що вивчається, від однієї чи декількох інших величин. Розглянемо спочатку залежність Y від однієї випадкової величини X .

Дві випадкові величини можуть бути зв'язані функціональною залежністю, або залежністю іншого роду, яку називають статистичною, або ж бути незалежними.

Строга функціональна залежність зустрічається рідко, так як обидві величини схильні до впливу випадкових чинників, причому серед них можуть бути і спільні для обох величин (під спільними розуміють фактори, які впливають і на Y і на X). У цьому випадку виникає статистична залежність.

Наприклад, якщо Y залежить від випадкових факторів

$$Z_1, Z_2, V_1, V_2,$$

а X залежить від випадкових факторів

$$Z_1, Z_2, U_1,$$

то між X і Y існує статистична залежність, так як серед випадкових факторів є спільні — Z_1 та Z_2 .

Статистичною називають залежність, при якій зміна однієї з величин приводить до зміни розподілу іншої. Зокрема, статистична залежність проявляється в тому, що при зміні однієї з величин змінюється середнє значення іншої; у цьому випадку статистичну залежність називають *кореляційною*.

Наведемо приклад випадкової величини Y , яка не зв'язана з величиною X функціонально, а зв'язана кореляційно. Нехай Y — урожай зерна, X — кількість добрив. З рівних за площею ділянок землі при рівних кількостях внесених добрив збирають різний врожай, тобто Y не є функцією від X . Це пояснюється дією випадкових факторів (температура повітря, тощо). Разом з тим, як показує досвід, середній урожай є функцією від кількості добрив, тобто Y зв'язана з X кореляційною залежністю.

Для уточнення поняття кореляційної залежності вводять поняття умовної середньої.

Припустимо, що вивчається зв'язок між випадковою величиною Y та випадковою величиною X . Нехай кожному значенню X відповідає декілька значень Y . Наприклад, нехай при $x_1 = 2$ величина Y прийняла значення $y_1 = 5, y_2 = 6, y_3 = 10$. Знайдемо середнє арифметичне цих чисел:

$$\overline{y_2} = \frac{5+6+10}{3} = 7.$$

Число $\overline{y_2}$ називають умовним середнім; риска над буквою y служить позначенням середнього арифметичного, а число 2 показує, що розглядаються ті значення Y , які відповідають $x_1 = 2$.

Стосовно прикладу з попереднього параграфу ці дані можна пояснити так: на кожну з трьох однакових ділянок землі внесли по 2 одиниці добрив і отримали відповідно 5, 6 та 10 одиниць зерна; середній врожай склав 7 одиниць.

Умовним середнім $\overline{y_x}$ називають середнє арифметичне значень Y , що відповідають значенню $X = x$.

Якщо кожному значенню x відповідає одне значення умовної середньої, то, очевидно, умовна середня є функція від x ; у цьому випадку говорять, що випадкова величина Y залежить від X кореляційно.

Кореляційною залежністю Y від X називають функціональну залежність умовної середньої $\overline{y}_x$ від x :

$$\overline{y}_x = f(x) \quad (*)$$

Рівняння $(*)$ називають *рівнянням регресії Y на X* ; функцію $f(x)$ називають *регресією Y на X* , а її графік — *лінією регресії Y на X* .

Аналогічно означається умовна середня $\overline{x}_y$ та кореляційна залежність X від Y .

Умовним середнім $\overline{x}_y$ називають середнє арифметичне значень X , що відповідають значенню $Y = y$.

Кореляційною залежністю X від Y називають функціональну залежність умовної середньої $\overline{x}_y$ від y :

$$\overline{x}_y = \varphi(y) \quad (**)$$

Рівняння $(**)$ називають *рівнянням регресії X на Y* ; функцію $\varphi(y)$ називають *регресією X на Y* , а її графік — *лінією регресії X на Y* .

Дві основні задачі теорії кореляції.

Перша задача теорії кореляції — встановити формулу кореляційного зв'язку, тобто вид функції регресії (лінійна, квадратична, показникова и т.д.). Найчастіше функції регресії виявляються лінійними. Якщо обидві функції регресії $f(x)$ та $\varphi(y)$ лінійні, то кореляцію називають *лінійною*, в іншому випадку — *нелінійною*. При лінійній кореляції обидві лінії регресії є прямими лініями.

Друга задача теорії кореляції — оцінити тісноту (силу) кореляційного зв'язку. Тіснота кореляційного зв'язку Y від X оцінюється по величині розсіювання значень навколо умовного середнього $\overline{y}_x$. Велике розсіювання говорить про слабку залежність Y від X або про відсутність зв'язку. Мале розсіювання вказує на присутність достатньо сильної залежності; можливо, навіть, що Y та X зв'язані функціонально, але під дією другорядних випадкових факторів цей зв'язок виявився розмитим, в результаті чого при одному і тому ж значенні x величина приймає різні значення.

Вибірковий коефіцієнт кореляції

Позначимо через r_B - вибірковий коефіцієнт кореляції наступну рівність:

$$\rho_{xy} \frac{\sigma_x}{\sigma_y} = r_B,$$

або

$$\rho_{xy} = r_B \frac{\sigma_y}{\sigma_x}.$$

Тоді вибіркове рівняння прямої лінії регресії Y на X матиме вигляд:

$$\bar{y}_x - \bar{y} = r_B \frac{\sigma_y}{\sigma_x} (x - \bar{x}).$$

Також вибірковий коефіцієнт кореляції визначається рівністю

$$r_B = \frac{\sum n_{xy} xy - n \bar{x} \bar{y}}{n \sigma_x \sigma_y}$$

де x, y - варіанти ознак X та Y ;

n_{xy} - частота спостережуваної пари ;

n - об'єм вибірки (сума всіх частот);

$\bar{x}, \bar{y}$ - вибіркові середні;

σ_x, σ_y - вибіркові середні квадратичні відхилення.

Властивості вибіркового коефіцієнта кореляції.

Наведемо властивості вибіркового коефіцієнта кореляції, з яких слідує, що він служить для оцінки тісноти лінійної кореляційної залежності.

Скористаємося формулами:

$$S_y = D_y(1 - r_B^2); \quad S_x = D_x(1 - r_B^2)$$

де S_y - дисперсія спостережуваних значень y довкола відповідних умовних середніх $\bar{y}_x$;

D_y - дисперсія спостережуваних значень y довкола загальної середньої $\bar{y}$.

1. Абсолютна величина вибіркового коефіцієнта кореляції не перевищує одиниці.

Доведення. Будь-яка дисперсія невід'ємна. Зокрема,

$$S_y = D_y(1 - r_B^2) \geq 0$$

Отже,

$$1 - r_B^2 \geq 0.$$

Звідси

$$-1 \leq r_B \leq 1,$$

або

$$|r_B| \leq 1.$$

2. Якщо вибірковий коефіцієнт кореляції дорівнює нулю і вибіркові лінії регресії – прямі, то X і Y не зв'язані лінійною кореляційною залежністю.

Доведення. При $r_B = 0$ рівняння вибіркової прямої регресії Y на X

$$\bar{y}_x - \bar{y} = r_B \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

має вигляд:

$$\bar{y}_x - \bar{y} = 0$$

або

$$\bar{y}_x = \bar{y}.$$

При $r_B = 0$ рівняння прямої регресії X на Y має вигляд:

$$\bar{x}_y = \bar{x}.$$

Таким чином, при $r_B = 0$ умовні середні зберігають постійне значення при зміні відповідних аргументів; в такому випадку можна вважати, що X та Y не зв'язані лінійною кореляційною залежністю. А прямі регресії паралельні відповідним координатним осям.

3. Якщо абсолютна величина вибіркового коефіцієнта кореляції дорівнює одиниці, то спостережувані значення ознак зв'язані лінійною функціональною залежністю.

Якщо $|r_B| = 1$, то $S_y = D_y(1 - r_B^2) = 0.$

Звідси слідує рівність:

$$y - \bar{y} - r_B \frac{\sigma_y}{\sigma_x} (x - \bar{x}) = 0$$

Будь-яка пара чисел (x, y) задовольняє це лінійне відносно x та y рівняння, тобто значення ознак у вибірці пов'язані лінійною функціональною залежністю.

4. Із зростанням абсолютної величини вибіркового коефіцієнта кореляції лінійна кореляційна залежність стає більш тісною і при $|r_B| = 1$ переходить у функціональну залежність.

Доведення. Із формул

$$S_y = D_y(1 - r_B^2); \quad S_x = D_x(1 - r_B^2)$$

видно, що зі зростанням абсолютної величини r_B дисперсії S_y і S_x зменшуються, тобто зменшується відстань спостережуваних значень ознак довкола умовних середніх, а це значить, що зв'язок між ознаками стає більш тісним і при $|r_B| = 1$, як слідує із властивості 3, переходить в функціональний.

Із наведених вище властивостей витікає зміст r_B :

вибірковий коефіцієнт кореляції характеризує тісноту лінійного зв'язку між кількісними ознаками у вибірці: чим ближче $|r_B|$ до 1, тим зв'язок сильніший; чим ближче $|r_B|$ до 0, тим зв'язок слабкіший.

Якщо вибірка має достатньо великий обсяг і добре представляє генеральну сукупність, то висновок про тісноту лінійної залежності між ознаками, отриманий за даними вибірки, може бути поширений і на генеральну сукупність. Наприклад, для оцінки коефіцієнта кореляції r_b нормально розподіленої генеральної сукупності (при $n \geq 50$) можна скористатися формулою:

$$r_B - 3 \frac{1-r_B^2}{\sqrt{n}} \leq r_b \leq r_B + 3 \frac{1+r_B^2}{\sqrt{n}}.$$

Вибіркове кореляційне відношення

Для оцінки тісноти лінійного кореляційного зв'язку між ознаками у вибірці служить вибірковий коефіцієнт кореляції. Для оцінки тісноти нелінійного кореляційного зв'язку вводять нові звідні характеристики:

η_{yx} - вибіркове кореляційне відношення Y до X ;

η_{xy} - вибіркове кореляційне відношення X до Y .

Вибірковим кореляційним відношенням Y до X називають відношення між групового середнього квадратичного відхилення до загального середнього квадратичного відхилення ознаки Y

$$\eta_{yx} = \frac{\sigma_{міжгр}}{\sigma_{заг}}$$

або в других позначеннях:

$$\eta_{yx} = \frac{\sigma_{\overline{yx}}}{\sigma_y}.$$

Тут

$$\sigma_{\bar{y}_x} = \sqrt{D_{\text{міжгр}}} = \sqrt{\frac{\sum n_x (\bar{y}_x - \bar{y})^2}{n}}$$

$$\sigma_y = \sqrt{D_{\text{заг}}} = \sqrt{\frac{\sum n_y (y - \bar{y})^2}{n}}$$

де n - об'єм вибірки (сума всіх частот);

n_x - частота значень x ознаки X ;

n_y - частота значень y ознаки Y ;

$\bar{y}$ - загальна середня ознаки Y ;

$\bar{y}_x$ - умовна середня ознаки Y .

Аналогічно визначається вибіркове кореляційне відношення X до Y :

$$\eta_{xy} = \frac{\sigma_{\bar{y}_x}}{\sigma_x}$$

Приклад. Знайти η_{yx} за даними кореляційної таблиці.

	10	20	30	n_y
15	4	28	6	38
25	6	-	6	12
n_x	10	28	12	$n = 50$
$\bar{y}_x$	21	15	20	

Розв'язання. Знайдемо загальну середню

$$\bar{y} = \frac{\sum n_y y}{n} = \frac{38 \cdot 15 + 12 \cdot 25}{50} = 17,4$$

Знайдемо загальне середнє квадратичне відхилення

$$\sigma_y = \sqrt{\frac{\sum n_y (y - \bar{y})^2}{n}} = \sqrt{\frac{38(15 - 17,4)^2 + 12(25 - 17,4)^2}{50}} = 4,27$$

Знайдемо між групове середнє квадратичне відхилення

$$\sigma_{\bar{y}_x} = \sqrt{\frac{\sum n_x (\bar{y}_x - \bar{y})^2}{n}} = \sqrt{\frac{10(21 - 17,4)^2 + 28(15 - 17,4)^2 + 12(20 - 17,4)^2}{50}} = 2,73$$

Тоді шукане кореляційне відношення

$$\eta_{yx} = \frac{\sigma_{\bar{y}_x}}{\sigma_y} = \frac{2,73}{4,27} = 0,64.$$

Властивості вибіркового кореляційного відношення.

Для спрощення далі вибіркве кореляційне відношення η_{yx} будемо позначати через η і називати «кореляційним відношенням».

1. Кореляційне відношення задовольняє подвійну нерівність:

$$0 \leq \eta \leq 1.$$

Доведення. Нерівність $\eta \geq 0$ слідує з того, що η - відношення невід'ємних чисел – середніх квадратичних відхилень (міжгрупового до загального).

Для доведення нерівності $\eta \leq 1$ скористаємося формулою

$$D_{\text{заг}} = D_{\text{внутрішньогр}} + D_{\text{міжгр}}.$$

Поділимо обидві частини рівності на $D_{\text{заг}}$, отримаємо:

$$1 = \frac{D_{\text{внутрішньогр}}}{D_{\text{заг}}} + \frac{D_{\text{міжгр}}}{D_{\text{заг}}},$$

або

$$1 = \frac{D_{\text{внутрішньогр}}}{D_{\text{заг}}} + \eta^2.$$

Так як обидва доданки невід'ємні і сума їх дорівнює одиниці, то кожне з них не більше одиниці; зокрема

$$\eta^2 \leq 1.$$

Враховуючи, що $\eta \geq 0$, отримаємо:

$$0 \leq \eta \leq 1.$$

2. Якщо $\eta = 0$, то ознака Y з ознакою X кореляційним зв'язком не зв'язані.

Доведення. За умовою

$$\eta = \frac{\sigma_{міжгр}}{\sigma_{заг}} = 0.$$

Звідси

$$\sigma_{міжгр} = 0$$

а, отже,

$$D_{міжгр} = 0.$$

Міжгруппова дисперсія – це дисперсія умовних (групових) середніх $\bar{y}_x$ відносно загальної середньої $\bar{y}$. Рівність нулю міжгруппової дисперсії означає, що при всіх значеннях X умовні середні зберігають постійне значення (що дорівнює загальній середній). Тобто, при $\eta = 0$ умовна середня не є функцією від X , а значить ознака Y не пов'язана кореляційною залежністю з ознакою X .

3. Якщо $\eta = 1$, то ознака Y пов'язана з ознакою X функціональною залежністю.

Доведення. За умовою

$$\eta = \frac{\sigma_{міжгр}}{\sigma_{заг}} = 1.$$

Звідси

$$\sigma_{міжгр} = \sigma_{заг}.$$

Піднесемо обидві частини рівності до квадрату і отримаємо:

$$D_{міжгр} = D_{заг}.$$

Так як

$$D_{заг} = D_{внутрішньогр} + D_{міжгр},$$

то

$$D_{внутрішньогр} = 0.$$

Оскільки внутрішньогрупова дисперсія є середнє арифметичне групових дисперсій, то із останньої рівності слідує, що дисперсія кожної групи дорівнює нулю. А це означає, що в групі містяться рівні значення Y , тобто кожному значенню X відповідає одне значення Y . Отже, при $\eta = 1$ ознака Y пов'язана з ознакою X функціональною залежністю.

4. Вибіркове кореляційне відношення не менше абсолютної величини вибіркового коефіцієнта кореляції:

$$\eta \geq |r_B|.$$

5. Якщо вибіркве кореляційне відношення дорівнює абсолютній величині вибіркового коефіцієнта кореляції, тобто $\eta \geq |r_B|$, то має місце точна лінійна кореляційна залежність.

Криволінійна кореляція

Якщо графік регресії $\overline{y_x} = f(x)$ або $\overline{x_y} = \varphi(y)$ зображується кривою лінією, то кореляцію називають *криволінійною*.

Наприклад, функції регресії Y на X можуть мати вигляд:

- $\overline{y_x} = ax^2 + bx + c$ (параболічна кореляція другого порядку);

- $\overline{y_x} = ax^3 + bx^2 + cx + d$ (параболічна кореляція третього порядку);

- $\overline{y_x} = \frac{a}{x} + b$ (гіперболічна кореляція).

Теорія криволінійної кореляції вирішує ті ж завдання, що й теорія лінійної кореляції (встановлення форми та частоти кореляційного зв'язку).

Невідомі параметри рівняння регресії шукають методом найменших квадратів. Для оцінки тісноти криволінійної кореляції використовують вибіркові кореляційні відношення.

Розглянемо параболічну кореляцію другого порядку. Припустимо, що дані n спостережень (вибірки) дозволяють вважати, що має місце саме така кореляція. У цьому випадку вибіркве рівняння регресії Y на X матиме вигляд:

$$\overline{y_x} = Ax^2 + Bx + C \quad (*)$$

де A, B, C - невідомі параметри.

Користуючись методом найменших квадратів, отримаємо систему лінійних рівнянь відносно невідомих параметрів:

$$\left(\sum n_x x^4\right) A + \left(\sum n_x x^3\right) B + \left(\sum n_x x^2\right) C = \sum n_x \overline{y_x} x^2;$$

$$\left(\sum n_x x^3\right) A + \left(\sum n_x x^2\right) B + \left(\sum n_x x\right) C = \sum n_x \overline{y_x} x;$$

$$\left(\sum n_x x^2\right) A + \left(\sum n_x x\right) B + nC = \sum n_x \overline{y_x}.$$

Знайдені параметри A, B, C підставляємо в рівняння регресії (*).

Приклад 1. Знайти вибіркове рівняння регресії $\overline{y_x} = Ax^2 + Bx + C$ за даними, наведеними в кореляційній таблиці.

	1	1,1	1,2	n_y
6	8	2		10
7		30		30
7,5		1	9	10
n_x	8	33	9	$n = 50$
$\overline{y_x}$	6	6,73	7,5	

Складемо розрахункову таблицю

x	n_x	$\overline{y_x}$	$n_x x$	$n_x x^2$	$n_x x^3$	$n_x x^4$	$n_x \overline{y_x}$	$n_x \overline{y_x} x$	$n_x \overline{y_x} x^2$
1	8	6	8	8	8	8	48	48	48
1,1	33	6,73	36,3	39,93	43,93	48,32	222,09	244,30	268,73
1,2	9	7,5	10,8	12,96	15,55	18,66	67,50	81	97,20
Σ	50		55,1	60,89	67,48	74,98	337,59	373,30	413,93

Підставивши числа в формулу (*), отримаємо систему рівнянь відносно невідомих коефіцієнтів A, B, C :

$$74,98A + 67,48B + 60,89C = 413,93$$

$$67,48A + 60,89B + 55,10C = 373,30$$

$$60,89A + 55,10B + 50C = 337,59$$

Розв'язавши цю систему, отримаємо $A = 1,94, B = 2,98, C = 1,10$.

Підставивши знайдені коефіцієнти в рівняння регресії $\overline{y_x} = Ax^2 + Bx + C$, отримаємо

$$\overline{y_x} = 1,94x^2 + 2,98x + 1,10.$$

Множинна кореляція

Якщо досліджується зв'язок між декількома ознаками, то кореляцію називають *множинною*.

У найпростішому випадку число ознак дорівнює трьом, і зв'язок між ними лінійний:

$$z = ax + by + cz.$$

У такому випадку виникають задачі:

1. знайти за даними спостережень вибіркоче рівняння вигляду

$$z = Ax + By + Cz \tag{1}$$

тобто потрібно знайти коефіцієнти регресії A, B та параметр C ;

2. оцінити тісноту зв'язку між Z та ознаками X, Y ;

3. оцінити тісноту зв'язку між Z та X (при сталій Y), між Z та Y (при сталій X).

Перша задача розв'язується методом найменших квадратів, причому замість рівняння (1) зручно шукати рівняння зв'язку вигляду

$$z - \bar{z} = A(x - \bar{x}) + B(y - \bar{y}),$$

де

$$A = \frac{r_{xz} - r_{yz}r_{xy}}{1 - r_{xy}^2} \cdot \frac{\sigma_z}{\sigma_x}; \quad B = \frac{r_{yz} - r_{xz}r_{xy}}{1 - r_{xy}^2} \cdot \frac{\sigma_z}{\sigma_y}$$

Тут r_{xz} , r_{yz} , r_{xy} - коефіцієнти кореляції відповідно між ознаками X та Z, Y та Z, X та Y; σ_x , σ_y , σ_z - середні квадратичні відхилення.

Тіснота зв'язку ознаки з ознаками оцінюється вибіркоvim сукупним коефіцієнтом кореляції:

$$R = \sqrt{\frac{r_{xz}^2 - 2r_{xy}r_{xz}r_{yz} + r_{yz}^2}{1 - r_{xy}^2}}$$

причому $0 \leq R \leq 1$.

Тіснота зв'язку між Z та X (при сталій Y), між Z та Y (при сталій X) оцінюється відповідно частковими вибілковими коефіцієнтами кореляції:

$$r_{xz(y)} = \frac{r_{xz} - r_{xy}r_{yz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{yz}^2)}}$$

$$r_{yz(x)} = \frac{r_{yz} - r_{xy}r_{xz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{xz}^2)}}$$

Ці коефіцієнти мають такі ж властивості та зміст, як і звичайний вибілковий коефіцієнт кореляції, тобто слугують для оцінки лінійного зв'язку між ознаками.

Знайти вибілкові рівняння регресії $\bar{y} = Ax^2 + Bx + c$ за даними кореляційної таблиці.

Задача 1.

	2	3	5	n_y
25	20	-	-	20
45	-	30	1	31
110	-	1	48	49
n_x	20	31	49	$n = 100$

Відповідь: $\bar{y} = 2,94x^2 + 7,27x - 1,25$

Задача 2.

	1	2	n_y
2	30	1	31
6	1	18	19
n_x	31	19	$n = 50$

Відповідь: $\bar{y} = 0,39x^2 + 2,49x - 0,75$

Основні поняття дисперсійного аналізу

Нехай генеральні сукупності $X_1, X_2, \dots, X_p$ розподілені нормально і мають однакову, хоч і невідому, дисперсію; математичні сподівання також відомі, але можуть бути різними. Потрібно при заданому рівні значимості за вибірковими середніми перевірити нульову гіпотезу:

$$H_0: M(X_1) = M(X_2) = \dots = M(X_p)$$

про рівність всіх математичних сподівань. Тобто потрібно встановити значимо чи не значимо відрізняються вибіркові середні. Здавалосьь, для порівняння декількох середніх ($p > 2$) можна порівняти їх попарно. Проте зі збільшенням числа середніх, збільшується і найбільша різниця між ними: середнє нової вибірки може виявитися більше найбільшого або менше найменшого із середніх, отриманих до нового випробування. По цій причині для порівняння декількох середніх користуються іншим

методом, яких оснований на порівнянні дисперсій і тому називається дисперсійним аналізом.

Дисперсійний аналіз використовує властивість адитивності дисперсії випадкової величини, що обумовлено дією незалежних факторів. У залежності від числа джерел дисперсії розрізняють однофакторний та багатофакторний дисперсійний аналіз.

На практиці дисперсійний аналіз застосовують, щоб установити чи істотно впливає деякий якісний фактор, який має рівнів на величину, що вивчається.

Основна ідея дисперсійного аналізу полягає в порівнянні «факторної дисперсії», породжуваної дією фактора та «залишкової дисперсії», обумовленої випадковими причинами. Якщо відмінність між цими дисперсіями значима, то фактор має істотний вплив на X ; в цьому випадку середні спостережуваних значень на кожному рівні (групові середні) будуть відрізнятися також істотно.

Якщо уже встановлено, що фактор істотно впливає на X , і потрібно установити який із рівнів надає найбільший вплив, то додатково проводять попарне порівняння середніх.

Іноді дисперсійний аналіз застосовують, щоб встановити однорідність декількох сукупностей (дисперсії цих сукупностей однакові за припущенням; якщо дисперсійний аналіз покаже, що і математичні сподівання однакові, то у цьому сенсі сукупності однорідні).

У найбільш складних випадках досліджують вплив декількох факторів на декількох постійних або випадкових рівнях і з'ясовують вплив окремих рівнів і їх комбінацій (*багатофакторний аналіз*).

Дисперсійний аналіз особливо ефективний при вивченні кількох факторів. При класичному методі вивчення змінюють тільки один фактор, а решту залишають постійними. При цьому для кожного фактору проводиться своя серія спостережень, що не використовується при вивченні інших факторів. Крім того, при такому методі досліджень не вдається визначити взаємодію факторів при одночасній їх зміні. При дисперсійному аналізі кожне спостереження служить для одночасної оцінки всіх факторів та їх взаємодії.

Дисперсійний аналіз полягає у виділенні й оцінюванні окремих факторів, що викликають зміну досліджуваної випадкової величини. При цьому проводиться

розклад сумарної вибіркової дисперсії на складові, обумовлені незалежними факторами. Кожна з цих складових є оцінкою дисперсії генеральної сукупності. Щоб дати оцінку дієвості впливу даного фактору, необхідно оцінити значимість відповідної вибіркової дисперсії у порівнянні з дисперсією відтворення, обумовленою випадковими факторами. Перевірка значимості оцінок дисперсії проводять з допомогою критерію Фішера.

Коли розрахункове значення критерію Фішера виявиться меншим табличного, то вплив досліджуваного фактору немає підстав вважати значимим. Коли ж розрахункове значення критерію Фішера виявиться більшим табличного, то цей фактор впливає на зміни середніх. Вважається, що виконуються наступні припущення:

1. Випадкові помилки спостережень мають нормальний розподіл.

2. Фактори впливають тільки на зміну середніх значень, а дисперсія спостережень залишається постійною.

Фактори, що розглядаються в дисперсійному аналізі, бувають трьох родів:

1. З випадковими рівнями, коли вибір рівнів проходить з безмежної сукупності можливих рівнів і рівні вибираються випадковим чином;

2. З фіксованими рівнями;

3. Змішаного типу - частина факторів розглядається на фіксованих рівнях, але рівні решти вибираються випадковим чином.

Дисперсійний аналіз застосовується в різних формах в залежності від структури об'єкту, що досліджується; вибір відповідної форми є однією з головних труднощів в практичному застосуванні аналізу.

Загальна, факторна дисперсія

Нехай на кількісну нормально розподілену ознаку X діє фактор F , який має p постійних рівнів. Вважатимемо, що число спостережень на кожному рівні однакове і дорівнює q .

Нехай спостерігалось pq значень x_{ij} ознаки X , де i - номер випробування ($i = 1, 2, \dots, q$), j - номер рівня фактору ($j = 1, 2, \dots, p$). Результати спостережень наведено в таблиці

Номер випробування	Рівні фактору F_j			
	F_1	F_2	...	F_p
1	x_{11}	x_{21}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2p}
...	...	...	...	...
q	x_{q1}	x_{q2}	...	x_{qp}
Групова середня	$\bar{x}_{ep1}$	$\bar{x}_{ep2}$	...	$\bar{x}_{ep p}$

Введемо за означенням:

$$S_{заг} = \sum_{j=1}^p \sum_{i=1}^q (x_{ij} - \bar{x})^2$$

(загальна сума квадратів відхилень спостережуваних значень від загальної середньої $\bar{x}$),

$$S_{факт} = q \sum_{j=1}^p (x_{epj} - \bar{x})^2$$

(факторна сума квадратів відхилень групових середніх від загальної середньої, яка характеризує розсіювання «між групами»),

$$S_{залиш} = \sum_{i=1}^q (x_{i1} - \bar{x}_{ep1})^2 + \sum_{i=1}^q (x_{i2} - \bar{x}_{ep2})^2 + \dots + \sum_{i=1}^q (x_{ip} - \bar{x}_{ep p})^2$$

(залишкова сума квадратів відхилень спостережуваних значень групи від своєї групової середньої, яка характеризує розсіювання «усередині груп»).

Практично залишкову суму знаходять з рівності:

$$S_{залиш} = S_{заг} - S_{факт}.$$

Елементарними перетвореннями можна отримати формули, більш зручні для розрахунків:

$$S_{заг} = \sum_{j=1}^p P_j - \frac{[\sum_{j=1}^p R_j]^2}{pq},$$

$$S_{факт} = \frac{\sum_{j=1}^p R_j^2}{q} - \frac{[\sum_{j=1}^p R_j]^2}{pq},$$

де $P_j = \sum_{i=1}^q x_{ij}^2$ — сума квадратів значень ознаки на рівні F_j ,

$R_j = \sum_{i=1}^q x_{ij}$ — сума значень ознаки на рівні F_j .

Для спрощення обчислень віднімають від кожного спостережуваного значення одне і теж число C , приблизно рівне загальній середній. Якщо зменшувати значення $y_{ij} = x_{ij} - 52$, то

$$S_{заг} = \sum_{j=1}^p Q_j - \frac{[\sum_{j=1}^p T_j]^2}{pq},$$

$$S_{факт} = \frac{\sum_{j=1}^p T_j^2}{q} - \frac{[\sum_{j=1}^p T_j]^2}{pq},$$

де $Q_j = \sum_{i=1}^q y_{ij}^2$ — сума квадратів зменшених значень ознаки на рівні F_j ,

$T_j = \sum_{i=1}^q y_{ij}$ — сума зменшених значень ознаки на рівні F_j .

Поділивши суми квадратів відхилень на відповідне число степенів свободи, отримаємо загальну, факторну та залишкову дисперсію:

$$S_{заг}^2 = \frac{S_{заг}}{pq - 1}, \quad S_{факт}^2 = \frac{S_{факт}}{p - 1}, \quad S_{залиш}^2 = \frac{S_{залиш}}{p(q - 1)}$$

де p - число рівнів фактора.

q — число спостережень на кожному рівні.

Якщо нульова гіпотеза про рівність середніх справедлива, то всі ці дисперсії є незміщеними оцінками генеральної дисперсії. Наприклад, якщо об'єм вибірки $n = pq$, то

$$S_{заг}^2 = \frac{S_{заг}}{pq - 1} = \frac{S_{заг}}{n - 1}$$

- виправлена вибіркова дисперсія, яка як відомо, є незміщеною оцінкою генеральної дисперсії.

Методи дисперсійного аналізу

Ставиться задача: перевірити при заданому рівні значимості нульову гіпотезу про рівність декількох ($p > 2$) середніх нормальних сукупностей з невідомими, але однаковими дисперсіями. Розв'язання цієї задачі зводиться до порівняння факторної та залишкової дисперсій.

1. Нехай нульова гіпотеза про рівність деяких середніх (будемо називати їх груповими) правильна. У цьому випадку факторна та залишкова дисперсії є незміщеними оцінками невідомої генеральної дисперсії і, відповідно, різняться несуттєво. Якщо порівняти ці оцінки за критерієм, то очевидно, критерій покаже, що нульову гіпотезу про рівність факторної та залишкової дисперсій слід прийняти.

Таким чином, якщо гіпотеза про рівність групових середніх правильна, то правильна і гіпотеза про рівність факторної та залишкової дисперсій.

2. Нехай нульова гіпотеза про рівність групових середніх помилкова. У цьому випадку зі зростанням розбіжності між груповими середніми буде збільшуватися факторна дисперсія, а разом з нею і відношення $F_{спост} = \frac{S_{факт}^2}{S_{залиш}^2}$. У результаті $F_{спост}$ виявиться більше $F_{кр}$ і, отже, гіпотеза про рівність дисперсій буде відхилена.

Таким чином, якщо гіпотеза про рівність групових середніх помилкова, то помилкова і гіпотеза про рівність факторної та залишкової дисперсій.

Для того, щоб перевірити нульову гіпотезу про рівність групових середніх нормальних сукупностей з однаковими дисперсіями, достатньо перевірити за критерієм F нульову гіпотезу про рівність факторної та залишкової дисперсій. Саме у цьому і полягає метод дисперсійного аналізу.

Порівняння двох дисперсій

Нехай генеральні сукупності X та Y розподілені нормально. По незалежним вибіркам об'ємом n_1, n_2 знайдені виправлені вибіркові дисперсії s_X^2, s_Y^2 . Необхідно по виправленим дисперсіям, при заданому рівні значимості α , перевірити нульову гіпотезу, яка полягає в тому, що генеральні дисперсії розглядуваних сукупностей рівні між собою:

$$H_0: D(X) = D(Y).$$

Правило 1. Для того, щоб при заданому рівні значимості α перевірити нульову гіпотезу $H_0: D(X) = D(Y)$ про рівність генеральних дисперсій двох нормальних сукупностей при конкуруючій гіпотезі $H_0: D(X) > D(Y)$, потрібно поррахувати відношення більшої виправленої дисперсії до меншої, тобто

$$F_{\text{набл}} = \frac{s_{\sigma}^2}{s_m^2}$$

і за таблицею критичних точок розподілу Фішера – Снедекора при заданому рівні значимості α та числам степенів свободи k_1, k_2 знайти критичну точку $F_{\text{кр}}(\alpha, k_1, k_2)$.

Якщо $F_{\text{набл}} < F_{\text{кр}}$ — немає підстав відкинути нульову гіпотезу.

Якщо $F_{\text{набл}} > F_{\text{кр}}$ — нульову гіпотезу відкидають.

Правило 2. Для того, щоб при заданому рівні значимості α перевірити нульову гіпотезу $H_0: D(X) = D(Y)$ про рівність генеральних дисперсій двох нормальних сукупностей при конкуруючій гіпотезі $H_0: D(X) \neq D(Y)$, потрібно поррахувати відношення більшої виправленої дисперсії до меншої, тобто

$$F_{\text{набл}} = \frac{s_{\sigma}^2}{s_m^2}$$

і за таблицею критичних точок розподілу Фішера – Снедекора при заданому рівні значимості $\frac{\alpha}{2}$ та числам степенів свободи k_1, k_2 знайти критичну точку $F_{\text{кр}}(\frac{\alpha}{2}, k_1, k_2)$.

Якщо $F_{\text{набл}} < F_{\text{кр}}$ — немає підстав відкинути нульову гіпотезу.

Якщо $F_{\text{набл}} > F_{\text{кр}}$ — нульову гіпотезу відкидають.

Приклад 1. За двома незалежними вибірками, об'єми яких $n_1 = 12, n_2 = 15$, вибраних з нормальних генеральних сукупностей X і Y , знайдено виправлені вибіркові дисперсії $s_X^2 = 11,41, s_Y^2 = 6,52$. Потрібно при рівні значимості $\alpha = 0,05$ перевірити нульову гіпотезу $H_0: D(X) = D(Y)$ про рівність генеральних дисперсій при альтернативній гіпотезі $H_0: D(X) > D(Y)$.

Розв'язання. Знайдемо відношення більшої виправленої дисперсії до меншої

$$F_{\text{набл}} = \frac{11,41}{6,52} = 1,75.$$

Так як конкуруюча гіпотеза має вигляд $H_0: D(X) > D(Y)$, критична область правостороння.

За рівнем значимості $\alpha = 0,05$ та числам степенів свободи $k_1 = 12 - 1 = 11, k_2 = 15 - 1 = 14$, знаходимо критичну точку $F_{кр}(0,05; 11; 14)$ за таблицею.

Так як $F_{\text{набл}} < F_{кр}$ - нема підстав відхилити нульову гіпотезу про рівність генеральних дисперсій.

Приклад 2 . За двома незалежними вибірками, об'єми яких $n_1 = 10, n_2 = 18$, вибраних з нормальних генеральних сукупностей X і Y , знайдено виправлені вибіркові дисперсії $s_X^2 = 1,23, s_Y^2 = 0,41$. Потрібно при рівні значимості $\alpha = 0,1$ перевірити нульову гіпотезу $H_0: D(X) = D(Y)$ про рівність генеральних дисперсій при конкуруючій гіпотезі $H_0: D(X) \neq D(Y)$.

Розв'язання. Знайдемо відношення більшої виправленої дисперсії до меншої

$$F_{\text{набл}} = \frac{1,23}{0,41} = 3.$$

За умовою конкуруюча гіпотеза має вигляд $H_0: D(X) \neq D(Y)$, тому критична область двостороння.

За рівнем значимості $\frac{\alpha}{2} = \frac{0,1}{2} = 0,05$ та числам степенів свободи $k_1 = 10 - 1 = 9, k_2 = 18 - 1 = 17$, знаходимо критичну точку $F_{кр}(0,05; 9; 17) = 2,50$ за таблицею.

Так як $F_{\text{набл}} > F_{кр}$ нульову гіпотезу про рівність генеральних дисперсій відхиляємо. Іншими словами, вибіркові виправлені дисперсії різняться значимо.

Висновки

Отже, кореляція вимірює ступінь зв'язку між двома змінними і допомагає встановити, наскільки сильно вони взаємозв'язані та може бути позитивною або негативною (коли збільшення). Величина кореляційного коефіцієнта не вказує на причинно-наслідковий зв'язок між змінними. Кореляція може бути наслідком інших факторів або просто випадковим збігом.

Важливо враховувати інші фактори і проводити подальші аналізи, щоб зробити висновки про взаємозв'язок між змінними належним чином.

Загалом, кореляція є корисним інструментом для дослідження статистичних зв'язків, але вимагає обережного інтерпретування та додаткового аналізу для отримання повної карти взаємозв'язків між змінними.

Основна ідея дисперсійного аналізу полягає в порівнянні внутрішньої дисперсії (варіації в межах кожної групи) зовнішньої дисперсії (варіації між групами) для встановлення наявності статистично значущих різниць. Дисперсійний аналіз допомагає встановити наявність статистично значущих різниць між групами, що сприяє кращому розумінню даних і підтримує прийняття обґрунтованих рішень.

Питання та завдання до контролю знань студентів

- 1 Що таке кореляція і як вона вимірюється?
- 2 Які види кореляції існують і в чому полягає їхнє використання?
- 3 Як можна інтерпретувати коефіцієнт кореляції?
- 4 Які фактори впливають на величину кореляційного коефіцієнта?
- 5 Як кореляція відрізняється від причинно-наслідкового зв'язку?
- 6 Що таке дисперсійний аналіз і в яких випадках він застосовується?
- 7 Які основні кроки виконуються під час проведення дисперсійного аналізу?
- 8 Які основні типи дисперсійного аналізу і в чому полягають їхні відмінності?
- 9 Як інтерпретувати результати дисперсійного аналізу і що означають F-статистика та p-значення?
- 10 Які переваги і обмеження має дисперсійний аналіз як статистичний метод?