

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ВНУТРІШНІХ СПРАВ
КРЕМЕНЧУЦЬКИЙ ЛЬОТНИЙ КОЛЕДЖ**

Циклова комісія природничих дисциплін

ТЕКСТ ЛЕКЦІЇ

з навчальної дисципліни «Теорія ймовірностей і математична статистика»
обов'язкових компонент освітньо-професійної програми
першого (бакалаврського) рівня вищої освіти
Облік і аудит

за темою №6— Генеральна та вибіркова сукупність. Варіаційні ряди.
Числові характеристики вибірки та їх статистичні оцінки

Харків 2022

ЗАТВЕРДЖЕНО

Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від 30.08.2022 №8

СХВАЛЕНО

Методичною радою Кременчуцького
льотного коледжу Харківського
національного університету
внутрішніх справ
Протокол від 22.08.2022 №1

ПОГОДЖЕНО

Секцією Науково-методичної ради
ХНУВС з технічних дисциплін
Протокол від 29.08.2022 №8

Розглянуто на засіданні циклової комісії природничих дисциплін, протокол
від 10.08.2022 № 1

Розробник: доцент циклової комісії природничих дисциплін, к.ф. –м.н.
Семенов В.О.

Рецензенти:

1. Доцент кафедри автомобілі та трактори Кременчуцького національного університету імені Михайла Остроградського, кандидат технічних наук Черниш А.А.
2. Ст.викладач циклової комісії економіки та управління Кременчуцького льотного коледжу Харківського національного університету внутрішніх справ, кандидат економічних наук Цимбалістова О.А.

План лекції

1. Завдання математичної статистики. Генеральна й вибіркова сукупності.
2. Емпірична функція розподілу. Полігон і гістограма.
3. Числові характеристики вибіркової сукупності.

Рекомендована література:

Основна

1. Барковський В. В. Теорія ймовірностей та математична статистика ; 5-те видання : навчальний посібник / В. В. Барковський, Н. В. Барковська, О. К. Лопатін. – К. : Центр учбової літератури, 2010. – 424 с.
2. Волощенко А. Б. Теорія ймовірностей і математична статистика : навчально-методичний посібник для самостійного вивчення дисципліни / А. Б. Волощенко, І. А. Джалладова. – К. : КНЕУ, 2003. – 256 с.
3. Жильцов О.Б. Теорія ймовірностей та математична статистика у прикладах і задачах: навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов ; за ред. Г.О. Михаліна. — К. : Київ. ун-т ім. Б. Грінченка, 2015. — 336 с.
4. Конспект лекцій.

Інформаційні ресурси в інтернеті

5. Жильцов О.Б. Теорія ймовірностей та математична статистика у прикладах і задачах: навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов ; за ред. Г.О. Михаліна. — К. : Київ. ун-т ім. Б. Грінченка, 2015. — 336 с.
[\[https://yandex.ua/search/?lr=143&oprnd=9346860898&text=жильцов%20теорія%20ймовірностей\]](https://yandex.ua/search/?lr=143&oprnd=9346860898&text=жильцов%20теорія%20ймовірностей)

Текст лекції

1. Завдання математичної статистики. Генеральна й вибіркова сукупності.

Математичною статистикою називається наука, що займається методами обробки дослідних даних, отриманих у результаті спостережень над випадковими явищами. Будь-який такий результат можна подати як сукупність значень, отриманих у результаті n випробувань над якоюсь випадковою величиною або системою випадкових величин.

Предмет математичної статистики складають методи реєстрації, опису й аналізу статистичних експериментальних даних, отриманих у результаті спостереження масових випадкових явищ. До основних задач математичної статистики відносяться:

- а) Задача визначення закону розподілу випадкової величини за статистичними даними.
- б) Задача перевірки статистичних гіпотез.
- в) Задача визначення невідомих параметрів розподілу.

Генеральною сукупністю в математичній статистиці називається множина однотипних об'єктів, кількісна чи якісна ознака яких підлягає вивченню.

Підмножина об'єктів, вилючених у відповідний спосіб із генеральної сукупності, називається **вибірковою сукупністю**. Якщо елементи вибірки записати в порядку зростання, отримаємо **варіаційний ряд**.

Попарно **різні** елементи x_i , $i = 1, 2, \dots, k$ варіаційного ряду називають **варіантами**; кількість значень варіанти x_i у вибірці — її **частотою** n_i (сума частот усіх варіант дорівнює об'єму вибірки), відношення частоти до об'єму вибірки — **відносною частотою** або **емпіричною ймовірністю**, або **статистичною ймовірністю**.

Статистичним розподілом частот або відносних частот по варіантах (коротко: **статистичним розподілом вибірки**) називають перелік варіант варіаційного ряду з відповідними частотами або відносними частотами. Статистичний розподіл також задають у вигляді послідовності проміжків $[a_{i-1}; a_i)$, $i = 1, 2, \dots, k$, у які попадають спостерігаємі значення випадкової величини X і відповідних їм частот або відносних частот (статистичних імовірностей). Частотою інтервалу вважають суму частот усіх варіант з даного інтервалу. Щоб підкреслити зазначені відмінності в першому випадку говорять про **дискретний**, а в другому — про **інтервальний** статистичний розподіл вибірки.

Приклад 1. Під час дослідження випадкової величини X із генеральної сукупності було отримано вибірку 4, 3, 6, 4, 7, 2, 5, 1, 2, 5, 4, 4, 3, 5, 6, 3, 4, 1, 3, 4. Знайти об'єм вибірки, побудувати варіаційний ряд вибірки та її статистичний розподіл.

Розв'язання. Оскільки вибірка складається з 20 значень, то об'єм вибірки $n = 20$. Побудуємо варіаційний ряд вибірки:

1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7.

У даній вибірці сім різних значень варіант: 1, 2, 3, 4, 5, 6, 7.

Знайдемо відповідні частоти:

$n_1 = 2; n_2 = 2; n_3 = 4; n_4 = 6; n_5 = 3; n_6 = 2; n_7 = 1$.

Запишемо шуканий статистичний розподіл у вигляді таблиці:

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

Приклад 2. Вибірка задана розподілом частот

x_i	3	5	7	10	15
n_i	2	4	7	4	3

Знайти розподіл відносних частот.

Розв'язання. Знайдемо об'єм вибірки: $n = 2 + 4 + 7 + 4 + 3 = 20$.

Визначимо відносні частоти:

$$w_1 = \frac{2}{20} = 0,1; \quad w_2 = \frac{4}{20} = 0,2; \quad w_3 = \frac{7}{20} = 0,35;$$

$$w_4 = \frac{4}{20} = 0,2; \quad w_5 = \frac{3}{20} = 0,15.$$

Отже, шуканий розподіл відносних частот має вигляд:

x_i	3	5	7	10	15
p_i	0,1	0,2	0,35	0,2	0,15

Сума відносних частот має дорівнювати 1.

$$0,1 + 0,2 + 0,35 + 0,2 + 0,15 = 1.$$

Розподіл побудовано правильно.

2. Емпірична функція розподілу. Полігон і гістограма.

Емпіричною функцією розподілу дискретної випадкової величини X або функцією дискретного розподілу вибірки називають функцію

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

де n_x — кількість елементів вибірки, менших від x (тобто сума частот усіх варіант, менших від x); n — об'єм вибірки. Легко побачити, що число $\frac{n_x}{n}$ є

сумою відносних частот усіх варіант x_i , менших за x . Емпірична функція розподілу $F^*(x)$ має такі самі властивості, як і функція розподілу випадкової величини X , зокрема, якщо якщо x_1 — найменша варіанта, а x_k — найбільша, то $F^*(x) = 0$ при $x \leq x_1$ і $F^*(x) = 1$ при $x > x_k$.

Приклад 3. Знайти емпіричну дискретну функцію розподілу за статистичним розподілом вибірки

x_i	3	5	7	10	15
n_i	2	4	7	4	3

Розв'язання. Знайдемо об'єм вибірки:

$$n = 2 + 4 + 7 + 4 + 3 = 20.$$

Оскільки найменша варіанта дорівнює 3, то $F^*(x) = 0$ при всіх $x \leq 3$.

Значення $X < 5$, а саме значення $x_1 = 3$, спостерігалось двічі, тому

$$F^*(x) = \frac{2}{20} = 0,1 \text{ при всіх } 3 < x \leq 5.$$

Значення $X < 7$, а саме значення $x_1 = 3$ і $x_2 = 5$, спостерігалось $2 + 4 = 6$

разів, тому $F^*(x) = \frac{6}{20} = 0,3$ при $5 < x \leq 7$.

Значення $X < 10$, а саме значення $x_1 = 3$, $x_2 = 5$ і $x_3 = 7$, спостерігалось $2 + 4 + 7 = 13$ разів, тому $F^*(x) = \frac{13}{20} = 0,65$ при $7 < x \leq 10$.

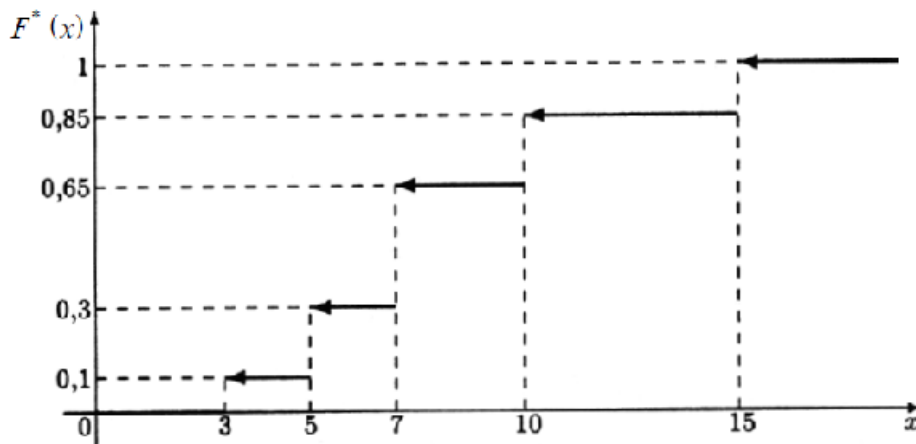
Значення $X < 15$, а саме значення $x_1 = 3$, $x_2 = 5$, $x_3 = 7$ і $x_4 = 10$, спостерігалось $2 + 4 + 7 + 4 = 17$ разів, тому $F^*(x) = \frac{17}{20} = 0,85$ при $10 < x \leq 15$.

Оскільки $x_5 = 15$ — найбільша варіанта, то $F^*(x) = 1$ при $x > 15$.

Отже, шукана емпірична функція має вигляд:

$$F^*(x) = \begin{cases} 0, & \text{якщо } x \leq 3; \\ 0,1, & \text{якщо } 3 < x \leq 5; \\ 0,3, & \text{якщо } 5 < x \leq 7; \\ 0,65, & \text{якщо } 7 < x \leq 10; \\ 0,85, & \text{якщо } 10 < x \leq 15; \\ 1, & \text{якщо } 15 < x. \end{cases}$$

Графік цієї функції зображено на рисунку:



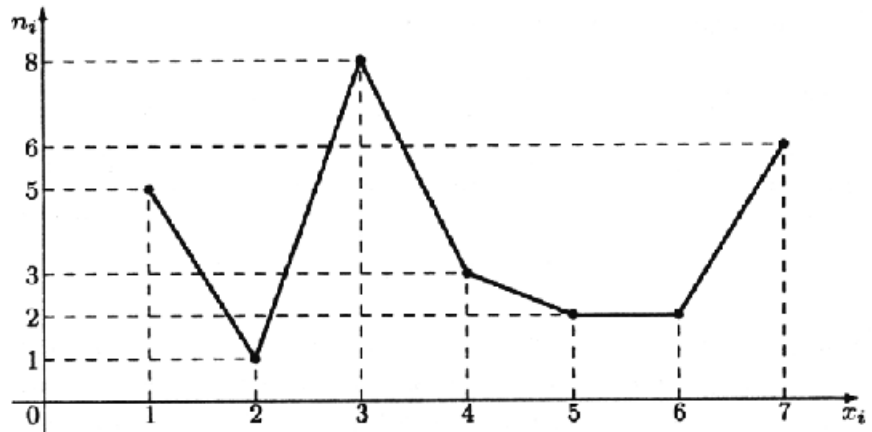
Якщо випадкова величина X розглядається як дискретна, то для неї вводиться поняття полігона частот або полігона відносних частот відповідної вибірки. *Полігоном частот* називають ламану, відрізки якої з'єднують точки (x_1, n_1) , (x_2, n_2) , ..., (x_k, n_k) , де x_i - упорядковані за величиною варіанти вибірки, а n_i - відповідні їм частоти, $i = 1, 2, \dots$

Полігоном відносних частот називають ламану, відрізки якої з'єднують точки (x_1, p_1) , (x_2, p_2) , ..., (x_k, p_k) , де x_i — варіанти вибірки, а p_i — відповідні їм статистичні ймовірності, $i = 1, 2, \dots, k$.

Приклад 4. Побудувати полігон частот за статистичним розподілом вибірки

x_i	1	2	3	4	5	6	7
n_i	5	1	8	3	2	2	6

Розв'язання. Відкладемо на осі абсцис значення варіант x_i , а на осі ординат — значення відповідних їм частот n_i . Послідовно з'єднуючи між собою точки (x_i, n_i) відрізками, дістаємо полігон частот:



У разі, коли X — неперервна величина і обсяг вибірки великий, результати вибірки подають інтервальним рядом. Для побудови інтервального статистичного розподілу вибірки у разі великої кількості спостережень увесь проміжок $[a; b)$, у якому розміщені спостережувані значення випадкової величини X , як правило, розбивають на декілька частинних проміжків $[a_{i-1}; a_i)$, $i = 1, 2, \dots, k$ однакової довжини h і знаходять n_i - суму частот)варіант, що потрапляють в i -й проміжок. Для вибору числа h рекомендовано використовувати формулу

$$h \approx \frac{x_{\max} - x_{\min}}{1 + 3,2 \lg n},$$

де x_{\max} , x_{\min} — відповідно найбільше і найменше значення вибірки; n — об'єм вибірки. Здобутий ряд геометрично подається ступінчатою фігурою — *гістограмою*. Для побудови її на осі абсцис відкладають інтервали, а на них як на основах будують прямокутники, висота яких про-порційна до частоти (відносної частоти) інтервалу. Гістограма дає певне уявлення про графік щільності розподілу.

Приклад 5. Вибірку задано інтервальним розподілом частот

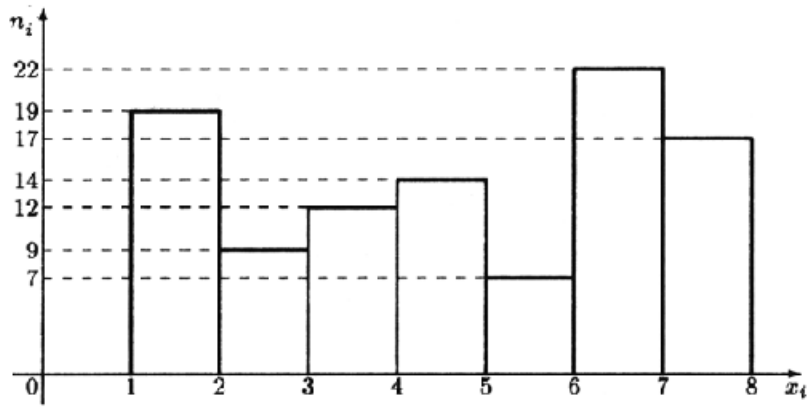
$[x_i; x_{i+1})$	$[1; 2)$	$[2; 3)$	$[3; 4)$	$[4; 5)$	$[5; 6)$	$[6; 7)$	$[7; 8)$
n_i	19	9	12	14	7	22	17

Побудувати гістограму частот.

Розв'язання. Спостережувана випадкова величина X розглядається як абсолютно неперервна. Знайдемо довжину інтервалів:

$$h = x_{i+1} - x_i = 1.$$

Далі побудуємо на осі абсцис задані частинні інтервали довжиною $h = 1$. Проведемо над цими інтервалами відрізки, що паралельні осі абсцис і розміщені на відстанях, які дорівнюють відповідним частотам n_i . Краї цих відрізків з'єднаємо з віссю абсцис відрізками, паралельними осі ординат. Отриману гістограму частот зображено на рисунку.



3. Числові характеристики вибіркової сукупності.

Середнє арифметичне значення вибірки називається *вибірковим середнім* \bar{x}_e і обчислюється за формулами

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

де x_i — значення i -ї варіанти; n_i — частота i -ї варіанти; n — об'єм вибірки; k — кількість варіант у вибірці; p_i — відносна частота i -ї варіанти.

Середній квадрат відхилення значень елементів вибірки від вибіркового середнього називається *вибірковою дисперсією* D_B і обчислюється за формулами

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

Після перетворень формули для знаходження вибіркової дисперсії дещо спрощуються:

$$D_e = \frac{\sum_{i=1}^k x_i^2 n_i}{n} - \bar{x}_e^2 = \sum_{i=1}^k x_i^2 p_i - \bar{x}_e^2,$$

тобто вибіркова дисперсія дорівнює різниці середнього квадрата елементів вибірки й квадрата вибіркового середнього.

Квадратний корінь з вибіркової дисперсії $\sigma = \sqrt{D_e}$ називається *середнім квадратичним відхиленням вибірки*.

Приклад 6. Задано статистичний розподіл вибірки

x_i	1	3	4	7	10	12	15
n_i	5	2	12	7	4	3	2

Знайти вибіркове середнє, вибірккову дисперсію та середнє квадратичне відхилення вибірки.

Розв'язання. Обчислимо об'єм вибірки:

$$n = 5 + 2 + 12 + 7 + 4 + 3 + 2 = 35.$$

Знайдемо відповідно вибіркове середнє, вибіркиму дисперсію та середнє квадратичне відхилення вибірки:

$$\bar{x}_g = \frac{\sum_{i=1}^k x_i n_i}{n} = \frac{1 \cdot 5 + 3 \cdot 2 + 4 \cdot 12 + 7 \cdot 7 + 10 \cdot 4 + 12 \cdot 3 + 15 \cdot 2}{35} = \frac{214}{35} = 6,1;$$

$$D_g = \frac{\sum_{i=1}^k x_i^2 n_i}{n} - \bar{x}_g^2 = \frac{1^2 \cdot 5 + 3^2 \cdot 2 + 4^2 \cdot 12 + 7^2 \cdot 7 + 10^2 \cdot 4 + 12^2 \cdot 3 + 15^2 \cdot 2}{35} - \left(\frac{214}{35} \right)^2 \approx 15,2$$

$$\sigma = \sqrt{D_g} = \sqrt{15,2} = 3,9.$$

Медіаною точкового статистичного розподілу вибірки називається число Me , визначене за правилом:

1) якщо об'єм вибірки $n = 2m + 1$ непарний, то медіаною буде значення елемента варіаційного ряду з номером $m + 1$, тобто $Me = x_{m+1}$;

2) якщо об'єм вибірки $n = 2m$ парний, то медіаною буде середнє значення елементів варіаційного ряду з номерами m і $m + 1$, тобто $M_e = \frac{x_{m+1} + x_m}{2}$.

Модом (Mo) точкового статистичного розподілу вибірки називається варіанта з найбільшою частотою.

Варіаційним розмахом (або розмахом варіації) називається різниця між найбільшим і найменшим значеннями вибірки:

$$R = x_{\max} - x_{\min}.$$