# Вибірки та їх представлення

Теорія ймовірностей, ймовірнісні процеси і математична статистика

# Предмет і методи математичної статистики

**Математична статистика** – це розділ математики, в якому вивчаються методи збору, систематизації і аналізу результатів спостережень масових випадкових явищ з метою виявлення існуючих закономірностей.

# Предмет і методи математичної статистики

**Математична статистика** – це розділ математики, в якому вивчаються методи збору, систематизації і аналізу результатів спостережень масових випадкових явищ з метою виявлення існуючих закономірностей.

З'ясування закономірностей, яким підпорядковані масові випадкові явища, здійснюється за допомогою методів теорії ймовірностей.

# Предмет і методи математичної статистики

**Математична статистика** – це розділ математики, в якому вивчаються методи збору, систематизації і аналізу результатів спостережень масових випадкових явищ з метою виявлення існуючих закономірностей.

З'ясування закономірностей, яким підпорядковані масові випадкові явища, здійснюється за допомогою методів теорії ймовірностей.

Аналіз статистичних даних може здійснюватися з метою:

# Предмет і методи математичної статистики

**Математична статистика** – це розділ математики, в якому вивчаються методи збору, систематизації і аналізу результатів спостережень масових випадкових явищ з метою виявлення існуючих закономірностей.

З'ясування закономірностей, яким підпорядковані масові випадкові явища, здійснюється за допомогою методів теорії ймовірностей.

Аналіз статистичних даних може здійснюватися з метою:

**а)** оцінки невідомої ймовірності події; оцінки невідомої функції розподілу; оцінки невідомих параметрів розподілу, загальний вигляд якого відомий; оцінки залежності випадкової величини від однієї або кількох випадкових величин;

# Предмет і методи математичної статистики

**Математична статистика** – це розділ математики, в якому вивчаються методи збору, систематизації і аналізу результатів спостережень масових випадкових явищ з метою виявлення існуючих закономірностей.

З'ясування закономірностей, яким підпорядковані масові випадкові явища, здійснюється за допомогою методів теорії ймовірностей.

Аналіз статистичних даних може здійснюватися з метою:

**а)** оцінки невідомої ймовірності події; оцінки невідомої функції розподілу; оцінки невідомих параметрів розподілу, загальний вигляд якого відомий; оцінки залежності випадкової величини від однієї або кількох випадкових величин;

**б)** перевірки статистичних гіпотез про вигляд невідомого розподілу або про величину параметрів розподілу, вигляд якого відомий.

# Генеральна і вибіркові сукупності

Нехай потрібно дослідити сукупність однорідних об'єктів відносно деякої якісної чи кількісної ознаки.

# Генеральна і вибіркові сукупності

Нехай потрібно дослідити сукупність однорідних об'єктів відносно деякої якісної чи кількісної ознаки.

**Якісні ознаки** характеризують деякі властивості або стан об'єкта, наприклад, стать, професія, якість продукції тощо.

Нехай потрібно дослідити сукупність однорідних об'єктів відносно деякої якісної чи кількісної ознаки.

**Якісні ознаки** характеризують деякі властивості або стан об'єкта, наприклад, стать, професія, якість продукції тощо.

**Кількісні ознаки** одержують в результаті вимірювання або спостереження і виражаються числами, наприклад, маса, об'єм, прибуток тощо.

# Генеральна і вибіркові сукупності

Нехай потрібно дослідити сукупність однорідних об'єктів відносно деякої якісної чи кількісної ознаки.

**Якісні ознаки** характеризують деякі властивості або стан об'єкта, наприклад, стать, професія, якість продукції тощо.

**Кількісні ознаки** одержують в результаті вимірювання або спостереження і виражаються числами, наприклад, маса, об'єм, прибуток тощо.

Кількісні ознаки позначаються через $X, Y, Z, \ldots$; вони можуть бути дискретними або неперервними, тобто набувати будь-яких числових значень в деяких межах.

# Генеральна і вибіркові сукупності

Можливі два способи дослідження ознаки сукупності: суцільне спостереження і вибіркове спостереження.

# Генеральна і вибіркові сукупності

Можливі два способи дослідження ознаки сукупності: суцільне спостереження і вибіркове спостереження.

При суцільному спостереженні вивчається кожний об'єкт сукупності, а при вибірковому — деякої частини, відібраної із всієї сукупності.

# Генеральна і вибіркові сукупності

Можливі два способи дослідження ознаки сукупності: суцільне спостереження і вибіркове спостереження.

При суцільному спостереженні вивчається кожний об'єкт сукупності, а при вибірковому — деякої частини, відібраної із всієї сукупності.

Вся сукупність об'єктів, яка підлягає дослідженню, називається **генеральною**, а та її частина, яка попала на перевірку або дослідження, **вибірковою сукупністю** або **вибіркою**.

Число об'єктів у генеральній сукупності (у вибірці) називається її **обсягом**.

## Генеральна і вибіркові сукупності

Можливі два способи дослідження ознаки сукупності: суцільне спостереження і вибіркове спостереження.

При суцільному спостереженні вивчається кожний об'єкт сукупності, а при вибірковому — деякої частини, відібраної із всієї сукупності.

Вся сукупність об'єктів, яка підлягає дослідженню, називається **генеральною**, а та її частина, яка попала на перевірку або дослідження, **вибірковою сукупністю** або **вибіркою**.

Число об'єктів у генеральній сукупності (у вибірці) називається її **обсягом**.

За результатами вивчення ознаки у вибірці робиться висновок про властивості генеральної сукупності.

# Генеральна і вибіркові сукупності

Розрізняють два способи відбору елементів генеральної сукупності у вибірку: **випадковий** і **невипадковий**.

# Генеральна і вибіркові сукупності

Розрізняють два способи відбору елементів генеральної сукупності у вибірку: **випадковий** і **невипадковий**.

**Випадковий відбір** означає, що кожен об'єкт сукупності має рівний шанс потрапити до вибірки.

# Генеральна і вибіркові сукупності

Розрізняють два способи відбору елементів генеральної сукупності у вибірку: **випадковий** і **невипадковий**.

**Випадковий відбір** означає, що кожен об'єкт сукупності має рівний шанс потрапити до вибірки.

Якщо сукупність містить групи об'єктів, то випадковий відбір повинен забезпечити представництво у вибірці об'єктів кожної із цих груп. Такий відбір називають **типовим**.

# Статистичний розподіл вибірки

Нехай $X$ — випадкова величина, яка уособлює певну кількісну ознаку об'єктів генеральної сукупності.

# Статистичний розподіл вибірки

Нехай $X$ — випадкова величина, яка уособлює певну кількісну ознаку об'єктів генеральної сукупності.

Із цієї сукупності одержано вибірку ознаки $X$ : $X_1, X_2, \ldots, X_n$, де $X_i$ $(i = \overline{1, n})$ — незалежні однаково розподілені випадкові величини, розподіл кожної із яких співпадає із розподілом $X$.

# Статистичний розподіл вибірки

Нехай $X$ — випадкова величина, яка уособлює певну кількісну ознаку об'єктів генеральної сукупності.

Із цієї сукупності одержано вибірку ознаки $X$ : $X_1, X_2, \ldots, X_n$, де $X_i$ $(i = \overline{1, n})$ — незалежні однаково розподілені випадкові величини, розподіл кожної із яких співпадає із розподілом $X$.

Різні значення, які приймає кількісна ознака, наприклад, $X = x_i$ $(i = \overline{1, k})$ називаються **варіантами** або **реалізаціями**.

# Статистичний розподіл вибірки

Нехай $X$ — випадкова величина, яка уособлює певну кількісну ознаку об'єктів генеральної сукупності.

Із цієї сукупності одержано вибірку ознаки $X$ : $X_1, X_2, \ldots, X_n$, де $X_i$ $(i = \overline{1, n})$ — незалежні однаково розподілені випадкові величини, розподіл кожної із яких співпадає із розподілом $X$.

Різні значення, які приймає кількісна ознака, наприклад, $X = x_i$ $(i = \overline{1, k})$ називаються **варіантами** або **реалізаціями**.

Послідовність варіант, розміщених у зростаючому порядку

$$x_1 < x_2 < \ldots < x_k$$

називають **варіаційним рядом**.

# Статистичний розподіл вибірки

Якщо варіанта $x_i$ зустрічається $n_i$ раз у вибірці, то число $n_i$, називається **частотою** варіанти $x_i$ $(i = \overline{1, k})$. При цьому $\sum_{i=1}^{k} n_i = n$ — обсяг вибірки.

# Статистичний розподіл вибірки

Якщо варіанта $x_i$ зустрічається $n_i$ раз у вибірці, то число $n_i$, називається **частотою** варіанти $x_i$ $(i = \overline{1,k})$. При цьому $\sum\limits_{i=1}^{k} n_i = n$ — обсяг вибірки.

Відношення $\dfrac{n_i}{n} = w_i$ називається **відносною частотою** варіанти, причому $\sum\limits_{i=1}^{k} w_i = 1$.

# Статистичний розподіл вибірки

Якщо варіанта $x_i$ зустрічається $n_i$ раз у вибірці, то число $n_i$, називається **частотою** варіанти $x_i$ $(i = \overline{1,k})$. При цьому $\sum\limits_{i=1}^{k} n_i = n$ — обсяг вибірки.

Відношення $\dfrac{n_i}{n} = w_i$ називається **відносною частотою** варіанти, причому $\sum\limits_{i=1}^{k} w_i = 1$.

В залежності від того, яких значень можуть набувати варіанти, варіаційні ряди поділяються на **дискретні** та **неперервні (інтервальні)**.

# Статистичний розподіл вибірки

## Означення

Перелік варіант варіаційного ряду і відповідних їм частот або відносних частот називають **статистичним розподілом** вибірки:

| | | | | |
|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | $\dots$ | $x_k$ |
| $n_i$ | $n_1$ | $n_2$ | $\dots$ | $n_k$ |

| | | | | |
|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | $\dots$ | $x_k$ |
| $w_i$ | $w_1$ | $w_2$ | $\dots$ | $w_k$ |

# Статистичний розподіл вибірки

## Означення

Перелік варіант варіаційного ряду і відповідних їм частот або відносних частот називають **статистичним розподілом** вибірки:

| $x_i$ | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
|---|---|---|---|---|
| $n_i$ | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |

| $x_i$ | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
|---|---|---|---|---|
| $w_i$ | $w_1$ | $w_2$ | $\ldots$ | $w_k$ |

## Приклад

Проведено обстеження величини річного прибутку (млн. грн.) на десяти підприємствах харчової промисловості: 2, 5, 1, 2, 2, 5, 1, 2, 1, 2. Побудувати статистичний розподіл вибірки.

# Статистичний розподіл вибірки

Для неперервного варіаційного ряду статистичний розподіл задається таблицею

| Інтервал | $[\alpha_0; \alpha_1)$ | $[\alpha_1; \alpha_2)$ | $\ldots$ | $[\alpha_{k-1}; \alpha_k)$ |
|---|---|---|---|---|
| $n_i$ | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |
| $w_i$ | $w_1$ | $w_2$ | $\ldots$ | $w_k$ |

тут $n_i$, $w_i$ — частота і відносна частота появи ознаки $X$ в інтервалі $[\alpha_{i-1}; \alpha_i)$.

# Статистичний розподіл вибірки

Для неперервного варіаційного ряду статистичний розподіл задається таблицею

| Інтервал | $[\alpha_0; \alpha_1)$ | $[\alpha_1; \alpha_2)$ | $\ldots$ | $[\alpha_{k-1}; \alpha_k)$ |
|---|---|---|---|---|
| $n_i$ | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |
| $w_i$ | $w_1$ | $w_2$ | $\ldots$ | $w_k$ |

тут $n_i$, $w_i$ — частота і відносна частота появи ознаки $X$ в інтервалі $[\alpha_{i-1}; \alpha_i)$.

Довжину $h_i = \alpha_i - \alpha_{i-1}$ цього інтервалу називають **кроком**.

**Щільність частоти** $f_i = \dfrac{n_i}{h_i}$

(**щільність відносної частоти** $g_i = \dfrac{w_i}{h_i}$)

характеризує число елементів (відносне число елементів), яке припадає на одиницю довжини $h_i$ інтервалу.

# Статистичний розподіл вибірки

**Щільність частоти** $f_i = \dfrac{n_i}{h_i}$

(**щільність відносної частоти** $g_i = \dfrac{w_i}{h_i}$)

характеризує число елементів (відносне число елементів), яке припадає на одиницю довжини $h_i$ інтервалу.

На практиці число інтервалів визначається за формулами:
$k = [1 + 3,32 \lg n]$ (Стерджес) або $k = [5 \lg n]$ (Брукс-Карузерс), або $k = [\sqrt{n}]$, тут $[a]$ — ціла частина числа $a$.

# Статистичний розподіл вибірки

**Щільність частоти** $f_i = \dfrac{n_i}{h_i}$

(**щільність відносної частоти** $g_i = \dfrac{w_i}{h_i}$)

характеризує число елементів (відносне число елементів), яке припадає на одиницю довжини $h_i$ інтервалу.

На практиці число інтервалів визначається за формулами:
$k = [1 + 3,32 \lg n]$ (Стерджес) або $k = [5 \lg n]$ (Брукс-Карузерс), або $k = [\sqrt{n}]$, тут $[a]$ — ціла частина числа $a$.

Шляхом заміни $x_i = \dfrac{\alpha_{i-1} + \alpha_i}{2}$ $(i = \overline{1, k})$ можна перейти від інтервального статистичного розподілу до дискретного.

# Емпірична функція розподілу

Нехай досліджується кількісна ознака $X$ генеральної сукупності. Розглянемо варіаційний ряд $x_1, x_2, \dots, x_k$ вибірки обсягу $n$, для якого відомий статистичний розподіл.

# Емпірична функція розподілу

Нехай досліджується кількісна ознака $X$ генеральної сукупності. Розглянемо варіаційний ряд $x_1, x_2, \dots, x_k$ вибірки обсягу $n$, для якого відомий статистичний розподіл.

## Означення

**Емпіричною функцією розподілу** випадкової величини $X$ називають функцію $F_n(x)$, яка визначає для кожного значення $x$ відносну частоту події $\{X < x\}$:

$$F_n(x) = W\{X < x\} = \frac{\mu(x)}{n}, \tag{7.1}$$

де $\mu(x)$ — кількість елементів вибірки, менших за $x$.

# Емпірична функція розподілу

Нехай досліджується кількісна ознака $X$ генеральної сукупності. Розглянемо варіаційний ряд $x_1, x_2, \ldots, x_k$ вибірки обсягу $n$, для якого відомий статистичний розподіл.

**Означення**

**Емпіричною функцією розподілу** випадкової величини $X$ називають функцію $F_n(x)$, яка визначає для кожного значення $x$ відносну частоту події $\{X < x\}$:

$$F_n(x) = W\{X < x\} = \frac{\mu(x)}{n}, \tag{7.1}$$

де $\mu(x)$ — кількість елементів вибірки, менших за $x$.

**Приклад**

Побудувати емпіричну функцію розподілу за даним статистичним розподілом вибірки:

| $X$ | 2 | 5 | 9 |
|---|---|---|---|
| $n_i$ | 12 | 18 | 30 |

# Графічне зображення статистичних розподілів

В залежності від вигляду варіаційного ряду і поставленої задачі будують різні графіки статистичних розподілів, зокрема **полігон**, **гістограму**, **кумулятивну криву**.

# Графічне зображення статистичних розподілів

В залежності від вигляду варіаційного ряду і поставленої задачі будують різні графіки статистичних розподілів, зокрема **полігон**, **гістограму**, **кумулятивну криву**.

Дискретний статистичний розподіл вибірки

# Графічне зображення статистичних розподілів

В залежності від вигляду варіаційного ряду і поставленої задачі будують різні графіки статистичних розподілів, зокрема **полігон**, **гістограму**, **кумулятивну криву**.

Дискретний статистичний розподіл вибірки

**Полігоном частот (або відносних частот)** називають ламану лінію, відрізки якої з'єднують точки з координатами $(x_i, n_i)$ (або $(x_i, w_i)$), $i = \overline{1, k}$.

## Графічне зображення статистичних розподілів

Неперервний статистичний розподіл вибірки

# Графічне зображення статистичних розподілів

Неперервний статистичний розподіл вибірки

**Гістограмою частот (або відносних частот)** називають східчасту фігуру у вигляді послідовності прямокутників з основами $h_i = \alpha_i - \alpha_{i-1}$ і висотами $f_i$ (або $g_i$).

Неперервний статистичний розподіл вибірки

**Гістограмою частот (або відносних частот)** називають східчасту фігуру у вигляді послідовності прямокутників з основами $h_i = \alpha_i - \alpha_{i-1}$ і висотами $f_i$ (або $g_i$).

Площа $S_i$ $i$-го прямокутника $S_i = h_i \cdot f_i = n_i$. Звідси, площа гістограми частот дорівнює обсягу вибірки: $\sum\limits_{i=1}^{k} S_i = n$.

# Графічне зображення статистичних розподілів

Неперервний статистичний розподіл вибірки

**Гістограмою частот (або відносних частот)** називають східчасту фігуру у вигляді послідовності прямокутників з основами $h_i = \alpha_i - \alpha_{i-1}$ і висотами $f_i$ (або $g_i$).

Площа $S_i$ $i$-го прямокутника $S_i = h_i \cdot f_i = n_i$. Звідси, площа гістограми частот дорівнює обсягу вибірки: $\sum\limits_{i=1}^{k} S_i = n$.

Аналогічно, для гістограми відносних частот маємо $S_i = h_i \cdot g_i = w_i$ і $\sum\limits_{i=1}^{k} S_i = 1$.

**Кумулятою дискретного статистичного розподілу** називають графік функції $F_n(x)$.

# Графічне зображення статистичних розподілів

**Кумулятою дискретного статистичного розподілу** називають графік функції $F_n(x)$.

**Кумулятивна крива неперервного статистичного розподілу** — ламана лінія, яка сполучає точки з координатами $(\alpha_i; F_n(\alpha_i))$, $i = \overline{0, k}$.

**Кумулятою дискретного статистичного розподілу** називають графік функції $F_n(x)$.

**Кумулятивна крива неперервного статистичного розподілу** — ламана лінія, яка сполучає точки з координатами $(\alpha_i; F_n(\alpha_i))$, $i = \overline{0, k}$.

**Приклад**

На кондитерській фабриці автомат розливає рідкий шоколад у форми для одержання шоколадних плиток. Для контролю навмання відібрано 20 плиток із партії готової продукції. Результати обстеження наведені в таблиці:

| Маса, г | 96–98 | 98–100 | 100-102 | 102–104 |
|---|---|---|---|---|
| К-ть плиток | 4 | 8 | 6 | 2 |

Побудувати: 1) полігон і гістограму відносних частот; 2) кумулятивну криву.

# Середні вибіркові характеристики

Розглянемо дискретний статистичний розподіл вибірки обсягу $n$:

| | | | | |
|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
| $n_i$ | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |

Розглянемо дискретний статистичний розподіл вибірки обсягу $n$:

| $x_i$ | $x_1$ | $x_2$ | $\dots$ | $x_k$ |
|---|---|---|---|---|
| $n_i$ | $n_1$ | $n_2$ | $\dots$ | $n_k$ |

Величина, яка обчислюється за формулою

$$\overline{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \tag{7.2}$$

називається **вибірковим середнім** такого розподілу (середнє арифметичне значення).

# Середні вибіркові характеристики

Розглянемо дискретний статистичний розподіл вибірки обсягу $n$:

| | | | | |
|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
| $n_i$ | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |

Величина, яка обчислюється за формулою

$$\overline{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \tag{7.2}$$

називається **вибірковим середнім** такого розподілу (середнє арифметичне значення).

Очевидно, що $\overline{x}_{\text{в}} = \sum\limits_{i=1}^{k} x_i w_i$, де $w_i$ — відносні частоти.

# Середні вибіркові характеристики

Розглянемо дискретний статистичний розподіл вибірки обсягу $n$:

| | | | | |
|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
| $n_i$ | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |

Величина, яка обчислюється за формулою

$$\overline{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \tag{7.2}$$

називається **вибірковим середнім** такого розподілу (середнє арифметичне значення).

Очевидно, що $\overline{x}_{\text{в}} = \sum\limits_{i=1}^{k} x_i w_i$, де $w_i$ — відносні частоти.

За вибіркове середнє неперервного статистичного розподілу приймається середнє арифметичне відповідного дискретного розподілу.

## Основні властивості вибіркового середнього

## Основні властивості вибіркового середнього

1°. Якщо всі варіанти збільшити (зменшити) в одне і те ж число раз $\lambda$, то вибіркове середнє збільшиться (зменшиться) у стільки ж разів:

$$\frac{1}{n}\sum_{i=1}^{k}(\lambda x_i)n_i = \lambda\overline{x}_{\text{в}} \text{ або } \overline{\lambda x}_{\text{в}} = \lambda\overline{x}_{\text{в}}.$$

## Основні властивості вибіркового середнього

1°. Якщо всі варіанти збільшити (зменшити) в одне і те ж число раз $\lambda$, то вибіркове середнє збільшиться (зменшиться) у стільки ж разів:

$$\frac{1}{n}\sum_{i=1}^{k}(\lambda x_i)n_i = \lambda\overline{x}_{\text{в}} \text{ або } \overline{\lambda x}_{\text{в}} = \lambda\overline{x}_{\text{в}}.$$

2°. Якщо всі варіанти збільшити (зменшити) на одне і те ж число $C$, то вибіркове середнє збільшиться (зменшиться) на таке ж число:

$$\frac{1}{n}\sum_{i=1}^{k}(x_i - C)n_i = \overline{x}_{\text{в}} - C \text{ або } \overline{(x-C)}_{\text{в}} = \overline{x}_{\text{в}} - C.$$

### Основні властивості вибіркового середнього

$3^\circ$. Сума всіх відхилень значень ознаки від вибіркового середнього дорівнює нулю:

$$\frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})n_i = 0 \text{ або } \overline{(x - \overline{x}_{\text{в}})}_{\text{в}} = 0.$$

## Основні властивості вибіркового середнього

$3^\circ$. Сума всіх відхилень значень ознаки від вибіркового середнього дорівнює нулю:

$$\frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})n_i = 0 \text{ або } \overline{(x - \overline{x}_{\text{в}})}_{\text{в}} = 0.$$

$4^\circ$. Вибіркове середнє алгебраїчної суми кількох ознак дорівнює сумі вибіркових середніх цих ознак:

$$\overline{(x + y)}_{\text{в}} = \overline{x}_{\text{в}} + \overline{y}_{\text{в}}.$$

Обчислення вибіркового середнього безпосередньо за формулою (7.2) часто приводить до громіздких арифметичних розрахунків. Цього можна уникнути, перейшовши від варіант $x_i$, до **умовних варіант**

$$u_i = \frac{x_i - C}{\lambda}, \tag{7.3}$$

де $C, \lambda$ — відповідно підібрані сталі.

Обчислення вибіркового середнього безпосередньо за формулою (7.2) часто приводить до громіздких арифметичних розрахунків. Цього можна уникнути, перейшовши від варіант $x_i$, до **умовних варіант**

$$u_i = \frac{x_i - C}{\lambda}, \tag{7.3}$$

де $C, \lambda$ — відповідно підібрані сталі.

Згідно властивостей $1^\circ - 2^\circ$ маємо формулу:

$$\overline{x}_{\text{в}} = \lambda \overline{u}_{\text{в}} + C. \tag{7.4}$$

# Середні вибіркові характеристики

**Груповим середнім** називають середнє арифметичне значень ознаки, що належать даній групі.

# Середні вибіркові характеристики

**Груповим середнім** називають середнє арифметичне значень ознаки, що належать даній групі.

Нехай сукупність обсягу $n$ складається із $m$ груп обсягами $N_j$ кожна, $j = \overline{1, m}$: $n = \sum\limits_{j=1}^{m} N_j$.

# Середні вибіркові характеристики

**Груповим середнім** називають середнє арифметичне значень ознаки, що належать даній групі.

Нехай сукупність обсягу $n$ складається із $m$ груп обсягами $N_j$ кожна, $j = \overline{1, m}$: $n = \sum\limits_{j=1}^{m} N_j$.

Якщо $\overline{x}_j$ — групові середні, то загальне вибіркове середнє $\overline{x}_{\text{в}}$ обчислюється за формулою

$$\overline{x}_{\text{в}} = \frac{1}{n} \sum_{j=1}^{m} \overline{x}_j N_j. \tag{7.5}$$

**Приклад**

За рівнем заробітної плати робітники двох цехів підприємства розподілені так:

| Зар. плата, грн. | Кількість робітників | |
|---|---|---|
| | Цех №1 | Цех №2 |
| 250–260 | – | 2 |
| 260–270 | 3 | 4 |
| 270–280 | 6 | 16 |
| 280–290 | 7 | 8 |
| 290–300 | 4 | – |
| Всього | 20 | 30 |

Обчислити середню заробітну плату робітників цехів і підприємства.

# Середні вибіркові характеристики

**Означення**

**Модою** $\overline{M}_0$ статистичного розподілу називається варіанта, яка має найбільшу частоту появи.

# Середні вибіркові характеристики

**Означення**

**Модою** $\overline{M}_0$ статистичного розподілу називається варіанта, яка має найбільшу частоту появи.

Для дискретного варіаційного ряду мода визначається безпосередньо за означенням.

**Означення**

**Модою** $\overline{M}_0$ статистичного розподілу називається варіанта, яка має найбільшу частоту появи.

Для дискретного варіаційного ряду мода визначається безпосередньо за означенням.

Для інтервального статистичного розподілу знаходиться інтервал $[\alpha_{i-1}, \alpha_i)$, який має найбільшу частоту (**модальний інтервал**). Значення моди обчислюється шляхом лінійної інтерполяції:

$$\overline{M}_0 = \alpha_{i-1} + \frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}}(\alpha_i - \alpha_{i-1}), \qquad (7.6)$$

де $n_i$ — частота модального інтервала; $n_{i-1}$ — частота домодального інтервала; $n_{i+1}$ — частота післямодального інтервала.

# Середні вибіркові характеристики

## Означення

**Медіаною** $\overline{M}_e$ статистичного розподілу називають варіанту, яка припадає на середину розподілу варіаційного ряду.

# Середні вибіркові характеристики

**Означення**

**Медіаною** $\overline{M}_e$ статистичного розподілу називають варіанту, яка припадає на середину розподілу варіаційного ряду.

Для дискретного варіаційного ряду із непарною кількістю елементів медіана дорівнює серединній варіанті, а для ряду із парною кількістю елементів — півсумі двох серединних варіант.

**Означення**

**Медіаною** $\overline{M}_e$ статистичного розподілу називають варіанту, яка припадає на середину розподілу варіаційного ряду.

Для дискретного варіаційного ряду із непарною кількістю елементів медіана дорівнює серединній варіанті, а для ряду із парною кількістю елементів — півсумі двох серединних варіант.

Для інтервального статистичного розподілу визначається медіанний інтервал $[\alpha_{i-1}, \alpha_i)$ із умови $F_n(\alpha_{i-1}) \leq 0,5$ і $F_n(\alpha_i) > 0,5$. Значення медіани на цьому інтервалі обчислюється шляхом лінійної інтерполяції за формулою:

$$\overline{M}_e = \alpha_{i-1} + \frac{0,5 - F_n(\alpha_{i-1})}{F_n(\alpha_i) - F_n(\alpha_{i-1})}(\alpha_i - \alpha_{i-1}). \tag{7.7}$$

**Приклад**

Наведено статистичний розподіл випадкової величини $X$ — кількість балів, одержаних абітурієнтами на вступному іспиті з математики:

| $X$ | 0–10 | 10–30 | 30–50 | 50–70 | 70–80 | 80–100 |
|---|---|---|---|---|---|---|
| $n_i$ | 15 | 22 | 28 | 18 | 12 | 5 |

Знайти моду і медіану розподілу.

# Характеристики розсіювання

Найпростішою оцінкою розсіювання варіант статистичного розподілу вибірки є **варіаційний розмах**

$$R = x_{\max} - x_{\min}, \tag{7.8}$$

де $x_{\max}, x_{\min}$ — найбільше і найменше значення ознаки.

# Характеристики розсіювання

Найпростішою оцінкою розсіювання варіант статистичного розподілу вибірки є **варіаційний розмах**

$$R = x_{\max} - x_{\min}, \tag{7.8}$$

де $x_{\max}, x_{\min}$ — найбільше і найменше значення ознаки.

З практичної точки зору найбільш важливою є характеристика розсіювання значень варіаційного ряду навколо вибіркового середнього.

# Характеристики розсіювання

Найпростішою оцінкою розсіювання варіант статистичного розподілу вибірки є **варіаційний розмах**

$$R = x_{\max} - x_{\min}, \tag{7.8}$$

де $x_{\max}, x_{\min}$ — найбільше і найменше значення ознаки.

З практичної точки зору найбільш важливою є характеристика розсіювання значень варіаційного ряду навколо вибіркового середнього.

**Означення**

**Дисперсією** вибірки $D_{\text{в}}$ називається середнє арифметичне квадратів відхилень варіант відносно $\overline{x}_{\text{в}}$:

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^2 n_i \quad \text{або} \quad D_{\text{в}} = \sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^2 w_i. \tag{7.9}$$

## Основні властивості дисперсії вибірки

### Основні властивості дисперсії вибірки

1°. Дисперсія сталої дорівнює нулю.

## Основні властивості дисперсії вибірки

1°. Дисперсія сталої дорівнює нулю.

2°. Якщо всі варіанти помножити на стале число $\lambda$, то дисперсія вибірки помножиться на число $\lambda^2$:

$$\frac{1}{n}\sum_{i=1}^{k}(\lambda x_i - \lambda\overline{x}_{\text{в}})^2 n_i = \lambda^2 D_{\text{в}}.$$

### Основні властивості дисперсії вибірки

1°. Дисперсія сталої дорівнює нулю.

2°. Якщо всі варіанти помножити на стале число $\lambda$, то дисперсія вибірки помножиться на число $\lambda^2$:

$$\frac{1}{n}\sum_{i=1}^{k}(\lambda x_i - \lambda \overline{x}_{\text{в}})^2 n_i = \lambda^2 D_{\text{в}}.$$

3°. Якщо всі варіанти збільшити (зменшити) на одне і те ж число $C$, то дисперсія вибірки не зміниться:

$$\frac{1}{n}\sum_{i=1}^{k}\left[(x_i + C) - (\overline{x}_{\text{в}} + C)\right]^2 n_i = D_{\text{в}}.$$

## Основні властивості дисперсії вибірки

1°. Дисперсія сталої дорівнює нулю.

2°. Якщо всі варіанти помножити на стале число $\lambda$, то дисперсія вибірки помножиться на число $\lambda^2$:

$$\frac{1}{n}\sum_{i=1}^{k}(\lambda x_i - \lambda\overline{x}_{\text{в}})^2 n_i = \lambda^2 D_{\text{в}}.$$

3°. Якщо всі варіанти збільшити (зменшити) на одне і те ж число $C$, то дисперсія вибірки не зміниться:

$$\frac{1}{n}\sum_{i=1}^{k}\left[(x_i + C) - (\overline{x}_{\text{в}} + C)\right]^2 n_i = D_{\text{в}}.$$

4°. Дисперсія вибірки дорівнює різниці середнього арифметичного квадратів варіант і квадрату вибіркового середнього:

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - (\overline{x}_{\text{в}})^2 \quad \text{або} \quad D_{\text{в}} = \overline{(x^2)}_{\text{в}} - (\overline{x}_{\text{в}})^2. \qquad (7.10)$$

Для спрощення обчислення дисперсії вибірки можна скористатись умовними варіантами (7.3); маємо формулу:

$$D_{\text{в}x} = \frac{\lambda^2}{n} \sum_{i=1}^{k} u_i^2 n_i - (\overline{x}_{\text{в}} - C)^2 \quad \text{або} \quad D_{\text{в}x} = \lambda^2 D_{\text{в}u}. \tag{7.11}$$

# Характеристики розсіювання

Нехай всю вибірку обсягу $n$ розбито на $m$ груп.

# Характеристики розсіювання

Нехай всю вибірку обсягу $n$ розбито на $m$ груп. Позначимо через $N_j$ — обсяг $j$-тої групи; $\overline{x}_j$ — групове середнє; $D_j$ — групову дисперсію $(j = \overline{1, m})$; $\overline{x}_\text{в}$ — загальне середнє; $D_\text{в}$ — загальну вибіркову дисперсію.

# Характеристики розсіювання

Нехай всю вибірку обсягу $n$ розбито на $m$ груп. Позначимо через $N_j$ — обсяг $j$-тої групи; $\overline{x}_j$ — групове середнє; $D_j$ — групову дисперсію ($j = \overline{1, m}$); $\overline{x}_{\text{в}}$ — загальне середнє; $D_{\text{в}}$ — загальну вибіркову дисперсію.

Величина

$$D_{\text{м}} = \frac{1}{n} \sum_{j=1}^{m} (\overline{x}_j - \overline{x}_{\text{в}})^2 N_j \tag{7.12}$$

називається **міжгруповою дисперсією** і характеризує дисперсію групових середніх.

# Характеристики розсіювання

Нехай всю вибірку обсягу $n$ розбито на $m$ груп. Позначимо через $N_j$ — обсяг $j$-тої групи; $\overline{x}_j$ — групове середнє; $D_j$ — групову дисперсію ($j = \overline{1, m}$); $\overline{x}_{\text{в}}$ — загальне середнє; $D_{\text{в}}$ — загальну вибіркову дисперсію.

Величина

$$D_{\text{м}} = \frac{1}{n} \sum_{j=1}^{m} (\overline{x}_j - \overline{x}_{\text{в}})^2 N_j \tag{7.12}$$

називається **міжгруповою дисперсією** і характеризує дисперсію групових середніх.

Загальна дисперсія вибірки обчислюється за формулою

$$D_{\text{в}} = \frac{1}{n} \sum_{j=1}^{m} D_j N_j + D_{\text{м}}. \tag{7.13}$$

# Характеристики розсіювання

**Середнє квадратичне відхилення** (або **стандартне відхилення**) вибірки:

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}}.$$

**Коефіцієнт варіації**:

$$V = \frac{\sigma_{\text{в}}}{\overline{x}_{\text{в}}} \cdot 100\%.$$

**Середнє квадратичне відхилення** (або **стандартне відхилення**) вибірки:

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}}.$$

**Коефіцієнт варіації**:

$$V = \frac{\sigma_{\text{в}}}{\overline{x}_{\text{в}}} \cdot 100\%.$$

**Приклад**

Обчислити коефіцієнт варіації наступного статистичного розподілу:

| $x_i$ | 0–10 | 10–30 | 30–50 | 50–70 | 70–80 | 80–100 |
|---|---|---|---|---|---|---|
| $n_i$ | 15 | 22 | 28 | 18 | 12 | 5 |

**Центральним емпіричним моментом $l$-го порядку** називають величину $\nu_l^*$, яка обчислюється за формулою:

$$\nu_l^* = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^l n_i. \qquad (7.14)$$

# Емпіричні моменти

**Центральним емпіричним моментом $l$-го порядку** називають величину $\nu_l^*$, яка обчислюється за формулою:

$$\nu_l^* = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^l n_i. \qquad (7.14)$$

Якщо $\overline{x}_в = 0$, то момент називається **початковим** і позначається через $\mu_l^*$, тобто

$$\mu_l^* = \frac{1}{n}\sum_{i=1}^{k} x_i^l n_i. \qquad (7.15)$$

# Емпіричні моменти

**Центральним емпіричним моментом $l$-го порядку** називають величину $\nu_l^*$, яка обчислюється за формулою:

$$\nu_l^* = \frac{1}{n} \sum_{i=1}^{k} (x_i - \overline{x}_{\text{в}})^l n_i. \quad (7.14)$$

Якщо $\overline{x}_{\text{в}} = 0$, то момент називається **початковим** і позначається через $\mu_l^*$, тобто

$$\mu_l^* = \frac{1}{n} \sum_{i=1}^{k} x_i^l n_i. \quad (7.15)$$

$$\mu_1^* = \overline{x}_{\text{в}}, \qquad \nu_1^* = 0, \qquad \nu_2^* = D_{\text{в}}.$$

# Емпіричні моменти

Центральні емпіричні моменти третього і четвертого порядку використовуються при дослідженні форми статистичного розподілу.

Центральні емпіричні моменти третього і четвертого порядку використовуються при дослідженні форми статистичного розподілу.

**Коефіцієнтом асиметрії** статистичного розподілу називається число $A$, яке обчислюється за формулою

$$A = \frac{\nu_3^*}{\sigma_{\text{в}}^3}. \tag{7.16}$$

# Емпіричні моменти

Центральні емпіричні моменти третього і четвертого порядку використовуються при дослідженні форми статистичного розподілу.

**Коефіцієнтом асиметрії** статистичного розподілу називається число $A$, яке обчислюється за формулою

$$A = \frac{\nu_3^*}{\sigma_{\text{в}}^3}. \tag{7.16}$$

Якщо $A = 0$, то розподіл має симетричну форму, тобто варіанти, рівновіддалені від $\overline{x}_{\text{в}}$, мають однакову частоту.

# Емпіричні моменти

Центральні емпіричні моменти третього і четвертого порядку використовуються при дослідженні форми статистичного розподілу.

**Коефіцієнтом асиметрії** статистичного розподілу називається число $A$, яке обчислюється за формулою

$$A = \frac{\nu_3^*}{\sigma_в^3}. \tag{7.16}$$

Якщо $A = 0$, то розподіл має симетричну форму, тобто варіанти, рівновіддалені від $\overline{x}_в$, мають однакову частоту.

При $A > 0$ ($A < 0$) має місце правостороння (лівостороння) асиметрія.

# Емпіричні моменти

**Ексцесом** статистичного розподілу називається число, яке обчислюється за формулою

$$E = \frac{\nu_4^*}{\sigma_{\text{в}}^4} - 3. \qquad (7.17)$$

**Ексцесом** статистичного розподілу називається число, яке обчислюється за формулою

$$E = \frac{\nu_4^*}{\sigma_{\text{в}}^4} - 3. \tag{7.17}$$

Ексцес є показником крутизни (загостреності) графіка статистичного розподілу у порівнянні з нормальним розподілом, для якого $E = 0$.

**Ексцесом** статистичного розподілу називається число, яке обчислюється за формулою

$$E = \frac{\nu_4^*}{\sigma_{\text{в}}^4} - 3. \tag{7.17}$$

Ексцес є показником крутизни (загостреності) графіка статистичного розподілу у порівнянні з нормальним розподілом, для якого $E = 0$.

Якщо $E > 0$ ($E < 0$), то статистичний розподіл має більш загострену (пологу) вершину порівняно із нормальною кривою.