МІРИ ЦЕНТРАЛЬНОЇ ТЕНДЕНЦІЇ

ЛЕКЦІЯ 4

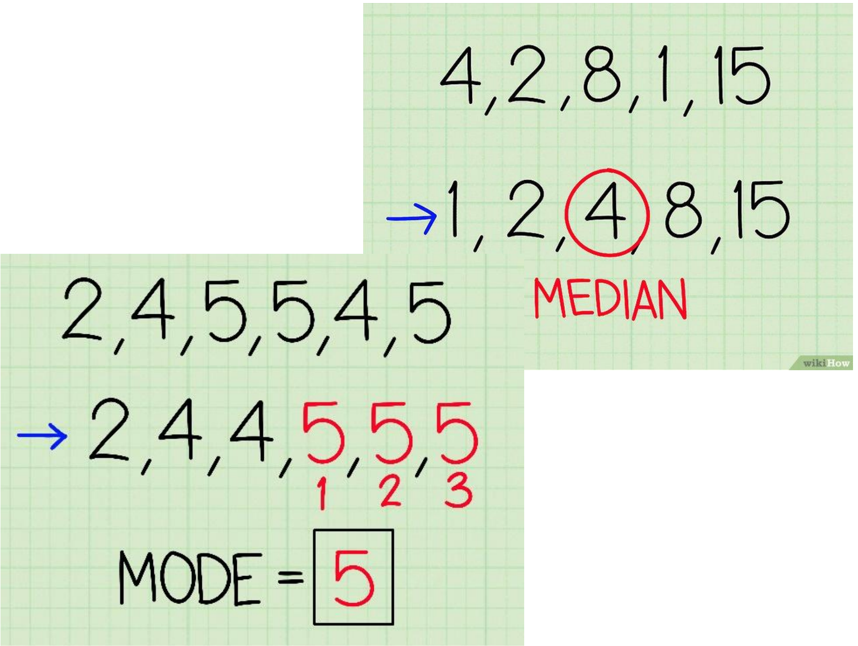
План

1. Мода та її обчислення.
2. Медіана та її обчислення.
3. Середнє арифметичне: обчислення та властивості.
4. Інтерпретація мір центральної тенденції. Вибір міри центральної тенденції.

Первинні методи кількісної обробки даних

До основних методів первинної статистичної обробки відносяться обчислення:

- **мір центральної тенденції**
- **мір розкиду** (мінливості) даних та квантилі розподілу.



1) яке значення найбільш характерне для вибірки?

Міри центрально ї тенденції

2) чи великий розкид даних щодо цього характерного значення, тобто яка варіативність даних?

Міри мінливості (розкиду)

Міри центральної тенденції

- мода
- медіана
- середнє арифметичне

Мода *– це значення у множині спостережень, яке зустрічається* ***найчастіше***

Аналізується сукупність статистичних даних

$X_1, X_2, \ldots, X_n$.

Модою цих даних називають значення, яке зустрічається в сукупності <u>найчастіше</u>. Позначається мода: **<u>Mo</u>**.

Мода – не завжди єдине значення. В окремих випадках мода може складатися з кількох чисел, які зустрічаються однакову кількість разів (але найчастіше).

При визначенні моди необхідно дотримуватись таких вимог:

1. Якщо в даних всі значення зустрічаються однаково часто, кажуть, що в них немає моди: (1, 2, 3, 4)

Наприклад:

2, 2, 3, 3, 4, 4, 5, 5 – *моди немає*

При визначенні моди необхідно дотримуватись таких вимог:

2. якщо варіанти <u>суміжні</u> і мають <u>однакову</u> частоту, моду визначають як *середнє значення сусідніх варіантів*:

Наприклад:

2, 2, 3, 4, 4, 4, 5, 5, 5 $\qquad \mathrm{Mo} = \frac{\mathbf{(4+5)}}{\mathbf{2}} = \mathbf{4,5}$

Mo (1, 2, 2, 3, 3, 4) = 2,5

Mo (1, 2, 2, 2, 5, 5, 5, 6) =(2+5)/2=3,5

При визначенні моди необхідно дотримуватись таких вимог:

3. Якщо два несусідні значення мають однакову частоту, то кажуть, що в даних є дві моди, а ряд даних називається ***бімодальним***:

Наприклад:

Мо (1, 1, 1, 2, 3, 4, 5, 5, 5) = 1 та 5

Mo (10 11 11 11 12 13 14 14 14 17) = 11 и 14

Приклад

Студенти академічної групи отримали настпуні оцінки на екзамені.

Номер студента у списку	1	2	3	4	5	6	7	8	9	10	11
Оцінка на екзамені	**«5»**	**«4»**	**«5»**	**«3»**	**«3»**	**«2»**	**«2»**	**«3»**	**«5»**	**«5»**	**«4»**

Оцінку “2” на екзамені у групі отримали 2 студенти, оцінку “3” – 3, оцінку “4” – 2, оцінку “5” – 4.

Отже, найбільше студентів отримали оцінку “5” і саме вона є модою цієї сукупності: **Мо = 5.**

Приклад

Припустимо, маємо тест з 40 питань для визначення типу темпераменту у людини за встановленою методикою.

Нехай за цією методикою опитано деяку людину.

Тип темпераменту	сангвінік	холерик	флегматик	меланхолік
Кількість відповідей	6	8	23	3

Модою цих даних є тип “флегматик”, оскільки він найчастіше зустрічається у відповідях. Отже, цей тип переважає в характері опитаного.

Приклад

Розглянемо результати соціологічного дослідження, здійсненого з метою встановлення середньої кількості дітей у сім’ї. Загалом було опитано 84 сім’ї. Отримані наступні результати опитування.

Кількість дітей у сім’ї	**0**	**1**	**2**	**3**	**4 і більше**
Кількість сімей	16	23	23	14	8

У цьому разі моду досліджуваної групи сімей утворюють значення 1 і 2: **Mo = {1; 2}.**

Медіана

значення, яке перебуває на середині упорядкованої послідовності емпіричних даних.

це значення, яке ділить упорядковану множину даних навпіл, так що одна половина даних виявляється меншою за медіану, а друга – більшою.

При визначенні медіани необхідно дотримуватись таких вимог:

1) якщо кількість спостережень у вибірці <u>непарне</u>, то медіана дорівнює значенню, розташованому точно посередині впорядкованої вибірки.

Наприклад: 11, 13, <u>18</u>, 19, 20 **Me=18**

Якщо обсяг вибірки невеликий, то медіану легко знайти за варіаційним рядом.

Якщо ж вибірка має великий обсяг, то можна розрахувати номер елемента, який є медіаною вибірки з непарною кількістю спостережень, за формулою

$$Me = X_i \qquad i = \frac{n+1}{2}$$

де n – об’єм вибірки

Наприклад:

n=127

$$i = \frac{127 + 1}{2} 64$$

$Me=X_{64}$, тобто медіаною є значення, розташоване на 64-му місці у впорядкованій вибірці

При визначенні медіани необхідно дотримуватись таких вимог:

2) якщо кількість спостережень у вибірці парне, то медіана дорівнює середньому значенню центральних сусідніх елементів.

$$Me = \frac{x_{n/2} + x_{(n/2)+1}}{2}$$

Наприклад:

n=12 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 7

$$Me = \frac{4+5}{2} = 4,5$$

Для вибірки великого обсягу номера двох елементів, розташованих в середині впорядкованої вибірки, що містить парну кількість значень, обчислюються за допомогою формул:

$$\boldsymbol{i = \frac{n}{2}} \quad \text{та} \quad \boldsymbol{i + 1 = \frac{n}{2} + 1}$$

Наприклад:

n=116

i=116:2=58

i+1=116:2+1=59

Me=(x_{58}+x_{59}):2, тобто медіана дорівнює середньому значенню величин, розташованих на 58-му і 59-му місцях впорядкованої вибірці.

Наприклад

У результаті тестування відомі IQ-індекси шістьох співробітників компанії.

Номер співробітника у списку	**1**	**2**	**3**	**4**	**5**	**6**
IQ-індекс	124	131	128	142	132	140

Визначимо медіану цих значень. Для цього, упорядкувавши список, отримаємо таку послідовність IQ-індексів:

124, 128, 131, 132, 140, 142.

Оскільки кількість значень у групі **парна** (6 індексів), для визначення медіани потрібно розглянути два числа, які містяться посередні списку – 131 та 132. Отже, обчислюємо медіану:

$$\boldsymbol{Me = \frac{131 + 132}{2} = 131,5}$$

Середнє арифметичне

Середнім значенням вибірки (позначають $\overline{X}$) називають середнє арифметичне всіх чисел ряду даних вибірки

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

n – об’єм вибірки

Наприклад

1, 3, 3, 5, 5, 9, 9

n=7

$$\overline{\boldsymbol{x}} = \frac{\mathbf{1+3+3+5+5+9+9}}{\mathbf{7}} = \frac{\mathbf{35}}{\mathbf{7}} = \mathbf{5}$$

Якщо значення властивості повторюються, то $\overline{x}$ розраховується за формулою:

$$\overline{\boldsymbol{x}} = \frac{\boldsymbol{x_1 n_1} + \boldsymbol{x_2 n_2} + \cdots + \boldsymbol{x_k n_k}}{\boldsymbol{n}}$$

де $n_1+n_2+\ldots+n_k=n$

для попереднього прикладу

$$\overline{x} = \frac{1+3*2+5*2+9*2}{7} = 5$$

Приклад

Обчислити моду, медіану і середнє значення вибірки, поданої у вигляді статистичного розподілу:

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

1) модою є таке значення x_i, частота котрого n_i є максимальною:

Mo=7, оскільки це значення зустрічається найчастіше (4 рази);

Розв’язання

2) для визначення медіани спочатку визначимо, скільки значень містить вибірка: $n=15$ – <u>непарне</u>. Тоді медіаною буде значення, що розташоване посередині ряду, номер якого дорівнює

$$\mathbf{i=(n+1)/2=(15+1)/2=8}$$

Починаємо послідовно складати частоти n_i, поки не дістанемось потрібного елемента:

$$\mathbf{n_1+n_2+n_3=3+1+3=7}$$

Значить наступним 8 елементом буде значення вибірки, що дорівнює 5.

$$\mathbf{Me=X_8=5}$$

Розв'язання

3) для визначення середнього значення потрібно кожне значення x_i помножити на його частоту n_i, добуток скласти і поділити на об'єм вибірки n:

$$\overline{x} = \frac{2*3+3*1+4*3+5*2+7*4+10*2}{15} = \frac{79}{15} = 5,26$$

Особливості мір центральної тенденції

- мода вибірки обчислюється просто, її можна визначити «на око». Для дуже великих груп даних мода є досить стабільною мірою центру розподілу;
- медіана займає проміжне положення між модою і середнім з погляду її підрахунку. Ця міра особливо легко визначається у разі ранжованих даних;
- середнє арифметичне передбачає використовування всіх значень вибірки, причому всі вони впливають на значення цієї міри. Зазвичай вибіркове середнє застосовується при прагненні до найбільшої точності у визначенні центральної тенденції.

Особливості мір центральної тенденції

- Медіана обчислюється в тому випадку, коли у серії є «нетипові» дані, що різко впливають на середнє.
- Мода використовується в ситуаціях, коли не потрібна висока точність, але важлива швидкість визначення міри центральної тенденції.
- Обчислення всіх трьох показників проводиться також для оцінки розподілу даних. При нормальному розподілі даних середнє арифметичне значення, медіана і мода однакові або дуже близькі.

Поради, щодо визначення мір центральної тенденції

1. Моду та медіану обчислити найпростіше.

2. В малих групах мода нестабільна:

$$Mo(1, 2, 3, 3, 3, 4, 4, 5)=3; \quad Mo(1, 2, 2, 3, 3, 4, 4, 4)=4$$

3. На медіану не впливають величини крайніх значень ряду даних.

Поради, щодо визначення мір центральної тенденції

4. На величину середнього арифметичного впливають значення кожного елементу ряду. Порівняння мір центральної тенденції в рядах, що відрізняються одним значенням

	Мода	Медіана	Середнє
1: 1,3,3,5,6,7,8	3	5	4,7
2: 1,3,3,5,6,7,16	3	5	5,9

Поради, щодо визначення мір центральної тенденції

5. Деякі множини даних можуть не мати реальної міри центральної тенденції:

На лавці сидить 5 чоловіків. Два жебраки з майном 25 копійок. Третій – робітник, його заощадження нараховували 5000 грн. Четвертий чоловік мав 100 000 грн. П'ятий – мультимільйонер з прибутком 5 000 000 грн.

Таким чином, ряд: 0,25, 0,25, 5000, 100 000, 5 000 000

Mo=0,25 Me=5000 $\overline{X}$=1021000

Поради, щодо визначення мір центральної тенденції

6. Вибір міри центральної тенденції може бути обумовлений типом змінних, які аналізуються:

Номінативна шкала (шкала найменувань)

Порядкова шкала

Інтервальна шкала або шкала відношень

МОДА

МЕДІАНА

СЕРЕДНЄ