ЛЕКЦІЯ 4
МІРИ ЦЕНТРАЛЬНОЇ ТЕНДЕНЦІЇ

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.

1. Мода та її обчислення

Всі методи кількісної обробки прийнято розділяти на *первинні* та *вторинні*.

Первинна статистична обробка націлена на впорядкування інформації про об'єкт і предмет вивчення. На цій стадії «сирі» відомості групуються за тими чи іншими критеріями, що заносяться у зведені таблиці. Первинно оброблені дані дають можливість дослідникові зрозуміти характер всієї сукупності даних в цілому: про їх однорідність-неоднорідність, компактність-розкиданість, чіткість-розмитість і т.д. Ця інформація чітко спостерігається в візуальних формах представлення даних і дає відомості про їх розподіл.

До основних методів первинної статистичної обробки відносяться:

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

Первинний статистичний аналіз всієї сукупності отриманих у дослідженні даних надає можливість охарактеризувати її в гранично стислому вигляді і відповісти на два головних запитання:

1) яке значення найбільш характерне для вибірки?;

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

Для вирішення першого запитання обчислюються міри центральної тенденції, для вирішення другого – міри мінливості (або розкиду).

Мірами центральної тенденції називають чисельні показники типових властивостей емпіричних даних. Ці показники дають відповіді на питання про те, наприклад, «який середній рівень інтелекту студентів педагогічного університету?», «яке типове значення показника відповідальності певної групи осіб?».

Існує порівняно невелика кількість таких показників-мір і в першу чергу:

- мода,
- медіана,
- середнє арифметичне.

Кожна конкретна міра центральної тенденції має свої особливості, що роблять її цінною для характеристики об'єкта дослідження в певних умовах

Завдання мір центральної тенденції – виявити найтиповіший показник для даної вибірки.

Мода *– це значення у множині спостережень, яке зустрічається найчастіше (Mo).*

Нехай аналізується сукупність статистичних даних

$$X_1, X_2, \dots, X_n.$$

Модою цих даних називають значення, яке зустрічається в сукупності найчастіше. Позначатимемо моду так: **Mo**.

Зауважимо, що 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*

2. якщо варіанти суміжні і мають однакову частоту, моду визначають як середнє значення сусідніх варіантів:

Наприклад:

2, 2, 3, 4, 4, 4, 5, 5, 5 мода $\text{Mo} = \frac{(4+5)}{2} = 4{,}5$

Мо (1, 2, 2, 3, 3, 4) = 2,5

1 2 2 2 5 5 5 6, Мо=(2+5)/2=3,5

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

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

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

Крім того, в ряду даних можуть бути найбільша і менші моди, при цьому найбільша мода – єдине значення, яке задовольняє визначення моди.

Приклад

Студенти академічної групи отримали такі 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»

Легко розрахувати, що 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

Відома класифікація психологічних типів людини визначає чотири ідеальних характери: сангвінік, холерик, меланхолік та флегматик. Припустимо, маємо тест з 40 питань для аналізу характеру людини за наведеною класифікацією.

Нехай на кожне питання є чотири варіанти відповіді, причому кожний варіант відповідає одному з типів. Нехай за цією методикою опитано деяку людину. Припустимо, 6 з її відповідей зараховано до типу “сангвінік”, 8 – “холерик”, 3 – “меланхолік”, 23 – “флегматик”. Тоді модою цих даних є тип “флегматик”, оскільки він найчастіше зустрічається у відповідях. Отже, цей тип переважає в характері опитаного.

Зауважимо також, що розглянуто дані, які вимірюються за **номінальною** шкалою. Отже, моду на відміну від середнього значення можна застосовувати навіть для аналізу нечислових значень.

Приклад

Розглянемо результати соціологічного дослідження, здійсненого з метою встановлення середньої кількості дітей у сім’ї. Загалом було опитано 84 сім’ї. Наведемо 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: Мо = {1; 2}.

2. Медіана та її обчислення

Медіана (Ме) *– значення, яке перебуває на середині упорядкованої послідовності емпіричних даних.*

Медіана (Ме) *– 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 непарне, то медіана дорівнює значенню, розташованому точно посередині впорядкованої вибірки.

Наприклад: 11, 13, 18, 19, 20, Ме=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

Ме = X_i

$$i = \frac{n+1}{2}$$

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.

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

$$\boldsymbol{Me} = \frac{\boldsymbol{x}_{n/2} + \boldsymbol{x}_{(n/2)+1}}{2}$$

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

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

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

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

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 в впорядкованої вибірці.

Наприклад:
Нехай 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:

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

3. Середнє арифметичне: обчислення та властивості

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

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

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

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

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

<u>Розв'язання:</u>

1) модою є таке значення x_i, частота котрого n_i є максимальною:
   Mo=7, оскільки це значення зустрічається найчастіше (4 рази);
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

i=(n+1)/2=(15+1)/2=8

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

$n_1+n_2+n_3=3+1+3=7$

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

$Me=X_8=5$;

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:

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

Таким чином, міри центральної тенденції дозволяють видити про концентрацію даних на числовій осі, тобто показують, де в основному розміщені значення ознаки.

4. Інтерпретація мір центральної тенденції. Вибір міри центральної тенденції.

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.

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.

Кожна міра має певні характеристики, які роблять її цінною в певних умовах. Вибір тієї чи іншої міри іноді вимагає певних роздумів. Наведемо декілька корисних порад та характеристик 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; Mo(1, 2, 2, 3, 3, 4, 4, 4)=4

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

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

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

Наводиться такий приклад *(Дж. Гласс, Дж. Стенлі, 1976)*:

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

Таким чином, ряд: (0,25, 0,25, 5000, 100 000, 5 000 000).

$Mo=0,25$ $Md=5000$	Отже, жодна з мір центральної тенденції для такого ряду не є адекватною
$\overline{X} = 1021000$	Другий висновок – для групи з 5 значеннями навряд чи потрібна взагалі якась статистика

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

Номінативна шкала (найменування)	Порядкова шкала	Інтервальна шкала або шкала відношень
МОДА		
	МЕДІАНА	
		СЕРЕДНЄ

Номінативна шкала (шкала найменувань)	**Порядкова шкала**	**Інтервальна шкала або шкала відношень**
МОДА		
	МЕДІАНА	
		СЕРЕДНЄ

7. У унімодальних вибірках, які симетричні, середнє, медіана і мода співпадають