

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ  
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ВНУТРІШНІХ СПРАВ**

*Факультет № 6  
Кафедра соціології та психології*

**ТЕКСТ ЛЕКЦІЇ**

з навчальної дисципліни «**Математичні методи в психології**»  
обов'язкових компонент  
освітньої програми першого (бакалаврського) рівня вищої освіти

*053 Психологія (практична психологія)*

**Тема №3. Первинна описова статистика**

**Харків 2023**

**ЗАТВЕРДЖЕНО**

Науково-методичною радою  
Харківського національного  
університету внутрішніх справ  
Протокол від 30.08.2023 № 7

**СХВАЛЕНО**

Вченою радою факультету № 6  
Протокол від 25.08.2023 № 7

**ПОГОДЖЕНО**

Секцією Науково-методичної  
ради ХНУВС гуманітарних та  
соціально- економічних дисциплін  
Протокол від 29.08.2023 № 7

Розглянуто на засіданні кафедри соціології та психології (протокол №8 від 15.08.2023)

**Розробник:**

Доцент кафедри соціології та психології, кандидат психологічних наук, доцент  
Твердохвалова Ю.Л.

**Рецензенти:**

1. Професор кафедри психології Харківського національного педагогічного університету імені Г.С. Сковороди, доктор психологічних наук, професор, Кузнєцов М.А.
2. Доцент кафедри соціології та психології факультету № 6 Харківського національного університету внутрішніх справ, кандидат психологічних наук, доцент Греса Н.В.

## План лекції

- 3.1. Міри центральної тенденції
- 3.2. Міри мінливості даних
- 3.3. Міри положення

### Рекомендована література (основна, допоміжна), інформаційні ресурси в Інтернеті

#### Основна:

1. Климчук В.О. Математичні методи у психології. Навчальний посібник для студентів психологічних спеціальностей. Київ : Освіта України. 2009. 288 с.
2. Телейко А.Б. Чорней Р.К. Математико-статистичні методи в соціології та психології : Навч. посібник. Київ : МАУП, 2007. 424 с.
3. Руденко В.М., Руденко Н.М. Математичні методи в психології : підручник. Київ : Академвидав, 2009. 384 с.

#### Допоміжна:

1. Літнарів Р.М. Основи математичної статистики у психології : Навчальний посібник. Ч.3. Рівне : МЕНУ, 2006. 49 с.
2. Татяничков А.О. Методичні рекомендації до виконання лабораторних робіт з курсу «Методи психологічного дослідження: математичні методи в психології». Одеса : Вид-во Університету Ушинського, 2019. 38 с.
3. Климчик В.О. Кластерний аналіз: використання в психологічних дослідженнях// Практична психологія та соціальна робота. 2006. №4. С. 30-36.
4. Циба В.Т. Математичні основи соціологічних досліджень: кваліметричний підхід. - К.: МАУП, 2002. - 248 с.
5. Климчук В.О. Викладання курсу “Математичні методи у психології” в умовах кредитно-модульної системи // Соціальна психологія. 2008. №2 (28). С. 180-189.

## Текст лекції

### 3.1. Міри центральної тенденції

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.

**Мода** (Mo) - це значення, яке найбільш часто зустрічається у вибірці, тобто число з найбільшою частотою. Якщо всі значення в групі зустрічаються з однаковою частотою, то вважається, що моди немає. Якщо два сусідніх значення мають однакову частоту і більше частоти будь-якого іншого значення, мода є середнє цих двох значень. Якщо те ж саме відноситься до двох несуміжних значень, то існує дві моди, а група оцінок є бімодальною.

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

Правила обрахування:

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

2. Якщо два сусідні значення мають однакову частоту, то модою називають їх середнє:  $Mo(1, 2, 2, 3, 3, 4) = 2,5$

3. Якщо два несусідні значення мають однакову частоту, то кажуть, що в даних є дві моди, а ряд даних називається бімодальним:  $Mo(1, 1, 1, 2, 3, 4, 5, 5, 5) = 1$  та  $5$

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

*Середнє арифметичне значень* (вибіркове середнє або середнє)

Важливою мірою центральної тенденції є середнє арифметичне.

**СЕРЕДНЄ АРИФМЕТИЧНЕ** –

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Середнє арифметичне обчислюється досить просто навіть вручну. Однак, цей процес можна ще спростити, якщо окремі значення в ряду даних повторюються :

Спрощене обчислення середнього арифметичного

Вихідні дані	Дані ( $X_i$ )	Частоти ( $f_i$ )	$X_i f_i$	Обчислення
2 6 10	2	1	2	$\bar{X} = \frac{\sum X_i f_i}{n} =$ $= \frac{166}{21} = 7,9$
3 6 10	3	2	6	
3 6 11	5	4	20	
5 8 11	6	3	18	
5 8 11	8	2	16	
5 9 15	9	2	18	
5 9 18	10	2	20	
	11	3	33	
	15	1	15	
	18	1	18	
	$n = \sum f_i = 21$		$\sum X_i f_i = 166$	

Можлива ситуація, коли, знаючи середні арифметичні декількох різних груп, вам треба знайти загальне середнє арифметичне для всіх груп разом. Наприклад, є 3 класи з різною кількістю учнів і середніми показниками успішності:

Середні показники успішності груп А, Б та В

Група	$\bar{X}$	$n$
А	11,9	24
Б	8,2	30
В	10,8	28

Якби ми мали вихідні дані, за якими були обраховані середні, ми б додали 82 значення ( $24+30+28=82$ ) і поділили їх на 82. Однак, цього ми зробити не можемо, а тому звернемося до формули середнього арифметичного:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$n$  нам відомо, треба лише знайти для кожної групи суму її показників. Виходячи з формули середнього арифметичного:

$$\sum_i X_i = n \cdot \bar{X}.$$

Тому **середнє арифметичне для об'єднаних груп матиме вигляд:**

$$\bar{X} = \frac{n_a X_a + n_b X_b + n_c X_c}{n_a + n_b + n_c}$$

**Медіана** (Me) - це значення, яке приходить на середину упорядкованої послідовності емпіричних даних. Медіана не обов'язково повинна збігатися з конкретним значенням. Збіг відбувається у випадку непарного числа значень (відповідей), розбіжність - при парному їх числі. В останньому випадку медіана обчислюється як середнє арифметичне двох центральних значень у впорядкованому ряді даних.

**МЕДІАНА** – це значення, яке ділить упорядковану множину даних навпіл, так що одна половина даних виявляється меншою за медіану, а друга – більшою (Md)

Правила обчислення:

1. Якщо ряд містить непарну кількість значень, то медіана є центральним значенням  $Md(11,13,25,48,49)=25$

2. Якщо ряд містить парну кількість значень, то медіана обчислюється як середнє двох центральних значень:  $Md(11,13,25,48)=19$

3. Якщо ряд достатньо великий, то щоб знайти місце медіани варто використати таку формулу:

$$N_{Md} = \frac{n+1}{2} \quad \begin{array}{l} \text{Де } N_{Md} - \text{місце медіани в ряду даних,} \\ n - \text{кількість елементів в ряду даних.} \end{array}$$

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 ранжированих даних;
- середнє арифметичне передбачає використання 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, різко впливаючі 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.

#### **ВИБІР МІРИ ЦЕНТРАЛЬНОЇ ТЕНДЕНЦІЇ**

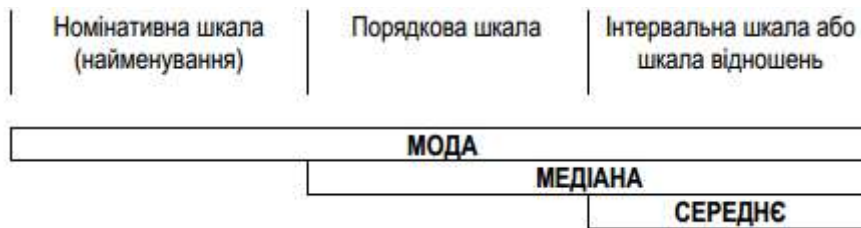
Кожна міра має певні характеристики, які роблять її цінною в певних умовах. Вибір тієї чи іншої міри іноді вимагає певних роздумів. Наведемо декілька корисних порад та характеристик мір центральної тенденції.

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

*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

6. Вибір міри центральної тенденції може бути обумовлений типом змінних, які аналізуються (Мартин Д., 2002):



### 3.2. Міри мінливості даних

Міри центральної тенденції говорять про концентрацію групи значень навколо певного показника. Такі міри дають показник, який до певної міри представляє всю вибірку. Однак, у цьому випадку ігноруються відмінності, що існують між окремими значеннями. Для визначення цих відмінностей івикористовують **міри мінливості**.

**Міри мінливості даних** - це статистичні показники, що дозволяють судити про ступінь однорідності отриманої множини даних, її компактності. Найбільш поширені в психологічних дослідженнях показники: розмах, дисперсія, стандартне відхилення.

**Розмах (R)** - це інтервал між максимальним і мінімальним значеннями ознаки. Він визначається легко і швидко, але чутливий до випадкових даних особливо при малій кількості досліджуваних. Це найпростіша міра відхилення.

Розмах визначається лише крайніми значеннями ознаки

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

На розмах, однак, не впливають дані, що лежать між максимальним та мінімальним показниками. Так:

$$R(10 \ 20 \ 30 \ 40 \ 50)=40 \text{ та } R(10 \ 10 \ 10 \ 10 \ 50)=40$$

Отже, розмах є досить грубою мірою мінливості. Однак, він є досить поширеною мірою, оскільки дозволяє обчислити **коефіцієнт осциляції**.

**КОЕФІЦІЄНТ ОСЦИЛЯЦІЇ** відображає відносні коливання крайніх значень ряду відносно середнього показника.

$$K_o = \frac{R}{\bar{X}} \cdot 100$$

Знайдемо коефіцієнти осциляції для двох вище наведених рядів:

$$K_{o1} = (40/30) \cdot 100 = 133\%;$$

$$K_{o2} = (40/18) \cdot 100 = 222\%$$

Таким чином, видно, що в другому ряду крайні значення коливаються відносно середнього більше, ніж в першому – **виникає питання про можливість порівнювати ці ряди між собою**.

Як ми відмітили, на розмах не впливають всі дані, що лежать між крайніми значеннями. Було б досить корисним знайти міру, яка б дозволяла при обрахуванні варіативності ряду враховувати всі його значення.

**Дисперсія (D)** обчислюються тільки для інтервальних та абсолютних шкал. Дисперсія є середнім арифметичним значенням квадратів відхилень окремих

значень ознаки від їхнього середнього арифметичного значення. Дисперсія має розмірність рівну квадратові розмірності ознаки, і обчислюється за формулою:

$$D = \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Обчислення дисперсії

Обчислення дисперсії			
Дані	Різниця ( $x_i - \bar{x}$ )	Квадрат різниці ( $x_i - \bar{x}$ ) <sup>2</sup>	Дисперсія ( $\sigma^2$ )
1	1-2=-1	1	$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{12}{6-1} = 2,4$
3	3-2=1	1	
3	3-2=1	1	
0	0-2=-2	4	
4	4-2=2	4	
1	1-2=-1	1	
$\bar{x}=2$		$\sum_{i=1}^n (x_i - \bar{x})^2 = 12$	

Дисперсія, подібно до середнього арифметичного, має декілька властивостей (доведення – самостійно).

1. Додавання константи (C) до кожного значення ряду не змінює дисперсію.
2. Множення кожного значення ряду на константу (C) збільшує дисперсію в C<sup>2</sup> разів.

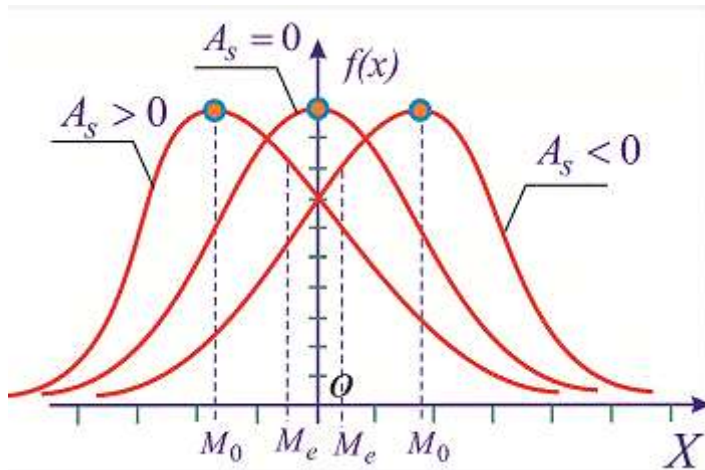
Стандартне відхилення ( $\delta$ ) застосовуються для інтервальних та абсолютних даних. Щоб її визначити необхідно з дисперсії розрахувати квадратний корінь. Його позитивне значення і приймається за міру мінливості, іменовану середньоквадратичним або стандартним відхиленням:

$$\delta = \sqrt{D}$$

Асиметрія (A) характеризує ступінь несиметричності розподілу відносно його середнього. Позитивна асиметрія вказує на відхилення вершини розподілу в бік від'ємних значень, негативна - у бік додатних.

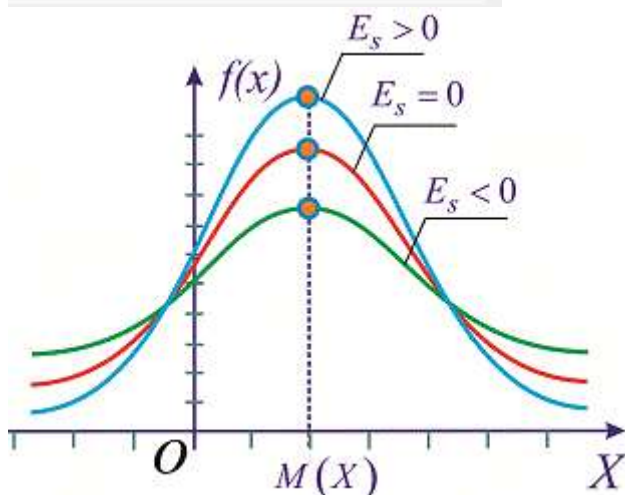
$$A = \frac{\sum (x_i - \bar{x})^3}{n \cdot \sigma^3}$$





Ексцес ( $E$ ) характеризує відносну опуклість або згладженість розподілу вибірки порівняно з нормальним розподілом. Позитивний ексцес позначає відносно загострений розподіл, негативний - відносно згладжений.

$$E = \frac{\sum (x_i - \bar{x})^4}{n \cdot \sigma^4} - 3$$



«Стандартом» розподілів служить нормальний розподіл з нульовою асиметрією і ексцесом.

### 3.3. Міри положення

У психології також широко використовуються *міри положення*, які називаються квантилями розподілу. *Квантиль* - це точка на числовій осі вимірної ознаки, яка ділить всю сукупність упорядкованих даних на дві групи з відомим співвідношенням їх чисельності. З одним з квантилів ми вже знайомі - це медіана. Це значення ознаки, яке ділить всю сукупність вимірювань на дві групи з рівною кількістю. Крім медіани часто використовуються процентилі та квартили.

*Процентилі* - це 99 точок - значень ознаки ( $P_1 \dots P_{99}$ ), які ділять упорядковану (за зростанням) множину даних на 100 частин, які рівні за чисельністю. Визначення конкретного значення процентиля аналогічно визначенню медіани. Наприклад, при визначенні 10-го процентиля,  $P_{10}$ ,

спочатку всі значення ознаки упорядковуються за зростанням. Потім відраховується 10% досліджуваних, що мають найменшу вираженість ознаки.  $P_{10}$  буде відповідати тому значенню ознаки, яке відокремлює ці 10% досліджуваних від решти 90%.

*Квартилі* - це 3 точки - значення ознаки ( $P_{25}$ ,  $P_{50}$ ,  $P_{75}$ ), які ділять упорядковану (за зростанням) множину даних на 4 рівні за чисельністю частини. Перший квартиль відповідає 25-му процентилю, другий - 50-му процентилю або медіані, третій квартиль відповідає 75-му процентилю.

Процентилі і квартилі використовуються для визначення частоти виникнення тих чи інших значень (або інтервалів) виміряної ознаки або для виділення підгруп і окремих досліджуваних, які найбільш типові або нетипові для даної множини спостережень.

### **Завдання на самостійну підготовку**

1. Поняття міри центральної тенденції. Поняття мода, середнє арифметичне значення, медіана.
2. Квантілі розподілу. Процентілі та квартилі.
3. Міри мінливості. Поняття розмах, дисперсія, стандартне відхилення. Властивості дисперсії. Стандартизація.