

«ЗАТВЕРДЖУЮ»  
Завідувач кафедрою  
інформатики та кінезіології

\_\_\_\_\_ І.П. Заневский

(підпис, ініціали, прізвище)

\_\_\_\_\_ 200\_\_р

**ЛЕКЦІЯ №3**  
**з навчальної дисципліни**  
**«КОМП'ЮТЕРНА ТЕХНІКА І МАТЕМАТИЧНА СТАТИСТИКА»**

(найменування навчальної дисципліни)

**Тема: Числові статистичні характеристики**

**Навчальний потік**

для студентів першого року навчання факультету спорту,  
фізичного виховання, здоров'я людини і туризму

**Навчальна мета:** Ознайомити студентів з характеристиками центральної тенденції та варіації, зокрема з середнім арифметичним, медіаною, модою і характеристиками розсіяння: розмахом варіації, середнім лінійним відхиленням, дисперсією, середнім квадратичним відхиленням, коефіцієнтом варіації.

**Виховна мета:** Розвивати пам'ять, логічне мислення, увагу, спостережливість, зацікавити математичною статистикою як наукою, розширювати кругозір з математичної статистики в галузі ФКіС, а також формувати систематизовані знання з математичної статистики.

**Навчальні питання і розподілення часу:**

Вступ \_\_\_\_\_ -10...хв.

1. Характеристики розподілу. \_\_\_\_\_ 10хв.

2. Характеристики центральної тенденції: середнє арифметичне, медіана, мода. \_\_\_ 30хв.

3. Характеристики розсіяння: розмах варіації, середнє лінійне відхилення, дисперсія, середнє квадратичне відхилення, коефіцієнт варіації. \_\_\_\_\_ 30хв.

Заключення та відповіді на запитання \_\_\_\_\_ - 10...хв.

## Навчальна література

1. Ільків О.С. Матвіїв В.І. Інформатика та комп'ютерна техніка (з елементами математичної статистики): Навч. посіб. – Львів: ЛДУФК, 2010.
  2. Дибкова Л.М. Інформатика та комп'ютерна техніка: Навч. Посіб. – Київ: Академвидав, 2005.
  3. Литвин І.І. Інформатика: теоретичні основи і практикум. – Львів: Новий світ, 2004.
  4. Глинський Я.М. Практикум з інформатики. – Львів: Деол, 2005.
  5. Русіло П.О., Заневський І.П. Математична статистика. Обробка і аналіз результатів спортивних вимірів. - Львів, 1995.
  6. Статистика. Підручник за ред. С.С. Герасименка. Київ: КНЕУ, 2000.
  7. Основы математической статистики. Под ред. В.С.Иванова, -М.: ФиС, 1990.
  8. Глинський Я.М. Інформатика: інформаційні технології. - Львів: Деол, 2003.
- 

### Додаткова

1. Петров П.К. Информационные технологии в физической культуре и спорте. – М.: ФКиС, 2008.
2. Височанський В.С., Кардаш А.І., Костів О.В., Черняхівський В.В. “Елементи інформатики”, - Львів: “Світ”, 1990.
3. Андреев А.Г. и др. Microsoft Windows XP: Home Edition и Professional. Русские версии / Под общ. ред. А.Н. Чекмарева. – СПб.: БХВ - Петербург, 2003.
4. Руденко В.Д., Макарчук О.М., Патланжоглу М.О. Практичний курс інформатики / За ред. Мадзігога В.М. - К.: Фенікс, 1997.
5. Фигурнов В.Е. IBM PC для пользователя. - М.: Финансы и статистика, 1992.
6. Зайцева Т.И., Смирнова О.Ю. В сб.: Информационные технологии в образовании. – М., 2000.
7. Kosmol A., Kosmol I. Komputery - nowoczesne technologie w sporcie. - Warszawa: AWF, 1999
8. Волков В.Ю. Компьютерные технологии в образовательном процессе по физической культуре в вузе: Монография. – СПб.: СПбГТУ, 1997.
9. Макарова М.В. Електронна комерція. Посібник. Київ. Видавничий центр "Академія". 2002.

### **Тема 3. Числові статистичні характеристики**

1. Характеристики розподілу.
2. Характеристики центральної тенденції: середнє арифметичне, медіана, мода.
3. Характеристики розсіяння: розмах варіації, середнє лінійне відхилення, дисперсія, середнє квадратичне відхилення, коефіцієнт варіації.

Статистичні характеристики вибіркової сукупності (ВС) називають **статистиками**, які розглядають як оцінки відповідних характеристик генеральної сукупності (ГС), які називають **параметрами**. Оскільки ВС не точно відтворює структуру ГС, то статистики також не збігаються з параметрами. Розбіжності між ними називають **похибками репрезентативності**. За причинами виникнення ці похибки поділяються на **систематичні і випадкові**. **Систематичні похибки** виникають внаслідок порушення принципу випадковості відбору або із-за неатестованих шкал вимірів. **Випадкові похибки** – наслідок випадковості відбору елементів сукупності для обстеження. Нехай  $\bar{X}$  - генеральне середнє,  $\bar{x}$  - вибіркоче середнє, тоді дія виключно випадкових похибок буде визначати розподіл результатів навколо  $\bar{x} = \bar{X}$ , рис. 2, а дія систематичних похибок означає зміщення центра розподілу  $\bar{x}$ , відносно  $\bar{X}$  на величину  $t$  (рис.3).

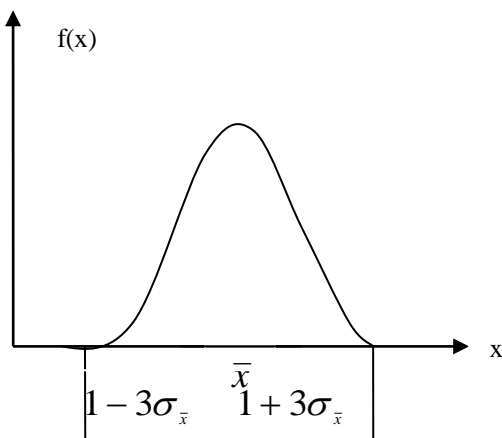


Рис. 2.  $\pm 3\sigma$  - область дії випадкових похибок

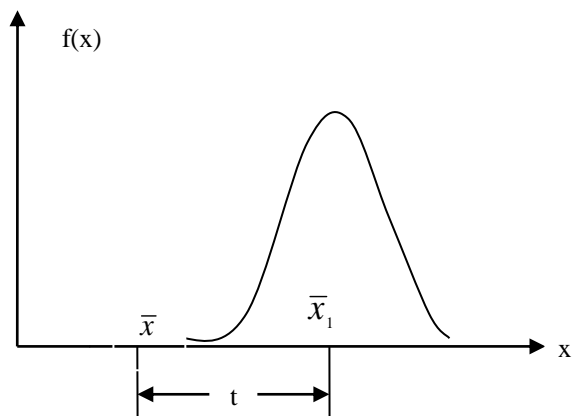


Рис. 3.  $t$  - систематична похибка

При організації вибіркового спостереження важливо запобігти появі систематичних похибок, як найбільш небезпечних. Що стосується випадкових помилок, то уникнути їх неможливо, проте вибіркочий метод дає строгий науковий метод їх оцінки. Так, відомо, що з імовірністю 0,997 випадкові похибки змінюються в межах  $\pm 3\sigma$ , тобто в даному разі довірче число  $t=3$ . В межах  $\pm 2\sigma$  випадкова похибка буде знаходитись з імовірністю 0,954, а в межах  $\pm \sigma$  - з ймовірністю 0,683.

Існує два основних способи добору: **повторний і безповторний**. При **повторному** відборі елемент ГС навмання відбирають, фіксують його номер, потім повертають цей елемент назад у ГС. При повторному відборі той самий елемент може потрапити у ВС ще раз, оскільки склад ГС не змінюється.

**Безповторним** є добір, при якому відібрані елементи вибірки назад у ГС не повертаються.

Щоб забезпечити принцип випадковості відбору, як повторний так і безповторний відбір здійснюють з використанням таблиці (генератора) рівномірно розподілених випадкових чисел (додаток 10). Допустимо, що необхідно провести 10% відбір із групи 120 спортсменів з метою оцінки середнього рівня їх фізичної підготовки. Тоді, пронумерувавши всіх спортсменів, ми виписуємо 12 послідовних тризначних цифр менших 120 із таблиці додатку 10, тобто, це будуть номери: 100, 135, 63, 8, 166, 90, 15, 33, 190, 153, 131, 165. Аналогічно здійснюють безповторний відбір лише з тією різницею, що із номерів, які можуть двічі повторюватись, один закреслюють і добирають із таблиці наступний номер, але так, щоб число номерів у даній вибірці не перевищувало 12.

В практиці вибіркового методу використовують два типи вибірових оцінок – точкові та інтервальні. **Точкова оцінка** – це оцінка параметра ГС за даними вибірки, наприклад, середнє по вибірці  $\bar{x}$  є вибірковою оцінкою параметра  $\bar{X}$  ГС. Для обчислення інтервальної **оцінки** потрібно знати граничну помилку вибірки  $\Delta_{\bar{x}} = t \cdot \sigma_{\bar{x}}$ , яка є максимально можливою похибкою для прийнятої імовірності  $P$ . В даному разі довірче число вказує  $t$  як співвідносяться гранична помилка і стандартна помилка вибірки  $\sigma_{\bar{x}}$ .

Стандартна похибка вибірки  $\sigma_{\bar{x}}$  обчислюється за формулою:

- при повторному відборі  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n - 1}}$ ;

- при безповторному  $\sigma_{\bar{x}} = \sigma \sqrt{\frac{N - n}{n(N - 1)}}$ ,

де  $\sigma^2$  - вибіркова дисперсія;  $n, N$  – відповідно обсяги ВС і ГС.

При використанні наведених формул слід враховувати, що дисперсія частки:

$$\sigma_p^2 = pq,$$

де  $p, q$  - частки певних елементів, а граничні похибки для частки в повторній і безповторній вибірках такі:

$$\Delta_p = t \sqrt{\frac{pq}{n-1}}; \quad \Delta_p = t \sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}}.$$

Отже, величини граничних похибок  $\Delta_{\bar{x}}$  чи  $\Delta_p$  залежить від варіації  $\sigma$  чи  $\sigma_p$ , обсягу вибірки  $n$ , та від  $\frac{n}{N}$ , а також прийнятого рівня імовірності  $P$ , якому відповідає квантиль  $t$ . Квантиль – це число, яке є коренем рівняння  $F(t) = P$ , де  $F(t)$ - функція розподілу.

Інтервальна оцінка – це інтервал, що з певною імовірністю накриває шуканий параметр ГС і, яка розраховується на основі вибіркової оцінки цього ж параметра. Цей інтервал називають довірчим і його межі визначаються за формулами:

- для середньої:  $P(\bar{x} - t\sigma_{\bar{x}} \leq \bar{X} \leq \bar{x} + t\sigma_{\bar{x}}) = 1 - \alpha$ ,
- для частки:  $P(p - t\sigma_p \leq P \leq p + t\sigma_p) = 1 - \alpha$ ,

де  $\alpha$ - рівень ризику, значення  $\bar{x} - t\sigma_{\bar{x}}$ ,  $p - t\sigma_p$  називають нижніми, а  $\bar{x} + t\sigma_{\bar{x}}$ ,  $p + t\sigma_p$  - верхніми межами довірчих інтервалів.

Результатом статистичного аналізу є статистичні характеристики та числові показники вибірки.

Варіаційні ряди і графіки емпіричних розподілів не дають повної характеристики вибірки. Числові показники дають кількісне уявлення про емпіричні дані та дозволяють порівняти їх між собою.

Числові показники вибірки поділяються на дві групи:

- *характеристики центральної тенденції (положення);*
- *характеристики розсіяння (варіації).*

Центральну тенденцію вибірки описують найбільш імовірні результати вимірів: *середнє арифметичне, медіана і мода.*

Середнє арифметичне – це одна з основних характеристик вибірки, що визначається діленням суми всіх результатів вимірів на обсяг вибірки:

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

де  $i = 1, 2, \dots, n$ .

Медіаною називається значення ознаки  $x$ , коли одна половина значень експериментальних даних менша її, а друга половина – більша.

Якщо даних небагато (обсяг вибірки малий), то медіана знаходиться за рангом медіани:

$$R_{Me} = \frac{n+1}{2}$$

При непарному числі членів вибірки ранг медіани вказує на номер результату виміру в ранжованому ряді, котрий і буде медіаною. Якщо вибірка має парне число членів, то медіана буде дорівнювати середньому арифметичному двох членів ранжованого ряду з рангами  $R_{Me} = n/2$  і  $R_{Me} = n/2 + 1$ .

Інтервал групування, в якому накопичена частота вперше буде більшим половини обсягу вибірки, називається **медіанним**.

Для згрупованих даних медіана визначається за такою формулою:

$$Me = x_{n_{Me}} + h \frac{0,5n - f_{Me-1}}{f_{Me}}$$

де  $x_{n_{Me}}$  – значення нижньої границі медіанного інтервалу;

$h$  – ширина (крок) інтервалів групування;

$f_{Me}$  – частота медіанного інтервалу;

$f_{Me-1}$  – накопичена частота інтервалу, який знаходиться перед медіанним;

$n$  – обсяг вибірки.

**Мода** – це значення ознаки, яка зустрічається у вибірці найбільш часто, тобто значення виміру з найбільшою імовірністю. Інтервал групування варіаційного ряду з найбільшою частотою називається модальним. Чисельне значення моди розраховується за такою формулою

$$M_o = x_{hM_o} + h \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})}$$

де  $x_{hM_o}$  – значення нижньої границі модального інтервалу,

$f_{M_o}$  – частота модального інтервалу,

$f_{M_o-1}$  і  $f_{M_o+1}$  – частоти інтервалів, які знаходяться перед і після модального.

Поруч із показниками положення (центральної тенденції) вираховують і характеристики варіації (розсіяння) вибірки.

Розмах **вибірки** (варіації) вираховується як різниця між максимальним і мінімальним значеннями вибірки, тобто різниця між крайніми значеннями ознаки ранжованого ряду:

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

Інформативність розмаху вибірки невелика, він враховує тільки крайні відхилення всіх результатів.

**Дисперсія** – це середнє значення квадратів відхилень кожного елемента ряду вимірів від середнього арифметичного.

Дисперсію, яку обчислюють за вибірковими даними, називають **вибірковою дисперсією**:

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

Вибіркова дисперсія характеризує систематичні похибки і її називають зміщеною оцінкою генеральної дисперсії. Незміщена оцінка є точковою оцінкою генеральної дисперсії, яку обчислюють за такою формулою:

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

Для великих обсягів вибірки різниця між зміщеною та незміщеною оцінкою генеральної дисперсії – невелика, в майбутньому будемо користуватися одним позначенням  $D$ .

Зміщену оцінку генеральної дисперсії використовують для характеристики вибірки, а незміщену оцінку – для визначення параметрів генеральної сукупності (інтервалу довіри середнього арифметичного, закону нормального розподілу).

**Середнє квадратичне відхилення** (стандартне відхилення) – це арифметичний квадратний корінь із дисперсії:

$$\sigma = \sqrt{D}$$

Розмірність стандартного відхилення співпадає з одиницями вимірювання змінної ознаки та характеризує ступінь відхилення результатів від середнього арифметичного значення в абсолютних одиницях.

Стандартне відхилення є абсолютною мірою розсіяння, яку записують у такому вигляді:

$$\bar{X} \pm \sigma$$

Наприклад:  $\bar{X} = 743,43$  см, а  $\sigma = 42,33$  см.

Запис оцінки точності  $X$  виглядатиме так:  $(743 \pm 42)$  см.

Для порівняння двох і більше сукупностей, які мають різні одиниці вимірів, використовують коефіцієнт варіації.

Коефіцієнт варіації знаходиться як відношення середнього квадратичного відхилення до середнього арифметичного і виражається у відсотках:

$$V = \frac{\sigma}{\bar{x}} 100\%$$

Коефіцієнт варіації є відносною мірою розсіяння ознаки та показує варіативність вибірки:

від 0 % до 10 % – варіація мала, вибірка однорідна;

від 10 % до 20 % – варіація середня, вибірка середньої однорідності;

більше 20 % – варіація велика, вибірка неоднорідна.



### **Питання для самоконтролю:**

1. Дайте визначення поняттю «генеральна сукупність».
2. Дайте визначення поняттю «вибіркова сукупність(вибірка)».
3. Дайте визначення поняттю «групування».
4. Назвіть характеристики центральної тенденції та варіації.
5. Що є результатом статистичного аналізу?
6. Означення статистичного розподілу, його основні елементи.
7. Характеристики центральної тенденції: середнє арифметичне, мода і медіана. Формули для їх обчислення.
8. З якою метою обчислюють характеристики варіації?
9. Назвати абсолютні характеристики розсіяння і навести формули для їх обчислення.
10. Яка з абсолютних характеристик варіації є найважливішою?
11. Як обчислюють характеристики форми розподілу: асиметрію і ексцес?

### **Навчально-матеріальне забезпечення**

Мультимедійний проектор

### ***Самостійна робота:***

1. Способи обчислення числових статистичних характеристик.
2. Характеристики емпіричних розподілів.
3. Скошеність та ексцес.

**Лекцію розробили:** к.пед. н., доц. О.С. Ільків.

(посада, вчений ступінь, вчене звання, підпис, ініціали, прізвище)

**Обговорено на засіданні кафедри:** інформатики та кінезіології

(назва кафедри)

Протокол від \_\_\_\_\_ 20\_\_\_\_ р. № \_\_\_\_\_