

«ЗАТВЕРДЖУЮ»  
Завідувач кафедри  
Інформатики, кінезіології та кіберспорту  
\_\_\_\_\_ І.П. Заневський  
(підпис, ініціали, прізвище)  
\_\_\_\_\_ 20\_\_ р

## ЛЕКЦІЯ №6 з навчальної дисципліни

### «КОМП'ЮТЕРНА ТЕХНІКА І МАТЕМАТИЧНА СТАТИСТИКА»

### Тема: Кореляційний та регресійний аналізи

#### Навчальний потік

для студентів першого року навчання факультетів: фізичної культури і спорту, педагогічної освіти, терапії та реабілітації

**Навчальна мета:** Ознайомити студентів з функціональним і статистичним взаємозв'язками, графічним методом аналізу статистичного взаємозв'язку, парним лінійним коефіцієнтом кореляції Браве-Пірсона, ранговим коефіцієнтом кореляції Спірмена, лінійною регресією, визначенням коефіцієнтів рівняння лінійної регресії.

**Виховна мета:** Розвивати пам'ять, логічне мислення, увагу, спостережливість, поглиблювати знання з математичної статистики 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. Функціональний і статистичний взаємозв'язки \_\_ 5хв.
2. Графічний метод аналізу статистичного взаємозв'язку \_\_ 10хв.
3. Форма, тіснота й спрямованість статистичного взаємозв'язку -10хв.
4. Парний лінійний коефіцієнт кореляції Браве-Пірсона. \_\_ 10хв.
5. Ранговий коефіцієнт кореляції Спірмена. \_\_\_\_ 5хв.
6. Лінійна регресія. \_\_\_\_\_ 5хв.
7. Визначення коефіцієнтів рівняння лінійної регресії. \_\_ 10хв.
8. Дослідження регресійних зв'язків між спортивними результатами. \_\_\_\_ 10хв.
9. Оцінка достовірності існування статистичного взаємозв'язку. \_\_\_\_ 5хв.

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

#### Навчально-матеріальне забезпечення

Мультимедійний проектор \_\_\_\_\_

## Навчальна література

### Основна:

1. Ільків О.С. Матвій В.І. Інформатика та комп'ютерна техніка (з елементами математичної статистики): Навч. посіб. – Львів: ЛДУФК, 2010.
2. Бакушевич Я.М., Капаціла Ю.Б. Інформатика та комп'ютерна техніка. -К.: Магнолія, 2024.
3. Буйницька О. Інформаційні технології та технічні засоби навчання. Навч. посіб. -К: цент навч. лі-ри, 2019.
4. Глинський Я.М. Інформатика: підручник.- Львів: Львівська політехніка, 2023.
5. Качан О.В. Упровадження інноваційних технологій у фізкультурно-оздоровчу та спортивну діяльність закладів освіти: навчально-методичний посібник Слов'янськ: Витоки, 2022.
6. Руденко В.М. Математична статистика. Навч. посіб. -К: цент навч. лі-ри, 2019.
7. Пасічник В.В., Пасічник О.В., Басюк Т.М., Думанський Н.О. Основи інформаційних технологій. Навч. посіб. 2020.
8. Windows 2010: навчальний посібник / Укладач: Дячук С. Ф. – Тернопіль: Вид-во ТНТУ імені Івана Пулюя, 2021.
9. О. Л. Тоцька. Сучасні інформаційні технології в професійній діяльності: лабор. практикум – Луцьк: Вежа-Друк, 2020.

### Допоміжна:

1. Заневський І. П., Заневська Л. Г. Комп'ютерні та інформаційні технології в активній рекреації й спортивно-оздоровчому туризмі: навч. посіб. для магістрів фіз. виховання. – Л.: ЛДУФК, 2010.
2. Є В. Павлиш, Л. Гліненко, Н. Шаховська Основи інформаційних технологій і систем. - Львів: Львівська політехніка, 2018.
3. Сусіденко В. Інформаційні системи і технології в обліку. Навч. посіб. –К.: центр навч. лі-ри, 2019.
4. Сорока П.М., Харченко В.В., Харченко Г.А. Інформаційні системи і технології в управлінні організацією: Навч. посіб. – К.: ЦП «Компринт», 2019.
5. Г. Кармелюк Теорія ймовірностей та математична статистика. Навч. посіб. –К.: центр навч. лі-ри, 2019.
6. Антомонов М.Ю. Математична обробка та аналіз медико-біологічних даних. 2-е видання- Київ: МІЦ «Медінформ», 2018.
7. Microsoft Access 2016: навчальний посібник в електронному вигляді / Укладачі В.О. Нелюбов, Ю.Ю. Білак. – Ужгород: ДВНЗ «УжНУ», 2019.
8. Нелюбов В. О., Куруца О. С. Основи інформатики. Microsoft Excel 2016: навчальний посібник. – Ужгород: ДВНЗ «УжНУ», 2018.
9. Основи інформаційних технологій: навч. посібник для здобувачів професійної освіти / А. М. Гуржій, Л. І. Возненко, Н. І. Поворознюк, В. В. Самсонов. -Київ: Літера ЛТД, 2023.

### 1. Інформаційні ресурси інтернет

1. <http://www.nbuv.gov.ua> – Національна бібліотека України ім. В.І. Вернадського.
2. Закон України «Про доступ до публічної інформації» (2022). Вилучено з <https://ips.ligazakon.net/document/T112939>
3. Основні положення статистичних досліджень у спорті <https://vseosvita.ua/.../osnovni-polozenna-statisticnih-doslidz...>
4. МАТЕМАТИЧНА СТАТИСТИКА [liber.onu.edu.ua/pdf/matem\\_stat.pdf](http://liber.onu.edu.ua/pdf/matem_stat.pdf)

## План

1. Функціональний і статистичний взаємозв'язки.
2. Графічний метод аналізу статистичного взаємозв'язку.
3. Форма, тіснота й спрямованість статистичного взаємозв'язку
4. Парний лінійний коефіцієнт кореляції Браве-Пірсона.
5. Ранговий коефіцієнт кореляції Спірмена
6. Лінійна регресія.
7. Визначення коефіцієнтів рівняння лінійної регресії.
8. Дослідження регресійних зв'язків між спортивними результатами.
9. Оцінка достовірності існування статистичного взаємозв'язку

Дослідження носять комплексний характер. При контролі за ходом тренувального процесу оцінюють цілий ряд фізіологічних, біомеханічних, біохімічних та інших параметрів. При цьому часто виникає питання про взаємозв'язок ознак.

Взаємозв'язок, при якому кожному значенню одного показника відповідає чітко визначене значення другого, називається функціональним. Наприклад: середня швидкість на відрізку шляху дистанції бігуна функціонально зв'язана з часом на даному відрізку, закон Фіхнера в психології і закон Хілла в фізіології, рівняння прямої на площині і т.д.

При дослідженнях у спорті частіше зустрічаються показники з іншим видом взаємозв'язку. Взаємозв'язок, при якому одному значенню однієї величини може відповідати декілька значень другої, називається *статистичним*. Наприклад: одному значенню довжини тіла (ріст) людини може відповідати декілька значень її маси і навпаки.

Статистичний метод, який використовується для дослідження взаємозв'язків, називається *кореляційним аналізом*.

Кореляційний аналіз полягає у визначенні ступеня зв'язку між двома випадковими величинами  $X$  і  $Y$ .

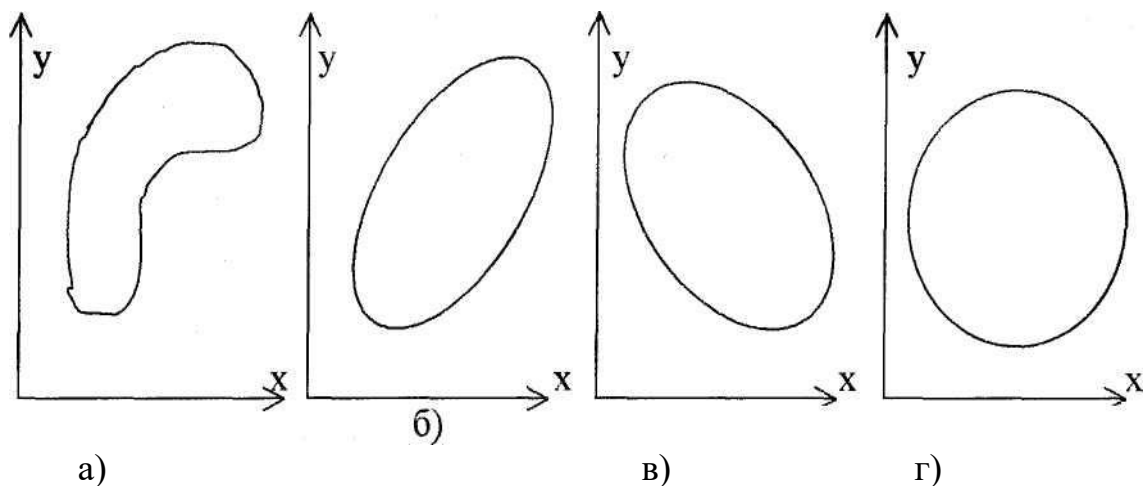
Дослідження характеру взаємозв'язку починається з побудови графічного зображення результатів вимірів в прямокутній системі координат, де кожна пара результатів буде відображатись точкою (рис. 2). Така графічна залежність називається *діаграмою розсіяння* або *кореляційним полем*.

**Кореляційне поле** відображає *статистичний взаємозв'язок* між результатами вимірів. Візуальний аналіз кореляційного поля дозволяє якісно оцінити форму, спрямованість і тісноту взаємозв'язку.

Форма визначається по виду кореляційного поля: якщо через кореляційне поле можна провести пряму лінію – форма зв'язку лінійна, якщо ні – нелінійна.

Спрямованість взаємозв'язку визначається із залежності між результатами вимірів. Якщо у випадку покращення одного показника

покращується другий – *пряма залежність, спрямованість додатня*. Збільшення одного результату викликає зменшення другого результату виміру – *обернена залежність, спрямованість від'ємна*.



**Рис. 8.** Приклади статистичних взаємозв'язків:

- а) нелінійна форма;
- б) лінійна форма, додатня спрямованість, тіснота середня;
- в) лінійна форма, від'ємна спрямованість, тіснота слабка;
- г) відсутність статистичного взаємозв'язку.

Для оцінки тісноти взаємозв'язку в кореляційному аналізі використовується абсолютна величина (модуль) спеціального показника – коефіцієнта кореляції. Абсолютне значення коефіцієнта кореляції знаходиться в межах від 0 до 1. Коефіцієнт кореляції дає кількісну оцінку статистичного взаємозв'язку між результатами вимірів.

#### Границі і значення коефіцієнта кореляції.

$r=0$	кореляція відсутня
$0 < r < 0,2$	кореляція відсутня дуже слабкий
$0,2 < r < 0,4$	слабкий взаємозв'язок
$0,4 < r < 0,7$	середній взаємозв'язок
$0,7 < r < 0,99$	сильний статистичний взаємозв'язок
$r = 1$	функціональний взаємозв'язок

Для оцінки статистичного взаємозв'язку, коли вимірювання проведено в шкалі відношень або інтервалів і форма взаємозв'язку лінійна, використовується коефіцієнт кореляції Браве-Пірсона, який обчислюється за наступною формулою:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \left[ n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}}$$

де  $x_i, y_i$  - значення  $i$ -го результату виміру;

Для практичних розрахунків більш зручна наступна формула:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{\sqrt{\left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \left[ n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}}$$

$X, Y$  – середнє арифметичне показників вимірів  $X$  і  $Y$ ;

$n$  – число вимірів (обсяг вибірки).

Знак коефіцієнта кореляції показує *спрямованість* зв'язку. Коефіцієнт кореляції зі знаком *мінус* показує *обернену* пропорційну залежність і навпаки – *додатній* знак коефіцієнта кореляції характеризує *прямую* залежність.

Коли виміри виконані в шкалі порядку, то визначення взаємозв'язку показників проводять з використанням *рангового коефіцієнта кореляції Спірмена*, який обчислюють за наступною формулою:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

де  $d_i$  – різниця рангів даної пари показників  $x_i$  і  $y_i$ ;

$i$  – місце виміру в ранжованому ряді;

$n$  – число вимірів.

Квадрат коефіцієнта кореляції називається **коефіцієнтом детермінації**:

$$d = r^2$$

*Коефіцієнт детермінації* визначає міру лінійної залежності, тобто ту частину загальної варіації одного показника, яка зумовлена варіацією іншого показника. Стверджують, що тільки  $d$  (%) взаємозв'язку спортивних результатів пояснюється їх взаємним впливом, а решта  $(100 - d)$  % варіативності залежить від інших незумовлених факторів.

При оцінці вірогідності коефіцієнтів взаємозв'язків потрібно встановити наступне:

- чи існує дійсно статистичний взаємозв'язок між двома явищами (параметрами, показниками)?
- в яких вірогідних межах (границях) лежить істинне значення коефіцієнта кореляції в генеральній сукупності?

Для лінійного парного коефіцієнта кореляції Браве-Пірсона це питання розв'язується за допомогою *таблиці 4 (Льків, Матвійв)*. Використання цієї таблиці дозволяє визначити значення коефіцієнта кореляції генеральної сукупності при заданому рівні значущості  $\alpha$  і числі ступенів свободи  $\nu$ .

Розрахункове значення  $t$ -критерію Стьюдента обчислюється за формулою:

$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}$$

Порівнюючи значення розрахункового і табличного коефіцієнта кореляції можна з вірогідністю  $P = (1 - \alpha) 100$  % говорити про існування статистичного взаємозв'язку:

$$r_{\text{роз}} < \Gamma_{\text{таб}}, \text{ — взаємозв'язок не існує}$$

$$r_{\text{роз}} > \Gamma_{\text{таб}} \text{ — взаємозв'язок існує.}$$

Оцінка *вірогідності* рангового *коефіцієнта кореляції Спірмена* виконується на основі  $t$ -критерію Стьюдента, який розраховується за наступною формулою:

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}}$$

де  $n$  – число пар вимірів

$r_s$  - ранговий коефіцієнт кореляції Спірмена.

Коли розрахункове значення  $t$ -критерію Стьюдента менше критичного  $t_{\text{роз}} < t_{\text{таб}}$  то статистичний взаємозв'язок відсутній, і навпаки. Число ступенів свободи, як і у випадку лінійного парного коефіцієнта кореляції, також на два менше числа обсягу вибірки  $\nu = n - 2$ . Критичне значення  $t$ -критерію Стьюдента визначається за таблицею 2 (Ільків, Матвіїв).

### Регресійний аналіз

В практичних дослідженнях виникає необхідність апроксимувати (описати наближено) діаграму розсіяння математичним рівнянням. Для лінійної залежності необхідно кореляційний еліпс замінити прямою лінією. В прямокутній системі координат рівняння прямої лінії записується в наступному вигляді:

$$y = a + b x$$

Такий наближений запис статистичної залежності називається *рівнянням регресії*.

**Регресія** – це залежність середніх значень випадкової величини  $y$  від величини  $x$ .

У рівнянні регресії статистичної залежності  $b$  називають коефіцієнтом регресії,  $a$  – вільним членом рівняння регресії.

Значення коефіцієнта регресії обчислюється шляхом розв'язку системи нормальних рівнянь:

$$n a + b \sum x = \sum y$$

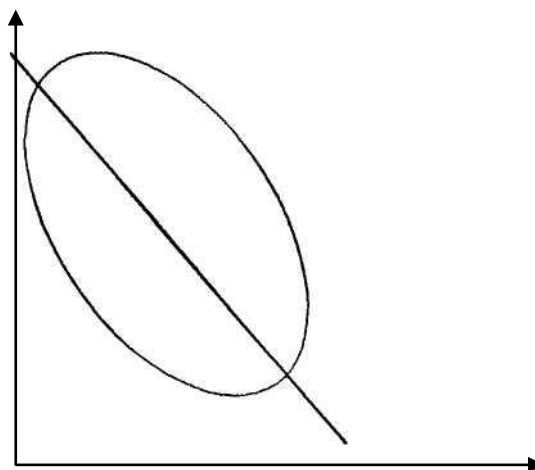
$$a \sum x + b \sum x^2 = \sum x y$$

Коефіцієнт регресії  $b$  має розмірність, яка дорівнює відношенню розмірностей показників, що вивчаються, і такий самий знак, що і коефіцієнт кореляції.

Лінію рівняння регресії будують на кореляційному полі по двох точках (наприклад, при  $X_{\min}$  і  $X_{\max}$ , отримуючи відповідні значення  $y_1$  і  $y_2$  :

$$y_1 = a + b x_{\max}$$

$$y_2 = a + b x_{\min}$$



**Рис. 9.** Кореляційне поле з нанесеною лінією рівняння регресії.

#### **Питання для самоконтролю:**

1. Дайте визначення поняттю «кореляційний аналіз»
2. Які параметри дозволяють якісно оцінити візуальний аналіз кореляційного поля?
3. Який показник використовують для оцінки статистичного взаємозв'язку?
4. Який показник використовують для оцінки статистичного взаємозв'язку коли виміри виконані в шкалі порядку?
5. На що вказує знак коефіцієнта кореляції?
6. Що таке регресія? Рівняння регресії

#### **Навчально-матеріальне забезпечення**

Мультимедійний проектор\_\_\_\_\_



**Самостійна робота:**

Основи прогнозування з використанням регресійних залежностей.

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

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

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

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