

«ЗАТВЕРДЖУЮ»  
Завідувач кафедри  
економіки, інформатики та  
кінезіології

\_\_\_\_\_ І.П. Заневский  
(підпис, ініціали, прізвище)

\_\_\_\_\_ 200\_\_р

## ЛЕКЦІЯ №6 з навчальної дисципліни

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

## Тема: Кореляційний та регресійний аналізи

### Навчальний потік

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

**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.

## 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., Макаруч 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.

### **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 (зріст) людини може відповідати декілька значень її маси і навпаки.

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

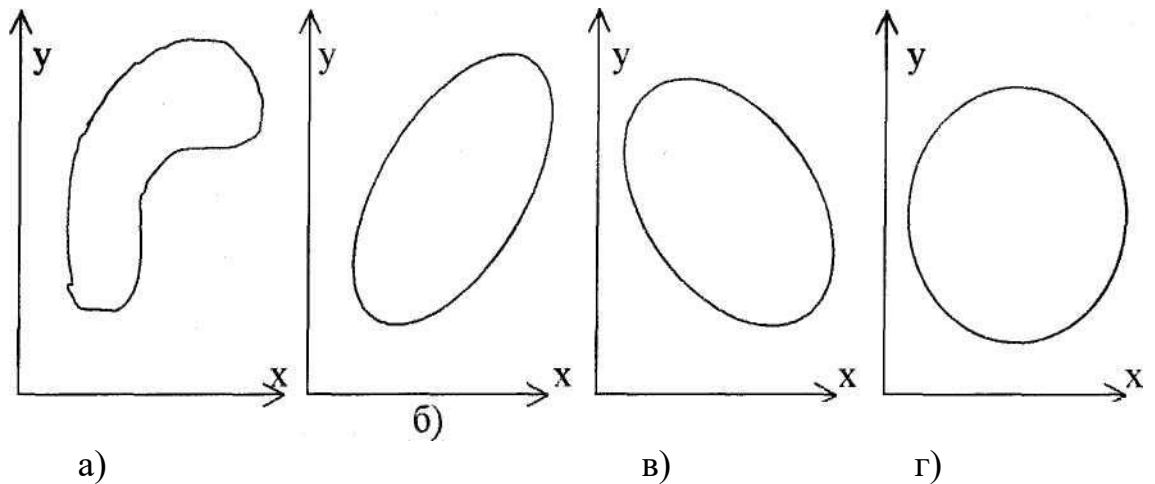
Кореляційний аналіз полягає у визначенні ступеня зв'язку між двома випадковими величинами  $X$  і  $Y$ .

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:

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

$\bar{X}, \bar{Y}$  – 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 – *додатний* знак коефіцієнта кореляції характеризує *пряму* 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 Браує-Пірсона це питання розв'язується за допомогою *таблиці 4 (Ільків, Матвійів)*. Використання цієї таблиці дозволяє визначити значення коефіцієнта кореляції генеральної сукупності при заданому рівні значущості  $\alpha$  і числі ступенів свободи  $\nu$ .

Розрахункове значення  $t$ -критерію Стюдента 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 Стюдента, 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, Матвійів*).

### 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:

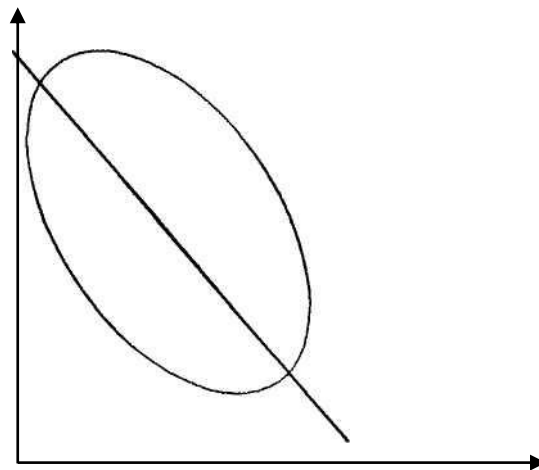
$$\begin{aligned} n a + b \sum x &= \sum y \\ a \sum x + b \sum x^2 &= \sum x y \end{aligned}$$

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 розробили:** к.пед. н., доц. О.С. Ільків к.фіз-мат. н., доц. А.П. Власов

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

**Обговорено на засіданні кафедри:** економіки, інформатики, кінезіології

(назва кафедри)

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