

ТЕМА 8. СТАТИСТИЧНЕ ВИВЧЕННЯ ЗВ'ЯЗКУ МІЖ ЯВИЩАМИ

8.1. Види зв'язків соціально-економічних явищ та завдання їх статистичного вивчення.

8.2. Кореляційно-регресійний аналіз зв'язку, його суть, завдання та основні етапи.

8.3. Оцінка щільності та істотності кореляційного зв'язку.

Автор Олександр Маценко, PhD

8.1. ВИДИ ЗВ'ЯЗКІВ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ЯВИЩ ТА ЗАВДАННЯ ЇХ СТАТИСТИЧНОГО ВИВЧЕННЯ

За характером залежності між факторними і результатними ознаками зв'язки поділяються на:

- функціональні;
- стохастичні.

Функціональним вважається зв'язок, при якому кожному значенню факторної ознаки (аргументу) відповідає одне або декілька строго визначених значень результатної ознаки (функції).

В економічній науці превалюють стохастичні зв'язки.

Стохастичні (ймовірнісні) – це зв'язки, при яких кожному значенню факторної ознаки відповідає множина значень результатної, які варіюють і не мають строго визначеної залежності від зміни значення факторної ознаки.

Окремим випадком стохастичних зв'язків є **кореляційні**. Це зв'язки, при яких кожному значенню факторної ознаки відповідає середнє значення результатної із множини його можливих значень.

З історії

Correlation – зв'язок. Саме слово «кореляція» ввів в статистику англійський біолог і статистик Френсіс Гальтон в кінці XIX століття. В перекладі з англійської воно означає «відповідність, співвідношення». Ф.Гальтон і К. Пірсон вважаються основоположниками теорії кореляції.

Існують наступні варіанти кореляційного зв'язку:

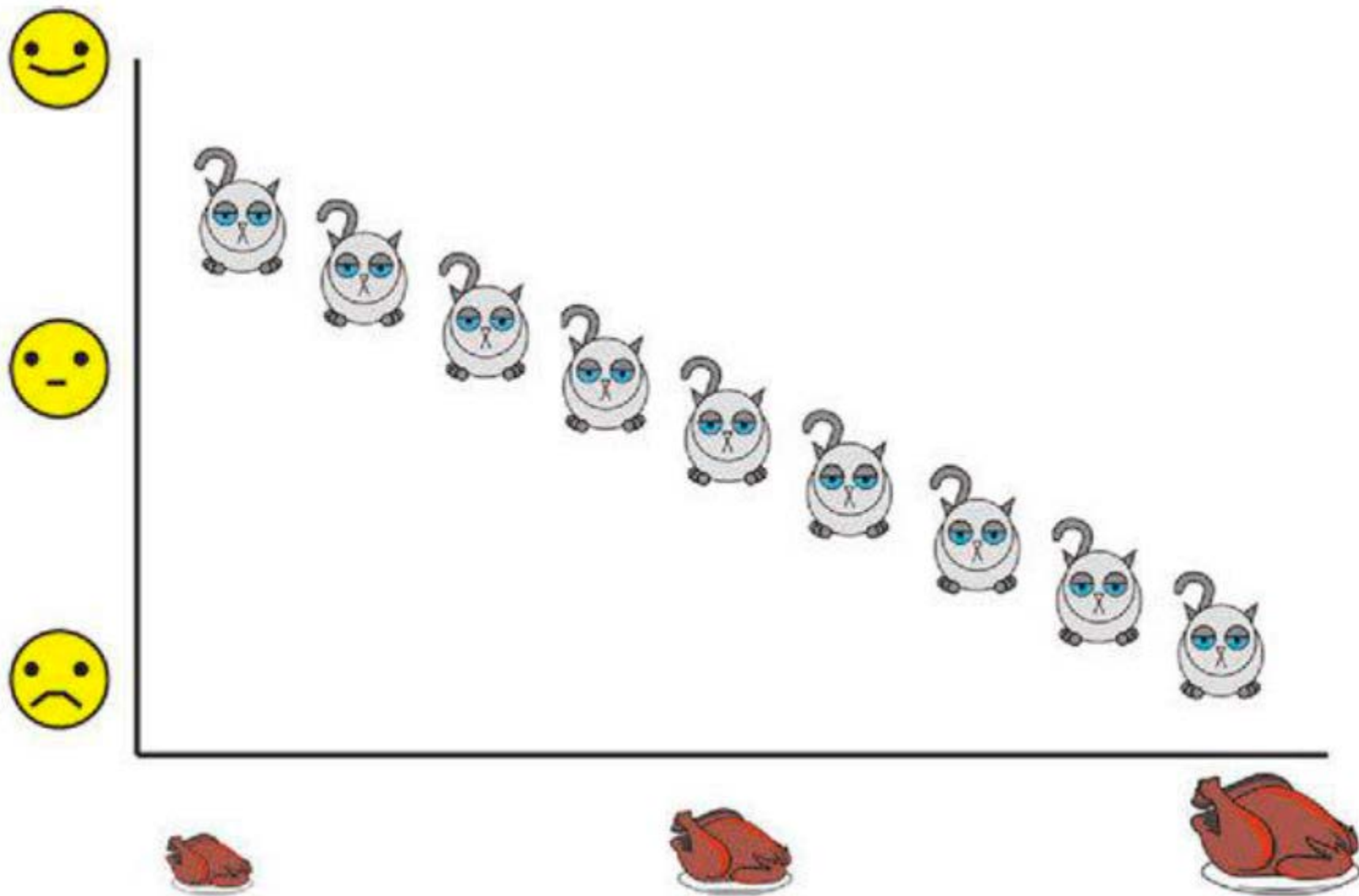
- 1) **Парна кореляція** – це залежність між двома ознаками (результатною та факторною) або двома факторними.
- 2) **Множинна кореляція** – це залежність між результатною і двома чи більше факторними ознаками.

ВИДИ ЗВ'ЯЗКІВ

За загальним напрямом зв'язки бувають прямими і зворотними.

Прямим називається зв'язок, при якому зі збільшенням факторної ознаки збільшується і результатна.

Зворотним називається зв'язок, при якому зі збільшенням факторної ознаки результатна зменшується.

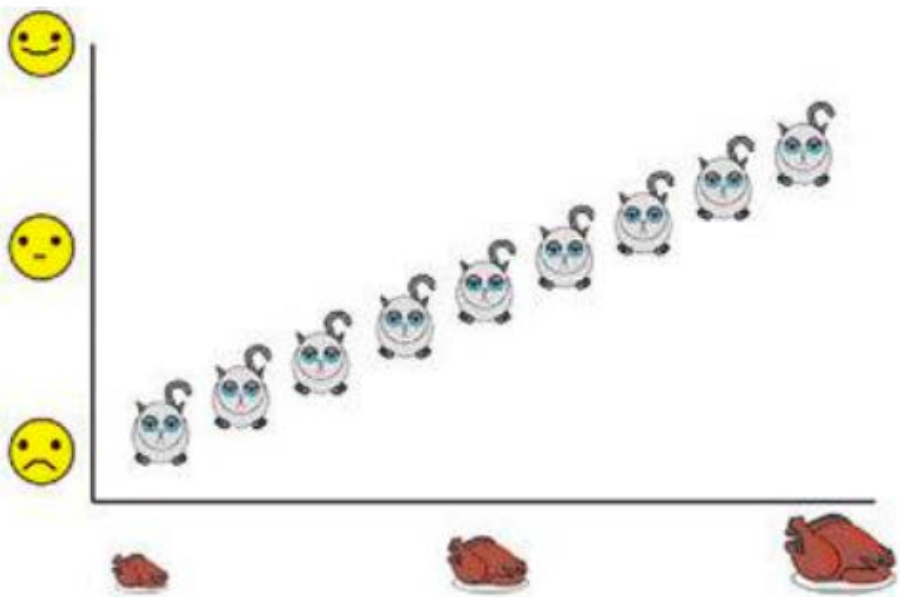


Лінійний зворотній зв'язок

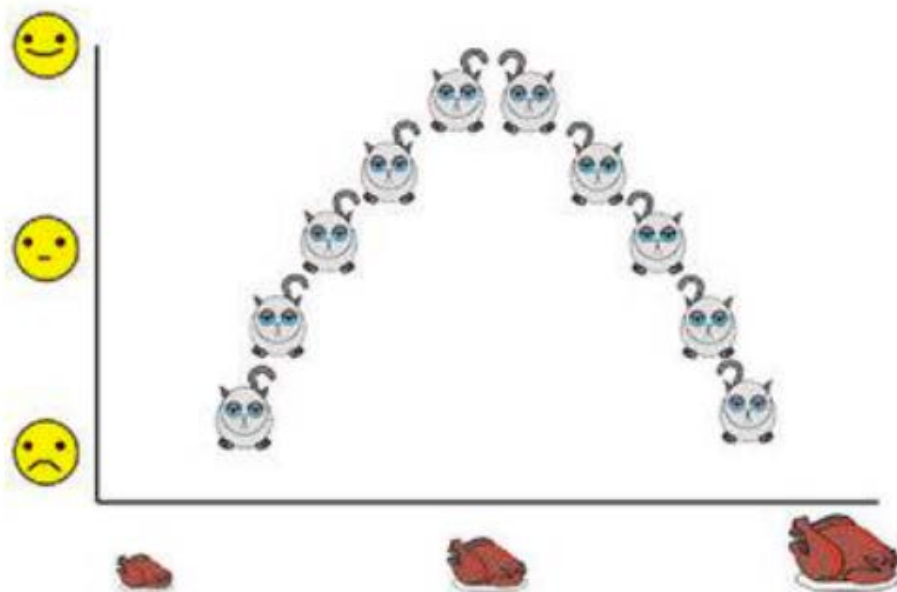
За аналітичним вираженням зв'язки бувають *лінійні* і *нелінійні*.

Лінійним називається зв'язок, який можна виразити рівнянням прямої лінії.

Нелінійним є зв'язок, який може бути виражений рівнянням якої-небудь кривої лінії (параболи, гіперболи, показникової, степеневої функції тощо.).



Лінійний зв'язок



Нелінійний зв'язок

8.2. Кореляційно-регресійний аналіз зв'язку, його суть, завдання та основні етапи

За допомогою КРА вирішуються дві основні задачі:

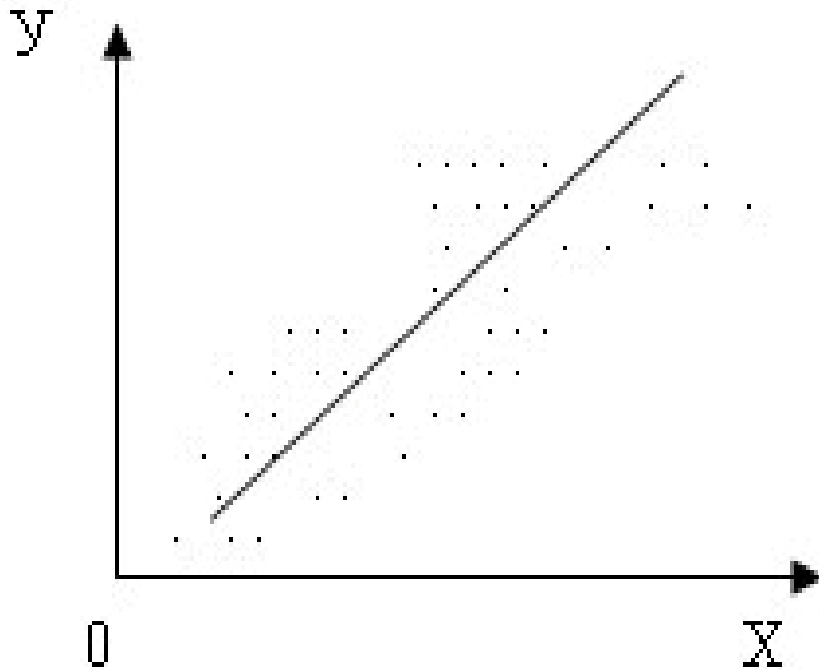
- 1) Виявлення форми зв'язку і параметрів рівняння зв'язку (це задача регресійного аналізу).
- 2) Визначення щільності зв'язку (задача кореляційного аналізу).

1-ша задача вирішується за допомогою регресійного аналізу і полягає у підборі математичного рівняння (моделі), яке виражає залежність між результатною і факторною ознаками. Для виявлення наявності зв'язку, його форми і на пряму використовуються методи: метод зіставлення паралельних рядів, графічний, метод аналітичних групувань.

2-га задача вирішується методами кореляційного аналізу шляхом розрахунку показників щільності зв'язку:

- лінійного коефіцієнта кореляції;
- кореляційного відношення;
- коефіцієнта детермінації;
- індекса кореляції.

Найбільш суттєвим етапом КРА є підбір рівняння, що характеризує сутність зв'язку.



$$\bar{y}_x = a + bx$$

де \bar{y}_x – у вирівняний по x ;
 x – незалежна змінна (фактор);
 a, b – параметри шуканої лінії:
 a – початок відліку або значення при $x = 0$;
 b – коефіцієнт регресії, який показує середню зміну y при кожній зміні x на одну одиницю.

Якщо: $b > 0$, то зв'язок прямий;

$b < 0$, зв'язок обернений;

$b = 0$, зв'язок відсутній.

Серед множин кривих необхідно знайти таку, яка дає найменшу суму квадратів відхилень фактичних значень результатної ознаки від вирівняних (теоретичних) значень:

$$\sum (y_i - \bar{y}_{x_i})^2 = \min$$

Параметри a і b визначають шляхом розв'язання системи нормальних рівнянь

$$\begin{cases} na + b\sum x = \sum y \\ a\sum x + b\sum x^2 = \sum xy \end{cases}$$

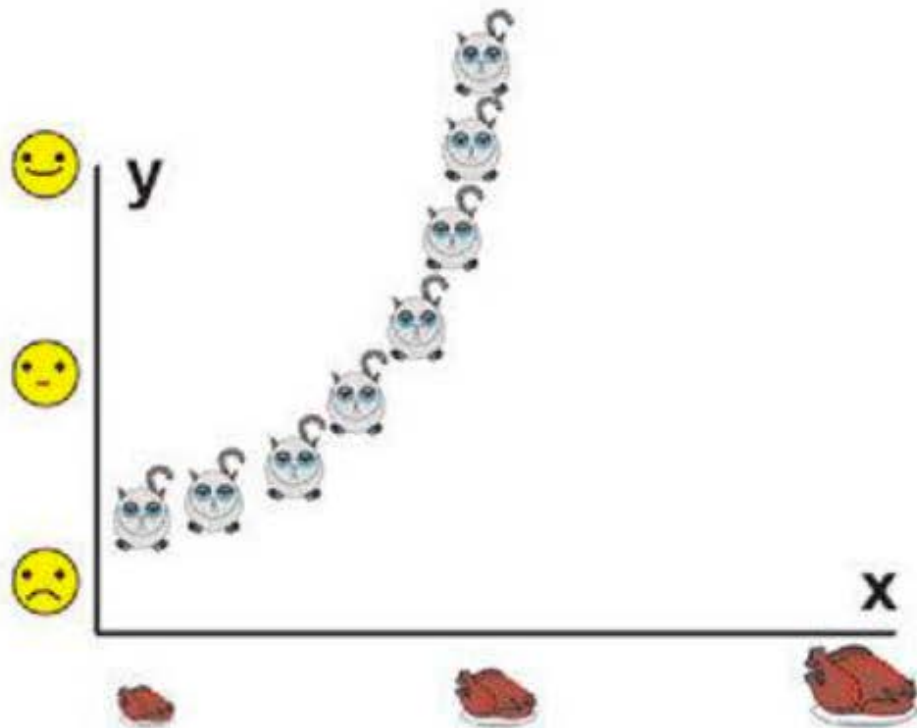
Також параметри a і b шуканої прямої можна визначити і за такими формулами:

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - (\bar{x})^2} \qquad a = \bar{y} - b\bar{x}$$

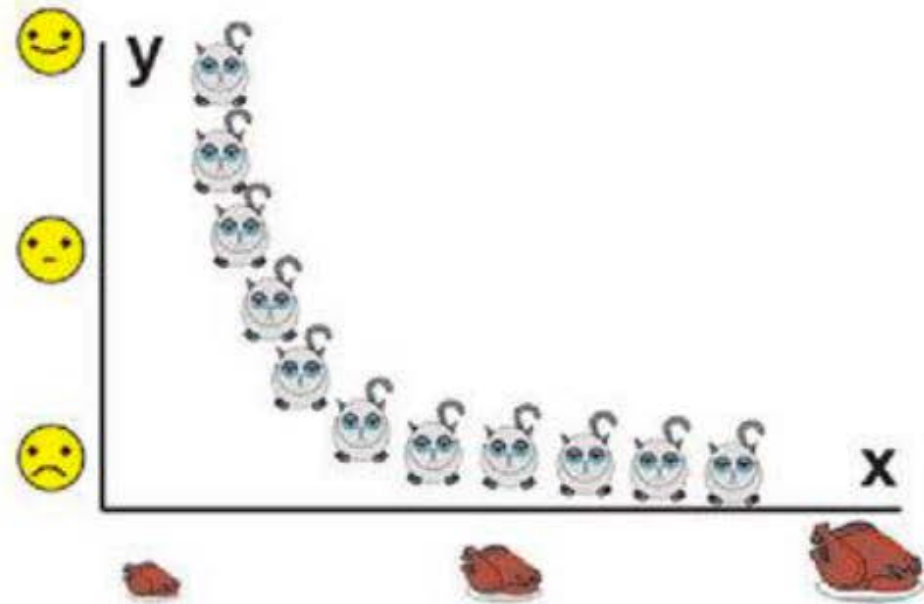
При цьому сума значень результатної ознаки фактичної та вирівняної повинна бути однаковою.



Нелінійна квадратична регресія



$$y = b_0 + b^x$$



$$y = b_0 + \frac{b}{x}$$

8.3 Оцінка щільності та істотності кореляційного зв'язку

Для парної лінійної кореляції щільність зв'язку визначають як правило за допомогою лінійного коефіцієнта кореляції (коефіцієнт Пірсона):

$$r = b \frac{\sigma_x}{\sigma_y}$$

де b – коефіцієнт регресії;

σ_x – середнє квадратичне відхилення факторної ознаки;

σ_y – середнє квадратичне відхилення результатної ознаки.

$$\sigma_x = \sqrt{\overline{X^2} - (\overline{X})^2}$$

$$\sigma_y = \sqrt{\overline{Y^2} - (\overline{Y})^2}$$

Якщо $r < 0,3$ – зв'язок слабкий;

$r = 0,3 \div 0,7$ – зв'язок середньої сили;

$r > 0,7$ – зв'язок сильний;

$r = 1$ – зв'язок функціональний.

Лінійний коефіцієнт кореляції можна визначити і за іншими формулами

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$

$$r = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \cdot \sum (y - \bar{y})^2}}$$

Для всіх інших форм зв'язку, в тому числі і лінійної, щільність може бути визначена за допомогою універсального показника зв'язку – **кореляційного відношення**:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}}$$

$$\sigma^2 = \frac{\sum (y_i - \bar{y})^2 f_i}{\sum f_i} = \bar{y}^2 - (\bar{y})^2$$

- **загальна дисперсія** (характеризує варіацію результатної ознаки під впливом усіх факторів і причин, у тому числі і випадкових).

$$\delta^2 = \frac{\sum (\bar{y}_i - \bar{y})^2 f_i}{\sum f_i}$$

- **міжгрупова дисперсія** (характеризує варіацію результативної ознаки за рахунок факторної або групувальної ознаки).

Індекс кореляції визначають зіставленням внутрішньогрупової і загальної дисперсії, і обчислюють за формулою:

$$R = \sqrt{1 - \frac{\bar{\sigma}^2}{\sigma^2}}$$

Величина R збігається з показником η . Чим ближче R до 1, тим щільніший зв'язок між ознаками.

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2 f_i}{\sum f_i} \qquad \sigma_i^2 = \frac{\sum (x_i - \bar{x}_i)^2}{n_i}$$

$$\sigma^2 = \delta^2 + \bar{\sigma}^2$$

Оцінка істотності кореляційного зв'язку

Для оцінки надійності зв'язку використовують різні критерії, найбільш поширеним з яких є F -критерій Фішера:

$$F = \frac{\delta^2 k_2}{\bar{\sigma}^2 k_1}$$

δ^2 – міжгрупова дисперсія;

$\bar{\sigma}^2$ – залишкова дисперсія;

k_1, k_2 – ступені вільності для великої і малої дисперсій;

$$k_1 = m - 1$$

$$k_2 = n - m,$$

де m – кількість груп, на які розбита сукупність;

n – число одиниць сукупності.

Оцінка істотності кореляційного зв'язку

Фішер знайшов розподіл відношень дисперсій і розробив відповідні математичні таблиці, в яких наводиться F -критерій теоретичний F_m при двох ймовірностях 0,95 і 0,99.

Якщо $F_{\phi} > F_m$, то з прийнятим ступенем ймовірності можна стверджувати про наявність впливу фактора, який досліджується.

Коли ж $F_{\phi} < F_m$, то різниця між дисперсіями зумовлена впливом випадкових факторів.

Для оцінки достовірності зв'язку використовують t -критерій Стюдента. Вважається, що кореляційний зв'язок є достовірним лише при достатньому числі спостережень (не менше 20 – 30). Перевірка надійності здійснюється за формулою:

$$t = \frac{\eta}{\mu}$$

де η – кореляційне відношення;

μ – середня похибка кореляційного відношення:

$$\mu = \frac{1 - \eta^2}{\sqrt{n}} \quad \text{де } n \text{ – число спостережень.}$$

Зв'язок вважається істотним та надійним, якщо коефіцієнт Стюдента $t \geq 3$.

Якщо $t < 3$, зв'язок не можна вважати достовірним і висновки про вірогідність зв'язку між досліджуваними явищами сумнівні.

ДЯКУЮ ЗА УВАГУ!