

Економетрія

*Кафедра інформаційної економіки,
підприємництва та фінансів*

Лектор:

ГЛУЩЕВСЬКИЙ В'ячеслав Валентинович

ЗМІСТОВИЙ МОДУЛЬ 2

ОДНОФАКТОРНА ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ

Тема №3-5

- 1. Парна лінійна регресія.**
- 2. Використання методу найменших квадратів (МНК) для оцінки параметрів парної лінійної регресії.**
- 3. Дослідження рівняння парної лінійної регресії.**

ПАРНА ЛІНІЙНА РЕГРЕСІЯ

Парною лінійною регресією Y на X називається одностороння стохастична лінійна залежність між випадковими величинами показника Y (результативного показника) і фактора X (факторного показника), які знаходяться в причинно-наслідкових відношеннях, причому зміна фактора викликає зміну показника.

ПАРНА ЛІНІЙНА РЕГРЕСІЯ

Зв'язок між показником Y та фактором X з урахуванням можливих (випадкових) відхилень l , де α, β – невідомі параметри, які підлягають оцінці, **МОЖНА ПРЕДСТАВИТИ ТАК:**

$$Y = \alpha X + \beta + l$$

**Систематична
складова**

$$\hat{Y} = \alpha X + \beta$$

характеризує середнє
значення показника $\bar{Y}$
для конкретного значення X

**Випадкова
складова**

$$l$$

характеризує відхилення
індивідуальних значень
показника Y від його середнього $\bar{Y}$
для конкретних значень X

МОДЕЛЬ ПАРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ



МЕТОД НАЙМЕНШИХ КВАДРАТІВ (МНК)

оцінювання невідомих параметрів МОДЕЛІ ЛІНІЙНОЇ ПАРНОЇ РЕГРЕСІЇ

СУТНІСТЬ МНК полягає в підборі таких оцінок параметрів a і b парної лінійної регресії, для яких **СУМА КВАДРАТІВ ВІДХИЛЕНЬ** спостережуваних значень показника від згладжувальних (теоретичних, модельних) **БУДЕ МІНІМАЛЬНОЮ**, а саме:

$$Q(a, b) = \sum_{i=1}^n l_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - ax_i - b)^2 \rightarrow \min$$

НЕОБХІДНА УМОВА
ІСНУВАННЯ ЕКСТРЕМУМУ
для $Q(a, b)$ – це система нормальних рівнянь, яка підлягає розв'язанню:

$$\begin{cases} \frac{\partial Q(a, b)}{\partial a} = 0, \\ \frac{\partial Q(a, b)}{\partial b} = 0. \end{cases}$$

Знайдені оцінки параметрів:

$$\begin{cases} a = \text{const}, \\ b = \text{const}. \end{cases}$$

ВИРАЖЕННЯ ПАРАМЕТРІВ ПАРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ через числові характеристики показника і фактора

$$\begin{cases} a = \frac{n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \\ b = \bar{y} - a \bar{x} \end{cases}$$

де

x_i, y_i

— фактичні дані: індивідуальні значення фактора X і показника Y відповідно

n

— кількість спостережень пар значень фактора X і показника Y

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

— середні значення фактора X і показника Y відповідно

ДОСЛІДЖЕННЯ РІВНЯННЯ ПАРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ

КОЕФІЦІЄНТ КОРЕЛЯЦІЇ $r[X, Y]$ -

безрозмірна статистична характеристика, яка описує ступінь щільності лінійної залежності між парою випадкових величин X і Y

$$-1 \leq r[X, Y] \leq 1$$

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

$$r[X, Y] = 0$$

— зв'язок між X і Y відсутній

$$r[X, Y] > 0$$

— зв'язок між X і Y існує: пряма залежність

$$r[X, Y] < 0$$

— зв'язок між X і Y існує: обернена залежність

$$r[X, Y] = \{-1; 1\}$$

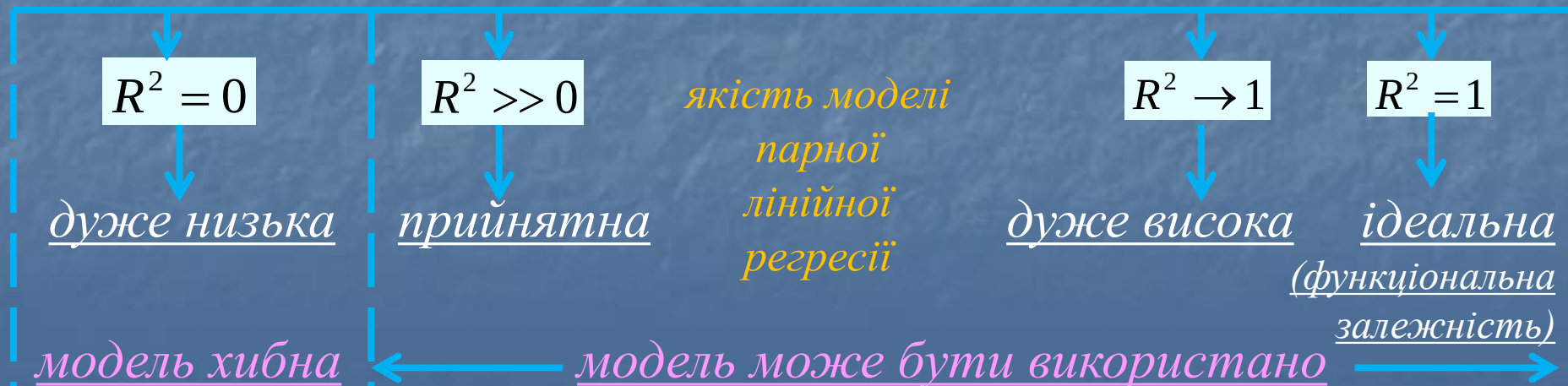
між X і Y дуже сильний зв'язок :
функціональна залежність

ДОСЛІДЖЕННЯ РІВНЯННЯ ПАРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ

ВИБІРКОВИЙ КОЕФІЦІЄНТ ДЕТЕРМІНАЦІЇ R^2 -

безрозмірна статистична характеристика, яка є оцінкою якості математичного опису існуючої залежності між двома явищами (показником Y і фактором X), що подано моделлю регресії:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \rightarrow R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad l_i^2 = (y_i - \hat{y}_i)^2 \rightarrow R^2 = 1 - \frac{\sum_{i=1}^n l_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$



ПЕРЕВІРКА СТАТИСТИЧНОЇ ЗНАЧУЩОСТІ ХАРАКТЕРИСТИК ПАРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ

БАЗОВЕ ЙМОВІРНІСНЕ ПРИПУЩЕННЯ:

усі випадкові величини розподілені
за нормальним законом

Коефіцієнт кореляції:

Критерій
Стьюдента: $|t| \geq t_{pk},$

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

Коефіцієнт детермінації:

$F \geq F_{pk},$ Критерій
Фішера:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}$$

← розрахункові значення →

$$t_{pk}, F_{pk}$$

- табличні значення критеріїв
Стьюдента та Фішера відповідно

p – заданий рівень ймовірності
 $k=n-m-1$ – кількість ступенів
вільності

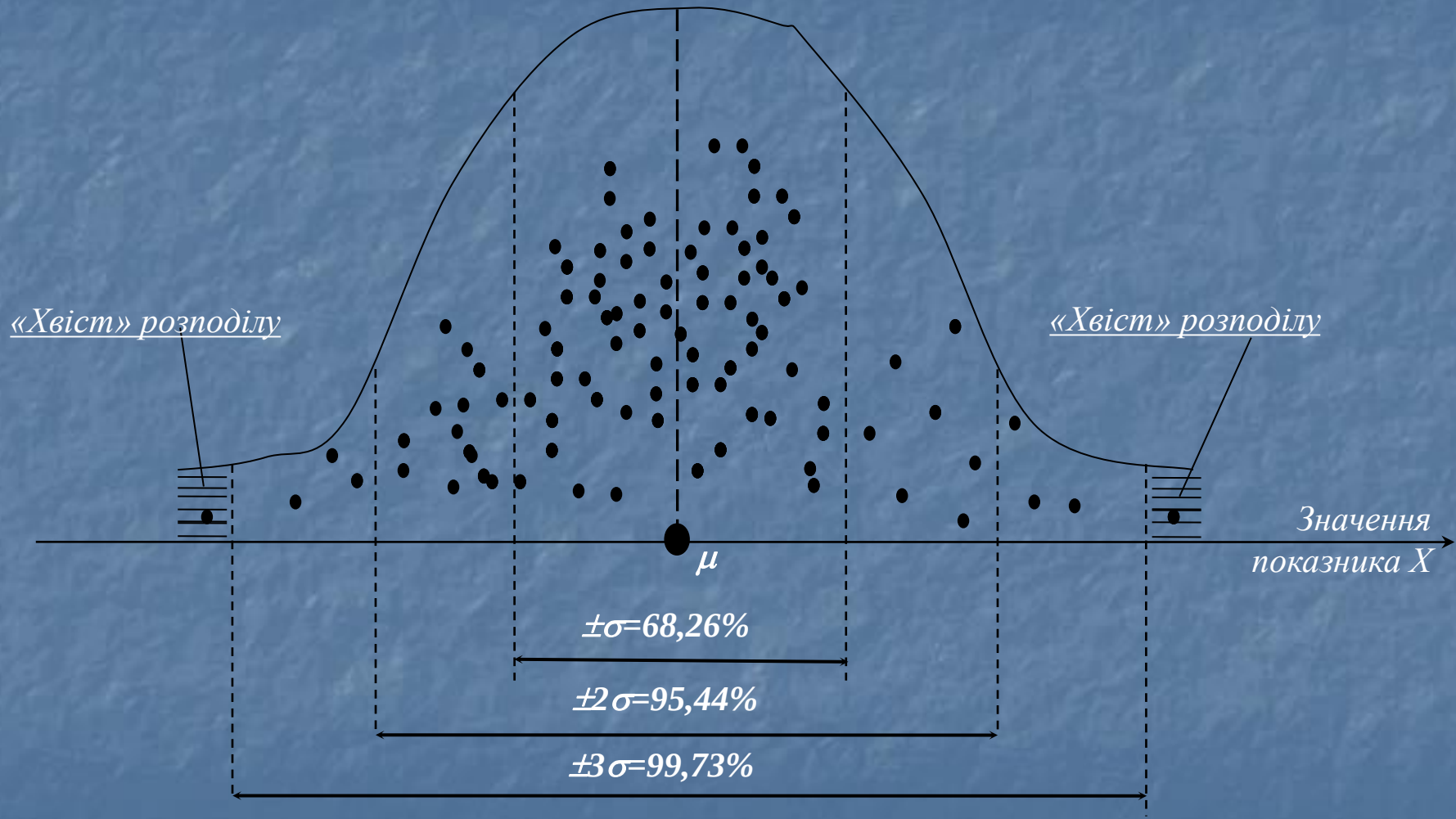
n – кількість спостережень
 m – кількість факторів ($m=1$)

ОСНОВНІ СКЛАДОВІ ПРОБЛЕМИ АНАЛІЗУ ВАРІАБЕЛЬНОСТІ БУДЬ-ЯКОГО ПРОЦЕСУ (ПОКАЗНИКА, ФАКТОРА)

1. Більша частина результатів будь-якого процесу визначається системою, в якій цей процес (показник, фактор) протікає; **лише незначна їх частина** обумовлюється зовнішніми по відношенню до цієї системи причинами

2. Для практики достатньо навчитися передбачати результати з тим ступенем упевненості, яка є логічно та/або економічно доцільною

Схематичне представлення значень певного показника у формі нормального закону розподілу



НАДІЙНІ ІНТЕРВАЛИ ПАРАМЕТРІВ ПАРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ

$$\hat{Y} = aX + b$$

для параметра **a**:

$$a - \Delta a \leq \alpha \leq a + \Delta a$$



$$\Delta a = t_{pk} \cdot S \cdot \frac{1}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

для параметра **b**:

$$b - \Delta b \leq \beta \leq b + \Delta b$$



$$\Delta b = t_{pk} \cdot S \cdot \sqrt{\frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1}}$$

t_{pk} - табличне значення критерію Стюдента

НАДІЙНІ ІНТЕРВАЛИ

для базисних значень регресійного показника Y

$$Y = \alpha X + \beta + l$$

надійний інтервал для y_i : $\hat{y}_i - \Delta\hat{y}_i \leq y_i \leq \hat{y}_i + \Delta\hat{y}_i \rightarrow \Delta\hat{y}_i = t_{pk} \cdot S[\hat{y}_i]$

t_{pk} - табличне значення критерію
Стюдента

p - заданий рівень ймовірності

n - кількість спостережень

m - кількість факторів ($m=1$)

$$S[\hat{y}_i] = S \cdot \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1}}$$

ПРОГНОЗУВАННЯ З ВИКОРИСТАННЯМ МОДЕЛІ РЕГРЕСІЇ

*Мир сложен, и точно
предсказать результат
большинства реальных процессов
невозможно в принципе*

*(Уолтер Шухарт, інженер-фізик, автор та
ідеолог теорії контрольних карт)*

ПРОГНОЗ ТА ЙОГО НАДІЙНИЙ ІНТЕРВАЛ

ПРОГНОЗУВАННЯ з використанням моделі регресії – наукове передбачення ймовірних шляхів розвитку соціально-економічних явищ і процесів на базі ретроспективних статистичних даних, яке ґрунтується на **припущенні про збереження загальної тенденції розвитку явищ у часі** (незмінність умов, які сформували цю тенденцію)

середнє значення прогнозу та його надійний інтервал:

$$\hat{y}_{\text{прогноз}} = ax_{\text{прогноз}} + b \rightarrow y_{\text{прогноз}} \in \left[\hat{y}_{\text{прогноз}} \pm \Delta \hat{y}_{\text{прогноз}} \right]$$

a, b

– обчислені значення параметрів моделі регресії

$x_{\text{прогноз}}$

– прогнозне значення фактора

$$\Delta \hat{y}_{\text{прогноз}} = t_{pk} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$S = \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / (n - m - 1)}$$

КОЕФІЦІЄНТ ЕЛАСТИЧНОСТІ для базисних даних і прогнозу

Коефіцієнт еластичності показує, на скільки відсотків зміниться показник $\hat{Y}$, якщо фактор X зміниться на 1%

$$K_{x_i} = \frac{x_i \cdot \hat{y}'(x_i)}{\hat{y}(x_i)} \xrightarrow{\hat{y}'(X) = (aX + b)'} K_X = \frac{(aX + b)' \cdot X}{\hat{Y}} = \frac{aX}{\hat{Y}}$$
$$K_{x_i} = \frac{a \cdot x_i}{\hat{y}(x_i)}$$
