

Економетрія

*Кафедра інформаційної економіки,
підприємництва та фінансів*

Лектор:

ГЛУЩЕВСЬКИЙ В'ячеслав Валентинович

ЗМІСТОВИЙ МОДУЛЬ 4

КЛАСИЧНА ЛІНІЙНА БАГАТОФАКТОРНА МОДЕЛЬ

Теми 8-9

- 1. Лінійна багаторфакторна економетрична модель (множинна регресія).**
- 2. Визначення факторного впливу на показник методом кореляційного аналізу.**

ПРО ПРИЧИННО-НАСЛІДКОВІ ЗВ'ЯЗКИ:

Результат діяльності економічних агентів залежить від складного переплетіння низки причин

↓
Причинами, які можуть суттєво впливати на результат діяльності, є інтенсивність залучення у бізнес-процеси економічних агентів ресурсів (матеріально-сировинних, трудових, фінансових, інформаційних тощо), раціональність управлінських рішень (внутрішні чинники) та різноманіття зовнішніх чинників

Усі дестабілізуючі чинники утворюють причинну базу, яка у сукупності й визначає реалістичність того чи іншого сценарію поведінки економічного агента та отримання ним певного результату діяльності

↓
Поміж усіх причин є такі фактори, що більше (суттєво більше) за інші впливають на результат діяльності, а також є такі, що дублюють вплив інших факторів, викривлюючи інформацію

ВИСНОВОК:

1. Результуючий показник діяльності економічних агентів слід моделювати з урахуванням певної кількості ключових причин (факторних показників) у вигляді множинної регресії.
2. Статистичну базу для побудови моделі множинної регресії утворюють синхронізовані за часом і об'єктом (об'єктами) ретроспективні дані спостережень для виділених з причинної бази деякої підмножини факторів.

МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ

Змістовний приклад:

Y – річний (квартальний, місячний, тижневий, денний) попит на товари першої необхідності на душу населення (у.г.о.);

X_1 – споживання м'ясних продуктів (у.г.о.);

X_2 – споживання молочних продуктів (у.г.о.);

X_3 – споживання зерна і продуктів його переробки (у.г.о.);

X_4 – споживання овочів і фруктів (у.г.о.);

X_5 – споживання цукру і кондитерських виробів (у.г.о.);

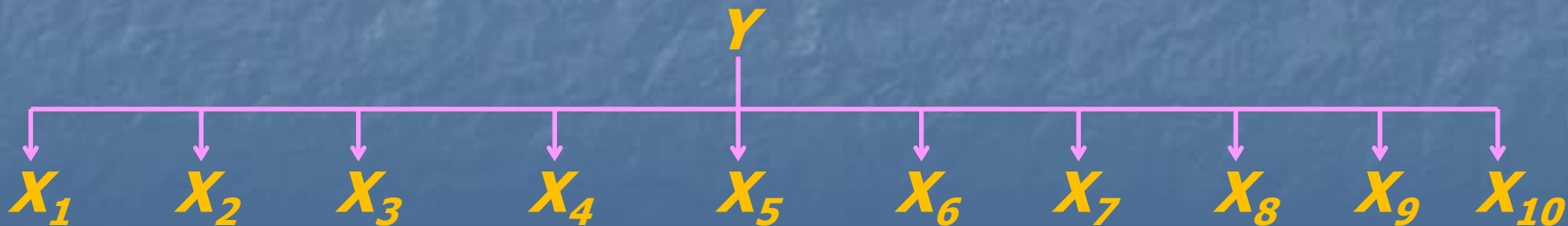
X_6 – споживання алкогольних і безалкогольних напоїв (у.г.о.);

X_7 – споживання смакових товарів (у.г.о.);

X_8 – купівля тканини та швейних виробів (у.г.о.);

X_9 – купівля галантерейних виробів (у.г.о.);

X_{10} – купівля взуття (у.г.о.).



МОДЕЛЬ МНОЖИННОЇ ЛІНІЙНОЇ РЕГРЕСІЇ

Змістовний приклад:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \alpha_6 X_6 + \alpha_7 X_7 + \alpha_8 X_8 + \alpha_9 X_9 + \alpha_{10} X_{10} + L$$

Загальний випадок:

$$m = 10$$

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_i X_i + \dots + \alpha_m X_m + L, \quad i = \overline{0, m}$$

Систематична складова

Випадкова складова

L

$$\hat{Y} = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_i X_i + \dots + \alpha_m X_m,$$

Стохастична залежність
 Y від X_i

$$\left\{ \begin{matrix} a_0 & a_1 & a_2 & \dots & a_i & \dots & a_m \end{matrix} \right\}$$

Статистичні оцінки
справжніх параметрів регресії

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_i X_i + \dots + a_m X_m,$$

Модель множинної регресії

МНК

Оцінка вектора
невдомих параметрів

$$\bar{a} = (a_0; a_1; a_2; \dots; a_i; \dots; a_m),$$

Оцінка адекватності
моделі факт. даним

$$F_{\text{розрах}} \geq F_{\text{табл}}(p; k_1; k_2)$$

ВИКОРИСТАННЯ
МОДЕЛІ

МАТЕМАТИЧНИЙ ІНСТРУМЕНТАРІЙ:

Матрична форма множинної лінійної регресії

$$\vec{Y} = [X] \vec{\alpha} + \vec{L}$$

або

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \\ \vdots \\ y_n \end{bmatrix} = [X] = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1i} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2i} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{j1} & \cdots & x_{ji} & \cdots & x_{jm} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{ni} & \cdots & x_{nm} \end{bmatrix} \vec{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_j \\ \vdots \\ \alpha_m \end{bmatrix} + \vec{L} = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_j \\ \vdots \\ l_n \end{bmatrix}$$

Матрична форма для
знаходження оцінок
невдомих параметрів
множинної лінійної регресії

$$\vec{a} = [X]^T [X]^{-1} [X]^T \vec{Y}$$

$$\frac{\partial(L^2)}{\partial a_i} = 0, i = \overline{0, m}$$

МНК

$$L^2 \xrightarrow{L = \vec{Y} - [X] \vec{\alpha}} \min$$

$$[X]^T = \begin{bmatrix} 1 & 1 & \cdots & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{j1} & \cdots & x_{n1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1i} & x_{2i} & \cdots & x_{ji} & \cdots & x_{ni} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{jm} & \cdots & x_{nm} \end{bmatrix}$$

АЛГОРИТМ ЗНАХОДЖЕННЯ НЕВІДОМИХ ПАРАМЕТРІВ:

1. Знайти добуток матриць: $[X]^T [X]$ — матриця, $(m+1) \times (m+1)$
2. Знайти обернену матрицю: $[[X]^T [X]]^{-1}$ — матриця, $(m+1) \times (m+1)$
3. Знайти добуток матриці на вектор: $[X]^T \vec{Y}$ — вектор-стовпець, $(m+1) \times 1$
4. Знайти добуток матриці на вектор:

$$\vec{a} = [[X]^T [X]]^{-1} [X]^T \vec{Y} \text{ — вектор-стовпець, } (m+1) \times 1$$

УЗАГАЛЬНЕНА СХЕМА ПЕРЕВІРКИ ПОБУДОВАНОЇ МОДЕЛІ НА АДЕКВАТНІСТЬ (МАТЕМАТИЧНИЙ АСПЕКТ):

Вибірковий коефіцієнт
множинної детермінації:

$$R^2 = 1 - \frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n (y_j - \bar{y})^2}$$

Вибірковий коефіцієнт
множинної кореляції:

$$R = \sqrt{1 - \frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n (y_j - \bar{y})^2}}$$

$$t_{\text{розрах}} = R \cdot \sqrt{\frac{n - m - 1}{1 - R^2}}$$

$$F_{\text{розрах}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$$

Оцінка значимості R за
критерієм Стюдента

$$|t_{\text{розрах}}| \geq t_{pk},$$

Оцінка адекватності
моделі факт. даним за
критерієм Фішера

$$F_{\text{розрах}} \geq F_{\text{табл}}(p; k_1; k_2)$$

p – задана надійність (ймовірність);
 $k_1 = m$ і $k_2 = n - m - 1$ – ступені вільності

Модель
адекватна
АНАЛІЗ

ПЕРЕВІРКА ПОБУДОВАНОЇ МОДЕЛІ НА АДЕКВАТНІСТЬ (ЗМІСТОВНА ІНТЕРПРЕТАЦІЯ):

Тіснота зв'язку загального впливу всіх незалежних змінних X (фактори, причини) на залежну змінну Y (регресант, результат) визначається коефіцієнтами множинної детермінації R^2 та множинної кореляції R , а також парними коефіцієнтами кореляції r_{ik} :

Коефіцієнт множинної детермінації R^2 характеризує, якою мірою варіація залежної змінної Y визначається варіацією сукупності незалежних змінних X

Коефіцієнт множинної кореляції R є оцінкою близькості математичної форми зв'язку до вибірових (фактичних) даних

Парні коефіцієнти кореляції r_{ik} дають оцінку тісноти зв'язку між парами змінних:

$$r_{Y;X_i}$$

- залежною змінною Y та незалежною X_i

$$r_{X_i;X_k}$$

- незалежними змінними X_i та незалежною X_k

СТАТИСТИЧНІ ГІПОТЕЗИ: ПЕРЕВІРКА ЗНАЧУЩОСТІ:

Перевірка статистичної значущості коефіцієнта множинної детермінації:

$H_0 : R^2 = 0$ — досліджуване рівняння множинної регресії НЕ ПОЯСНЮЄ змінювання регресанта Y під впливом сукупності факторів X_i

$H_0 : a_1 = a_2 = \dots = a_i = \dots = a_m = 0$ — УСІ КОЕФІЦІЄНТИ при незалежних змінних X_i дорівнюють НУЛЮ

$H_A : a_i \neq 0, i \in (1, \dots, m)$ — Хоча б ОДИН КОЕФІЦІЄНТ при незалежній змінній X_i відмінний від НУЛЯ, тобто хоча б один з факторів X_i ВПЛИВАЄ на змінювання регресанта Y

$F_{\text{розрах}} \geq F_{\text{табл}}(p; k_1; k_2)$ — H_0 відхиляється та приймається H_1 : модель множинної регресії є адекватною статистичним даним

Перевірка статистичної значущості коефіцієнта множинної кореляції:

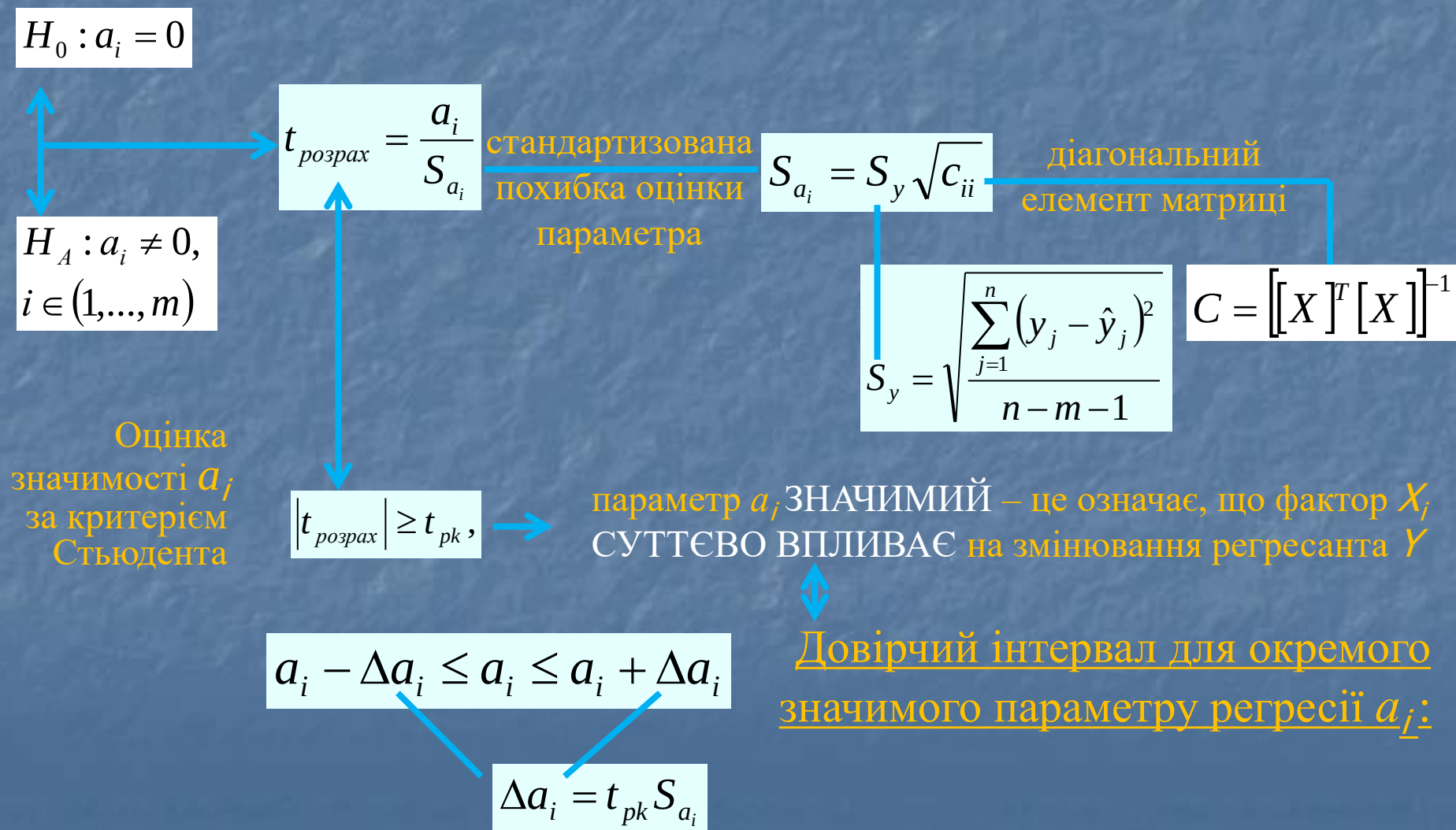
$H_0 : R = 0$ — ЗВ'ЯЗОК між залежною змінною Y (регресантом) та всіма незалежними факторами X_i ВІДСУТНІЙ

$|t_{\text{розрах}}| \geq t_{pk}$ — H_0 відхиляється та приймається H_1 : коефіцієнт множинної кореляції достовірний, тобто зв'язок між Y та $\{X_i\}$ є суттєвим

СТАТИСТИЧНІ ГІПОТЕЗИ:

ПЕРЕВІРКА ЗНАЧУЩОСТІ ТА ДОВІРЧІ ІНТЕРВАЛИ:

Перевірка статистичної значущості параметрів множинної регресії:



РОЗКЛАД КОЕФІЦІЄНТА МНОЖИННОЇ ДЕТЕРМІНАЦІЇ НА КОЕФІЦІЄНТИ ОКРЕМОЇ ДЕТЕРМІНАЦІЇ:

Для з'ясування частки впливу кожного окремого фактора X_i на показник Y використовують коефіцієнти окремої детермінації (частинні коефіцієнти детермінації):



ДОВІРЧІ ІНТЕРВАЛИ ДЛЯ БАЗИСНИХ ДАНИХ

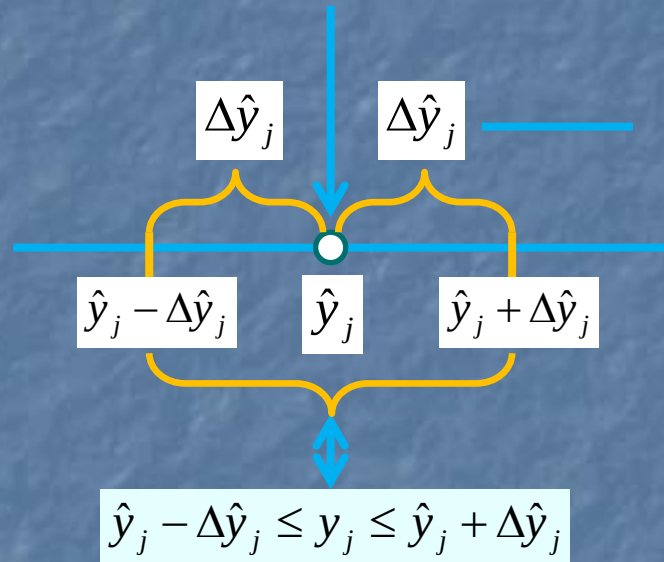
Вектор значень факторів для деякого j -го спостереження:

$$\vec{X}_j = [1; x_{j1}; \dots; x_{ji}; \dots; x_{jm}] \quad j = \overline{1, n}$$

Модельне значення регресанта:

$$\hat{Y}_j = \vec{a} \vec{X}_j = a_0 + a_1 x_{j1} + a_2 x_{j2} + \dots + a_i x_{ji} + \dots + a_m x_{jm}$$

Довірчий інтервал для базисних даних:



$$\Delta \hat{y}_j = S_y \sqrt{(\vec{X}_j [X]^T [X])^{-1} [\vec{X}_j]^T}$$

$$S_y = \sqrt{\frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{n - m - 1}}$$

Реальне значення залежної змінної Y
із заданим рівнем ймовірності p
потрапить до заданого інтервалу

$$[\vec{X}_j]^T = \begin{bmatrix} 1 \\ x_{j1} \\ \vdots \\ x_{ji} \\ \vdots \\ x_{jm} \end{bmatrix}$$

ПРОГНОЗУВАННЯ ЗА МНОЖИННОЮ ЛІНІЙНОЮ РЕГРЕСІЄЮ: ДОВІРЧІ ІНТЕРВАЛИ ДЛЯ ПРОГНОЗУ

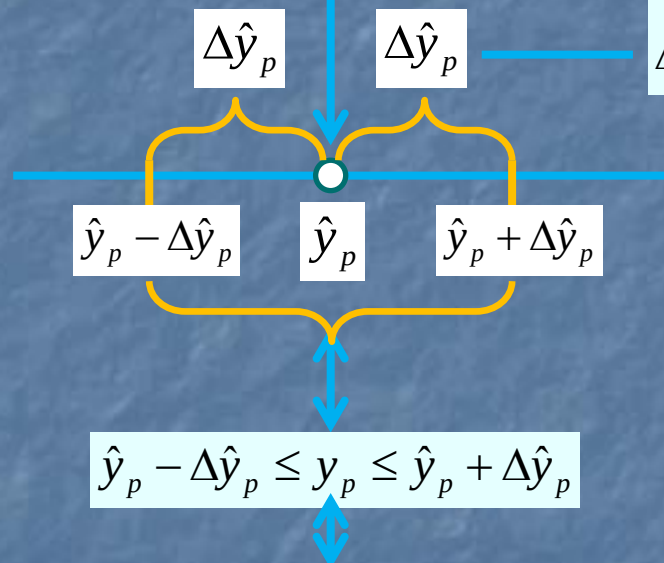
Прогнозні значення факторів:

$$\vec{X}_p = [1; x_{p1}; \dots; x_{pi}; \dots; x_{pm}]$$

Прогнозне значення регресанта:

$$\hat{Y}_p = \vec{a} \vec{X}_p = a_0 + a_1 x_{p1} + a_2 x_{p2} + \dots + a_i x_{pi} + \dots + a_m x_{pm}$$

Довірчий інтервал
для прогнозу:



$$\Delta\hat{y}_p = S_y \sqrt{(\vec{X}_p [X]^T [X]^{-1} \vec{X}_p)^T + 1}$$

$$S_y = \sqrt{\frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{n - m - 1}}$$

Прогнозне значення залежної змінної Y
із заданим рівнем ймовірності p
потрапить до заданого інтервалу

$$[\vec{X}_p]^T = \begin{bmatrix} 1 \\ x_{p1} \\ \vdots \\ x_{pi} \\ \vdots \\ x_{pm} \end{bmatrix}$$