

Тема 4.

Регрессионный анализ

Группы методов

Две группы методов: методы корреляционного анализа и методы регрессионного анализа

Корреляционный анализ

Существует ли связь между явлениями?

Насколько сильная связь между явлениями?

Регрессионный анализ

Каков характер связи между явлениями?

Построение регрессионной модели явлений.

**Парная регрессия. Метод
наименьших квадратов.**

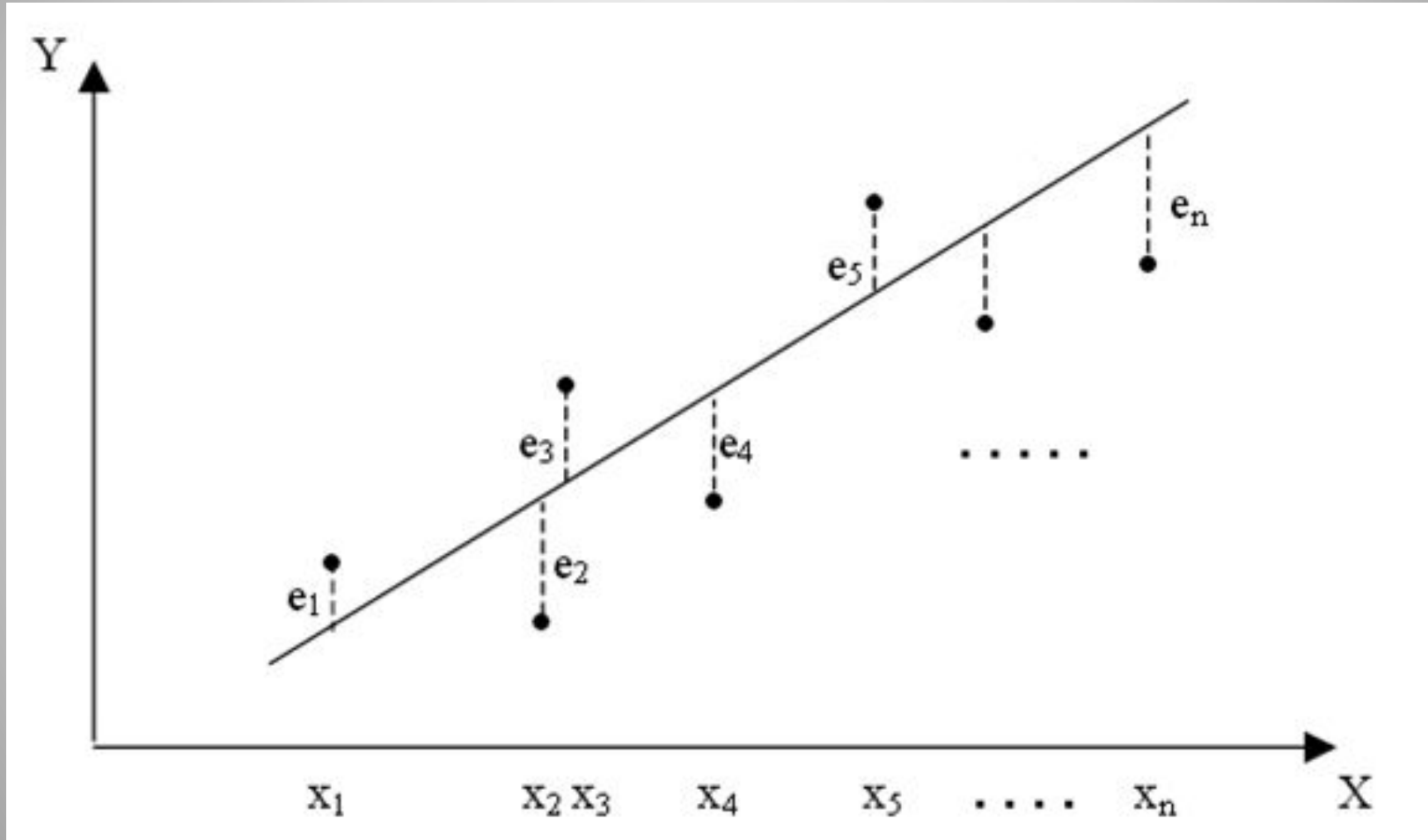
Построение уравнения регрессии

Задача построения уравнения регрессии для одного факторного и одного результативного признака формулируется следующим образом:

Пусть имеется набор значений двух переменных: результативного признака y_i и факторного признака x_i . Между этими переменными существует объективная связь вида: $y_i = f(x_i) + \varepsilon_i$.

Необходимо по данным наблюдения $y_i, x_i, i = \overline{1, n}$ подобрать функцию $\hat{y} = F(x)$, наилучшим образом описывающую существующую связь.

Графическая интерпретация



$$y_i = a + bx_i + e_i \quad i = 1, 2, \dots, n$$

Метод наименьших квадратов

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$$

Применение МНК для расчёта параметров регрессии рассмотрим на примере

$$\hat{y} = a + bx$$

$$S = \sum_{i=1}^n (y_i - a - bx_i)^2 \rightarrow \min$$

Вычисление коэффициентов

$$a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} = \bar{y} - b\bar{x}$$

$$b = \frac{\frac{\sum x \sum y}{n} - \sum xy}{\frac{(\sum x)^2}{n} - \sum x^2} \quad \text{или} \quad b = r_{xy} \frac{\sigma_y}{\sigma_x}$$

Матричное представление

$$y_i = a + b x_i + e_i, i = 1, 2, \dots, m$$

$$Y = X \beta + U$$

$$Y = \begin{pmatrix} y_1 \\ \dots \\ y_m \end{pmatrix}$$

$$U = \begin{pmatrix} u_1 \\ \dots \\ u_m \end{pmatrix}$$

$$\beta = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_m \end{pmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$