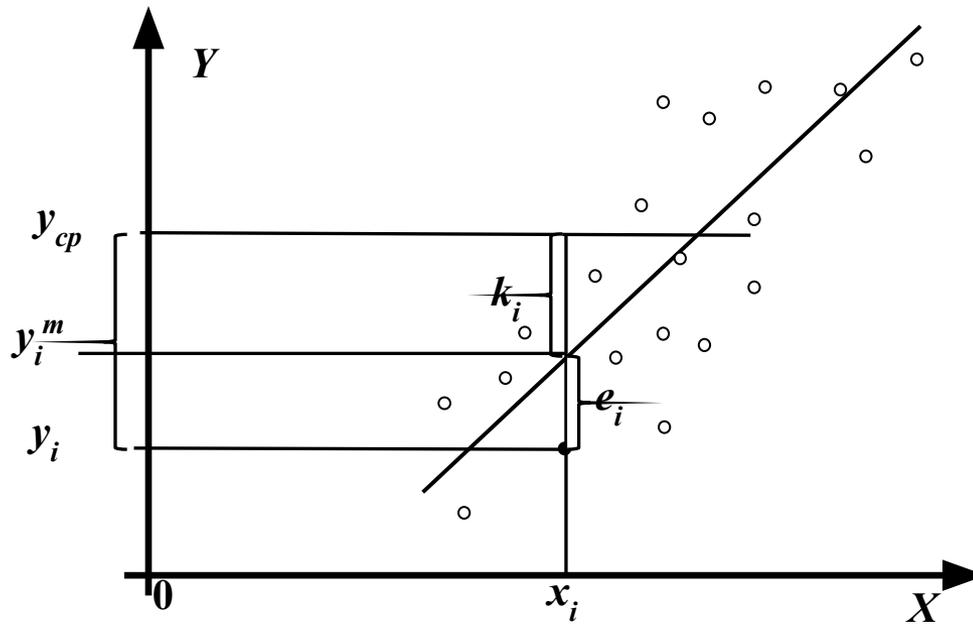


# 1. Корреляционный анализ

Корреляционный анализ для линейных моделей:



$$y_i - \bar{y} = (y_i^m - \bar{y}) + (y_i - y_i^m) = k_i + e_i$$

# 1. Корреляционный анализ

$$[(y_i - \bar{y})^2] = [k^2] + 2[k \cdot e] + [e^2]$$

$$[(y - \bar{y})^2] = [k^2] + [e^2]$$

$[(y - \bar{y})^2]$  - мера общего **разброса** переменной  $y$  относительно средней;

$[(y_i^m - \bar{y})^2] = [k^2]$  - мера **разброса** модели переменной  $y$  относительно средней (часть объясненная регрессией);

$[(y_i - y_i^m)^2] = [e^2]$  - мера остаточного (не объясненного) **разброса** переменной  $y$  относительно модели

# 1. Корреляционный анализ

Доля объясненного разброса

$$\frac{[k^2]}{[(y - \bar{y})^2]} = R^2 \quad \text{коэффициент детерминации}$$

$$1 = \frac{[k^2]}{[(y - \bar{y})^2]} + \frac{[e^2]}{[(y - \bar{y})^2]} = R^2 + \frac{[e^2]}{[(y - \bar{y})^2]}$$

$$R^2 = 1 - \frac{[e^2]}{[(y - \bar{y})^2]} \quad \begin{array}{l} \text{все минус необъясненная} \\ \text{доля} \end{array}$$

# 1. Корреляционный анализ

*Коэффициент детерминации:*

- мера в какой степени найденная линия **лучше** для объяснения поведения зависимой переменной  $Y$  чем горизонтальная прямая, проходящая через среднее значение;
- процент объясненных данных;
- совпадает с квадратом парного (множественного) коэффициента корреляции для линейной функции.

# 1. Корреляционный анализ

$$\frac{[(y_i - \bar{y})^2]}{n} = \sigma_Y^2 = \frac{[k^2]}{n} + \frac{[e^2]}{n} = \sigma_k^2 + \sigma_e^2$$

- дисперсия результирующего признака равна:  
факторная дисперсия плюс остаточная дисперсия

$$R^2 = \frac{\sigma_k^2}{\sigma_Y^2} = 1 - \frac{\sigma_e^2}{\sigma_Y^2} = I_{Y \cdot X}^2 \text{ - корреляционное соотношение}$$

$$e_i = y_i - y_i^m$$

# 1. Корреляционный анализ

$n$ -мерное линейное уравнение:

$$y_i^m = a_1 \cdot x_1 + \dots + a_{n-1} \cdot x_{n-1} + a_n$$

дисперсии:

$$\sigma_k^2 = D_f = a^T \cdot K \cdot a$$

$$\sigma_e^2 = D_e = \tilde{a}^T \cdot \tilde{K} \cdot \tilde{a}$$

Индекс корреляции (коэффициент детерминации)

$$I_{Y \cdot X}^2 = \frac{a^T K a}{\sigma_y^2} = 1 - \frac{\tilde{a}^T \tilde{K} a}{\sigma_y^2}$$

# 1. Корреляционный анализ

Другие виды коэффициентов корреляции:

- множественный

Дисперсии

$$D_e = \frac{1}{\tilde{K}_{nn}^{-1}};$$

$$D_Y = \tilde{K}_{nn}$$

Из коэффициента детерминации

$$R_{n|\dots}^2 = 1 - \frac{D_e}{D_y} = 1 - \frac{1}{\tilde{K}_{nn} \cdot \tilde{K}_{nn}^{-1}}$$

Для любого  $i$  – множественный (совокупный) коэффициент. Суть.

# 1. Корреляционный анализ

- частный (вклад в множественный каждого)

$\sigma_{y1\dots m}^2$  - факторная дисперсия при всех переменных

$\sigma_{y1\dots m-1}^2$  - факторная дисперсия без одной переменной

$\sigma_{y(1\dots m-1)}^2$  - остаточная дисперсия без одной переменной

$$R_{ym(12\dots m-1)} = \sqrt{\frac{\sigma_{y1\dots m}^2 - \sigma_{y1\dots m-1}^2}{\sigma_{y(1\dots m-1)}^2}} = \sqrt{\frac{\sigma_{y1\dots m}^2 - \sigma_{y1\dots m-1}^2}{\sigma_y^2 - \sigma_{y1\dots m-1}^2}}$$

# 1. Корреляционный анализ

Суть коэффициента.

При наличии ковариационной матрицы (матрицы моментов):

$$r_{ij|\dots} = \frac{\tilde{K}_{ij}^{-1}}{\sqrt{\tilde{K}_{ii}^{-1} \cdot \tilde{K}_{jj}^{-1}}}$$

Элементы анализа и выбора модели на основе введенных коэффициентов.

Контроль для трех рядов.

# 1. Корреляционный анализ

-коэффициент корреляции рангов Спирмена.

Необходимость. Ранжировка. Порядковый номер – *ранг*.

Качество  $A$  – *ранги*  $A_1, A_2, \dots, A_n$      $B$  –  $B_1, B_2, \dots, B_n$

разность между ними  $A_k - B_k = d_k$  - мера тесноты соответствия между качествами  $A$  и  $B$ .

Нуль - соответствие полное. По парному:

$$r_{12} = \rho = \frac{\sum (A_k - (n+1)/2) \cdot (B_k - (n+1)/2)}{\sqrt{\sum (A_k - (n+1)/2)^2 \cdot \sum (B_k - (n+1)/2)^2}} = \frac{\sum a_k \cdot b_k}{\sqrt{\sum a_k \cdot \sum b_k}}$$

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n^3 - n}$$

Свойства, использование

# 1. Корреляционный анализ

-коэффициент ассоциации (тетрафорический показатель 2 групп Д. Юла )

2 признака (например, появились – не появились) и две категории в каждом признаке (например, положительные значения – отрицательные значения). По результатам подсчетов составляется  $2 \times 2$  таблица вида

категория признак	1(+)	2(-)
A(да)	$a$	$b$
B(нет)	$c$	$d$

средний размер связи по Юлу между категориями 1 и 2 по признакам  $A$  и  $B$  будет

$$K_a = \frac{ad - cb}{ad + cb}$$