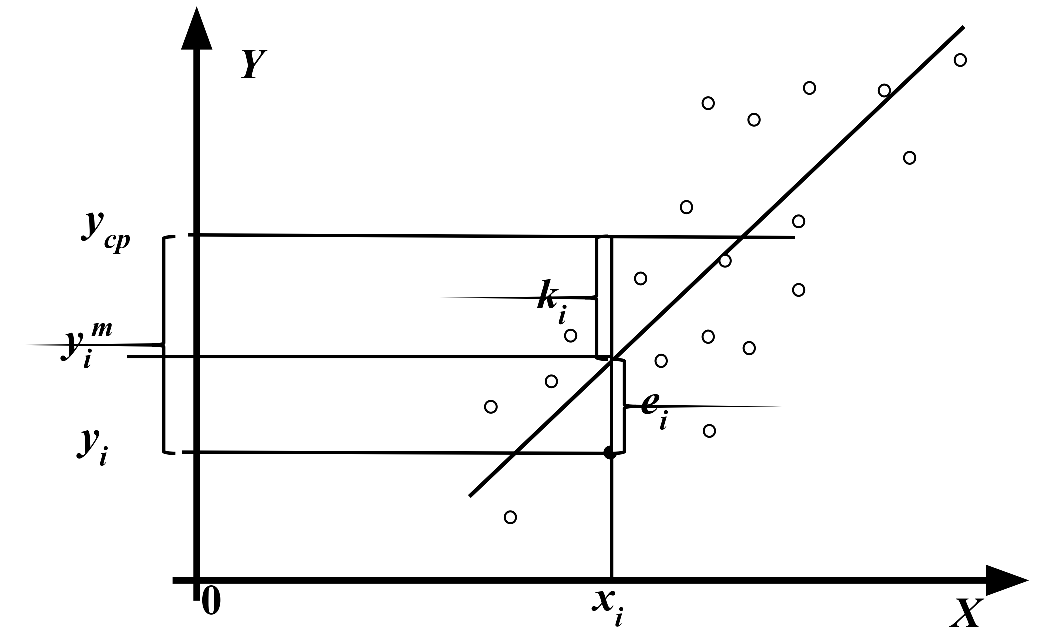
1. Корреляционный анализ

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

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

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

$\left[(y - \bar{y})^2\right]$ - 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

$$\frac{\left[k^2\right]}{\left[\left(y-\overline{y}\right)^2\right]}=R^2$$

коэффициент детерминации

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

$$R^2=1-\frac{\left[e^2\right]}{\left[\left(y-\overline{y}\right)^2\right]}$$

все минус необъясненная доля

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

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

$$e_i = y_i - y_i^m$$

1. Корреляционный анализ

n-мерное линейное уравнение:

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

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)

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

1. Корреляционный анализ

Другие виды коэффициентов корреляции:

- множественный

Дисперсии

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

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

Из коэффициента детерминации

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

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

1. Корреляционный анализ

- частный (вклад в множественный каждого)

$\sigma^2_{y1...m}$ - факторная дисперсия при всех переменных

$\sigma^2_{y1...m-1}$ - факторная дисперсия без одной переменной

$\sigma^2_{y(1...m-1)}$ - остаточная дисперсия без одной переменной

$$R_{ym(12...m-1)} = \sqrt{\frac{\sigma^2_{y1...m} - \sigma^2_{y1...m-1}}{\sigma^2_{y(1...m-1)}}} = \sqrt{\frac{\sigma^2_{y1...m} - \sigma^2_{y1...m-1}}{\sigma^2_{y} - \sigma^2_{y1...m-1}}}$$

1. Корреляционный анализ

Суть коэффициента.

При наличии ковариационной матрицы (матрицы моментов):

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

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:

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

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

Свойства, использование

1. Корреляционный анализ

-коэффициент *ассоциации* (*тетрахорический 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 2×2 таблица вида

признак \ категория	1(+)	2(–)
А(да)	a	b
В(нет)	c	d

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

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