Оценка качества моделей

Оценка качества показывает, насколько теоретические вычисления по построенной модели отклоняются от экспериментальных данных. Наличие связи двух переменных называется **корреляцией**.

Если оценка качества применяется до исследования, то она решает задачу: есть ли связь между входом X и выходом Y и оценивает силу этой связи.

1. **Линейный коэффициент корреляции**
2. **Нелинейный коэффициент корреляции**
3. **Коэффициент корреляции двух динамических рядов**
4. **Коэффициент множественной корреляции**
5. **Коэффициент ассоциаций**

Линейный коэффициент корреляции

Линейный коэффициент корреляции указывает, есть ли между двумя рядами X и Y линейная зависимость и какой силы. Вычисляется по следующей формуле:

$$\boxed{\mathrm{KR} = \frac{m_{xy} - m_x m_y}{\sigma_x \sigma_y}}$$

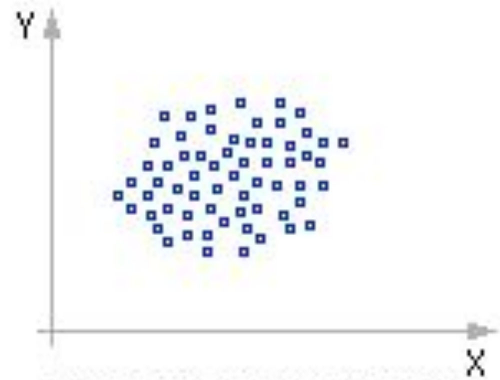
mx , my , mxy — математическое ожидание x, y, xy:

$$\boxed{m_x = \frac{1}{n}\sum_{i=1}^{n} x_i} \qquad \boxed{m_y = \frac{1}{n}\sum_{i=1}^{n} y_i} \qquad \boxed{m_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i}$$

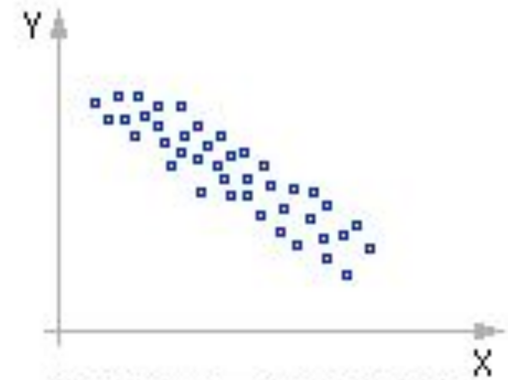
Линейный коэффициент корреляции может иметь знак плюс или минус. Положительная его величина свидетельствует о прямой связи между X и Y. Чем ближе KR к +1, тем связь более тесная. Отрицательная величина его свидетельствует об обратной связи; в этом случае границей является –1. Близость KR к нулю свидетельствует о слабой связи между X и Y

Графическая интерпретация корреляции

Сильная положительная корреляция (KR стремится к 1)



Корреляция отсутствует (KR близко к 0)

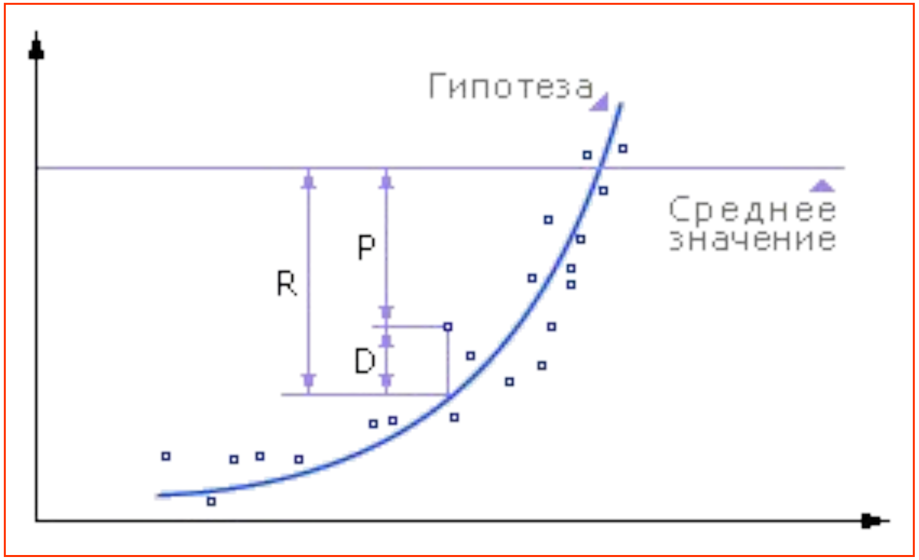


Сильная отрицательная корреляция (KR стремится к –1)

Дисперсия σx^2 и σy^2 показывает, насколько разбросаны точки от средней величины:

$$\sigma_x^2 = \sum_{i=1}^{n} \frac{(x_i - m_x)^2}{n} \qquad \sigma_y^2 = \sum_{i=1}^{n} \frac{(y_i - m_y)^2}{n}$$

Нелинейный коэффициент корреляции



$$\mathrm{KNR} = \sqrt{P - \frac{D}{P}}$$

$$P = \sum_{i=1}^{n} \frac{(y_i^{\mathrm{P.}} - y_i^{\mathrm{C.}})^2}{n}$$

P – разброс между реальными точками и средним значением

$$D = \sum_{i=1}^{n} \frac{(y_i^{g} - y_i^{c})}{n}$$

D – разброс между теоретическим значением точек и средним значением экспериментальных значений

Коэффициент корреляции двух динамических рядов

X и **Y** представляются в виде рядов **zi** и **ui** для того, чтобы исключить постоянную составляющую:

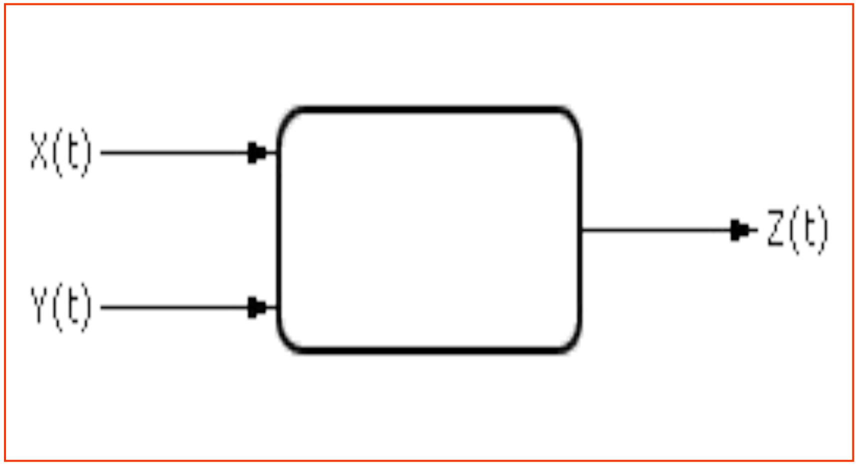
zi = xi – mx

ui = yi – my

$$m_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

$$r = \frac{\sum_{i=1}^{n} z_i u_i}{\sqrt{\sum_{i=1}^{n} z_i^2 \cdot \sum_{i=1}^{n} u_i^2}}$$



Коэффициент множественной корреляции

Коэффициент множественной корреляции R

$$R=\sqrt{\frac{r_{xz}^2+r_{xy}^2-2\,r_{xy}\cdot r_{xz}\cdot r_{yz}}{1-r_{xy}^2}}$$

$$\begin{cases} r_{xz}=\dfrac{m_{xz}-m_x m_z}{\sigma_x\,\sigma_z} \\ r_{xy}=\dfrac{m_{xy}-m_x m_y}{\sigma_x\,\sigma_y} \\ r_{yz}=\dfrac{m_{yz}-m_y m_z}{\sigma_y\,\sigma_z} \end{cases}$$

Связь двух признаков

Коэффициент ассоциаций позволяет выяснить, имеется ли какая-либо **связь между двумя признаками**. Если данный коэффициент близок к единице, то в этом случае можно говорить о существовании такой связи.

$$K = \frac{ad - bc}{ad + bc}$$

Пример. Попытаемся с помощью данной формулы выяснить, есть ли связь между ростом и весом человека? Пусть в нашем распоряжении имеются данные о весе и росте 500 человек:

	Вес < 67 кг.	**Вес > 67 кг.**
Рост < 167 см.	***a*** **= 304 чел.**	***b*** **= 17 чел.**
Рост > 167 см.	***c*** **= 112 чел.**	***d*** **= 67 чел.**

По формуле:

K = (304 · 67 – 17 · 112)/(304 · 67 + 17 · 112) = 0.83.

Так как величина 0.83 близка к 1, то можно говорить о существовании определенной связи между весом и ростом.

The end