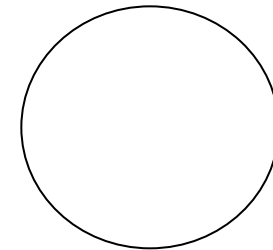
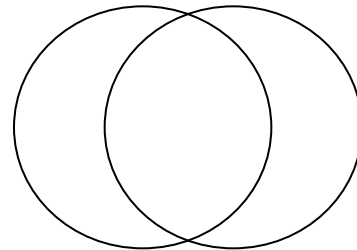
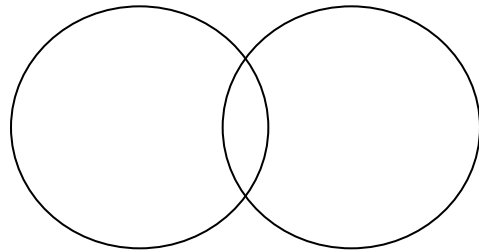


1. Статистический анализ

Процесс из нескольких рядов – связаны ряды между собой или нет?

Как с изменением одного ряда меняются другие?

Диаграммы Венна:



Задачи:

Есть ли связь, насколько значима, какой величиной охарактеризовать.

Количественная сторона - коэффициент

Качественная сторона – вид связи.

1. Статистический анализ

Из ТВ и МС теснота связи между СВ – второй смешанный центральный момент:

$$K_{XY} = MO((X - MO(X)) \cdot (Y - MO(Y))^T)$$

или ковариация.

Не удобно, разные дисперсии (масштабы)

Корреляция – масштабируемая ковариация –

Одинаковые единичные масштабы - количество.

Качество - регрессия – вид связи, как условное математическое ожидание $MO(X|y) = f(X, Y)$.

1. Статистический анализ

Отношения между несколькими величинами:

- функциональные $y = f(x, \dots)$
- стохастические (случайные, вероятностные).

Наибольшее распространение в реальных процессах - вероятностные связи.

Что отслеживают:

- изменение МО - связь *корреляционная*.
- изменение дисперсии - связь *скедастичная*.

Наиболее распространенная – корреляционная.

1. Статистический анализ

Статистические выводы и *статистический анализ* – ядро математической статистики.

Основные задачи:

- выявление в количественной мере статистической **связи** и её значимости;
- выявление в качественной мере статистической **связи** и её значимости;
- выявление главных факторов формирующих процесс.

Это *корреляционный анализ* - *регрессионный анализ*, *дисперсионный анализ* (и многие другие)

1. Корреляционный анализ

Основные задачи корреляционного анализа:

- **выявление связи** между исследуемыми переменными вообще;
- **оценивание тесноты** связей количественно;
- **оценивание значимости** связей;

Для этого:

- выбрать подходящую характеристику статистической связи;
- оценить его численное значение по имеющимся выборочным данным (точечная оценка, интервальная оценка);
- проверить гипотезу о значимости статистической связи.

1. Корреляционный анализ

1 шаг: *Выборочная ковариация*: ряды [$X Y Z \dots$]

$$\hat{K}_{XY} = \text{cov}(X, Y) = \frac{[(X - \bar{X}) \cdot (Y - \bar{Y})]}{n} = \frac{[XY] - n \cdot [X] \cdot [Y]}{n}$$

$$\hat{K}_{XZ} = \text{cov}(X, Z) = \frac{[(X - \bar{X}) \cdot (Z - \bar{Z})]}{n} = \frac{[XZ] - n \cdot [X] \cdot [Z]}{n}$$

$$\hat{K}_{YZ} = \text{cov}(Y, Z) = \frac{[(Y - \bar{Y}) \cdot (Z - \bar{Z})]}{n} = \frac{[YZ] - n \cdot [Y] \cdot [Z]}{n}$$

и т.д.

1. Корреляционный анализ

Масштабированная выборочная ковариация -
выборочный (парный) *коэффициент корреляции*

$$r_{XY} = \frac{\text{cov}(X, Y)}{m_X \cdot m_Y} = \frac{[(X - \bar{X}) \cdot (Y - \bar{Y})]}{\sqrt{[(X - \bar{X})^2] \cdot [(Y - \bar{Y})^2]}} = \frac{([XY] \cdot n - [X] \cdot [Y])}{\sqrt{(n \cdot [X^2] - [X]^2) \cdot (n \cdot [Y^2] - [Y]^2)}}$$

$$r_{XZ} = \frac{\text{cov}(X, Z)}{m_X \cdot m_Z} = \frac{[(X - \bar{X}) \cdot (Z - \bar{Z})]}{\sqrt{[(X - \bar{X})^2] \cdot [(Z - \bar{Z})^2]}} = \frac{([XZ] \cdot n - [X] \cdot [Z])}{\sqrt{(n \cdot [X^2] - [X]^2) \cdot (n \cdot [Z^2] - [Z]^2)}}$$

$$r_{YZ} = \frac{\text{cov}(Y, Z)}{m_Y \cdot m_Z} = \frac{[(Y - \bar{Y}) \cdot (Z - \bar{Z})]}{\sqrt{[(Y - \bar{Y})^2] \cdot [(Z - \bar{Z})^2]}} = \frac{([YZ] \cdot n - [Y] \cdot [Z])}{\sqrt{(n \cdot [Y^2] - [Y]^2) \cdot (n \cdot [Z^2] - [Z]^2)}}$$

1. Корреляционный анализ

Основа анализа - выборочная ковариационная матрица K (или матрица моментов λ) как совокупная характеристика рассеивания и связей рядов

$$K = \frac{1}{n} \cdot \begin{bmatrix} [(X - \bar{X})^2] & [(X - \bar{X})(Y - \bar{Y})] & \dots \\ [(X - \bar{X})(Y - \bar{Y})] & [(Y - \bar{Y})^2] & \dots \\ \dots & \dots & \dots \end{bmatrix}$$

Основной способ – матрица отклонений

$$V_{k \times n} = \begin{bmatrix} x_1 - \bar{X} & y_1 - \bar{Y} & \dots \\ x_2 - \bar{X} & y_2 - \bar{Y} & \dots \\ \dots & \dots & \dots \end{bmatrix} \longrightarrow K = \lambda = \frac{1}{n} \cdot V^T \cdot V$$

1. Корреляционный анализ

Выборочная корреляционная матрица

$$R = \begin{bmatrix} 1 & \frac{\text{cov}(X,Y)}{m_X \cdot m_Y} & \dots \\ \frac{\text{cov}(X,Y)}{m_X \cdot m_Y} & 1 & \dots \\ \dots & \dots & \dots \end{bmatrix} = \begin{bmatrix} 1 & r_{XY} & \dots \\ r_{YX} & 1 & \dots \\ \dots & \dots & \dots \end{bmatrix}$$

Получение по одному или через нормированную матрицу отклонений V_n

$$V_n = \begin{bmatrix} \frac{x_1 - X}{m_X} & \frac{y_1 - Y}{m_Y} & \dots \\ \frac{x_2 - X}{m_X} & \frac{y_2 - Y}{m_Y} & \dots \\ \dots & \dots & \dots \end{bmatrix} \longrightarrow R = \frac{1}{n} \cdot V_n^T \cdot V_n$$

1. Корреляционный анализ

Матрица плана A

$$A = \begin{bmatrix} x_1 & y_1 & \dots & 1 \\ x_2 & y_2 & \dots & 1 \\ \dots & \dots & \dots & 1 \end{bmatrix}$$

Нормальная матрица

$$N = A^T \cdot A$$

Её развернутый вид . Использование.

$$N = \begin{bmatrix} [X^2] & [XY] \dots & [X] \\ [XY] & [Y^2] \dots & [Y] \\ [X] & [Y] \dots & n \end{bmatrix}$$