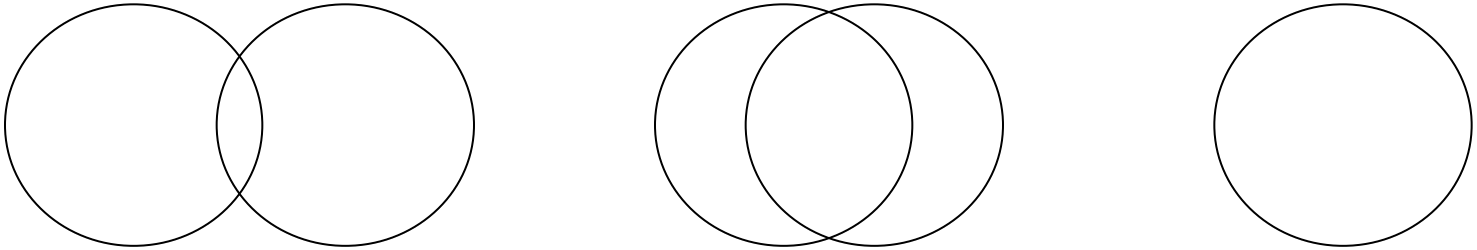
1. Статистический анализ

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 $\boldsymbol{MO(X|y) = f(X, Y)}$.

1. Статистический анализ

Отношения между несколькими величинами:

– *функциональные* $\boldsymbol{y = f(x, \ldots)}$

– *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*** …]

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

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

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

и т.д.

1. Корреляционный анализ

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

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

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

$$r_{XZ} = \frac{\operatorname{cov}(Y,Z)}{m_X \cdot m_Z} = \frac{\left[(Y-\overline{Y})\cdot(Z-\overline{Z})\right]}{\sqrt{\left[(Y-\overline{Y})^2\right]\cdot\left[(Z-\overline{Z})^2\right]}} = \frac{\left([YZ]\cdot n-[Y]\cdot[Z]\right)}{\sqrt{\left(n\cdot[Y^2]-[Y]^2\right)\cdot\left(n\cdot[Z^2]-[Z]^2\right)}}$$

1. Корреляционный анализ

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

$$K = \frac{1}{n} \cdot \begin{bmatrix} \left[(X-\overline{X})^2\right] & \left[(X-\overline{X})(Y-\overline{Y})\right] & \ldots \\ \left[(X-\overline{X})(Y-\overline{Y})\right] & \left[(Y-\overline{Y})^2\right] & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}$$

Основной способ – матрица отклонений

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

1. Корреляционный анализ

Выборочная корреляционная матрица

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

Получение по одному или через нормированную матрицу отклонений $\boldsymbol{V_n}$

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

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 развертнутый вид . Использование.

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