

4. Статистический анализ

Уравнение регрессии для многомерной однооткликковой линейной модели $y = f(x_1, x_2, \dots, x_n)$, \rightarrow вычисление основных корреляционных характеристик разными способами:

- множественный коэффициент корреляции
- частный коэффициент корреляции.

Основа вычислений:

- теорема о характеристиках многомерного условного закона распределения
- разложение общей дисперсии на составляющие
- использование формулы для парного коэффициента корреляции после определенного преобразования данных.

4. Статистический анализ

Теорема о характеристиках многомерного условного закона распределения для 1-отклика: имеется выборочная ковариационная матрица

$$K_0 = \begin{pmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{pmatrix} = \begin{pmatrix} K_{xx} & k_{xy} \\ k_{yx} & k_y \end{pmatrix}$$

для процесса с факторными переменными $X = (x_1, x_2, \dots, x_k)$ и результирующей переменной y на последнем месте ($X \ y$). Закон распределения процесса описывается условным многомерным нормальным законом распределения вероятностей характеристиками:

4. Статистический анализ

– условным математическим ожиданием (линейной формой множественной регрессии)

$$MO(y | x_1, \dots, x_n) = \overset{\boxtimes}{\bar{y}} = \bar{y} + k_{yx} \cdot K_{xx}^{-1} (x - \bar{x}) = \bar{y} + a' (x - \bar{x})$$

или в нормальном виде

$$\overset{\boxtimes}{\bar{y}} = \bar{y} + a' (x - \bar{x}) = a' \cdot x + (\bar{y} - a\bar{x}) = a' \cdot x + d$$

– условной дисперсией (дисперсией модели)

$$D(y | x_1, \dots, x_n) = \overset{\boxtimes}{\sigma_0^2} = k_y - k_{yx} \cdot K_{xx}^{-1} \cdot k_{xy}.$$

или

$$\overset{\boxtimes}{\sigma_0^2} = D_y (1 - r_{y|\dots}^2)$$

4. Статистический анализ

Общую дисперсию $D_{об}$ результирующей переменной y при многомерной 1-откликковой регрессии разлагают на факторную $D_{ф}$ (объясненную) и остаточную $D_{ос}$ (необъясненную) дисперсии

$$\begin{aligned} s_0 &= y - \bar{y} \\ [(y - \bar{y})^2] &= [(y - \hat{y})^2] + [(\hat{y} - \bar{y})^2] = \\ &= S_{об} = S_{ос} + S_{ф} \end{aligned}$$

$$D_{об} = D_{ос} + D_{ф}$$

Теорем: отношение объясненной части дисперсии $D_{ф}$ к общей части $D_{об}$ есть коэффициент детерминации, или квадрат множественного коэффициента корреляции

$$R_y^2 = r_{y|\dots}^2 = \frac{D_{\hat{o}}}{D_{\hat{a}}} = 1 - \frac{D_{\hat{n}}}{D_{\hat{a}}}$$

4. Статистический анализ

Так как, $\sigma_0^2 = D_{oc}$, $D_{об} = D$ имеем

$$\sigma_0^2 = D_{ост} (1 - r_{y|...}^2) = D_y = D (1 - r_{y|...}^2)$$

Сопоставив $\sigma_0^2 = D_{ост} = k_y - k_{yx} \cdot K_{xx}^{-1} \cdot k_{xy}$ и $D_{об} = D_{oc} + D_{\phi}$

имеем

$$D_{\hat{a}} = D_y = k_y$$

$$D_{\hat{\delta}} = k_{yX} \cdot K_{XX}^{-1} \cdot k_{Xy} = a' \cdot k_{Xy}$$

откуда из

$$R_y^2 = r_{y|...}^2 = \frac{D_{\phi}}{D_{об}} = 1 - \frac{D_{oc}}{D_{об}}$$

простой подстановкой получаем ряд формул для множественного коэффициента корреляции:

$$1. \quad D_{об} = k_y, \quad D_{ост} = \frac{1}{c_y} \rightarrow r_{y|...} = \sqrt{1 - \frac{1}{k_y \cdot c_y}}$$

4. Статистический анализ

2. Из $D_{\hat{a}} = D_y = k_y$

$$D_{\hat{o}} = k_{yX} \cdot K_{XX}^{-1} \cdot k_{Xy}$$

и $r_{y|\dots}^2 = \frac{D_{\hat{o}}}{D_{\hat{a}}}$, имеем

$$r_{y|\dots} = \sqrt{\frac{k_{yX} \cdot K_{XX}^{-1} \cdot k_{Xy}}{k_y}}$$

3. Из $D_{o\hat{o}} = D_y = k_y$, имеем

$$D_{\phi} = a' \cdot k_{Xy}$$

$$r_{y|\dots} = \sqrt{\frac{a' \cdot k_{Xy}}{k_y}}$$

4. Статистический анализ

4. Из $D_{o\hat{o}} = D_y = k_y$

$$D_{\phi} = k_{yX} \cdot K_{XX}^{-1} \cdot k_{Xy} = a' \cdot k_{Xy} = a' \cdot K_{XX} \cdot K_{XX}^{-1} k_{Xy} = a' \cdot K_{XX} \cdot a'^T$$

и $r_{y|\dots}^2 = \frac{D_{\hat{o}}}{D_{\hat{i}\hat{a}}}$, имеем

$$r_{y|\dots} = \sqrt{\frac{a' \cdot K_{XX} \cdot a'^T}{k_y}}$$

5. Домножив числитель и знаменатель п.4 на числитель получим

$$r_{y|\dots} = \frac{a' \cdot K_{XX} \cdot a'^T}{\sqrt{k_y \cdot a' \cdot K_{XX} \cdot a'^T}}$$

Отсюда - множественный коэффициент корреляции между результирующим признаком y и матрицей факторных переменных X - парный коэффициент корреляции Пирсона между вектором y и его модельным значением \hat{y} .

4. Статистический анализ

Учитывая связь между ковариационной матрицы K и корреляционной матрицы R , ковариации и корреляции

$$K = D \cdot R \cdot D; \quad D = \text{diag}(K); \quad R = D^{-1} \cdot K \cdot D^{-1}$$

$$\text{cov}(x, y) = k_{xy} = r_{xy} \cdot \sigma_x \cdot \sigma_y$$

из второй формулы после перехода от ковариационной матрицы K к корреляционной R имеем

$$r_{y|\dots} = \sqrt{r_{\cdot y} \cdot R_{\hat{o}}^{-1} \cdot r_{\cdot y}^T}$$

с $r_{\cdot y}$ – вектор-строкой из парных коэффициентов корреляции между y и всеми x_i , $R_{\hat{o}}$ – укороченной корреляционной матрицей без результирующего признака.

Выразив в основных формулах ковариации через корреляции, коэффициенты уравнения регрессии, их комбинации и т.д. можно получить и ряд других полезных формул.

4. Статистический анализ

При вычислении частных коэффициентов корреляции:

- условная дисперсия - преобразованная дисперсия результирующего признака D_y ,
- получена путем учета и *исключения* влияния на результирующий признак всех факторных переменных.
- для исключения и учёта в исследуемом процессе влияния на два любых вектора всех остальных, в ковариационной матрице они должны быть вместе в конце (или в начале),
- преобразованная часть ковариационной матрицы, в которой исключено и учтено влияние на 2 последних вектора всех остальных будет

$$K_{22 \cdot 1} = K_{22} - K_{21} \cdot K_{11}^{-1} \cdot K_{12} = D(Y | X)$$

4. Статистический анализ

Теорема: частный коэффициент корреляции между 2 векторами есть парный коэффициент корреляции из преобразованной части

$$K_{22 \cdot 1} = K_{22} - K_{21} \cdot K_{11}^{-1} \cdot K_{12}$$

при определенных выше условиях.

Из теоремы легко следует известная формула

$$r_{ij|\dots} = \frac{C_{ij}}{\sqrt{C_{ii} \cdot C_{jj}}}$$

через элементы обратной ковариационной матрицы

$$C = K^{-1} = \Lambda$$

4. Статистический анализ

Вычисления на основе моделирования.

1. Результирующая переменная z смоделирована через факторную переменную x в виде $z_1 = f(x)$ с остаточной дисперсией D_1 . Добавив в модель новую факторную переменную y имеем новую модель $z_2 = f(x, y)$ с остаточной дисперсией D_2 .

Теорема: частный коэффициент корреляции между z и y при исключении влияния переменной x будет

$$r_{zy|x} = \sqrt{\frac{D_1 - D_2}{D_1}} = \sqrt{1 - \frac{D_2}{D_1}}$$

т.е. равен относительному изменению дисперсии при добавлении новой переменной в модель.

4. Статистический анализ

Вычисления на основе моделирования.

2. При определении для векторов x , y , z частного коэффициента корреляции между z и y при исключении влияния переменной x :

- моделируем парной линейной регрессией ряды $y = f(x)$ и $z = f(x)$.

- получаем два вектора поправок (остатков) v_1 и v_2 от моделирования парной линейной регрессией ряда $\hat{y} = f(x)$ и ряда $z = f(x)$.

- парный коэффициент корреляции между рядами остатков v_1 и v_2 есть частный коэффициент корреляции между z и y при исключении влияния переменной x .

4. Статистический анализ

Трансформация систем координат как задача регрессии

Формулировка задачи:

Есть координаты (x, y) для n точек в старой системе координат K_c и есть координаты (X, Y) для этих же точек в новой системе K_n . Используя линейное преобразование старой системы в новую, получить коэффициенты перехода (трансформации) старой системы в новую, отвечающие заданному критерию качества.

4. Статистический анализ

Алгебраический вид линейной трансформации для 1 точки

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} a \cdot x_i + b \cdot y_i + c \\ d \cdot x_i + e \cdot y_i + f \end{bmatrix}$$

Матричный вид для n точек с отдельным сдвигом

$$\begin{bmatrix} X_1 & X_2 & \dots & X_n \\ Y_1 & Y_2 & \dots & Y_n \end{bmatrix}_n = \begin{bmatrix} a & b \\ d & e \end{bmatrix} \times \begin{bmatrix} X_1 & X_2 & \dots & X_n \\ Y_1 & Y_2 & \dots & Y_n \end{bmatrix}_c + \begin{bmatrix} c \\ f \end{bmatrix} =$$

$$= K_n^T = X \cdot K_c^T + d$$

а матричный вид для n точек с внутренним сдвигом

$$\begin{bmatrix} X_1 & X_2 & \dots & X_n \\ Y_1 & Y_2 & \dots & Y_n \end{bmatrix}_n = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \times \begin{bmatrix} X_1 & X_2 & \dots & X_n \\ Y_1 & Y_2 & \dots & Y_n \\ 1 & 1 & \dots & 1 \end{bmatrix}_c =$$

$$= K_n^T \cdot K' \cdot c^T$$

4. Статистический анализ

Имеем классическую модель 2-факторной регрессии с 2-мерным откликом. Моделируется K_n , по стандартному МНК-модель

$$K_n^T \cdot KV = c \cdot v^T$$

-уравнения поправок (V – матрица)

$$K = K \cdot c \cdot K_n^T - \begin{matrix} T \\ n \end{matrix}$$

-целевая функция

$$\Phi = v'^T \cdot v' = \text{Trace}(V^T \cdot V) = \min$$

где $v' = \text{vec}(V^T)$

4. Статистический анализ

Минимизация Φ приводит к правосторонней трансформации Гаусса

$$X = K' \cdot K_c^T - \bar{V} \rightarrow X' \cdot X_c = K' \cdot K_c'^T \cdot K_c' - K_n^T \cdot c'$$

или на основе обобщенной (матричной) леммы Гаусса

$$X' \cdot \underbrace{(K_c'^T \cdot K_c')}_{N} = \underbrace{(K_n^T \cdot K_c')}_{D}$$
$$X' \cdot N = D,$$

Решение на основе обращения матрицы $N^{-1} = Q$ - полная матрица оценок коэффициентов преобразования X'

$$X' = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} = D \cdot Q$$

4. Статистический анализ

Модельные значения

$$\hat{K}_n^T = K' \cdot c^T$$

Поправки

$$K = \hat{K}_n^T - K_n^T = K' \cdot K_c^T - Q_n^T = E_n^T \cdot \left(c^T - \right)$$

Тогда оценка точности модели

$$\hat{\sigma}_0 = \sqrt{\frac{\Phi}{2n - 6}}$$

а коэффициентов

$$K_{X'} = \hat{\sigma}_0^2 \cdot Q'$$

с

$$Q' = E \otimes (X'^T \cdot X')^{-1} = E \otimes Q$$

4. Статистический анализ

Имеем Q размера 3×3 и из

$$X' = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} = D \cdot Q$$

Получаем что погрешности для столбцов одинаковы

$$a = d, b = e, c = f$$

Поэтому считают только Q , учитывают правило и

$$K_{X'} = \sigma_0^2 \cdot Q$$

4. Статистический анализ

При решении задачи трансформации на основе условного математического ожидания для 2-факторной регрессии с 2-мерным откликом по общей теореме о характеристиках многомерного условного нормального закона распределения имеем для расчета в девиатах общую матрицу плана

$$K = \left(K_c \mid K_i \right) = \begin{pmatrix} X_{c1} & Y_{\hat{n}1} \mid & X_{i1} & Y_{i1} \\ \boxtimes & \boxtimes \mid & \boxtimes & \boxtimes \end{pmatrix}$$

общую блочную ковариационную матрицу

$$S = C^T \cdot C = \begin{pmatrix} C_c^T C_c & C_c^T C_H \\ C_H^T C_c & C_H^T C_H \end{pmatrix} = \begin{pmatrix} S_{cc} & S_{cH} \\ S_{Hc} & S_{HH} \end{pmatrix}$$

с центрированными блоками C

4. Статистический анализ

условное математическое ожидание

$$MO(K_n | K_c) = \bar{K}_n = \bar{S}_n + S_{nc} \cdot K_{cc}^{-1} (K_c - \bar{K}_c) = \bar{X}_n \hat{K} (K_c - \bar{K}_c)$$

или

$$\bar{K}_n = \bar{X} \cdot K_c + (\bar{K}_n - \bar{X} \cdot \bar{K}_c) = \bar{X} \cdot K_c + D$$

с искомыми коэффициентами трансформации

$$\begin{cases} \bar{X} = S_{nc} \cdot S_{cc}^{-1} \\ D = \bar{K}_n - \bar{X} \cdot \bar{K}_c \end{cases}$$

4. Статистический анализ

Условная ковариационная матрица

$$C_{K_n|K_c} = S_{nn} - S_{nc} \cdot S_{cc}^{-1} \cdot S_{cn}$$

след которой равен целевой функции Φ . Оценка модели

$$\hat{\sigma}_0 = \sqrt{\frac{\Phi}{2n - 6}}$$

Оценка коэффициентов сложна – минус для метода.

4. Статистический анализ

Контрольные вопросы по модулю 4

1. Многомерная многооткликковая регрессия – основные положения
2. Многомерная однооткликковая регрессия – оценка Гаусса (МНК)
3. Многомерная однооткликковая регрессия – оценка Байеса (метод средних)
4. Многомерная многооткликковая регрессия – матричный МНК
5. Получение корреляций на основе регрессии
6. Задача трансформации и методы её решения