4. Статистический анализ

Уравнение регрессии для многомерной однооткликовой 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 $X = (x_1, x_2, \ldots, x_k)$ и результирующей переменной y на последнем месте $(X \quad y)$. Закон распределения процесса описывается условным многомерным нормальным законом распределения вероятностей характеристиками:

4. Статистический анализ

– условным математическим ожиданием (линейной формой множественной регрессии)

$$MO\left(y \mid x_1,\ldots,x_n\right) = \hat{y} = \overline{y} + k_{yx} \cdot K_{xx}^{-1}\left(x - \overline{x}\right) = \overline{y} + a'\left(x - \overline{x}\right)$$

или в нормальном виде

$$\hat{y} = \overline{y} + a'\left(x - \overline{x}\right) = a' \cdot x + \left(\overline{y} - a\overline{x}\right) = a' \cdot x + d$$

– условной дисперсией (дисперсией модели)

$$D\left(y \mid x_1,\ldots,x_n\right) = \hat{\sigma}_0^2 = k_y - k_{yx} \cdot K_{xx}^{-1} \cdot k_{xy}.$$

или

$$\hat{\sigma}_0^2 = D_y\left(1 - r_{y|\ldots}^2\right)$$

4. Статистический анализ

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

$$s_0 = y - \overline{y}$$

$$[(y-\overline{y})^2] = [(y-\hat{y})^2] + [(\hat{y}-\overline{y})^2] =$$

$$= S_{об} = S_{ос} + S_{ф}$$

$$D_{об} = D_{ос} + D_{ф}$$

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

$$R_y^2 = r_{y|\ldots}^2 = \frac{D_{ф}}{D_{об}} = 1 - \frac{D_{ос}}{D_{об}}$$

4. Статистический анализ

Так как, $\hat{\sigma}_0^2 = D_{ос}$, $D_{об} = D_y$ имеем

$$\hat{\sigma}_0^2 = D_{ост}\left(1 - r^2_{y|общ}\right) = D_y = D\left(1 - r^2_{|\ldots}\right)$$

Сопоставив $\hat{\sigma}_0^2 = D_{ост} = k_y - k_{yx} \cdot K_{xx}^{-1} \cdot k_{xy}$. и $D_{об} = D_{ос} + D_{ф}$ имеем

$$D_{\text{îá}} = D_y = k_y$$

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

откуда из

$$R_y^2 = r^2_{y|\ldots} = \frac{D_{ф}}{D_{об}} = 1 - \frac{D_{ос}}{D_{об}}$$

простой подстановкой получаем ряд формул для множественного коэффициента корреляции:

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

4. Статистический анализ

2. Из $D_{\text{îá}} = D_y = k_y$

$$D_{\text{ô}} = k_{yX} \cdot K_{XX}^{-1} \cdot k_{Xy}$$

и $r_{y|\ldots}^2 = \dfrac{D_{\text{ô}}}{D_{\text{îá}}}$, имеем

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

3. Из $D_{об} = D_y = k_y$, имеем

$$D_{ф} = a' \cdot k_{Xy}$$

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

4. Статистический анализ

4. Из $$D_{об} = D_y = k_y$$

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

и $r_{y|\ldots}^2 = \dfrac{D_{\hat{o}}}{D_{\hat{\imath}\,\acute{a}}}$, имеем

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

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

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

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

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

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

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

$$r_{y|\ldots} = \sqrt{r_{\cdot y} \cdot R_{ó}^{-1} \cdot r_{\cdot y}^{\mathrm{T}}}$$

с r_y – вектор-строкой из парных коэффициентов корреляции между y и всеми x_i, R_y – 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 *Dy*,

- 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

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

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

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

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 $y = f(x)$ 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 строй системе координат K_c и есть координаты (X, Y) для этих же точек в новой системе $K_н$. 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

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

$$= K_{н}^{\mathrm{T}} = X \cdot K_{c}^{\mathrm{T}} + d$$

а матричный вид для n точек с внутренним сдвигом

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

$$= K_{н}^{\mathrm{T}} = K\ ' \cdot\ '^{\mathrm{T}}_{c}$$

4. Статистический анализ

Имеем классическую модель 2-факторной регресии с 2-мерным откликом. Моделируется $K_н$, по стандартному МНК

-модель

$$K_{н}^{\mathrm{T}} + KV = \; ' \cdot \; '^{\mathrm{T}}_{c}$$

-уравнения поправок (V – матрица)

$$V = K \; ' \cdot \; K_{c}'^{\mathrm{T}} - \; ^{\mathrm{T}}_{н}$$

-целевая функция

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

где $v' = vec(V^{T})$

4. Статистический анализ

Минимизация Ф приводит к правосторонней трансформации Гаусса

$$V = X' \cdot K_c'^{\mathrm{T}} - K_н^{\mathrm{T}} \rightarrow X' \cdot K_c'^{\mathrm{T}} \cdot K_c' = K_н^{\mathrm{T}} \cdot K_c'$$

или на основе обобщенной (матричной) леммы Гаусса

$$X' \cdot K_c'^{\mathrm{T}} \cdot K_c' = K_н^{\mathrm{T}} \cdot K_c'$$

$$X' \cdot N = D,$$

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

$$\hat{K}_{н}^{\mathrm{T}} = K' \cdot \; {}'^{\mathrm{T}}_{c}$$

Поправки

$$V = \hat{K}_{н}^{\mathrm{T}} - K_{н}^{\mathrm{T}} = K' \cdot K_{c}'^{\mathrm{T}} - Q_{н}^{\mathrm{T}} = E_{н}^{\mathrm{T}} \cdot \left(\; {}'^{\mathrm{T}}_{c} \cdot \; - \; \right)$$

Тогда оценка точности модели

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

а коэффициентов

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

с

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

4. Статистический анализ

Имеем Q размера 3x3 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

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

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

$$\underset{n\times 4}{K}=\left(K_c \middle|\ K_{í}\right)=\left(\begin{array}{cc|cc} X_{c1} & Y_{ñ1} & X_{í1} & Y_{í1} \\ \vdots & \vdots & \vdots & \vdots \end{array}\right)$$

общую блочную ковариационную матрицу

$$S=C^{\mathrm{T}}\cdot C=\begin{pmatrix} C_c^{\mathrm{T}}C_c & C_c^{\mathrm{T}}C \\ C_н^{\mathrm{T}}C_c & C_н^{\mathrm{T}}C_н \end{pmatrix}=\begin{pmatrix} S_{cc} & S_{cн} \\ S_{нc} & S_{нн} \end{pmatrix}$$

с центрированными блоками C

4. Статистический анализ

условное математическое ожидание

$$M\left(K_н \mid K_с\right) = \hat{K}_н = \bar{S}_н + S_{нс} \cdot K_{сс}^{-1}\left(K_с - \bar{K}_с\right) = \bar{X}_н + \hat{K}\left(K_с - \bar{\ }_с\right)$$

или

$$\hat{K}_н = \hat{X} \cdot K_с + \left(\bar{K}_н - X \cdot \bar{K}_с\right) = \hat{X} \cdot K_с + D$$

с искомыми коэффициентами трансформации

$$\begin{cases} \hat{X} = S_{нс} \cdot S_{сс}^{-1} \\ D = \bar{K}_н - \ \cdot \bar{\ }_с \end{cases}$$

4. Статистический анализ

Условная ковариационная матрица

$$C_{K_н|K_с} = S_{нн} - S_{нс} \cdot S_{сс}^{-1} \cdot S_{сн}$$

след которой равен целевой функции Ф. Оценка модели

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

Оценка коэффициентов сложна – минус для метода.

4. Статистический анализ

Контрольные вопросы по модулю 4

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