统计方法与机器学习作业2

10215501435 杨茜雅

Q_1:

原始数据 x_i, $i=1、2、3\cdots n$ $\qquad \bar{x}=\frac{\sum_{i=1}^{n} x_i}{n} \qquad \bar{y}=\frac{\sum_{i=1}^{n} y_i}{n}$

y_i, $i=1、2、3\cdots n$

变换后 $$\bar{\tilde{x}}=\frac{\sum_{i=1}^{n}\tilde{x}_i}{n}=\frac{\sum_{i=1}^{n}\left(\frac{x_i-C_2}{d_2}\right)}{n}=\frac{\bar{x}-C_2}{d_2}$$

$$\bar{\tilde{y}}=\frac{\sum_{i=1}^{n}\tilde{y}_i}{n}=\frac{\sum_{i=1}^{n}\left(\frac{y_i-C_1}{d_1}\right)}{n}=\frac{\bar{y}-C_1}{d_1}$$

$$l_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \qquad l_{\tilde{x}\tilde{y}}=\sum_{i=1}^{n}(\tilde{x}_i-\bar{\tilde{x}})(\tilde{y}_i-\bar{\tilde{y}})$$

$$=\sum_{i=1}^{n}\left(\frac{x_i-C_2}{d_2}-\frac{\bar{x}-C_2}{d_2}\right)\left(\frac{y_i-C_1}{d_1}-\frac{\bar{y}-C_1}{d_1}\right)$$

$$=\sum_{i=1}^{n}\frac{1}{d_1d_2}(x_i-\bar{x})(y_i-\bar{y})$$

$$=\frac{1}{d_1d_2}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{d_1d_2}l_{xy}$$

$$l_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad l_{\tilde{x}\tilde{x}}=\sum_{i=1}^{n}(\tilde{x}_i-\bar{\tilde{x}})^2=\sum_{i=1}^{n}\left(\frac{x_i-C_2}{d_2}-\frac{\bar{x}-C_2}{d_2}\right)^2=\frac{1}{d_2^2}l_{xx}$$

$$l_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2 \qquad l_{\tilde{y}\tilde{y}}=\sum_{i=1}^{n}(\tilde{y}_i-\bar{\tilde{y}})^2=\sum_{i=1}^{n}\left(\frac{y_i-C_1}{d_1}-\frac{\bar{y}-C_1}{d_1}\right)^2=\frac{1}{d_1^2}l_{yy}$$

$$\frac{l_{xy}}{l_{xx}}=\hat{\beta}_1 \qquad \tilde{\beta}_1=\frac{l_{\tilde{x}\tilde{y}}}{l_{\tilde{x}\tilde{x}}}=\frac{\frac{1}{d_1d_2}l_{xy}}{\frac{1}{d_2^2}l_{xx}}=\frac{d_2}{d_1}\frac{l_{xy}}{l_{xx}}=\frac{d_2}{d_1}\hat{\beta}_1$$

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x} \qquad \tilde{\beta}_0=\bar{\tilde{y}}-\tilde{\beta}_1\bar{\tilde{x}}=\frac{\bar{y}-C_1}{d_1}-\frac{\cancel{d_2}}{d_1}\hat{\beta}_1\left(\frac{\bar{x}-C_2}{\cancel{d_2}}\right)=\frac{\bar{y}-C_1-\hat{\beta}_1\bar{x}+\hat{\beta}_1C_2}{d_1}$$

$$=\frac{1}{d_1}(\hat{\beta}_0+\hat{\beta}_1C_2-C_1)$$

∴ 变换后数据的最小二乘估计和原始数据的最小二乘估计关系是 $\begin{cases}\tilde{\beta}_0=\frac{1}{d_1}(\hat{\beta}_0+C_2\hat{\beta}_1-C_2)\\ \tilde{\beta}_1=\frac{d_2}{d_1}\hat{\beta}_1\end{cases}$

总偏差平方和

$$S_T=\sum_{i=1}^{n}(y_i-\bar{y})^2=l_{yy} \qquad \tilde{S}_T=\sum_{i=1}^{n}(\tilde{y}_i-\bar{\tilde{y}})^2=\sum_{i=1}^{n}\left(\frac{y_i-\cancel{c_1}}{d_1}-\frac{\bar{y}-\cancel{c_1}}{d_1}\right)^2$$

$$=\frac{1}{d_1^2}\sum_{i=1}^{n}(y_i-\bar{y})^2=\frac{1}{d_1^2}S_T$$

回归平方和

$$S_R=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\sum_{i=1}^{n}(\cancel{\bar{y}}+\hat{\beta}_1(x_i-\bar{x})\cancel{-\bar{y}})^2=\hat{\beta}_1^2\sum_{i=1}^{n}(x_i-\bar{x})^2=\hat{\beta}_1^2\,l_{xx}$$

$$\tilde{S}_R=\sum_{i=1}^{n}(\hat{\tilde{y}}_i-\bar{\tilde{y}})^2=\sum_{i=1}^{n}(\quad+\hat{\tilde{\beta}}_1(\tilde{x}_i-\bar{\tilde{x}}))^2=\hat{\tilde{\beta}}_1^2\sum_{i=1}^{n}(\tilde{x}_i-\bar{\tilde{x}})^2=\hat{\tilde{\beta}}_1^2\,l_{\tilde{x}\tilde{x}}=\hat{\tilde{\beta}}_1^2\frac{1}{d_2^2}l_{xx}$$

$$=\frac{\cancel{d_2^2}}{d_1^2}\,\frac{1}{\cancel{d_2^2}}\,\cancel{\hat{\beta}_1^2}\,\frac{S_R}{\cancel{\hat{\beta}_1^2}}=\frac{S_R}{d_1^2}$$

残差平方和：

$$S_e=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}((y_i-\bar{y})-(\hat{y}_i-\bar{y}))^2=S_T-S_R$$

$$\tilde{S}_e=\sum_{i=1}^{n}(\tilde{y}_i-\hat{\tilde{y}}_i)^2=\sum_{i=1}^{n}\left(\frac{y_i-\cancel{c_1}}{d_1}-\frac{\hat{y}_i-\cancel{c_1}}{d_1}\right)^2=\frac{1}{d_1^2}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\frac{1}{d_1^2}S_e$$

$$\therefore\quad \tilde{S}_T=\frac{1}{d_1^2}S_T \qquad \tilde{S}_R=\frac{1}{d_1^2}S_R \qquad \tilde{S}_e=\frac{1}{d_1^2}S_e$$

(2)

原始 $F=\dfrac{S_R}{\frac{S_e}{n-2}}=(n-2)\dfrac{S_R}{S_e}=(n-2)\dfrac{\cancel{d_1^2}\,\tilde{S}_R}{\cancel{d_1^2}\,\tilde{S}_e}=\dfrac{\tilde{S}_R}{\frac{\tilde{S}_e}{n-2}}=\tilde{F}$

∴ F统计量的值保持不变

Q_2:

$$\begin{cases} \hat{y} = a+bx \quad b = \frac{l_{xy}}{l_{xx}} \quad a = \bar{y} - b\bar{x} \\ \hat{x} = c+dy \quad d = \frac{l_{xy}}{l_{yy}} \quad c = \bar{x} - d\bar{y} \end{cases}$$

$y = a+b(c+dy) = a+bc+bdy$ 若重合，则该式子恒成立

$y - bdy = \bar{y} - \cancel{b\bar{x}} + \cancel{b\bar{x}} - bd\bar{y} = (1-bd)\bar{y}$

即 $1-bd=0 \Rightarrow \frac{l_{xy}^2}{l_{xx}l_{yy}} = 1 = r^2 \Rightarrow r = \pm 1$

$|r|=1$ 表示 (x_i, y_i) 在一条直线上，这概率不大

∴ 我认为这两条直线不重合

交点 $y = \bar{y}$ $x = \frac{\bar{y}-a}{b} = \frac{\cancel{a}+\cancel{b}\bar{x}-\cancel{a}}{\cancel{b}} = \bar{x}$

∴ 交点为 $(\bar{x}, \bar{y})$

Q3:

$H = X(X'X)^{-1}X'$

I-H 对称

$$(I-H)^T = (I - X(X^TX)^{-1}X^T)^T = I - X((X^TX)^{-1})^T X^T = I - X(((X^TX)^T)^{-1}X^T$$

$$= I - X(X^TX)^{-1}X^T = I - H$$

∴ I-H 是一个对称矩阵

I-H 幂等

$$(I-H)^2 = I^2 - IH - HI + H^2 = I^2 - 2H + H^2$$

$$= I - 2X(X^TX)^{-1}X^T + X\underline{\cancel{(X^TX)^{-1}}}\,\underline{\cancel{X^TX}}\,\underline{(X^TX)^{-1}}\,\underline{X^T}$$

$$= I - 2X(X^TX)^{-1}X^T + X(X^TX)^{-1}X^T$$

$$= I - X(X^TX)^{-1}X^T = I - H$$

∴ I-H 是一个幂等矩阵

∴ I-H 是一个对称且幂等的矩阵

$$\therefore \operatorname{rank}(I-H) = tr(I-H) = tr(I) - tr(H) = n - (p+1) = n-p-1$$

↓

$$tr(X(X^TX)^{-1}X^T)$$

$$= tr((X^TX)^{-1}X^TX)$$

$$= tr(I_{p+1}) = p+1$$ （p为自变量个数）

$$\therefore \operatorname{rank}(I-H) = n-p-1$$

Q4:

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$ X 满秩

证 $\sum_{i=1}^{n} (y_i - \hat{y}_i) = 0$

① $\Big\downarrow$ 矩阵形式

$\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty$

$\hat{\beta} = (X^TX)^{-1}X^Ty$

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}^T (y - \hat{y}) = 0$$

$\therefore 1^T(y-\hat{y}) = 1^T(y - X\hat{\beta}) = 1^T(y - X(X^TX)^{-1}X^Ty)$

② 要证 $\longrightarrow$ 为0

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

即 $1^T(y - X(X^TX)^{-1}X^Ty) = 0$

$(1^T - 1^TX(X^TX)^{-1}X^T)y = 0$

$(\underline{1^T - (X(X^TX)^{-1}X^T)^T})y = 0$

证

③ $(X(X^TX)^{-1}X^T)^T = 1^T$

③ 令 $a = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ 则 $Xa = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1$ (列向量)

$\therefore X\cancel{(X^TX)^{-1}}\cancel{X^TX}a = Xa = 1$

$\therefore (X(X^TX)^{-1}X^T)^T = 1^T$

$\therefore 1^T(y - X\hat{\beta}) = 1^T(y - \hat{y}) = \sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$ 得证

Q5:

(1)

对 $y_1, y_2 \cdots y_n$ 进行中心化

对 $X_{1j}, X_{2j} \cdots X_{nj}$ 进行标准化　$j = 1、2、3 \cdots p$

记中心化之后的 y 为 $y^* = (y_1^*, y_2^*, \cdots y_n^*)^T$

记标准化之后的 X 为 X^{**} 矩阵 $\begin{pmatrix} 0 & \\ 0 & \\ \vdots & X_s \\ 0 & \end{pmatrix} = (0 \ \ X_s)$

① $X_s = X_c L = (I_n - H_{1_n}) X_o L$, $I_n - H_{1_n}$ 是一个对称的幂等矩阵

$$1_n^T (I_n - H_{1_n}) = 1_n^T - 1_n^T H_{1_n} = 1_n^T - \cancel{1_n^T 1_n} (\cancel{1_n^T 1_n})^{-1} 1_n^T = 0$$

② ∴ $1_n^T X_s = \underbrace{1_n^T (I_n - H_{1_n})}_{0} X_o L = 0 = X_s^T 1_n$

令 $\Omega_s = (X_s^T X_s - n^{-1} X_s^T 1_n 1_n^T X_s)^{-1} = (X_s^T (I_n - H_{1_n}) X_s)^{-1}$

∵ $I_n - H_{1_n}$ 对称幂等

∴ ③ $\Omega_s = (X_s^T (I_n - H_{1_n}) (I_n - H_{1_n}) X_s)^{-1} = (X_s^T (I_n - H_{1_n})^3 X_o L)^{-1}$

$= (X_s^T (I_n - H_{1_n}) X_o L)^{-1} = (X_s^T X_s)^{-1}$

$$\tilde{\beta} = ((X^{**})^T X^{**})^{-1}(X^{**})^T y^*$$

$$= \begin{pmatrix} n^{-1}1_n^T + n^{-2}\underbrace{\cancel{1_n^T X_s}}_{0}\cancel{\delta_s X_s^T} 1_n 1_n^T - n^{-1}\underbrace{\cancel{1_n^T X_s}}_{0}\delta_s X_s^T \\ \\ -n^{-1}\delta_s X_s^T 1_n 1_n^T + \delta_s X_s^T \end{pmatrix} y^*$$

$$= \begin{pmatrix} n^{-1}1_n^T \\ -n^{-1}\delta_s X_s^T 1_n 1_n^T + \delta_s X_s^T \end{pmatrix}(I_n - H_{1_n})y$$

$$= \begin{pmatrix} n^{-1}\underbrace{\cancel{1_n^T (I_n - H_{1_n})}}_{0} y \\ (-n^{-1}\delta_s \underbrace{\cancel{X_s^T 1_n}}_{0} 1_n^T + \delta_s X_s^T)(I_n - H_{1_n})y \end{pmatrix} = \begin{pmatrix} 0 \\ \delta_s X_s^T (I_n - H_{1_n})y \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ (X_s^T X_s)^{-1} X_s^T y^* \end{pmatrix} = \begin{pmatrix} 0 \\ \sqrt{L_{yy}}\,(X_s^T X_s)^{-1} X_s^T y^{**} \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ \sqrt{L_{yy}}\ \hat{\beta}_{s,\,slope} \end{pmatrix} = \begin{pmatrix} 0 \\ \cancel{\sqrt{L_{yy}}}\ \cancel{\frac{1}{\sqrt{L_{yy}}}} L^{-1}\underbrace{(X_c^T X_c)^{-1} X_c^T y^*}_{\hat{\beta}_{c,\,slope} = \hat{\beta}_{slope}} \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ L^{-1}\hat{\beta}_{slope} \end{pmatrix} \qquad \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_{slope} \end{pmatrix}$$

$$\therefore \tilde{\beta} = \begin{pmatrix} 0 & 0 \\ 0 & L^{-1} \end{pmatrix}\hat{\beta}$$

---

(2)

$$E(\tilde{\beta}) = E\left(\begin{pmatrix} 0 & 0 \\ 0 & L^{-1} \end{pmatrix}\hat{\beta}\right) = \begin{pmatrix} 0 & 0 \\ 0 & L^{-1} \end{pmatrix} E(\hat{\beta}) = \begin{pmatrix} 0 & 0 \\ 0 & L^{-1} \end{pmatrix} \beta$$

$$Var(\tilde{\beta}) = Var\left(\begin{pmatrix} 0 & 0 \\ 0 & L^{-1} \end{pmatrix}\hat{\beta}\right) = \begin{pmatrix} 0 & 0 \\ 0 & L^{-1} \end{pmatrix} var(\hat{\beta}) \begin{pmatrix} 0 & 0 \\ 0 & L^{-1} \end{pmatrix}^{T}$$

$$= \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & L^{-1} \end{pmatrix} (X^T X)^{-1} \begin{pmatrix} 0 & 0 \\ 0 & L^{-T} \end{pmatrix}$$

Q6:

(1) $\varepsilon_{ij}\ \underline{\text{iid}}\ N(0,\sigma^2)$

$y_{ij}=\mu_i+\varepsilon_{ij}$, $i=1,2,\cdots a$ | $j=1,2,\cdots,m$

响应变量 $y=(y_{11},y_{12},\cdots y_{1m},y_{21},y_{22}\cdots y_{2m}\cdots y_{am})^T$

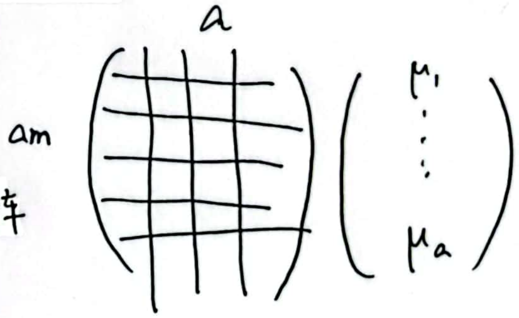
参数向量 $\beta=(\mu_1,\mu_2,\cdots\mu_a)^T$

误差矩阵 $e=(\varepsilon_{11},\varepsilon_{12}\cdots\varepsilon_{1m},\varepsilon_{21},\varepsilon_{22}\cdots\varepsilon_{2m}\cdots\varepsilon_{am})^T$

引入 x_i , $\{x_1\ x_2\cdots x_a\}$

$$\underbrace{y}_{am\times1}=\underbrace{x}_{}\underbrace{\beta}_{a\times1}+\underbrace{e}_{am\times1}$$

∴ x需要是一个 am×a 大小的矩阵



$$\begin{pmatrix} y_{11}\\ y_{12}\\ \\ y_{1m}\\ \\ \\ \end{pmatrix}_{am\times1}\quad \begin{array}{c} m\\ \\ m \end{array}\begin{pmatrix} \mu_1\\ \mu_1\\ \\ \mu_1\\ \mu_2\\ \vdots\\ \mu_2 \end{pmatrix}_{am\times1}+\begin{pmatrix} \varepsilon_{11}\\ \varepsilon_{12}\\ \\ \\ \\ \end{pmatrix}_{am\times1}$$

$$\therefore X=\begin{array}{c|c} & a\\ \hline a & \begin{matrix}1&0&0&0&\cdots&0\\1&0&0&\cdots&\cdots&0\\1&0&0&0&\cdots&0\end{matrix}\\ \hline 2a & \begin{matrix}0&1&0&0&\cdots&0\\0&1&0&0&\cdots&0\\0&1&0&0&\cdots&0\end{matrix}\\ \hline & \vdots\\ \hline am & \begin{matrix}0&0&0&\cdots&\cdots&1\\0&0&0&\cdots&\cdots&1\\0&0&0&\cdots&\cdots&1\end{matrix} \end{array}$$

$$\therefore X为\begin{pmatrix}1&&&\\&1&&\\&&\ddots&\\&&&1\end{pmatrix}_{am\times a}$$

---

(2)

$Y = X\beta + e$ 模型

$Y = (y_{11}, y_{12} \cdots y_{1m}, y_{21}, y_{22} \cdots y_{am})^T$ 响应变量

$\beta = (\mu_1, \mu_2 \cdots \mu_a)^T$ 参数向量

$X = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}_{am \times a}$ 设计矩阵X

误差向量 $e = (\varepsilon_{11}, \varepsilon_{12}, \cdots \varepsilon_{1m}, \varepsilon_{21}, \varepsilon_{22} \cdots \varepsilon_{am})^T$

(3)

$$\hat{\beta} = (\underset{a\times am}{X^T}\underset{am\times a}{X})^{-1} X^T Y = (\underset{\cancel{a\times a}}{m \cdot I_a})^{-1} \underset{\cancel{a\times am}}{X^T} \underset{\cancel{am\times 1}}{Y} = \frac{1}{m}\begin{pmatrix} y_{11} + y_{12} + \cdots y_{1m} \\ \vdots \\ y_{a1} + y_{a2} + \cdots y_{am} \end{pmatrix} = \begin{pmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_a \end{pmatrix}$$

$$\frac{1}{m} I_a^{-1} = \frac{1}{m} I_a$$

(单位矩阵逆是本身)

$$\hat{\mu}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij} = \overline{y_i}$$

(4)

$H_0: \mu_1 = \mu_2 = \cdots \mu_a = 0$ vs H_1: 至少存在1个 $\mu_i \neq 0$ $i \in [1, a]$

$$F_0 = \frac{SS_R / (a-1)}{SS_E / (n-a)}$$ = 单因素方差分析的检验统计量

$$SS_R = \sum_{i=1}^{a}\sum_{j=1}^{m} (\hat{y_{ij}} - \overline{y}_{..})^2 \qquad SS_E = \sum_{i=1}^{a}\sum_{j=1}^{m} (y_{ij} - \hat{y_{ij}})^2$$

$F_0 > F_{1-\alpha}(a-1, am-a)$ 和 $P_0 = P(F \geq F_0) < \alpha$ 时，显著