D-Day: Oct 13 9:30 AM

Q1:

在回归分析中，对数据进行变换

$$\tilde{y}_i = \frac{y_i - c_1}{d_1}, \quad \tilde{x}_i = \frac{x_i - c_2}{d_2}, \quad i = 1, 2, \cdots, n,$$

其中，选取c_1, c_2, d_1, d_2为适当的常数。请回答：

- 试建立由原始数据和变换后数据得到的最小二乘估计、总偏差平方和、回归平方和以及残差平方和之间的关系；
- 证明：由原始数据和变换后数据得到的F统计量的值保持不变。

解:

1) 由题, 对原数据: 最小二乘估计 $\begin{cases} \hat{\beta}_0 = \bar{y} - \beta_1 \bar{x} \\ \hat{\beta}_1 = l_{xx}^{-1} l_{xy} \end{cases}$

回归平方和 $SS_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$.

残差平方和 $SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

总偏差平方和 $SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$

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

∵ $\tilde{y}_i = \frac{y_i - c_1}{d_1}$ $\quad \tilde{x}_i = \frac{x_i - c_2}{d_2}$

∴代入可得

① 最小二乘估计 $\begin{cases} \hat{\beta}_0' = \frac{\bar{y} - c_1}{d_1} - \frac{\bar{x} - c_2}{d_2}\beta_1 = \frac{\bar{y} - c_1}{d_1} - \frac{\bar{x} - c_2}{d_2} \cdot \frac{d_2}{d_1}\hat{\beta}_1 = \frac{1}{d_1}\beta_0 - \frac{c_2}{d_1}\hat{\beta}_1 - \frac{c_1}{d_1} \\ \hat{\beta}_1' = \frac{d_2}{d_1}\hat{\beta}_1 \end{cases}$

② 回归平方和 $SS_R' = \frac{1}{d_1^2} SS_R$

③ 残差平方和 $SS_E' = \frac{1}{d_1^2} SS_E$

④ 总偏差平方和 $SS_T' = \frac{1}{d_1^2} SS_T$

2) 原始数据的F统计量

$$f_A = \frac{SS_R/(a-1)}{SS_E/(n-a)} = \frac{SS_R/1}{SS_E/(n-2)} \quad (*)$$

由1)得 $SS_R' = \frac{1}{d_1^2} SS_R$ $\quad SS_E' = \frac{1}{d_1^2} SS_E$

代入(*)得 $f_A' = \frac{SS_R/1}{SS_E/(n-2)} = f_A$

⇒ 原始数据和变换数据的F统计量的值保持不变。

Q2:

对给定的n组数据$(x_i, y_i), i = 1, 2, \cdots, n$, 若我们关心的是$y$如何依赖$x$的取值而变动，则可以建立回归方程

$$\hat{y} = a + bx.$$

反之，若我们关心的是x如何依赖y的取值而变动,则可以建立另一个回归方程

$$\hat{x} = c + dy.$$

试问这两条直线在直角坐标系中是否重合？为什么？若不重合,它们有无交点？若有,试给出交点的坐标。

解：由题，假设两条直线在直角坐标系中重合，则

y关于x：$\hat{y} = a + bx$

$\therefore a = \bar{y} - b\bar{x} \quad b = l_{xx}^{-1} l_{xy}$

x关于y：$\hat{x} = c + dy$

$\therefore c = \bar{x} - d\bar{y} \quad d = l_{yy}^{-1} l_{xy}$

联立得 $y = bdy + a + bc$

$(1-bd)y = (1-bd)\bar{y}$

$\because$ 重合

$\therefore 1 - bd = 0 \Rightarrow 1 - \dfrac{l_{xy}}{l_{xx}} \cdot \dfrac{l_{xy}}{l_{yy}} = 0 \Rightarrow \dfrac{l_{xy}^2}{l_{xx} l_{yy}} = 0$

$\Rightarrow r^2 = 1$

$r = \pm 1$

即当$r = \pm 1$时，两直线重合.

$\therefore$ 当$r \neq \pm 1$时，存在交点，$(1-bd)y = (1-bd)\bar{y}$

$y = \bar{y}$

$\therefore x = \bar{x}$

$\therefore$ 交点为$(\bar{x}, \bar{y})$

Q3:

令$\mathbf{H}=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$是一个帽子矩阵（如定理1-1），$\mathbf{I}$ 为单位阵。证明：$\mathbf{I}-\mathbf{H}$ 是一个对称且幂等的矩阵。并计算这个矩阵的秩。

证明:

① 对称阵 $A^T=A$

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

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

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

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

$=I-H$

$\therefore (I-H)^T=(I-H)$，是对称阵.

② 幂等阵 $H^2=H$

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

$\because$ H是幂等阵

$\therefore H^2=H \quad I^2=I$.

$\therefore$ 上式 $=I-2IH+H=I-2H+H=I-H$

$\therefore (I-H)^2=I-H$，是幂等矩阵.

③ $\because I-H$ 是幂等阵

$\therefore rank(I-H)=tr(I-H)=n-p-1$

即 $(I-H)$ 的秩为 $(n-p-1)$.

Q4:

在一个多元线性回归模型中，响应变量y_i的回归值为

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

$\mathbf{X}$是一个满秩矩阵，证明：$\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$。

证明：由题

$$\begin{cases} \hat{y} = X\hat{\beta} \\ \hat{\beta} = (X^T X)^{-1} X^T y \end{cases}$$

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

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

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

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

$\therefore \exists\, k = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, s.t $1 = kX$，即提取X的第1列 $(*)$

将$(*)$代入原式可得 $= [1^T - (X(X^TX)^{-1}X^T 1)^T]y$

$$= (1^T - \underbrace{k(X}_{1}\underbrace{(X^TX)^{-1}X^TX}_{I_n})^T]y$$

$$= (1^T - 1^T)y = 0$$

$\therefore \sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$，证毕.

Q5:

在多元线性回归模型

$$y=\beta_0+\beta_1x_1+\cdots+\beta_px_p+\varepsilon$$

中，我们有数据$\{(y_i,x_{i1},x_{i2},\cdots,x_{ip})\}_{i=1}^n$。我们可以得到最小二乘估计，记为$\hat\beta=(\hat\beta_0,\hat\beta_1,\cdots,\hat\beta_p)'$。

如果我们对$y_1,y_2,\cdots,y_n$进行中心化，对每一维自变量$x_{1j},x_{2j},\cdots,x_{nj}$均进行了标准化，$j=1,2,\cdots,p$，那么，我们得到的最小二乘估计为$\tilde\beta=(\tilde\beta_0,\tilde\beta_1,\cdots,\tilde\beta_p)'$。

请回答：

- 这两个估计$\tilde\beta$和$\hat\beta$之间有什么关系？
- 求$\tilde\beta$的期望和方差。

解:

1) ∵对 $y_1,y_2,\cdots,y_n$进行中心化，$x_{1j},x_{2j},\cdots,x_{nj}$标准化

∴ 中心化后为 y^* 标准化后为 X^{**}

∴ $X^{**}=(0\ \ X_s)$ $X_s=(I_n-H_{1n})X_0L$

令 $A_s=(X_s^T(I_n-H_{1n})X_s)^{-1}$ 幂等矩阵

$=(X_s^T(I_n-H_{1n})(I_n-H_{1n})X_0L)^{-1}$

$=(X_s^T(I_n-H_{1n})X_0L)^{-1}$

$=(X_s^TX_s)^{-1}$

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

将结论代入得

$$\tilde\beta=\begin{pmatrix}n^{-1}1_n^T\\ -n^{-1}A_sX_s^T1_n1_n^T+A_sX_s^T\end{pmatrix}(I_n-H_{1n})y$$

$1_n^TX_s=0$

$$=\begin{pmatrix}0\\ A_sX_s^T(I_n-H_{1n})y\end{pmatrix}$$

$$=\begin{pmatrix}0\\ (X_s^TX_s)^{-1}X_s^Ty^*\end{pmatrix}$$

$$=\begin{pmatrix}0\\ \sqrt{L_{yy}}\hat\beta_{slope}\end{pmatrix}$$

$$\hat\beta=\begin{pmatrix}\beta_0\\ \hat\beta_{slope}\end{pmatrix}\quad \tilde\beta=\begin{pmatrix}0\\ \sqrt{L_{yy}}\hat\beta_{slope}\end{pmatrix}\Rightarrow\tilde\beta=\begin{pmatrix}0&0\\0&\sqrt{L_{yy}}\end{pmatrix}\hat\beta$$

2) $\sqrt{L_{yy}}$ 是常数

∴ 由期望的性质 $E(\tilde\beta)=\begin{pmatrix}0&0\\0&\sqrt{L_{yy}}\end{pmatrix}\hat\beta$

Q6:

已知单因子方差分析模型

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

其中，ε_{ij}是独立同分布的随机变量，其分布为$N(0, \sigma^2)$，我们观测到的数据为$\{y_{ij}\}$。

证明：单因子方差分析模型可以看作一种多元线性回归模型。提示：

- 构造一个合适的设计矩阵~~\bm{X}~~ X；
- 定义响应变量向量、回归参数向量、设计矩阵、误差向量，并写出"数据版"的多元线性回归模型；
- 最小二乘法估计回归参数向量，并与μ_i进行比较；
- 利用F检验，对所构造对多元线性回归模型进行模型显著性检验，并与方差分析的结果进行比较。

证明：令 $\bar{\mu} = \frac{\sum_{i=1}^{a} \mu_i}{a}$ $\quad \alpha_i = \mu_i - \bar{\mu} \quad N = a \cdot m$

∴ $y_{ij} = \mu_i + \varepsilon_{ij} \Longrightarrow y_{ij} = \bar{\mu} + \alpha_i + \varepsilon_{ij}$

则离均差和为 $\sum_{i=1}^{a} \alpha_i = 0$

其中随机误差 ε_{ij} 独立同分布 $\varepsilon_{ij} \sim N(0, \sigma^2)$

(后面不是忘了写，是不会，其实这个提示也看得模棱两可，
前面的也只是看PPT和查资料照猫画虎了一下)