2014.4.10

第十六讲. 最小二乘与投影(续)

回顾：求$\beta_{(2)}$的LS估计

划分$X=(X_1,X_2)$, $\beta=\begin{pmatrix}\beta_{(1)}\\ \beta_{(2)}\end{pmatrix}\begin{matrix}(p-q)\times 1\\ q\times 1\end{matrix}$， 设$\mathbf{1}$在$X_1$中

$$Y = X\beta+\varepsilon = X_1\beta_{(1)}+X_2\beta_{(2)}+\varepsilon,$$

令$X_2^\perp = X_2 - P_{X_1}X_2$，$\beta^*_{(1)} = \beta_{(1)} + (X_1'X_1)^{-1}X_1'X_2\beta_{(2)}$，改写模型：

$$Y = X_1\beta^*_{(1)} + X_2^\perp\beta_{(2)} + \varepsilon = X^*\beta^*+\varepsilon$$

$$\Rightarrow \hat\beta^*_{(1)} = (X_1'X_1)^{-1}X_1'Y$$

$$\hat\beta_{(2)} = (X_2^{\perp\prime}X_2^\perp)^{-1}X_2^{\perp\prime}Y$$

$$\Rightarrow \hat\beta_{(1)} = \hat\beta^*_{(1)} - (X_1'X_1)^{-1}X_1'X_2\hat\beta_{(2)} = .. = (X_1^{\perp\prime}X_1^\perp)^{-1}X_1^{\perp\prime}Y$$

特例1. 单个回归系数的LS估计 (*q*=1)

$k>0$, $\hat\beta_k = (\mathbf{x}_k^{\perp\prime}\mathbf{x}_k^\perp)^{-1}\mathbf{x}_k^{\perp\prime}Y$，其中$\mathbf{x}_k^\perp = \mathbf{x}_k - \hat{\mathbf{x}}_k$， $\hat{\mathbf{x}}_k = P_{X_{(-k)}}\mathbf{x}_k$为$\mathbf{x}_k$在X其它列上的投影。

注1：$\hat\beta_k = \left(x_k^{\perp\prime}x_k^\perp\right)^{-1}x_k^{\perp\prime}Y \propto r_{yx_k\bullet 其它x}$(偏相关系数),

注2：$\mathrm{var}(\hat\beta_k \mid X) = \dfrac{\sigma^2}{\|x_k^\perp\|^2} = \dfrac{1}{1-R_k^2}\times\dfrac{\sigma^2}{\|x_k - \mathbf{1}\bar{x}_k\|^2}$,

其中$VIF = \dfrac{1}{1-R_k^2} = \dfrac{\|x_k - \mathbf{1}\bar{x}_k\|^2}{\|x_k^\perp\|^2} \ge 1$

$R_k^2 = 1 - \dfrac{\|x_k^\perp\|^2}{\|x_k - \mathbf{1}\bar{x}_k\|^2} = \dfrac{\|\hat{x}_k - \mathbf{1}\bar{x}_k\|^2}{\|x_k - \mathbf{1}\bar{x}_k\|^2}$为$x_k$与其它自变量的复相关系数平方.

这是因为$x_k^\perp \perp \hat{x}_k, x_k^\perp \perp \mathbf{1} \;\Rightarrow\; (x_k - \mathbf{1}\bar{x}_k) = (\hat{x}_k - \mathbf{1}\bar{x}_k) \oplus x_k^\perp$

注3：注意到 $\hat{\mathbf{x}}_k = P_{X_{(-k)}}\mathbf{x}_k$，为$\mathbf{x}_k$在其它自变量$X_{(-k)}$生成空间上的投影（拟合值）， 所以

$x_k^\perp = x_k - \hat{\mathbf{x}}_k$为“$x_k$对设计阵其它列回归后 的残差”，

而$\hat\beta_k = (x_k^{\perp\prime}x_k^\perp)^{-1}x_k^{\perp\prime}Y$，因此$\hat\beta_k$可由如下两步求得：

(1) $x_k \sim X_{(-k)}$ （即$x_k \sim x_1+\ldots+x_{k-1}+x_{k+1}+\ldots.+x_{p-1}$）

x_k对其它自变量回归，得 到残差$x_k^\perp$

(2) $Y \sim x_k^\perp$ 或 $Y^\perp \sim x_k^\perp$

简单回归，得到的斜率 即$\hat\beta_k$,

特例2. 中心化 ($q=p-1$)

划分$X=(\mathbf{1},Z)$,对应的$\beta=\begin{pmatrix}\beta_0\\ \gamma\end{pmatrix}$, γ为所有自变量回归系数，模型:

$$Y=X\beta+\varepsilon=\mathbf{1}\beta_0+Z\gamma+\varepsilon,$$

令 $Z^{\perp}=Z-P_1Z=Z-\mathbf{1}\frac{\mathbf{1}'Z}{n}=Z-\mathbf{1}\bar{\mathbf{x}}'$, 其中$\bar{\mathbf{x}}=\frac{Z'\mathbf{1}}{n}$

改写模型:

$$Y=\mathbf{1}\beta_0+Z\gamma+\varepsilon=\mathbf{1}(\beta_0+\bar{\mathbf{x}}'\gamma)+Z^{\perp}\gamma+\varepsilon$$

$$\Rightarrow\hat{\gamma}=(Z^{\perp}{}'Z^{\perp})^{-1}Z^{\perp}{}'Y，\widehat{(\beta_0+\bar{\mathbf{x}}'\gamma)}=\bar{Y}\Rightarrow\hat{\beta}_0=\bar{Y}-\bar{\mathbf{x}}'\hat{\gamma}$$

一些注解

- $\hat{Y}=P_1Y+P_{Z^{\perp}}Y=\mathbf{1}\bar{Y}+Z^{\perp}\hat{\gamma}$

  $P_{Z^{\perp}}Y=Z^{\perp}\hat{\gamma}$ 可看作是所有自变量对 $\hat{Y}$的贡献,其方差在总方差中所占百分比 (注意$Z^{\perp}\hat{\gamma}\perp\mathbf{1}$)

$$\frac{\|Z^{\perp}\hat{\gamma}\|^2}{SS_{总}}=\frac{\|\hat{Y}-\mathbf{1}\bar{Y}\|^2}{|Y-\mathbf{1}\bar{Y}\|^2}=R^2$$

  即复相关系数平方.

- 注意到

  $Z^{\perp}{}'Z^{\perp}/(n-1)=S_{zz}$: 所有自变量的样本协方差阵,

  $Z^{\perp}{}'Y/(n-1)\ =S_{zy}$: 所有自变量与响应变量的样本协方差阵.

$$所以\hat{\gamma}=\left(Z^{\perp}{}'Z^{\perp}\right)^{-1}Z^{\perp}{}'Y=S_{zz}^{-1}S_{zy}$$

$$\mathrm{var}(\hat{\boldsymbol{\gamma}}\mid X)=\sigma^2\left(Z^{\perp}{}'Z^{\perp}\right)^{-1}=\frac{\sigma^2}{n}(S_{zz})^{-1}$$

  所以自变量回归系数LS估计及其方差由样本协方差矩阵完全确定.

  但截距项β_0的LS估计依赖于样本均值。

- 若所有变量都已经标准化(减去均值,除以标准差),即

  $Y=X\gamma+\varepsilon$中所有变量的均值为0,方差为1

  则$\hat{\gamma}=(X'X)^{-1}X'Y=R_{XX}^{-1}R_{XY}$, R代表相关系数矩阵

例1. 为了估计两个物体的重 量α,β,用天平测量3次,
三次测量值为y_1,y_2,y_3,假设(可加)测量误差服从 $(0,\sigma^2)$,
针对下属每种情况，求 α,β的LS估计及其方差，以及 σ^2的估计。
(1) 三次测量的都是第一 个物体
(2) 前两次测量第一个物 体，第三次测第二个物 体
(3) 第一二次分别测 α,β,第三次测 $\alpha+\beta$
(4) 第一、二、三次分别 测$\alpha,\alpha-\beta,\alpha+\beta$
(5) 前两次测 $\alpha\pm\beta$，第三次测 $\alpha\mp\beta$

(1)
$$y_1=\alpha+\varepsilon_1,\quad y_2=\alpha+\varepsilon_2,\quad y_3=\alpha+\varepsilon_3,\qquad Y=\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}=\mathbf{1}\alpha+\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{pmatrix}$$

$$\hat{\alpha}=(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'Y=\bar{y},\quad \mathrm{var}(\hat{\alpha})=\sigma^2/3,\ \hat{\sigma}^2=s^2$$

（2）$y_1=\alpha+\varepsilon_1, y_2=\alpha+\varepsilon_2, y_3=\beta+\varepsilon_3$，

$$Y=\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}=\begin{pmatrix}1&0\\1&0\\0&1\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix}+\begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\end{pmatrix}$$

$$\text{记}\mathrm{x}_1=\begin{pmatrix}1\\1\\0\end{pmatrix},\mathrm{x}_2=\begin{pmatrix}0\\0\\1\end{pmatrix},\ \mathrm{x}_1\perp\mathrm{x}_2$$

$$\hat\alpha=\mathrm{x}_1'Y/\mathrm{x}_1'\mathrm{x}_1=(y_1+y_2)/2,$$

$$\hat\beta=\mathrm{x}_2'Y/\mathrm{x}_2'\mathrm{x}_2=y_3$$

$$\mathrm{var}(\hat\alpha)=\sigma^2/2,\mathrm{var}(\hat\beta)=\sigma^2$$

$$\hat\sigma^2=(y_1-\hat\alpha)^2+(y_2-\hat\alpha)^2+(y_3-\hat\beta)^2=(y_1-y_2)^2/2$$

（3）

$$y_1=\alpha+\varepsilon_1,\quad y_2=\beta+\varepsilon_2,\quad y_3=\alpha+\beta+\varepsilon_3,\qquad Y=\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}=\begin{pmatrix}1&0\\0&1\\1&1\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix}+\begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\end{pmatrix}$$

$$\text{记}\mathrm{x}_1=\begin{pmatrix}1\\0\\1\end{pmatrix},\mathrm{x}_2=\begin{pmatrix}0\\1\\1\end{pmatrix},\text{则 }\mathrm{x}_1^\perp=\mathrm{x}_1-\mathrm{x}_2(\mathrm{x}_2'\mathrm{x}_2)^{-1}\mathrm{x}_2'\mathrm{x}_1=\begin{pmatrix}1\\-1/2\\1/2\end{pmatrix}$$

$$\hat\alpha=\mathrm{x}_1^{\perp\prime}Y/\mathrm{x}_1^{\perp\prime}\mathrm{x}_1^\perp=(2y_1-y_2+y_3)/3,$$

$$\text{同理}\hat\beta=\mathrm{x}_2^{\perp\prime}Y/\mathrm{x}_2^{\perp\prime}\mathrm{x}_2^\perp=(-y_1+2y_2+y_3)/3,$$

$$\mathrm{var}(\hat\alpha\mid X)=\mathrm{var}(\hat\beta\mid X)=\frac{6}{9}\sigma^2.$$

$$\hat Y=\begin{pmatrix}\hat\alpha\\\hat\beta\\\hat\alpha+\hat\beta\end{pmatrix},\ \mathrm{e}=Y-\hat Y=\begin{pmatrix}(y_1+y_2-y_3)/3\\(y_1+y_2-y_3)/3\\-(y_1+y_2-y_3)/3\end{pmatrix},\ \Rightarrow\hat\sigma^2=(y_1+y_2-y_3)^2/3$$

（4）

$$y_1=\alpha+\varepsilon_1,\quad y_2=\alpha-\beta+\varepsilon_2,\quad y_3=\alpha+\beta+\varepsilon_3,\qquad Y=\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}=\begin{pmatrix}1&0\\1&-1\\1&1\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix}+\begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\end{pmatrix}$$

$$\text{记}\mathrm{x}_1=\begin{pmatrix}1\\1\\1\end{pmatrix},\mathrm{x}_2=\begin{pmatrix}0\\-1\\1\end{pmatrix},\text{则 }\mathrm{x}_1'\mathrm{x}_2=0$$

$$\hat\alpha=\mathrm{x}_1'Y/\mathrm{x}_1'\mathrm{x}_1=(y_1+y_2+y_3)/3,\ \hat\beta=\mathrm{x}_2'Y/\mathrm{x}_2'\mathrm{x}_2=(-y_2+y_3)/2,$$

$$\mathrm{var}(\hat\alpha\mid X)=\sigma^2/3,\mathrm{var}(\hat\beta\mid X)=\sigma^2/2.$$

$$\hat Y=\begin{pmatrix}\hat\alpha\\\hat\alpha-\hat\beta\\\hat\alpha+\hat\beta\end{pmatrix}=\begin{pmatrix}(y_1+y_2+y_3)/3\\(2y_1+5y_2-y_3)/6\\(2y_1-y_2+5y_3)/6\end{pmatrix},\mathrm{e}=Y-\hat Y=\begin{pmatrix}(2y_1-y_2-y_3)/3\\(-2y_1+y_2+y_3)/6\\(-2y_1+y_2+y_3)/6\end{pmatrix},$$

$$\hat\sigma^2=(2y_1-y_2-y_3)^2/6$$

不同的试验设计导致不同的估计精度，**最优设计?**

例2 (简单线性模型) 数据$(x_i,y_i),i=1,2,\dots,n.$ iid $\sim y=a+bx+\varepsilon$，

即$y_i=a+bx_i+\varepsilon_i$，ε_i iid $\sim(0,\sigma^2)$，ε_i与x_i独立。求LS估计。

$$\mathrm{Y}=\mathrm{X}\beta+\boldsymbol{\varepsilon}=(\mathbf{1},\mathbf{x})\begin{pmatrix}a\\b\end{pmatrix}+\boldsymbol{\varepsilon}=\mathbf{1}a+\mathbf{x}\mathrm{b}+\boldsymbol{\varepsilon}$$

$$\text{其中Y}=(y_1,\dots,y_n)',\mathbf{1}=(1,\dots,1)',\mathbf{x}=(x_1,\dots,x_n)',\quad\beta=\begin{pmatrix}a\\b\end{pmatrix}$$

$$\text{记 x}^\perp=\mathrm{x}-\mathrm{P}_{\mathbf{1}}\mathrm{x}=\mathrm{x}-\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathrm{x}=\mathrm{x}-\bar x\mathbf{1}$$

$$\mathrm{Y}=\mathbf{1}a+\mathbf{x}\mathrm{b}+\boldsymbol{\varepsilon}=\mathbf{1}(a+\bar xb)+\mathbf{x}^\perp\mathrm{b}+\boldsymbol{\varepsilon}\,\hat{=}\,\mathbf{1}a^*+\mathbf{x}^\perp\mathrm{b}+\boldsymbol{\varepsilon}$$

$$\text{所以}\hat b=(\mathrm{x}^{\perp\prime}\mathrm{x}^\perp)^{-1}\mathrm{x}^{\perp\prime}Y=s_{xx}^{-1}s_{xy},$$

$$\hat a^*=\bar y\Rightarrow\hat a=\bar y-\bar x\hat b$$

例3. 数据$(x_i, y_i, z_i), i=1,2,...,n.$

模型: $y_i = a + bx_i + cz_i + \varepsilon_i$, ε_i iid ~ $(0, \sigma^2)$, ε_i 与 x_i, z_i 独立。

求b的LS估计及其方差。

模型为

$$\mathrm{Y} = \mathbf{1}a + \mathrm{x}\mathrm{b} + \mathbf{z}c + \boldsymbol{\varepsilon} = (\mathbf{1}, \mathbf{x}, \mathbf{z})\begin{pmatrix} a \\ b \\ c \end{pmatrix} + \boldsymbol{\varepsilon} = \mathrm{X}\beta + \boldsymbol{\varepsilon}$$

其中$\mathrm{Y} = (y_1, ..., y_n)', \mathbf{1} = (1, ..., 1)', \mathbf{x} = (x_1, ..., x_n)'$, $\mathbf{z} = (z_1, ..., z_n)'$,

$\hat{b} = (\mathrm{x}^{\perp}{}'\mathrm{x}^{\perp})^{-1}\mathrm{x}^{\perp}{}'Y$, 其中$\mathrm{x}^{\perp} = \mathrm{x} - \mathrm{P}_{(\mathbf{1},\mathbf{z})}\mathrm{x}$, $\mathrm{P}_{(\mathbf{1},\mathbf{z})} = ?$

$\because L(\mathbf{1}, \mathbf{z}) = L(\mathbf{1}, \mathbf{z} - \mathbf{1}\bar{z}) = L(\mathbf{1}, \mathbf{z}^{\perp})$, 其中 $\mathbf{z}^{\perp} = \mathbf{z} - \mathrm{P}_{\mathbf{1}}\mathbf{z} = \mathbf{z} - \mathbf{1}\bar{z}$

$\therefore \mathrm{P}_{(\mathbf{1},\mathbf{z})} = \mathrm{P}_{\mathbf{1}} + \mathrm{P}_{\mathbf{z}^{\perp}} = \mathbf{1}\mathbf{1}'/n + \mathbf{z}^{\perp}\mathbf{z}^{\perp}{}'/\|\mathbf{z}^{\perp}\|^2$

$\therefore \mathrm{x}^{\perp} = \mathrm{x} - \mathrm{P}_{(\mathbf{1},\mathbf{z})}\mathrm{x} = \mathrm{x} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathrm{x} - \mathbf{z}^{\perp}\{\mathbf{z}^{\perp}{}'\mathbf{z}^{\perp}\}\mathbf{z}^{\perp}{}'\mathrm{x}$

$= \mathrm{x} - \mathbf{1}\bar{x} - \mathbf{z}^{\perp}\hat{\gamma}$, 其中$\hat{\gamma} = \{\mathbf{z}^{\perp}{}'\mathbf{z}^{\perp}\}\mathbf{z}^{\perp}{}'\mathrm{x} = s_{zx}/s_{zz}$

$$\therefore \hat{b} = (\mathrm{x}^{\perp}{}'\mathrm{x}^{\perp})^{-1}\mathrm{x}^{\perp}{}'Y = \frac{\sum\left(x_i - \bar{x} - (z_i - \bar{z})\hat{\gamma}\right)y_i}{\sum\left(x_i - \bar{x} - (z_i - \bar{z})\hat{\gamma}\right)^2}$$

$$\mathrm{var}(\hat{b}) = \frac{\sigma^2}{\sum\left(x_i - \bar{x} - (z_i - \bar{z})\hat{\gamma}\right)^2} = \frac{\sigma^2}{s_{xx} - s_{xz}^2/s_{zz}} = \frac{\sigma^2}{s_{xx}} \times \frac{1}{1 - r_{xz}^2}$$