2014.4.8

第十五讲. 最小二乘与投影

最小二乘(续)：拟合值与残差

模型：$\mathbf{Y}_{n\times 1}=X_{n\times p}\beta_{p\times 1}+\varepsilon_{n\times 1}$，$\varepsilon\sim(0,\sigma^2 I_n),\varepsilon\perp X$

假设X列满秩，最小二乘法估 计：$\hat{\beta}=\arg\min_{\beta}\|\mathbf{Y}-X\boldsymbol{\beta}\|^2$

正则方程 ：$X'(\mathbf{Y}-X\boldsymbol{\beta})=0\Rightarrow\hat{\boldsymbol{\beta}}=(X'X)^{-1}X'\mathbf{Y}$

- 拟合值 ：$\hat{\mathbf{Y}}=X\hat{\boldsymbol{\beta}}=X(X'X)^{-1}X'\mathbf{Y}=H\mathbf{Y}$

其中$H=X(X'X)^{-1}X'=P_X$,也称为帽子矩阵(hat matrix)

- 残差 ：$\mathbf{e}=\mathbf{Y}-\hat{\mathbf{Y}}=\mathbf{Y}-X\hat{\boldsymbol{\beta}}=(I_n-H)\mathbf{Y}$

(1) $X'\mathbf{e}=0$, $\mathbf{e}\perp\hat{Y}$; (2) $\mathrm{E}(\mathbf{e}\,|\,\mathrm{X})=0$, $\mathrm{var}(\mathbf{e}\,|\,X)=(I_n-H)\sigma^2$

- 残差平方和：$RSS=\|\mathbf{e}\|^2=\|Y-X\hat{\beta}\|^2=Y'(I_n-H)Y$
- 取$\hat{\sigma}^2=\frac{1}{n-p}\sum e_i^2=\frac{1}{n-p}\|\mathbf{e}\|^2=\frac{\mathrm{RSS}}{n-p}$称为$\sigma^2$的LS估计

总结 最小二乘法：从模型方程到拟合方程

模型方程	$Y=X\boldsymbol{\beta}+\varepsilon$, $\varepsilon\sim(0,\sigma^2 I_n)$, $\varepsilon\perp\!\!\!\perp X$
最小二乘法	$\min_{\beta}\sum(y_i-x_i'\boldsymbol{\beta})^2=\min_{\beta}\|\mathbf{Y}-X\boldsymbol{\beta}\|^2$
正则方程	$X'(\mathbf{Y}-X\hat{\boldsymbol{\beta}})=0$
LS 估计	$\hat{\boldsymbol{\beta}}=(X'X)^{-1}X'\mathbf{Y}$
拟合值 / 投影	$\hat{\mathbf{Y}}=X\hat{\boldsymbol{\beta}}=X(X'X)^{-1}X'\mathbf{Y}=\mathrm{P}_X\mathbf{Y}$
残差	$\mathbf{e}=\mathbf{Y}-\hat{\mathbf{Y}}=\mathbf{Y}-X\hat{\boldsymbol{\beta}}=(I_n-H)\mathbf{Y}$
残差平方和	$RSS=\|\mathbf{e}\|^2=Y'(I_n-H)Y$
误差方差估计	$\hat{\sigma}^2=\frac{\mathrm{RSS}}{n-p}$
拟合方程 / 正交分解	$Y=X\hat{\boldsymbol{\beta}}+\mathbf{e}$, $\mathbf{e}\sim(0,\sigma^2(I_n-H))$, $\mathbf{e}\perp X$

利用最小二乘法求解LS估计

记设计阵$\mathrm{X}=(\mathbf{x}_0\equiv\mathbf{1},\mathbf{x}_1,\ldots,\mathbf{x}_{p-1})$, 即X各列记为 $\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_{p-1}$

模型：$Y=X\beta+\varepsilon=\mathbf{x}_0\beta_0+\ldots+\mathbf{x}_{\mathrm{p-1}}\beta_{\mathrm{p-1}}+\varepsilon$, $\varepsilon\sim(0,\sigma^2 I_n),\varepsilon\perp X$

注意到$X\beta=\mathbf{x}_0\beta_0+\ldots+\mathbf{x}_{\mathrm{p-1}}\beta_{\mathrm{p-1}}$为$X$各列向量的线性组合，

最小二乘 $\min_{\beta}\|Y-X\beta\|^2\Leftrightarrow\min_{\mathrm{u}\in\mathrm{L}(X)}\|Y-\mathrm{u}\|^2$

定理8(第七讲)：记$\hat{\mathrm{Y}}=\mathrm{P_X Y}\in\mathrm{L(X)}$为Y在L(X)上的投影，

则$\|\mathrm{Y}-\hat{\mathrm{Y}}\|^2=\min_{\mathrm{u}\in\mathrm{L}(X)}\|\mathrm{Y}-\mathrm{u}\|^2$

$\hat{Y}$作为最优的X各列的线性组合,$\hat{Y}=X\hat{\beta}$,

其组合系数$\hat{\beta}$即为LS解,事实上:

由$\hat{Y}=\mathrm{P_X Y}=\mathrm{X(X'X)^{-1}X'Y}=\mathrm{X}\underline{\underline{\{\mathrm{(X'X)^{-1}X'Y}\}}}$,

所以β的最小二乘估计$\hat{\beta}=\underline{\underline{\mathrm{(X'X)^{-1}X'Y}}}$

注：从解方程的观点来看:

$\mathrm{Y}=\mathrm{X}\beta$一般无解 (因为$n\ge p$,一般$\mathrm{Y}\notin \mathrm{L(X)}$),

但

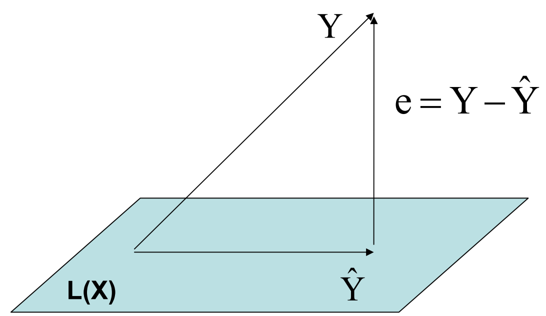
$\hat{\mathrm{Y}}=\mathrm{X}\beta$有解 (因为$\hat{\mathrm{Y}}\in \mathrm{L(X)}$),

而且当X列满秩时解唯一

(若 $\hat{\mathrm{Y}}=\mathrm{X}\beta=\mathrm{X}\gamma$,则$\mathrm{X}(\beta-\gamma)=0$,因为X列满秩,则$\beta=\gamma$)

拟合值向量 $\hat{\mathrm{Y}}=\mathrm{X}\hat{\beta}=\mathrm{P_X Y}$ 即投影

残差向量$\mathrm{e}=\mathrm{Y}-\hat{\mathrm{Y}}=(I-\mathrm{P_X})\mathrm{Y}$为$Y$在$\mathrm{L(X)}^{\perp}$上的投影,



正交分解：$\mathbf{Y}=\hat{\mathbf{Y}}\oplus\mathbf{e}$, $\mathbf{e}\perp\hat{\mathbf{Y}}$

拟合优度: 复相关系数平方$\mathbf{R^2}$

$\mathrm{Y}=\hat{\mathrm{Y}}\oplus\mathbf{e}$ (正交分解，$\mathbf{e}\perp\hat{Y}$)

因为X的第一列为**1**

$\Rightarrow \mathrm{Y}-\mathbf{1}\bar{Y}=(\hat{\mathrm{Y}}-\mathbf{1}\bar{Y})\oplus\mathbf{e}$ ($\mathbf{e}\perp\hat{\mathrm{Y}}-\mathbf{1}\bar{Y}$)

$\Rightarrow \|\mathrm{Y}-\mathbf{1}\bar{Y}\|^2=\|\hat{\mathrm{Y}}-\mathbf{1}\bar{Y}\|^2+\|\mathbf{e}\|^2$

记为 $\mathrm{SS}_Y = \mathrm{SS}_{\hat{Y}} + \mathrm{RSS}$

或 $\mathrm{SS}_{总} = \mathrm{SS}_{回} + \mathrm{RSS}$

定义：复相关系数平方 $R^2=\dfrac{\mathrm{SS}_{\hat{Y}}}{\mathrm{SS}_Y}=\dfrac{\|\hat{Y}-\mathbf{1}\bar{Y}\|^2}{\|Y-\mathbf{1}\bar{Y}\|^2}$

Adjusted R-squared: $\bar{R}^2=1-(1-R^2)\dfrac{n-1}{n-p}$

命题. 复相关平方为响应向量Y与拟合值向量$\hat{Y}$的相关系数的平方，即$R^2=[r_{Y\hat{Y}}]^2$

证明：$[r_{Y\hat{Y}}]^2=\dfrac{\left(\sum(Y_i-\bar{Y})(\hat{Y}_i-\bar{Y})\right)^2}{\sum(Y_i-\bar{Y})^2\sum(\hat{Y}_i-\bar{Y})^2}=\dfrac{\left((Y-\mathbf{1}\bar{Y})'(\hat{Y}-\mathbf{1}\bar{Y})\right)^2}{\|Y-\mathbf{1}\bar{Y}\|^2\|\hat{Y}-\mathbf{1}\bar{Y}\|^2}$.

因为$(\hat{Y}-\mathbf{1}\bar{Y})\in L(X)$,所以$\mathbf{e}'(\hat{Y}-\mathbf{1}\bar{Y})=0$,

$(Y-\mathbf{1}\bar{Y})'(\hat{Y}-\mathbf{1}\bar{Y})=(\hat{Y}+\mathbf{e}-\mathbf{1}\bar{Y})'(\hat{Y}-\mathbf{1}\bar{Y})$

$=(\hat{Y}-\mathbf{1}\bar{Y})'(\hat{Y}-\mathbf{1}\bar{Y})=\|\hat{Y}-\mathbf{1}\bar{Y}\|^2$

$$\Rightarrow [r_{Y\hat{Y}}]^2=\frac{\|\hat{Y}-\bar{Y}\mathbf{1}\|^2}{\|Y-\bar{Y}\mathbf{1}\|^2}=R^2$$

部分回归系数估计的表达

为了求解LS估计，可先投影得到投影$\hat{Y}$，其X后面的系数(唯一确定)即是$\hat{\beta}$. 由于 $\hat{Y}=X\hat{\beta}=\sum_{j=0}^{p-1}x_j\hat{\beta}_j$, x_j(X的第j列)的系数即为$\hat{\beta}_j$

命题1：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**1**在$X_1$中。

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

令$X_2^{\perp}=X_2-P_{X_1}X_2$，则$\beta_{(2)}$的LS估计为$\hat{\beta}_{(2)}=(X_2^{\perp}{}'X_2^{\perp})^{-1}X_2^{\perp}{}'Y$

且 $\mathrm{var}(\hat{\beta}_{(2)}\mid X)=\sigma^2(X_2^{\perp}{}'X_2^{\perp})^{-1}$

证明1：利用分块矩阵求逆公式（上节课）。

证明2：

令$X_2^{\perp}=X_2-P_{X_1}X_2$是$X_2$的正交化，则 $\hat{Y}=P_XY=P_{X_1}Y+P_{X_2^{\perp}}Y$，

其中X_2的系数，即为$\hat{\beta}_{(2)}$.

$$\begin{aligned}
\text{由}\hat{Y}&=P_{X_1}Y+P_{X_2^{\perp}}Y=X_1(X_1'X_1)^{-1}X_1Y+X_2^{\perp}(X_2^{\perp}{}'X_2^{\perp})^{-1}X_2^{\perp}{}'Y\\
&=X_1(X_1'X_1)^{-1}X_1Y+\{X_2-X_1(X_1'X_1)^{-1}X_1X_2\}(X_2^{\perp}{}'X_2^{\perp})^{-1}X_2^{\perp}{}'Y\\
&=X_1\underline{\{(X_1'X_1)^{-1}X_1Y-(X_1'X_1)^{-1}X_1X_2(X_2^{\perp}{}'X_2^{\perp})^{-1}X_2^{\perp}{}'Y\}}+X_2\underline{\underline{(X_2^{\perp}{}'X_2^{\perp})^{-1}X_2^{\perp}{}'Y}}\\
&=X_1\underline{\hat{\beta}_{(1)}}+X_2\underline{\underline{\hat{\beta}_{(2)}}}\text{，所以}\hat{\beta}_{(2)}=(X_2^{\perp}{}'X_2^{\perp})^{-1}X_2^{\perp}{}'Y
\end{aligned}$$

证明3：为了求解$\beta_{(2)}$的LS估计，分解$X_2=X_2^{\perp}+P_{X_1}X_2$，改写模型：

$$\begin{aligned}
Y&=X_1\beta_{(1)}+X_2\beta_{(2)}+\varepsilon=X_1\beta_{(1)}+(X_2^{\perp}+P_{X_1}X_2)\beta_{(2)}+\varepsilon\\
&=X_1\left[\beta_{(1)}+(X_1'X_1)^{-1}X_1'X_2\beta_{(2)}\right]+X_2^{\perp}\beta_{(2)}+\varepsilon
\end{aligned}$$

令$\beta_{(1)}^*=\beta_{(1)}+(X_1'X_1)^{-1}X_1'X_2\beta_{(2)}$，模型等价地表示为：

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

(1) 由$\hat{Y}=P_XY=P_{X_1}Y+P_{X_2^{\perp}}Y=X_1(X_1'X_1)^{-1}X_1'Y+X_2^{\perp}\underline{(X_2^{\perp}{}'X_2^{\perp})^{-1}X_2^{\perp}{}'Y}$

$X_2^{\perp}$后面的系数 $(X_2^{\perp}{}'X_2^{\perp})^{-1}X_2^{\perp}{}'Y$ 即为$\hat{\beta}_{(2)}$，或更简单地

(2) 由于$X^*{}'X^*$是对角分块矩阵

$$\begin{pmatrix}\hat{\beta}_{(1)}^*\\ \hat{\beta}_{(2)}\end{pmatrix}=(X^*{}'X^*)^{-1}X^*{}'Y=\begin{pmatrix}X_1'X_1 & 0\\ 0 & X_2^{\perp}{}'X_2^{\perp}\end{pmatrix}^{-1}\begin{pmatrix}X_1'Y\\ X_2^{\perp}{}'Y\end{pmatrix}=\begin{pmatrix}(X_1'X_1)^{-1}X_1'Y\\ (X_2^{\perp}{}'X_2^{\perp})^{-1}X_2^{\perp}{}'Y\end{pmatrix}$$

注1：$P_{X_2^{\perp}}Y=X_2^{\perp}\hat{\beta}_{(2)}$可看作是$X_2$对$\hat{Y}$的单独的贡献。

比如检验$H0:\beta_{(2)}=0$，将主要依据$\|X_2^{\perp}\hat{\beta}_{(2)}\|^2$的大小，

F检验$\propto\|X_2^{\perp}\hat{\beta}_{(2)}\|^2/\mathrm{RSS}$

注2：设**1**在X_1中，$Y-\mathbf{1}\bar{Y}=(\hat{Y}-\mathbf{1}\bar{Y})\oplus\mathbf{e}=(X_1\hat{\beta}_{(1)}^*-\mathbf{1}\bar{Y})\oplus X_2^{\perp}\hat{\beta}_{(2)}\oplus\mathbf{e}$

$\|Y-\mathbf{1}\bar{Y}\|^2=\|\hat{Y}-\mathbf{1}\bar{Y}\|^2+\mathrm{RSS}=\|X_1\hat{\beta}_{(1)}^*-\mathbf{1}\bar{Y}\|^2+\|X_2^{\perp}\hat{\beta}_{(2)}\|^2+RSS$

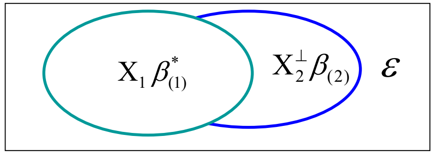
注3：若$X_1'X_2=\mathbf{0}$ 则$\hat{\beta}_{(1)}=(X_1'X_1)^{-1}X_1'Y$，$\hat{\beta}_{(2)}=(X_2'X_2)^{-1}X_2'Y$

若 $X=(x_0,x_1,\ldots,x_{p-1})$的各列正交，则 $\hat{\beta}_k=(x_k'x_k)^{-1}x_k'Y$

总结: $X_2^{\perp}=X_2-P_{X_1}X_2$

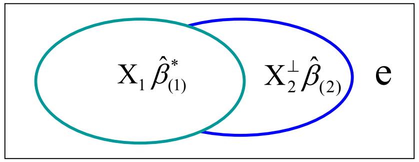
模型：$Y=X_1\beta_{(1)}+X_2\beta_{(2)}+\varepsilon=X_1\beta_{(1)}^{*}+X_2^{\perp}\beta_{(2)}+\varepsilon, X_1'X_2^{\perp}=0$

$Y=$



拟合方程：$Y=\hat{Y}\oplus e=X_1\hat{\beta}_{(1)}^{*}\oplus X_2^{\perp}\hat{\beta}_{(2)}\oplus e$

$Y=$



特例1. 单个回归系数的LS估计

命题2.令$X_{(-k)}$为X除了第k列$\mathbf{x}_k$之外的其它列组成的矩阵,

$X=(X_{(-k)},\mathbf{x}_k)$，令$\mathbf{x}_k^{\perp}=\mathbf{x}_k-\hat{\mathbf{x}}_k, \hat{\mathbf{x}}_k=P_{X_{(-k)}}\mathbf{x}_k$，则

$\hat{\beta}_k=(\mathbf{x}_k^{\perp}{}'\mathbf{x}_k^{\perp})^{-1}\mathbf{x}_k^{\perp}{}'Y=s_{x_k^{\perp}y}/s_{x_k^{\perp}x_k^{\perp}}$，且$\mathrm{var}(\hat{\beta}_k\mid X)=\sigma^2(\mathbf{x}_k^{\perp}{}'\mathbf{x}_k^{\perp})^{-1}$

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

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

其中$\dfrac{1}{1-R_k^2}=\dfrac{\|x_k-\mathbf{1}\bar{x}_k\|^2}{\|x_k^{\perp}\|^2}$称为方差扩大因子($VIF$, variance inflation factor),

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