回归分析 (01714601)
课程主页：http://staff.ustc.edu.cn/~ynyang/lm2020

第十二讲 最小二乘估计

2020.3.27

$$\hat{\boldsymbol{\beta}}_J = (X_J^{\perp\top} X_J^{\perp})^{-1} X_J^{\perp\top} \mathbf{y}$$

内容

- LS估计的方差
- LS估计的分量
- 中心化
- 复相关系数平方

线性回归模型：$\mathbf{y} = X_{n\times p}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2 I_n)$，$\boldsymbol{\varepsilon}$ 与X 独立。

最小二乘法：$\min\limits_{\boldsymbol{\beta}\in R^p} \sum \left(y_i - \mathbf{x}_i{}^\top \boldsymbol{\beta}\right)^2 = \min\limits_{\boldsymbol{\beta}\in R^p} \| \mathbf{y} - X\boldsymbol{\beta} \|^2 = \min\limits_{\mathbf{u}\in L(\mathrm{X})} \| \mathbf{y} - \mathbf{u} \|^2$

正则方程：$X^\top (\mathbf{y} - X\hat{\boldsymbol{\beta}}) = \sum\limits_{i=1}^{n} \mathbf{x}_i \left(y_i - \mathbf{x}_i{}^\top \hat{\boldsymbol{\beta}}\right) = 0$

LS估计：$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y} = \left(\sum\limits_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i{}^\top\right)^{-1} \sum\limits_{i=1}^{n} \mathbf{x}_i y_i$

最小二乘估计的方差

$$\operatorname{var}(\hat{\boldsymbol\beta} \mid X) = \sigma^2 \left(X^\top X\right)^{-1}$$

第四讲命题6. $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} > 0$, 则

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11\bullet 2}^{-1} & -\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22\bullet 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \Sigma_{22\bullet 1}^{-1} \end{pmatrix}$$

其中$\Sigma_{11\bullet 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$，$\Sigma_{22\bullet 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$，

这是因为$\operatorname{var}(\hat{\boldsymbol\beta} \mid X) = \operatorname{var}(\left(X^\top X\right)^{-1} X^\top \mathbf{y} \mid X) = \left(X^\top X\right)^{-1} X^\top [\operatorname{var}(\mathbf{y} \mid X)] X \left(X^\top X\right)^{-1}$

$= \left(X^\top X\right)^{-1} X^\top \left[\sigma^2 I_n\right] X \left(X^\top X\right)^{-1} = \sigma^2 \left(X^\top X\right)^{-1}$

注：$\left(X^\top X\right)^{-1}$基本上是一个协方差阵的逆。

设X第一列为$\mathbf{1}$，划分$\boldsymbol\beta = \begin{pmatrix} \beta_0 \\ \boldsymbol\gamma \end{pmatrix}$, $X = (\mathbf{1}, Z)$，$X^\top X = \begin{pmatrix} \mathbf{1}^\top \\ Z^\top \end{pmatrix} (\mathbf{1}, Z) = \begin{pmatrix} \mathbf{1}^\top \mathbf{1} & \mathbf{1}^\top Z \\ Z^\top \mathbf{1} & Z^\top Z \end{pmatrix}$

$\operatorname{var}(\hat{\boldsymbol\gamma} \mid X)/\sigma^2 = \left(X^\top X\right)^{-1}$右下角$= \left(Z^\top Z - Z^\top \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top Z\right)^{-1} = \left(Z^\top (I_n - P_1) Z\right)^{-1}$

$= \left(Z_{(c)}{}^\top Z_{(c)}\right)^{-1} = \left((n-1)S\right)^{-1}$, S为自变量的样本协方差矩阵.

Plug - in σ^2的LS估计, 即得到$\hat{\boldsymbol{\beta}}$方差的估计：

$$\widehat{\operatorname{var}(\hat{\boldsymbol{\beta}} \mid X)} = \hat{\sigma}^2 \left(X^\top X\right)^{-1}$$

LS估计的分量

记号：$\{0,\ldots,p-1\}$为$\boldsymbol{\beta}_{p\times 1}=(\beta_0,\ldots,\beta_{p-1})^T$分量和$X_{n\times p}=(1,\mathbf{x}_{(1)},\ldots,\mathbf{x}_{(p-1)})$列的标号集合，
设$J\subset\{0,\ldots,p-1\}$为任何一个非空标号集合，
$\boldsymbol{\beta}_J$为下标属于J的$\boldsymbol{\beta}$的分量构成，其它分量构成$\boldsymbol{\beta}_{-J}$；
X_J为标号属于J的X列组成的矩阵，其它列构成X_{-J}.

命题1：设$J\subset\{0,\ldots,p-1\}$为任何一个非空标号集合，模型为

$$\mathbf{y}_{n\times 1}=X_{n\times p}\boldsymbol{\beta}_{p\times 1}+\boldsymbol{\varepsilon}_{n\times 1}=X_J\boldsymbol{\beta}_J+X_{-J}\boldsymbol{\beta}_{-J}+\boldsymbol{\varepsilon}$$

X的第一列为$\mathbf{1}$，记$X_J^{\perp}=X_J-P_{X_{-J}}X_J$，则

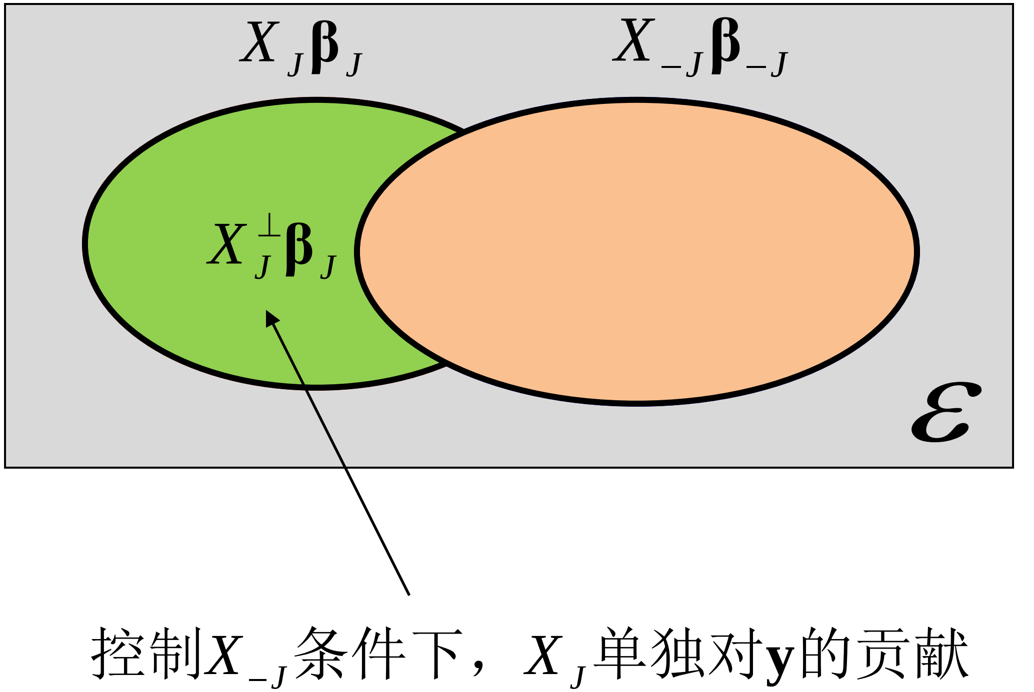
$$\hat{\boldsymbol{\beta}}_J=(X_J^{\perp\mathsf{T}}X_J^{\perp})^{-1}X_J^{\perp\mathsf{T}}\mathbf{y},$$

其方差

$$\operatorname{var}(\hat{\boldsymbol{\beta}}_J\mid X)=\sigma^2(X_J^{\perp\mathsf{T}}X_J^{\perp})^{-1}.$$

注：$\boldsymbol{\beta}$的部分分量的LS 估计及其方差与$\boldsymbol{\beta}$的LS估计具有相同的形式，
但其中是涉及到的“设计阵”需正交化。

$$\mathbf{y} = X_J \boldsymbol{\beta}_J + X_{-J} \boldsymbol{\beta}_{-J} + \boldsymbol{\varepsilon}$$

$\mathbf{y} =$ 

不妨设$J=\{0,\dots,q\}$,划分$\boldsymbol{\beta}=\begin{pmatrix}\boldsymbol{\beta}_{(1)}\\ \boldsymbol{\beta}_{(2)}\end{pmatrix}=\begin{pmatrix}\boldsymbol{\beta}_{J}\\ \boldsymbol{\beta}_{-J}\end{pmatrix}$, $X=(X_1,X_2)$

<u>证明1</u>（分块求逆）：$\hat{\boldsymbol{\beta}}=\begin{pmatrix}\hat{\boldsymbol{\beta}}_{(1)}\\ \hat{\boldsymbol{\beta}}_{(2)}\end{pmatrix}=\begin{pmatrix}X_1^\top X_1 & X_1^\top X_2\\ X_2^\top X_1 & X_2^\top X_2\end{pmatrix}^{-1}\begin{pmatrix}X_1^\top\mathbf{y}\\ X_2^\top\mathbf{y}\end{pmatrix}$

由分块矩阵求逆公式知$(X^\top X)^{-1}=\begin{pmatrix}X_1^\top X_1 & X_1^\top X_2\\ X_2^\top X_1 & X_2^\top X_2\end{pmatrix}^{-1}$的左上角元为：

$$\left(X_1^\top X_1-X_1^\top X_2(X_2^\top X_2)^{-1}X_2^\top X_1\right)^{-1}=\left(X_1^\top(I_n-\mathrm{P}_{X_2})X_1\right)^{-1}=\left(X_1^{\perp\top}X_1^{\perp}\right)^{-1}$$

所以$(X^\top X)^{-1}=\begin{pmatrix}\left(X_1^{\perp\top}X_1^{\perp}\right)^{-1} & -\left(X_1^{\perp\top}X_1^{\perp}\right)^{-1}X_1^\top X_2(X_2^\top X_2)^{-1}\\ *, & *\end{pmatrix}$

所以 $\hat{\boldsymbol{\beta}}_{(1)}=\left(X_1^{\perp\top}X_1^{\perp}\right)^{-1}X_1^\top\mathbf{y}-\left(X_1^{\perp\top}X_1^{\perp}\right)^{-1}X_1^\top X_2(X_2^\top X_2)^{-1}X_2^\top\mathbf{y}$

$=\left(X_1^{\perp\top}X_1^{\perp}\right)^{-1}X_1^\top\left[I-P_{X_2}\right]\mathbf{y}=\left(X_1^{\perp\top}X_1^{\perp}\right)^{-1}X_1^{\perp\top}\mathbf{y}$

证明2：注意如果$X_1^\top X_2 = 0$（列正交），则$\hat{\boldsymbol\beta}_{(1)}$和$\hat{\boldsymbol\beta}_{(2)}$容易得到：

$$\hat{\boldsymbol\beta} = \begin{pmatrix} \hat{\boldsymbol\beta}_{(1)} \\ \hat{\boldsymbol\beta}_{(2)} \end{pmatrix} = \begin{pmatrix} X_1^\top X_1 & 0 \\ 0 & X_2^\top X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1^\top \mathbf{Y} \\ X_2^\top \mathbf{Y} \end{pmatrix} = \begin{pmatrix} (X_1^\top X_1)^{-1} X_1^\top \mathbf{Y} \\ (X_2^\top X_2)^{-1} X_2^\top \mathbf{Y} \end{pmatrix}$$

对于$X_1^\top X_2 \neq 0$的情况，我们先将X_1正交化：$X_1^\perp = X_1 - \mathrm{P}_{X_2} X_1$，

则$X_2^\top X_1^\perp = 0$. 改写模型：

$$\mathbf{y} = X_1\boldsymbol\beta_{(1)} + X_2\boldsymbol\beta_{(2)} + \boldsymbol\varepsilon = (X_1^\perp + P_{X_2} X_1)\boldsymbol\beta_{(1)} + X_2\boldsymbol\beta_{(2)} + \boldsymbol\varepsilon$$

$$= X_1^\perp\boldsymbol\beta_{(1)} + X_2(X_2^\top X_2)^{-1} X_2^\top X_1\boldsymbol\beta_{(1)} + X_2\boldsymbol\beta_{(2)} + +\boldsymbol\varepsilon$$

$$= X_1^\perp\boldsymbol\beta_{(1)} + X_2\left\{(X_2^\top X_2)^{-1} X_2^\top X_1\boldsymbol\beta_{(1)} + \boldsymbol\beta_{(2)}\right\} + \boldsymbol\varepsilon$$

$$\triangleq X_1^\perp\boldsymbol\beta_{(1)} + X_2\boldsymbol\beta^*_{(2)} + +\boldsymbol\varepsilon = X^*\boldsymbol\beta^* + \boldsymbol\varepsilon,\text{（其中 } \boldsymbol\beta^*_{(2)} = \boldsymbol\beta_{(2)} + (X_2^\top X_2)^{-1} X_2^\top X_1\boldsymbol\beta_{(1)}\text{）}$$

由于$X^{*\top}X^{*}$是对角分块矩阵

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

$$=\begin{pmatrix}(X_1^{\perp\top}X_1^{\perp})^{-1}X_1^{\perp\top}\mathbf{y}\\ (X_2^{\top}X_2)^{-1}X_2^{\top}\mathbf{y}\end{pmatrix}$$

所以$\hat{\boldsymbol{\beta}}_{(1)}=(X_1^{\perp\top}X_1^{\perp})^{-1}\ X_1^{\perp\top}\mathbf{y}$

同时，我们有$\hat{\boldsymbol{\beta}}^{*}_{(2)}=(X_2^{\top}X_2)^{-1}X_2^{\top}\mathbf{y}.$

注意$\boldsymbol{\beta}_{(2)}=\boldsymbol{\beta}^{*}_{(2)}-(X_2^{\top}X_2)^{-1}X_2^{\top}X_1\boldsymbol{\beta}_{(1)}$

受证明2的启发，我们有下面更简单的证明：

证明3：由正则方程

$$0 = X^{\mathsf{T}}(\mathbf{y} - X\hat{\boldsymbol{\beta}}) = X^{\mathsf{T}}(\mathbf{y} - X_1\hat{\boldsymbol{\beta}}_{(1)} - X_2\hat{\boldsymbol{\beta}}_{(2)})$$

即$\mathbf{y} - X\hat{\boldsymbol{\beta}}$与$X$列向量张成的空间$L(X)$正交，特别地与$X_1^{\perp}$各列正交：

$$X_1^{\perp\mathsf{T}}(\mathbf{y} - X_1\hat{\boldsymbol{\beta}}_{(1)} - X_2\hat{\boldsymbol{\beta}}_{(2)}) = 0$$

因为$X_1^{\perp\mathsf{T}}X_2 = 0,\ X_1^{\perp\mathsf{T}}X_1 = X_1^{\perp\mathsf{T}}X_1^{\perp},$ 所以上式为

$$X_1^{\perp\mathsf{T}}\mathbf{y} - X_1^{\perp\mathsf{T}}X_1^{\perp}\hat{\boldsymbol{\beta}}_{(1)} = 0$$

$$\Rightarrow \hat{\boldsymbol{\beta}}_{(1)} = (X_1^{\perp\mathsf{T}}X_1^{\perp})^{-1}X_1^{\perp\mathsf{T}}\mathbf{y}$$

中心化

矩阵中心化：

对于数据矩阵$Z_{n\times m} = \begin{pmatrix} \mathbf{x}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{x}_n^{\mathsf{T}} \end{pmatrix}$，其中$n$行代表样本，$m$列为变量。

$\bar{\mathbf{x}} = Z^{\mathsf{T}}\mathbf{1}/n = (\mathbf{x}_1 + \ldots + \mathbf{x}_n)/n$，$Z$的中心化矩阵为

$$Z_{(c)} = Z - \mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}} = Z - \mathbf{1}\mathbf{1}^{\mathsf{T}}Z/n = (I_n - P_{\mathbf{1}})Z$$

其中$P_{\mathbf{1}} = \mathbf{1}\mathbf{1}^{\mathsf{T}}/n = \mathbf{1}(\mathbf{1}^{\mathsf{T}}\mathbf{1})^{-1}\mathbf{1}^{\mathsf{T}}$，样本协方差矩阵为

$$S = \frac{1}{n-1} Z_{(c)}^{\mathsf{T}} Z_{(c)} = \frac{1}{n-1} Z^{\mathsf{T}}(I_n - P_{\mathbf{1}})Z.$$

- 归一化：Z每列除以该列的模长。
- 标准化：Z每列减去该列的均值除以标准差