2014.4.3

第十四讲. 最小二乘

1. 多重线性回归模型

(总体)模型：

$y = \mathrm{x}'\beta + \varepsilon = \beta_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$, 其中

(1) $\varepsilon \sim (0, \sigma^2)$,

(2) ε与x独立。

有时也采用如下模型假设

y是响应变量，$\mathrm{x} = (1, x_2, \ldots, x_p)'$是$p \times 1$向量(自变量),

多重线性模型 (总体模型)

均值线性: $E(y \mid \mathrm{x}) = \mathrm{x}'\beta = \beta_1 + \beta_2 x_2 + \ldots + \beta_p x_p$,

方差齐性: $\mathrm{var}(y \mid \mathrm{x}) = \sigma^2$

模型蕴含了如下事实:

(1)均值线性：各个自变量对响应变量的效应是可加的,
且回归系数不受其它自变量取值的影响,
（否则考虑高阶模型，引入交互作用).

(2) 误差均值为0：可识别,否则截距项无法估计；

(3) 方差齐性:数据的一致性（不是必须的假设）

(4) 误差项与自变量独立: 保证回归系数估计的无偏性

回归系数正比于偏相关系数

以$x_{(-k)}$表示x向量中除了x_k之外其他分量组成的向量,

$\beta_{(-k)}$表示β向量中除了β_k之外其他分量组成的向量,

令 $x_k^{\perp} = x_k - \Sigma_{k(-k)}\Sigma_{(-k)(-k)}^{-1} x_{(-k)}$, 改写

$y = \beta_k x_k + \beta_{(-k)} x_{(-k)} + \varepsilon = \beta_k x_k^{\perp} + \tilde{\beta}_{(-k)} x_{(-k)} + \varepsilon$

其中$\tilde{\beta}_{(-k)} = \beta_{(-k)} + \beta_k \Sigma_{k(-k)} \Sigma_{(-k)(-k)}^{-1}$

所以$\mathrm{cov}(y, x_k^{\perp}) = \mathrm{cov}(\beta_k x_k^{\perp} + \tilde{\beta}_{(-k)} x_{(-k)} + \varepsilon, x_k^{\perp}) = \beta_k\, \mathrm{var}(x_k^{\perp})$

所以$\beta_k = \dfrac{\mathrm{cov}(y, x_k^{\perp})}{\mathrm{var}(x_k^{\perp})} = \dfrac{\mathrm{cov}(y^{\perp}, x_k^{\perp})}{\mathrm{var}(x_k^{\perp})} = \rho_{y^{\perp}x_k^{\perp}} \sqrt{\dfrac{\mathrm{var}(y^{\perp})}{\mathrm{var}(x_k^{\perp})}} \propto \rho_{y^{\perp}x_k^{\perp}}$

模型的数据形式

数据：$(y_i, \tilde{\mathbf{x}}_i)$，$i = 1,2\ldots,n$独立，其中 $p\times 1$向量$\tilde{\mathbf{x}}_i$为自变量的第 i 个观察值，y_i为响应变量,$i = 1,2\ldots,n$.

假设$(y_i, \tilde{\mathbf{x}}_i)$，$i = 1,2\ldots,n$,独立，满足多重线性回归模型：

$$y = \mathbf{x}'\beta + \varepsilon,\quad \varepsilon \sim (0,\sigma^2),\quad \varepsilon与\mathbf{x}独立$$

其中β为 $p\times 1$回归系数列向量。即

线性回归模型：

$$y_i = \tilde{\mathbf{x}}_i'\beta + \varepsilon_i, i = 1,2\ldots,n$$

其中$\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ iid $\sim (0,\sigma^2)$，且ε_i与$\tilde{\mathbf{x}}_i$独立。

记

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix},\quad X = \begin{pmatrix} \tilde{\mathrm{x}}_1' \\ \tilde{\mathrm{x}}_2' \\ \cdots \\ \tilde{\mathrm{x}}_n' \end{pmatrix},\quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_p \end{pmatrix},\quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix},$$

其中 X 称为设计阵(第i行为$\tilde{\mathrm{x}}_i'$)。

线性回归模型（矩阵-向量形式）：

$$\mathbf{Y}_{n\times 1} = X_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\varepsilon}_{n\times 1},\quad \boldsymbol{\varepsilon} \sim (\mathbf{0},\sigma^2 I_n),\quad \boldsymbol{\varepsilon}与X独立。$$

或有时用前两阶矩定义 模型：

(a) $\mathrm{E}(\mathbf{Y}|X) = X\boldsymbol{\beta}$,

(b) $\mathrm{var}(\mathbf{Y}|X) = \sigma^2 I_n$

2. 最小二乘

最小二乘法：

假设 X列满秩 (必要条件 $:n \ge p$)，最小二乘法 (LS):

$$\min_\beta \sum (y_i - \tilde{\mathrm{x}}_i'\beta)^2 = \min_\beta \|\mathbf{Y} - X\boldsymbol{\beta}\|^2$$

求导 (见下页)得正则方程 ：

$$X'(\mathbf{Y} - X\boldsymbol{\beta}) = 0$$

$$\Rightarrow X'\mathbf{Y} = X'X\boldsymbol{\beta}$$

$$\Rightarrow \text{LS估计：}\quad \hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{Y}$$

$$\text{且 } \mathrm{var}(\hat{\boldsymbol{\beta}}\,|\,X) = \sigma^2(X'X)^{-1}$$

矩方法：$0 = X'\varepsilon = X'(\mathbf{Y} - X\boldsymbol{\beta})$
由ε与X的各列独立

pre-conditioning:
方程$Y = X\beta + \varepsilon$两边同乘X'

关于矩阵/向量求导：

设$y = f(X)$为一矩阵函数,
其中$X = (x_{ij},\ {\scriptstyle 1\le i\le n,\,1\le j\le m})$为$n\times m$矩阵，
定义 $\dfrac{\partial y}{\partial X} = \left(\dfrac{\partial y}{\partial x_{ij}},\ 1\le i\le n,\ 1\le j\le m\right)$

特别地

（1）若 $\mathrm{a}, \mathrm{x} \in \mathrm{R}^n$，$y = \mathrm{a}'\mathrm{x}$，则 $\dfrac{\partial(\mathrm{a}'\mathrm{x})}{\partial \mathrm{x}} = \mathrm{a}$

（2）若 A为$n\times n$对称矩阵 ，$\mathrm{x} \in \mathrm{R}^n$,则 $\dfrac{\partial(\mathrm{x}'\mathrm{A}\mathrm{x})}{\partial \mathrm{x}} = 2\mathrm{A}\mathrm{x}$

$$LS: \|\mathbf{Y} - X\boldsymbol{\beta}\|^2 = \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'X\boldsymbol{\beta} + \boldsymbol{\beta}'X'X\boldsymbol{\beta}$$

$$\text{则}\frac{\partial(\|\mathbf{Y} - X\boldsymbol{\beta}\|^2)}{\partial\boldsymbol{\beta}} = -2(\mathbf{Y}'X)' + 2X'X\boldsymbol{\beta} = -2X'(\mathbf{Y} - X\boldsymbol{\beta})$$

部分回归系数的LS估计：

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

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

$X=(X_{(-k)},\mathbf{x}_k)$，则第$k$个回归系数$\beta_k$的估计：

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

其中$\mathbf{x}_k^{\perp}=\mathbf{x}_k-\mathrm{P}_{X_{(-k)}}\mathbf{x}_k$

若$\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}>0$，则$\Sigma^{-1}=\begin{pmatrix}\Sigma_{11\bullet2}^{-1} & -\Sigma_{11\bullet2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\\ -\Sigma_{22\bullet1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \Sigma_{22\bullet1}^{-1}\end{pmatrix}$

证明：由于$X=(X_1,X_2)$，由分块矩阵求逆公式知

$$(X'X)^{-1}=\begin{pmatrix}X_1'X_1 & X_1'X_2\\ X_2'X_1 & X_2'X_2\end{pmatrix}^{-1}=\begin{pmatrix}*, & *\\ *, & (X_2'X_2-X_2'X_1(X_1'X_1)^{-1}X_1'X_2)^{-1}\end{pmatrix}$$

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

另外$X'Y=\begin{pmatrix}X_1'Y\\ X_2'Y\end{pmatrix}$，所以$\hat\beta=\begin{pmatrix}\hat\beta_{(1)}\\ \hat\beta_{(2)}\end{pmatrix}=\begin{pmatrix}X_1'X_1 & X_1'X_2\\ X_2'X_1 & X_2'X_2\end{pmatrix}^{-1}\begin{pmatrix}X_1'Y\\ X_2'Y\end{pmatrix}$

所以 $\hat\beta_{(2)}=-\left(X_2^{\perp}{}'X_2^{\perp}\right)^{-1}X_2'X_1(X_1'X_1)^{-1}X_1Y+\left(X_2^{\perp}{}'X_2^{\perp}\right)^{-1}X_2'Y$

$=\left(X_2^{\perp}{}'X_2^{\perp}\right)^{-1}\left[X_2'-X_2'X_1(X_1'X_1)^{-1}X_1\right]Y=\left(X_2^{\perp}{}'X_2^{\perp}\right)^{-1}X_2^{\perp}{}'Y$