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

第十五讲 LS估计的优良性

2020.4.8

$$R^2_{J\bullet I} = \frac{R^2_{I\cup J} - R^2_I}{1 - R^2_I}, \quad X = (\mathbf{1}, X_I, X_J)$$

回顾：复偏相关系数平方

第14讲命题4已更新，强调 $R^2_{J\bullet I}$，删除了记号$R^2_{J\backslash I}$

$X=(\mathbf{1},X_I,X_J)$

(第14讲)命题4. 对于模型 $\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}=\mathbf{1}\beta_0+X_I\boldsymbol{\beta}_I+X_J\boldsymbol{\beta}_J+\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim(0,\sigma^2)$

其中$I\subset\{1,2,\ldots,p-1\}$, $J=\{1,2,\ldots,p-1\}\backslash I$. 则$R^2=R^2_{I\cup J}$ 可分解为

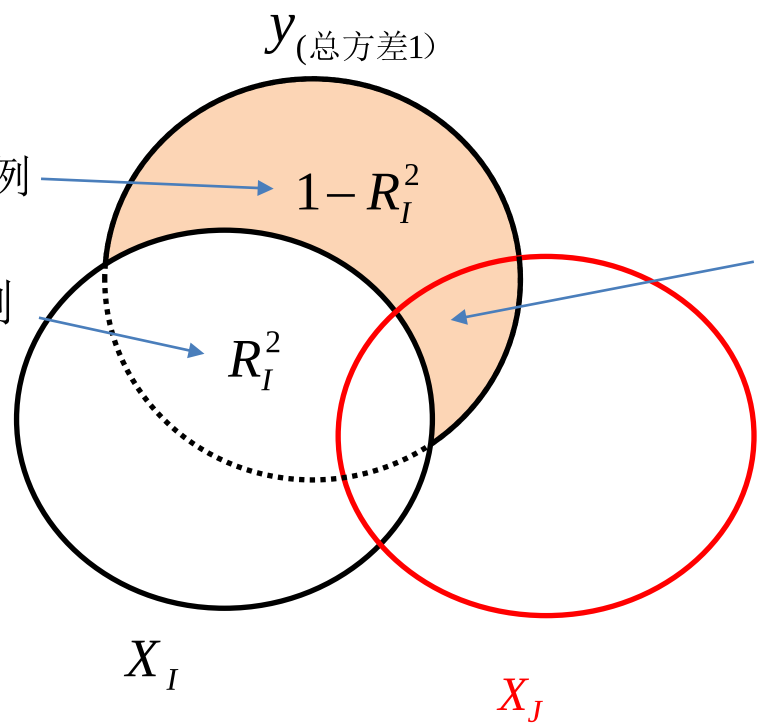
$$R^2_{I\cup J}=\frac{\hat{\Sigma}_{y,I}\hat{\Sigma}^{-1}_{I,I}\hat{\Sigma}_{I,y}}{\hat{\Sigma}_{yy}}+\frac{\hat{\Sigma}_{yJ\bullet I}\hat{\Sigma}^{-1}_{JJ\bullet I}\hat{\Sigma}_{Jy\bullet I}}{\hat{\Sigma}_{yy}}\triangleq R^2_I+(1-R^2_I)R^2_{J\bullet I}$$

其中$R^2_{J\bullet I}=\dfrac{\hat{\Sigma}_{yJ\bullet I}\hat{\Sigma}^{-1}_{JJ\bullet I}\hat{\Sigma}_{Jy\bullet I}}{\hat{\Sigma}_{yy\bullet I}}$,称为复偏相关系数平方，其中$\hat{\Sigma}_{y,I}$ 为y与$\mathbf{x}_I$的样本协方差阵,$\hat{\Sigma}_{I,I}$为$\mathbf{x}_I$的样本协方差矩阵等等.

定义：$R^2_{J\bullet I}=\dfrac{\hat{\Sigma}_{yJ\bullet I}\hat{\Sigma}^{-1}_{JJ\bullet I}\hat{\Sigma}_{Jy\bullet I}}{\hat{\Sigma}_{yy\bullet I}}$ 称为(复)偏相关系数平方

$y \sim X_I$

$y \sim X_I + X_J$



X_I不能解释的比例

X_I能解释的比例

X_J的加入能进一步解释$R^2_{J\setminus I}$，

$\Delta R^2 = R^2_{J\cup I} - R^2_I$

ΔR^2在$1-R_I^2$中的占比$\dfrac{\Delta R^2}{1-R_I^2}$

度量了控制X_I之后，

X_J对解释y方差的贡献。

$$R^2_{J\bullet I} \triangleq \frac{R^2_{J\cup I} - R^2_I}{1-R^2_I} = \frac{\hat{\Sigma}_{yJ\bullet I}\hat{\Sigma}^{-1}_{JJ\bullet I}\hat{\Sigma}_{Jy\bullet I}}{\hat{\Sigma}_{yy\bullet I}}$$ 称为(复)偏相关系数平方.

$R^2_{J\bullet I}$应该记为$R^2_{yJ\bullet I}$

命题4的证明基于正交分解,证明过程表明$R_{J\bullet I}^2$有如下表示

$$\boxed{R_{J\bullet I}^2 = \frac{\| P_V \mathbf{y} - P_{V_0}\mathbf{y} \|^2}{\| \mathbf{y} - P_{V_0}\mathbf{y} \|^2}, V = L(X) = L(\mathbf{1}, X_I, X_J), \quad V_0 = L(\mathbf{1}, X_I).}$$

该式子给出了$R_{J\bullet I}^2$的几何解释. 后面我们将会看到，检验X_J是否应该出现在模型中的F检验主要依赖$R_{J\bullet I}^2$：

$$F = \frac{n-p}{|J|} \times \frac{R_{J\bullet I}^2}{1-R_{J\bullet I}^2}$$

而卡方检验$W = (n-p) \times R_{J\bullet I}^2$.

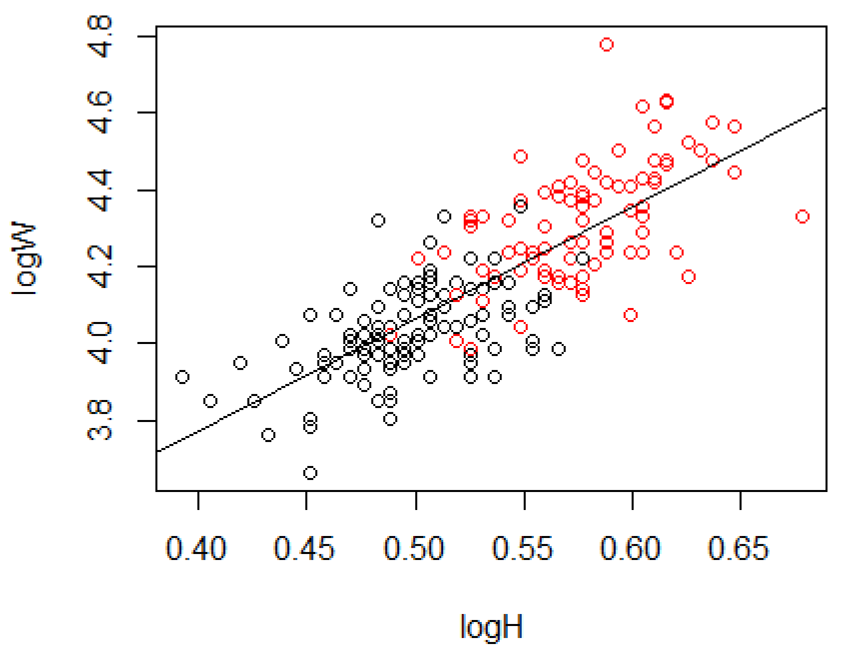
当只检验一个回归系数时$J = \{j\}, |J| = 1,$ 记$r = R_{j\bullet I}$

$$t = \pm\sqrt{F} = (n-p)\frac{r}{\sqrt{1-r^2}}.$$

案例：体重与身高的关系

例1. 数据集（http://staff.ustc.edu.cn/~ynyang/lm2020/lab/height-weight.txt）包含三个变量为 W (weight,kg), H (height,m), sex(1男,0女)，我们研究W与H的关系，制定一个标准化体重指数。

方法1：不考虑性别，拟合简单回归模型 $:\log(\mathrm{W}) = a + b \times \log(\mathrm{H}) + \varepsilon$



假设误差正态，$\log(\mathrm{W}) = a + b \times \log(\mathrm{H}) + \varepsilon$，$\varepsilon \sim N(0, \sigma^2) \Rightarrow$

$$\varepsilon / \sigma = \left(\log(\mathrm{W}) - a - b \times \log(\mathrm{H})\right) / \sigma \sim N(0,1)$$

ε / σ是一个标准化的量，我们用它作为体重指数：

如果某个人的ε / σ在值超过80%的人，则属于偏胖，即

$$\varepsilon / \sigma > 0.842 \quad \Leftrightarrow \quad \mathrm{W} / \mathrm{H}^b > \exp(a + 0.842\sigma)$$

简单回归得到：$\hat{a} = 2.599, \hat{b} = 2.930, \hat{\sigma} = 0.122$，判别准则为

$$\mathrm{W} / \mathrm{H}^{2.933} > \exp(2.599 + 0.842 * 0.122) = 14.9$$

上述方式是否合理？若$W = 80\mathrm{kg}, H = 1.78m$,则$\mathrm{W} / \mathrm{H}^{2.933} = 14.7$,这表明这个人的体重指数没有超过80%的人，比较正常。这显然不太合理。

方法2：在回归分析中控制性别

$$\log\mathrm{W} = a + b \times \log\mathrm{H} + c \times \mathrm{sex} + \varepsilon$$

$$\mathrm{sex} = 1 : \log\mathrm{W} = (a + c) + b \times \log\mathrm{H} + \varepsilon, \varepsilon \sim (0, \sigma^2)$$

$$\mathrm{sex} = 0 : \log\mathrm{W} = \quad (a) + b \times \log\mathrm{H} + \varepsilon, \varepsilon \sim (0, \sigma^2)$$

相关系数矩阵：
性别与体重、身高都有关

$$\begin{array}{c c} & \begin{array}{ccc} \log\mathrm{H} & \mathrm{sex} & \log\mathrm{W} \end{array} \\ \begin{array}{c} \log\mathrm{H} \\ \mathrm{sex} \\ \log\mathrm{W} \end{array} & \begin{pmatrix} 1 & 0.74 & 0.78 \\ 0.74 & 1 & 0.72 \\ 0.78 & 0.72 & 1 \end{pmatrix} \end{array}$$

这表明两变量线性模型蕴含了logH与sex效应的的可加性*additivity* :

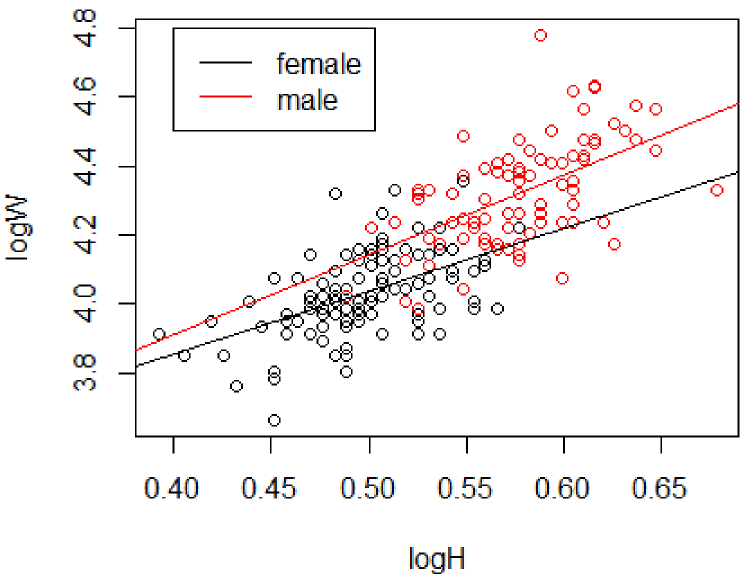
logH的效应/回归系数与sex取值无关，反之亦然.

另外，误差方差是常数，与自变量的具体取值无关，特别地，男女组内的误差方差相同。

为此，我们对男、女分别回归$\log W = a + b \times \log H + \varepsilon$,

两组的$\hat{b}, \hat{\sigma}$可以认为是近似相同的:

	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$
male	2.98	2.31	0.13
female	3.12	1.83	0.10



因此我们认为前述两变量模型$\log W = a + b \times \log H + c \times \text{sex} + \varepsilon$是恰当的

拟合模型$\log W = a + b \times \log H + c \times \mathrm{sex} + \varepsilon$,得到如下结果

```
Call:
lm(formula = logW ~ logH + sex, data = hw2)

Coefficients:
                 Estimate  Std. Error t value   Pr(>|t|)
(Intercept)      3.00871    0.11560  26.028    < 2e-16 ***
logH             2.05716    0.23092   8.909    3.55e-16 ***
sex              0.12408    0.02427   5.113    7.51e-07 ***
---

Residual standard error: 0.1145 on 196 degrees of freedom
Multiple R-squared:  0.6601,    Adjusted R-squared:  0.6567
F-statistic: 190.4 on 2 and 196 DF,  p-value: < 2.2e-16
```

$\varepsilon/\sigma = \left[\log \mathrm{W} - a - b \times \log \mathrm{H} - c \times \mathrm{sex}\right]/\sigma \sim N(0,1)$

ε/σ是一个标准化的量，与自变量无关.

偏胖准则：$\varepsilon/\sigma > 0.842 \Leftrightarrow W/H^{b} > C = \exp(a + 0.842\sigma + c \times \mathrm{sex})$

代入LS估计$\hat{a} = 3.0, \hat{b} = 2, \hat{c} = 0.124, \hat{\sigma} = 0.1145$，80%阈值为

$$C = \exp(3 + 0.842 \times 0.1145 + 0.124 \times \mathrm{sex}) = \begin{cases} 22.118 & female \\ 25.038 & male \end{cases}$$

所以80%偏胖准则为：

男性：$BMI = W/H^{2} > 25$

女性：$BMI = W/H^{2} > 22$

例2. 我们从身高体重数据的样本协方差和样本均值求解LS估计和R^2

记$x_1=\log \mathrm{H}$, $x_2=\mathrm{sex}$, $y=\log \mathrm{W}$, 样本均值为: $\bar{x}_1=0.533, \bar{x}_2=0.442, \bar{y}=4.160$,

协方差矩阵为

$$\begin{array}{c} \\ x_1 \\ x_2 \\ y \end{array}\begin{array}{c} \begin{array}{ccc} x_1 & x_2 & y \end{array} \\ \begin{pmatrix} 0.00274 & 0.01925 & 0.00802 \\ 0.01925 & 0.24791 & 0.07036 \\ 0.00802 & 0.07036 & 0.03822 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cc} \hat{\Sigma}_{\mathbf{xx}} & \hat{\Sigma}_{\mathbf{x}y} \end{array} \\ \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{1y} \\ \sigma_{21} & \sigma_{22} & \sigma_{2y} \\ \sigma_{y1} & \sigma_{y2} & \sigma_{yy} \end{pmatrix} \end{array}$$

解: (1) 求解LS 估计$\hat{b}, \hat{c}, \hat{a}$:

$$\begin{pmatrix} \hat{b} \\ \hat{c} \end{pmatrix} = \hat{\Sigma}_{\mathbf{xx}}^{-1}\hat{\Sigma}_{\mathbf{x}y} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{1y} \\ \sigma_{2y} \end{pmatrix} = \begin{pmatrix} 2.053 \\ 0.124 \end{pmatrix},$$

$$\hat{b} = \hat{\Sigma}_{11\bullet 2}^{-1}\hat{\Sigma}_{1y\bullet 2} = (\sigma_{11} - \sigma_{12}{}^2/\sigma_{22})^{-1}(\sigma_{1y} - \sigma_{12}\sigma_{22}^{-1}\sigma_{2y}) = 2.05$$

$$\hat{c} = \hat{\Sigma}_{22\bullet 1}^{-1}\hat{\Sigma}_{2y\bullet 1} = (\sigma_{22} - \sigma_{21}{}^2/\sigma_{11})^{-1}(\sigma_{2y} - \sigma_{21}\sigma_{11}^{-1}\sigma_{1y}) = 0.124$$

$$\hat{a} = \bar{y} - (\bar{x}_1\hat{b} + \bar{x}_2\hat{c}) = 4.16 - 0.533\times 2.053 - 0.422\times 0.124 = 3.0$$

$$\begin{array}{cccc} & x_1 & x_2 & y \\ x_1 & 0.00274 & 0.01925 & 0.00802 \\ x_2 & 0.01925 & 0.24791 & 0.07036 \\ y & 0.00802 & 0.07036 & 0.03822 \end{array} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{1y} \\ \sigma_{21} & \sigma_{22} & \sigma_{2y} \\ \sigma_{y1} & \sigma_{y2} & \sigma_{yy} \end{pmatrix}$$

(2) 求解LS 估计$\hat{\sigma}^2$

$$\hat{\Sigma}_{yy\bullet\mathbf{x}} = \hat{\Sigma}_{yy} - \hat{\Sigma}_{y\mathbf{x}}\hat{\Sigma}_{\mathbf{xx}}^{-1}\hat{\Sigma}_{\mathbf{x}y} = \hat{\Sigma}_{yy} - \hat{\Sigma}_{y\mathbf{x}}\begin{pmatrix}\hat{b} \\ \hat{c}\end{pmatrix}$$

$$= 0.0382 - (0.00802, \quad 0.07036)\begin{pmatrix}\hat{b} \\ \hat{c}\end{pmatrix} = 0.0382 - 0.0252 = 0.0130$$

$$n = 200,\ p = 3, \text{所以} \hat{\sigma}^2 = \frac{n-1}{n-p}\hat{\Sigma}_{yy\bullet\mathbf{x}} = 0.0131$$

(3) 求R^2 $= \hat{\Sigma}_{y\mathbf{x}}\hat{\Sigma}_{\mathbf{xx}}^{-1}\hat{\Sigma}_{\mathbf{x}y} / \hat{\Sigma}_{yy} = 0.0252 / 0.0382 = 0.66$

$$R_1^2 = \frac{\hat{\Sigma}_{y1}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{1y}}{\hat{\Sigma}_{yy}} = \frac{0.00802^2}{0.00274 \times 0.03822} = 0.615, \quad r_{1y} = \sqrt{0.615} = 0.78$$

$$\Rightarrow R^2_{2\bullet 1} = \frac{R^2 - R_1^2}{1 - R_1^2} = \frac{0.66 - 0.615}{1 - 0.615} = 0.117,$$

$$\Rightarrow y, x_2\text{的偏相关系数} r_{2y\bullet 1} = \sqrt{0.117} = 0.342(\text{与}\hat{c}\text{同号})$$

$$\text{或如下直接计算：} r_{2y\bullet 1} = \frac{r_{2y} - r_{21} r_{y1}}{\sqrt{1 - r_{21}^2}\sqrt{1 - r_{y1}^2}} = \frac{0.72 - 0.74 * 0.78}{\sqrt{1 - 0.74^2}\sqrt{1 - 0.78^2}}$$

$$\text{另外，} R_2^2 = \frac{\hat{\Sigma}_{y2}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{2y}}{\hat{\Sigma}_{yy}} = \frac{0.07036^2}{0.24791 \times 0.03822} = 0.523,$$

$$R^2_{1\backslash 2} = R^2 - R_2^2 = 0.137, R^2_{1\bullet 2} = \frac{R^2_{1\backslash 2}}{1 - R_2^2} = \frac{0.137}{1 - 0.523} = 0.287$$

$$y, x_1\text{的偏相关系数} r_{1y\bullet 2} = \sqrt{0.287} = 0.536$$

最小二乘估计的性质

$E(x^2) = \mathrm{var}(x) + \left(E(x)\right)^2,$

x是$\mu = E(x)$的无偏估计，但x^2不是μ^2的无偏估计

引理1. 设$\mathbf{x}$为$n\times 1$随机向量, $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$，$\mathrm{cov}(\mathbf{x}) = \Sigma$,

A为$n\times n$常数矩阵，则

$$E\left(\mathbf{x}^{\top} A\mathbf{x}\right) = \boldsymbol{\mu}^{\top} A\boldsymbol{\mu} + \mathrm{tr}(A\Sigma)$$

证明：$\mathbf{x}^{\top} A\mathbf{x} = \mathrm{tr}(\mathbf{x}^{\top} A\mathbf{x}) = \mathrm{tr}(A\mathbf{x}\mathbf{x}^{\top}) \quad \Rightarrow E(\mathbf{x}^{\top} A\mathbf{x}) = \mathrm{tr}(AE(\mathbf{x}\mathbf{x}^{\top}))$

而$\Sigma = E(\mathbf{x}\mathbf{x}^{\top}) - \boldsymbol{\mu}\boldsymbol{\mu}^{\top}$

所以$\mathrm{tr}(AE(\mathbf{x}\mathbf{x}^{\top})) = \mathrm{tr}(A\Sigma + A\boldsymbol{\mu}\boldsymbol{\mu}^{\top})$

$= \mathrm{tr}(A\Sigma) + \mathrm{tr}(A\boldsymbol{\mu}\boldsymbol{\mu}^{\top}) = \mathrm{tr}(A\Sigma) + \boldsymbol{\mu}^{\top} A\boldsymbol{\mu}$

模型假设： $\mathbf{y}_{n\times 1} = X_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$， $\boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2 I_n)$， $\boldsymbol{\varepsilon}$ 与 X 独立

LS估计： $\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y}$，其方差 $\mathrm{var}(\hat{\boldsymbol{\beta}} \mid X) = \sigma^2 \left(X^\top X\right)^{-1}$

σ^2的LS估计: $\hat{\sigma}^2 = \dfrac{\mathrm{RSS}}{n-p} = \dfrac{\|\mathbf{y} - X\hat{\boldsymbol{\beta}}\|^2}{n-p}$

定理1. 无偏性: $\mathrm{E}(\hat{\boldsymbol{\beta}} \mid X) = \boldsymbol{\beta}$， $\mathrm{E}(\hat{\sigma}^2 \mid X) = \sigma^2$

证明：(1) $\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y} = (X^\top X)^{-1} X^\top (X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (X^\top X)^{-1} X^\top \boldsymbol{\varepsilon}$

所以$E(\hat{\boldsymbol{\beta}} \mid X) = \boldsymbol{\beta} + (X^\top X)^{-1} X^\top E\left(\boldsymbol{\varepsilon} \mid X\right) = \boldsymbol{\beta}$

最后一步是因为$\boldsymbol{\varepsilon}$与X独立$\Rightarrow$ 所以$E\left(\boldsymbol{\varepsilon} \mid X\right) = E(\boldsymbol{\varepsilon}) = 0$。

15

(2) 因为$\mathbf{e}=\mathbf{y}-X\hat{\boldsymbol{\beta}}=(I_n-P_X)\mathbf{y}=(I_n-P_X)(X\beta+\boldsymbol{\varepsilon})=(I_n-P_X)\boldsymbol{\varepsilon}$，

所以$E\left(\mathbf{e}^{\top}\mathbf{e}\mid X\right)=E\left(\boldsymbol{\varepsilon}^{\top}(I_n-P_X)\boldsymbol{\varepsilon}\mid X\right)$

$E(\boldsymbol{\varepsilon}\mid X)=0, \mathrm{var}(\boldsymbol{\varepsilon}\mid X)=\sigma^2 I_n$

$$\overset{\text{引理1}}{=}\mathrm{tr}\left((I_n-P_X)\,\mathrm{var}(\boldsymbol{\varepsilon}\mid X)\right)=\sigma^2\mathrm{tr}(I_n-P_X)=(n-p)\sigma^2$$

所以$E(\hat{\sigma}^2\mid X)=E(\mathbf{e}^{\top}\mathbf{e}/(n-p)\mid X)=\sigma^2$

推论. 由定理1，对任何p列常数矩阵A，线性组合$A\hat{\boldsymbol{\beta}}$是$A\boldsymbol{\beta}$ 的无偏估计，特别地，$\hat{\beta}_k$是β_k的无偏估计

定义(Loewner's 偏序) : 对任何两个对称 $n\times n$矩阵A, B, $A\geq B$

定义为$A-B\geq 0$(非负定）$\Leftrightarrow$ 对任何$\mathbf{x}\neq \mathbf{0}\in R^n$，$\mathbf{x}^\top A\mathbf{x}\geq \mathbf{x}^\top B\mathbf{x}$

性质：若$A\geq B$，则对任何$k\times n$矩阵C，$CAC^\top \geq CBC^\top$

定理2(Gauss - Markov定理). 最小二乘估计是BLUE

(best linear unbiased estimate, 最优无偏线性估计), 即在所有$\boldsymbol{\beta}$的

线性无偏估计中，LS估计$\hat{\boldsymbol{\beta}}$的方差最小(Loewner意义下)。

<u>Gauss - Markov定理的证明:</u>

设$\tilde{\boldsymbol{\beta}} = C_{p\times n}\mathbf{y}$是$\boldsymbol{\beta}$的任一线性无偏估计，所以

$$\boldsymbol{\beta} = \mathrm{E}(\tilde{\boldsymbol{\beta}} \mid X) = E(C\mathbf{y} \mid X) = CX\boldsymbol{\beta},$$

上式对任何$\boldsymbol{\beta}$成立，故$CX = I_p$。

因为$I_n - H \ge 0 \Updownarrow I_n \ge H$,所以

$$CC^\top \ \ge\ CHC^\top = CX(X^\top X)^{-1}X^\top C^\top = (X^\top X)^{-1},$$

最后一步是因为$CX = I_p$.

所以 $\mathrm{var}(\tilde{\boldsymbol{\beta}} \mid X) = C\,\mathrm{var}(\mathbf{y} \mid X)C^\top = \sigma^2 CC^\top$

$$\ge \sigma^2 (X^\top X)^{-1} = \mathrm{var}(\hat{\boldsymbol{\beta}} \mid X).$$

18

因为矩阵$A \geq 0 \Leftrightarrow$ 对任何向量$\mathbf{x}, \mathbf{x}^T A\mathbf{x} \geq 0.$

所以GM 定理通常叙述为：

GM定理：在所有$\mathbf{c}^\top \boldsymbol{\beta}$的线性无偏估计中$(\mathbf{c} \in R^p)$,

$\mathbf{c}^\top \hat{\boldsymbol{\beta}} = \mathbf{c}^\top (X^\top X)^{-1} X^\top \mathbf{y}$的方差最小。

特别地，$\hat{\beta}_k$是β_k的BLUE，$\hat{\beta}_k - \hat{\beta}_j$是$\beta_k - \beta_j$的BLUE，等等。

GM定理：假设$A_{k\times p}(k \leq p)$行满秩, 记参数$\theta_{k\times 1} = A\boldsymbol{\beta}$, 则

$A\hat{\boldsymbol{\beta}}$ 是$A\boldsymbol{\beta}$的BLUE

特别地，若$A = (0, I_k), A\boldsymbol{\beta} = (0, I_k)\begin{pmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{pmatrix} = \boldsymbol{\beta}_2$, 划分$X = (X_1, X_2)$,

则$\hat{\boldsymbol{\beta}}_2 = (X_2^{\perp\top} X_2^{\perp})^{-1} X_2^{\perp\top} \mathbf{y}$是$\boldsymbol{\beta}_2$的BLUE

定理3. 如果误差服从正态分布 $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$,则

(1) $\boldsymbol{\beta}$的极大似然估计$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$（LS估计）;

σ^2的极大似然估计为$\tilde{\sigma}^2_{\text{mle}} = \dfrac{1}{n}\text{RSS} = \dfrac{n-p}{n}\hat{\sigma}^2$.

(2) 在所有$\boldsymbol{\beta}$的无偏估计中(未必是线性估计),

LS估计的方差最小

证明：(1) $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$, $\mathbf{y} \mid X \sim N(X\boldsymbol{\beta}, \sigma^2 I_n)$，$\mathbf{y}$的概率密度即似然函数为$(c = (2\pi)^{-n/2})$:

$$L(\boldsymbol{\beta}, \sigma^2) = c(\sigma^2)^{-n/2} \exp\left(-\left(\| \mathbf{y} - X\boldsymbol{\beta} \|^2 / 2\sigma^2\right)\right),$$

log-似然函数为:

$$l(\boldsymbol{\beta}, \sigma^2) = \log L = \log(c) - n\log(\sigma^2)/2 - \| \mathbf{y} - X\boldsymbol{\beta} \|^2 / 2\sigma^2$$

求偏导数：

$$\frac{\partial l(\boldsymbol{\beta},\sigma^2)}{\partial \boldsymbol{\beta}} = X^\top(\mathbf{y}-X\boldsymbol{\beta})/\sigma^2 = 0$$

$$\Rightarrow X^\top(\mathbf{y}-X\boldsymbol{\beta}) = 0, \text{此即正则方程}$$

$$\Rightarrow \text{极大似然估计}\tilde{\boldsymbol{\beta}} = (X^\top X)^{-1}X^\top\mathbf{y} = \hat{\boldsymbol{\beta}}$$

$$\frac{\partial l(\boldsymbol{\beta},\sigma^2)}{\partial \sigma^2} = -n/(2\sigma^2) + \|\mathbf{y}-X\boldsymbol{\beta}\|^2/2\sigma^4 = 0$$

$$\Rightarrow \tilde{\sigma}^2 = \|\mathbf{y}-X\tilde{\boldsymbol{\beta}}\|^2/n = \mathrm{RSS}/n$$

(2) 记$\boldsymbol{\theta}=\begin{pmatrix}\boldsymbol{\beta}\\ \sigma^2\end{pmatrix}$, Fisher 信息阵为

$$I(\boldsymbol{\theta})=E\left(-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\right)=E\begin{pmatrix} X^{\top}X/\sigma^2 & \dfrac{X^{\top}(\mathbf{y}-X\boldsymbol{\beta})}{\sigma^4} \\ \dfrac{(\mathbf{y}-X\boldsymbol{\beta})^{\top}X}{\sigma^4} & -\dfrac{n}{2\sigma^4}+\dfrac{\|\mathbf{y}-X\boldsymbol{\beta}\|^2}{\sigma^6}\end{pmatrix}$$

$$=\begin{pmatrix} X^{\top}X/\sigma^2 & 0 \\ 0 & n/2\sigma^4\end{pmatrix}$$

由C-R不等式，任意$\boldsymbol{\theta}$的无偏估计$\tilde{\boldsymbol{\theta}}=\begin{pmatrix}\tilde{\boldsymbol{\beta}}\\ \tilde{\sigma}^2\end{pmatrix}$的方差(给定$X$)

$$\mathrm{var}(\tilde{\boldsymbol{\theta}})\geq \mathrm{I}(\boldsymbol{\theta})^{-1}=\begin{pmatrix}\sigma^2(X^{\top}X)^{-1} & 0 \\ 0 & 2\sigma^4/n\end{pmatrix}$$

即$\mathrm{var}(\tilde{\boldsymbol{\beta}})\geq\sigma^2(X^{\top}X)^{-1}=\mathrm{var}(\hat{\boldsymbol{\beta}})$