2014.4.15

第十七讲. LS估计的性质及F检验

最小二乘估计的性质

性质1. 无偏性：$\mathrm{E}(\hat{\beta}\mid X)=\beta$，方差：$\mathrm{var}(\hat{\beta}\mid X)=\sigma^2(X'X)^{-1}$

证明：$\mathrm{E}(\hat{\beta}\mid X)=E\{(X'X)^{-1}X'Y\mid X\}$

$=(X'X)^{-1}X'E(Y\mid X)=(X'X)^{-1}X'E\{X\beta+\varepsilon\mid X\}$

$=\beta+(X'X)^{-1}X'E\{\varepsilon\mid X\}$

因为ε与X独立，所以$E\{\varepsilon\mid X\}=E\{\varepsilon\}=0$,所以$\mathrm{E}(\hat{\beta}\mid X)=\beta$

$$\mathrm{var}(\hat{\beta}\mid X)=(X'X)^{-1}X'\mathrm{var}(\varepsilon\mid X)X(X'X)^{-1}=\sigma^2(X'X)^{-1}$$

注1：分量$\hat{\beta}_k$是β_k的无偏估计,线性组合$\mathrm{c}'\hat{\beta}$是$c'\beta$的无偏估计

性质2(相合性及渐近正态性). 设$(\tilde{\mathrm{x}}_i,y_i)$ iid满足线性模型

$y_i=\tilde{\mathrm{x}}_i'\beta+\varepsilon_i,\varepsilon_i\sim(0,\sigma^2)$与$\tilde{\mathrm{x}}_i$独立(不假设$\varepsilon_i$正态）。

设$\hat{\beta}$为LS估计，则当$n\to\infty$时,

(1) 相合性：$\hat{\beta}\xrightarrow{\mathrm{p}}\beta$;

(2)渐近正态性：$\sqrt{n}\left(\hat{\beta}-\beta\right)\xrightarrow{\mathrm{d}}N(0,\sigma^2\mathrm{M}^{-1})$，$\mathrm{M}=\mathrm{E}(\tilde{\mathrm{x}}_i\tilde{\mathrm{x}}_i')$

证明:(1) 由大数定律$\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathrm{x}}_i\tilde{\mathrm{x}}_i'\xrightarrow{\mathrm{p}}\mathrm{M}$, $\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathrm{x}}_i\varepsilon_i'\xrightarrow{\mathrm{p}}\mathrm{E}(\tilde{\mathrm{x}}_i\varepsilon_i)=0$

故$\hat{\beta}=(X'X)^{-1}X'Y=\beta+(X'X)^{-1}X'\varepsilon=\beta+\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathrm{x}}_i\tilde{\mathrm{x}}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathrm{x}}_i\varepsilon_i\right)\xrightarrow{\mathrm{p}}\beta$

(2)由中心极限定理$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mathrm{x}}_i\varepsilon_i\xrightarrow{\mathrm{d}}N(0,\sigma^2\mathrm{M})$,而$\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathrm{x}}_i\tilde{\mathrm{x}}_i'\xrightarrow{\mathrm{p}}\mathrm{M}$

故$\sqrt{n}\left(\hat{\beta}-\beta\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathrm{x}}_i\tilde{\mathrm{x}}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mathrm{x}}_i\varepsilon_i\right)\xrightarrow{\mathrm{d}}N(0,\sigma^2\mathrm{M}^{-1})$

性质 3. $\hat{\sigma}^2=\frac{\mathrm{RSS}}{n-p}$，则 $\mathrm{E}(\hat{\sigma}^2\mid X)=\sigma^2$

证明：$\hat{\sigma}^2=\frac{1}{n-p}\|\mathbf{e}\|^2=\frac{1}{n-p}\mathbf{e}'\mathbf{e}.$

因为$\mathbf{e}=(I_n-H)Y=(I_n-H)(X\beta+\varepsilon)=(I_n-H)\varepsilon$,

且$E(\varepsilon\varepsilon'\mid X)=\sigma^2I_n$,所以

$$\mathrm{E}\{\mathbf{e}'\mathbf{e}\mid X\}=E\{\varepsilon'(I_n-H)\varepsilon\mid X\}$$

$$=E\{\mathrm{tr}[(I_n-H)\varepsilon\varepsilon']\mid X\}$$

$$=\mathrm{tr}[(I_n-H)E\{\varepsilon\varepsilon'\mid X\}=\sigma^2\mathrm{tr}(I_n-H)=(n-p)\sigma^2$$

Gauss-Markov定理

定义(Loewner's 偏序):对任何两个对称 $m\times m$矩阵A,B，$A\ge B$ 定义为$A-B\ge 0$(非负定)，等价地对任何x，$x'Ax\ge x'Bx$

性质：若 $A\ge B$，则对任何 $k\times m$矩阵C，$CAC'\ge CBC'$

性质4 (Gauss-Markov定理). 最小二乘估计是最优无偏线性估计（BLUE: best linear unbiased estimate), 即在所有β的线性无偏估计中，LS估计 $\hat\beta=(X'X)^{-1}X'Y$的方差最小(Loewner偏序意义下)。

Gauss-Markov定理的证明：

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

$$\beta=\mathrm{E}(\tilde\beta\mid X)=E(CY\mid X)=CX\beta,$$

上式对任何β成立，故$CX=I_p$。

$\mathrm{var}(\tilde\beta\mid X)=C\,\mathrm{var}(Y\mid X)C'=\sigma^2CC'$，而$\mathrm{var}(\hat\beta\mid X)=\sigma^2(X'X)^{-1}$

所以需要证明因为$CC'\ge(X'X)^{-1}$

因为$I_n-H\ge 0$，所以

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

Gauss-Markov定理也表述为（课本）：

GM定理:

在所有$c'\beta$的线性无偏估计中，LS估计$c'\hat\beta=c'(X'X)^{-1}X'Y$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$A_{q\times p}(q\le p)$行满秩,记参数$\theta_{q\times1}=A\beta$,则LS估计$\hat\theta=A\hat\beta$ 是$A\beta$的BLUE

特别地，若$A=(0,I_q)$, $A\beta=(0,I_q)\begin{pmatrix}\beta_{(1)}\\ \beta_{(2)}\end{pmatrix}=\beta_{(2)}$,划分$\mathrm{X}=(\mathrm{X}_1,\mathrm{X}_2)$,

则$\hat\beta_{(2)}=(\mathrm{X}_2^{\perp\prime}\mathrm{X}_2^{\perp})^{-1}\mathrm{X}_2^{\perp\prime}Y$是$\beta_{(2)}$的BLUE

性质5. 如果误差服从正态分布$\varepsilon_i\sim N(0,\sigma^2)$,则

(1) β的极大似然估计即LS估计;

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

(3) $\hat\beta\mid \mathrm{X}\sim N\left(\beta,\sigma^2(X'X)^{-1}\right)$, $\dfrac{(n-p)\hat\sigma^2}{\sigma^2}\sim\chi^2_{n-p}$,且两者独立。

证明：$\boldsymbol{\varepsilon}\sim N(0,\sigma^2I_n)$,

(3) $\hat\beta=(X'X)^{-1}X'Y=\beta+\underline{(X'X)^{-1}X'}\boldsymbol{\varepsilon}\sim N(\beta,\sigma^2(X'X)^{-1})$

$\mathbf{e}=(I_n-H)Y=(I_n-H)(X\beta+\boldsymbol{\varepsilon})=(I_n-H)\boldsymbol{\varepsilon}$,

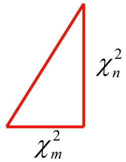
$$\Rightarrow\frac{(n-p)\hat\sigma^2}{\sigma^2}=\frac{\|\mathbf{e}\|^2}{\sigma^2}=(\varepsilon/\sigma)'\underline{(I_n-H)}(\varepsilon/\sigma)\sim\chi^2_{n-p}.$$

因为$\underline{(X'X)^{-1}X'}\ \underline{(I_n-H)}=0$,所以$\hat\beta$与$\hat\sigma^2$独立

F 检验

F-分布也叫做**Snedecor's F-**分布或者**Fisher-Snedecor**分布，
其定义为两个独立平均卡方随机变量之比的分布:

$$F = \frac{\chi_m^2 / m}{\chi_n^2 / n} \sim F_{m,n}$$



为了表示对**Fisher**的敬意，**G.W. Snedecor** 将其称作**F-**分布。

部分回归系数的显著性检验

全模型: $Y_{n\times1} = X_{n\times p}\beta_{p\times1} + \varepsilon_{n\times1} = X_1\beta_{(1)} + X_2\beta_{(2)} + \varepsilon, \quad \varepsilon \sim N(\mathbf{0}, \sigma^2 I_n)$

其中$X = (X_1, X_2)$, X_1为$n\times(p-q)$,第一列为$\mathbf{1}$.X_1为$n\times q$;

$\beta = \begin{pmatrix}\beta_{(1)}\\ \beta_{(2)}\end{pmatrix}$, $\beta_{(1)}$长度为$p-q$(含截距)，$\beta_{(2)}$长度为q.

原假设H_0：$\beta_{(2)} = 0_{q\times1}$，

子模型: $Y = X_1\beta_{(1)} + \varepsilon$, $\varepsilon \sim N(\mathbf{0}, \sigma^2 I_n)$（原假设下的模型）

令 $X_2^\perp = X_2 - P_{X_1}X_2$，$\beta_{(1)}^* = \beta_{(1)} + (X_1'X_1)^{-1}X_1'X_2\beta_{(2)})$

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

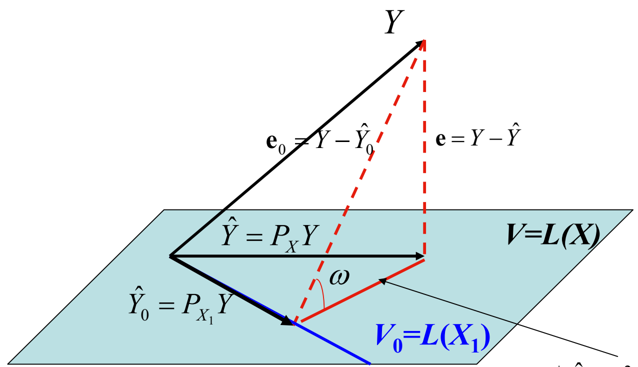
记$\hat{Y} = P_X Y$为Y在$V = L(X)$上的投影

记$\hat{Y}_0 = P_{X_1}Y$为Y在$V_0 = L(X_1) \subset V$上的投影

$$\hat{Y} = P_X Y = P_{X_1}Y + P_{X_2^\perp}Y = X_1\hat{\beta}_{(1)}^* + X_2^\perp\hat{\beta}_{(2)} \overset{记为}{=} \hat{Y}_0 + X_2^\perp\hat{\beta}_{(2)}$$

如果原假设成立时($\beta_{(2)} = 0$), 则$V = L(X)$和$V_0 = L(X_1)$差异不大，

特别地，Y在V，V_0上的投影之差 $\hat{Y} - \hat{Y}_0 = X_2^\perp\hat{\beta}_{(2)}$ 的模长差别不大。



$\| X_2^\perp\hat{\beta}_{(2)} \|^2 = \| \hat{Y} - \hat{Y}_0 \|^2$ 反映了全模型与子模型的差异

考虑到刻度以及q的大小，F检验取为

$$F = \frac{\| X_2^\perp\hat{\beta}_{(2)} \|^2 / q}{\hat{\sigma}^2} = \frac{\| \hat{Y} - \hat{Y}_0 \|^2 / q}{\| Y - \hat{Y} \|^2 / (n-p)} = \frac{n-p}{q} ctg(\omega)^2$$

定理：原假设下 $F \sim F_{q,n-p}$

证明1：利用 $F = \dfrac{\| X_2^{\perp}\hat{\beta}_{(2)} \|^2}{q\hat{\sigma}^2}$。

由 $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$，知 $\hat{\beta}_{(2)} \sim N(\beta_{(2)}, \sigma^2 (X_2^{\perp}{}'X_2^{\perp})^{-1})$

H_0成立时, $\beta_{(2)} = 0$，所以 $A = \| X_2^{\perp}\hat{\beta}_{(2)} \|^2 = \hat{\beta}_{(2)}{}'X_2^{\perp}{}'X_2^{\perp}\hat{\beta}_{(2)} \sim \sigma^2\chi_q^2$

另外, $B = (n-p)\hat{\sigma}^2 \sim \sigma^2\chi_{n-p}^2$，且与$\hat{\beta}$独立,

所以 $$\frac{A/q}{B/(n-p)} = \frac{\| X_2^{\perp}\hat{\beta}_{(2)} \|^2}{q\hat{\sigma}^2} = F \sim F_{k,n-p}$$

证明2：利用 $F = \dfrac{n-p}{q} \times \dfrac{\| \hat{Y} - \hat{Y}_0 \|^2}{\| Y - \hat{Y} \|^2}$。

(1) 因为 $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$, 以及 $Y - \hat{Y} = (I_n - P_X)\boldsymbol{\varepsilon}$

$$\Rightarrow \frac{(n-p)\hat{\sigma}^2}{\sigma^2} = \frac{\| Y - \hat{Y} \|^2}{\sigma^2} \sim \chi_{n-p}^2 (\text{不论}H_0\text{成立与否}).$$

(2) 原假设下 $Y = X_1\beta_{(1)} + \boldsymbol{\varepsilon}$，$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$

所以 $\hat{Y} - \hat{Y}_0 = P_X Y - P_{X_1} Y = P_{X_2^{\perp}} Y = P_{X_2^{\perp}}(X_1\beta_{(1)} + \boldsymbol{\varepsilon}) = P_{X_2^{\perp}}\boldsymbol{\varepsilon}$

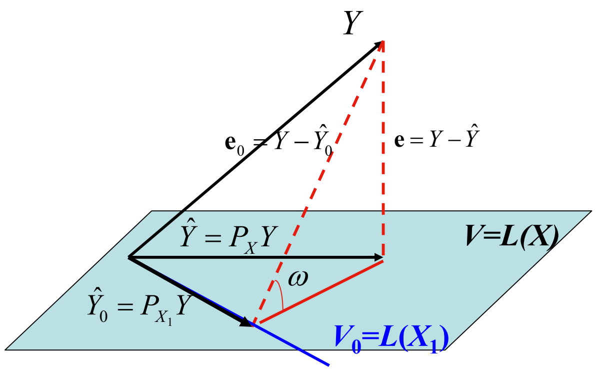
$$\Rightarrow \frac{\| \hat{Y} - \hat{Y}_0 \|^2}{\sigma^2} \sim \chi_q^2$$

(3) 因为 $(I_n - P_X)P_{X_2^{\perp}} = 0$, 所以 $\| \hat{Y} - \hat{Y}_0 \|^2$ 与 $\| Y - \hat{Y} \|^2$ 独立,

$$\Rightarrow F = \frac{\| \hat{Y} - \hat{Y}_0 \|^2 / q}{\| Y - \hat{Y} \|^2 / (n-p)} \sim F_{q,n-p}$$

F的其它表达

$$F = \frac{\| X_2^{\perp}\hat{\beta}_{(2)} \|^2 / q}{\hat{\sigma}^2} = \frac{\| \hat{Y} - \hat{Y}_0 \|^2 / q}{\| Y - \hat{Y} \|^2 / (n-p)} = \frac{n-p}{q} ctg(\omega)^2$$



$$F = \frac{n-p}{q} \times \frac{RSS_0 - RSS}{RSS}$$

这是因为 $\| \hat{Y} - \hat{Y}_0 \|^2 = \| \mathbf{e}_0 - \mathbf{e} \|^2 = \| \mathbf{e}_0 \|^2 - \| \mathbf{e} \|^2 = RSS_0 - RSS$

其中 $RSS = \| \mathbf{e} \|^2$, $RSS_0 = \| \mathbf{e}_0 \|^2$ 为全,子模型下的残差平方和

$$F = \frac{n-p}{q} \times \frac{R^2 - R_0^2}{1 - R^2}$$

这是因为 $\| \hat{Y} - \hat{Y}_0 \|^2 = \| \hat{Y} - \mathbf{1}\bar{y} \|^2 - \| \hat{Y}_0 - \mathbf{1}\bar{y} \|^2$ 及

$R^2 = \| \hat{Y} - \mathbf{1}\bar{y} \|^2 / SS_{总}$，$R_0^2 = \| \hat{Y}_0 - \mathbf{1}\bar{y} \|^2 / SS_{总}$

方差分析

$$F = \frac{n-p}{q} \times \frac{RSS_0 - RSS}{RSS}$$

上述形式表明**F-**检验可看作是是比较全模型和子模型的拟合程度。比较残差平方和也称作**ANOVA (ANalysis Of Variance**，方差分析)

$$\begin{aligned} SS_{总} &= \| Y - \mathbf{1}\bar{y} \|^2 \\ &= \| \hat{Y} - \mathbf{1}\bar{y} \|^2 + RSS \\ &= \| \hat{Y}_0 - \mathbf{1}\bar{y} \|^2 + \| \hat{Y} - \hat{Y}_0 \|^2 + RSS \\ &= \| \hat{Y}_0 - \mathbf{1}\bar{y} \|^2 + (RSS_0 - RSS) + RSS \end{aligned}$$

X1解释的部分	X1不能解释,但x2能解释的部分	X1,x2都不能解释的部分

```
> anova ( sub.model, full.model )
```

Y～X1　　**Y～X1+X2**

特例1. 单个回归系数的t-检验 $H_0:\beta_k=0$

$$F=\frac{\|\mathrm{x}_k^{\perp}\hat{\beta}_k\|^2}{\hat{\sigma}^2}=\frac{\hat{\beta}_k^{\ 2}}{\hat{\sigma}^2/\|\mathrm{x}_k^{\perp}\|^2}\sim_{H_0}F_{1,n-p}$$

$$t=\pm\sqrt{F}=\frac{\hat{\beta}_k}{\hat{\sigma}\ \ /\|\mathrm{x}_k^{\perp}\|}\sim_{H_0}t_{n-p}$$

这可由$\hat{\beta}_k\sim N(\beta_k,\sigma^2/\|\mathrm{x}_k^{\perp}\|^2)$直接得到。

特例2. 回归方程的显著性检验

$Y=\mathbf{1}\beta_0+Z\gamma+\varepsilon,\beta_0$是截距

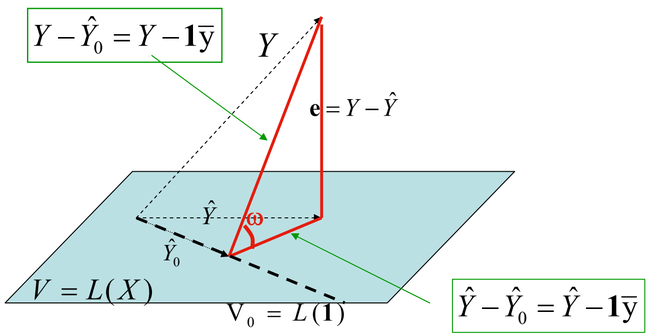
$\mathrm{H}_0:\gamma=(\beta_1,...,\beta_{\mathrm{p-1}})'=0$

记$Z^{\perp}=Z-P_1Z=Z-\mathbf{1}\bar{\mathrm{x}}'$为自变量的中心化矩阵。

回归方程显著性的 F检验：

$$F=\frac{n-p}{p-1}\times\frac{\|\hat{Y}-\hat{Y}_0\|^2}{\|Y-\hat{Y}\|^2}=\frac{\|Z^{\perp}\hat{\gamma}\|^2}{(p-1)\hat{\sigma}^2}\sim_{H_0}F_{p-1,n-p}$$

其中$\hat{Y}_0=\mathbf{1}\bar{\mathrm{y}}$，所以 $F=\frac{n-p}{p-1}\times\frac{R^2}{1-R^2}$



$$\cos^2\omega=\frac{\|\hat{Y}-\mathbf{1}\bar{\mathrm{y}}\|^2}{\|Y-\mathbf{1}\bar{\mathrm{y}}\|^2}=R^2$$

$$F=\frac{n-p}{p-1}\times\frac{\|\hat{Y}-\hat{Y}_0\|^2}{\|Y-\hat{Y}\|^2}=\frac{n-p}{p-1}\times\frac{R^2}{1-R^2}=\frac{n-p}{p-1}\times(ctg\omega)^2$$

$\mathrm{Salary}=\beta_0+\beta_1\mathrm{Sex}+\beta_2\mathrm{Rank}+\beta_3\mathrm{Year}+\varepsilon$

```
Call:
lm(formula = Salary ~ Sex + Rank + Year, data = salary)

Coefficients:
             Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)  11011.76   966.95      11.388    3.03e-15 ***
Sex          603.77     811.20      0.744     0.46
Rank         4747.18    452.58      10.489    5.18e-14 ***
Year         393.86     74.53       5.285     3.04e-06 ***
---
Residual standard error: 2398 on 48 degrees of freedom
Multiple R-squared: 0.8454,     Adjusted R-squared: 0.8358
F-statistic: 87.51 on 3 and 48 DF,  p-value: < 2.2e-16
```

单个回归系数的t检验 $\sqrt{F}$

回归方程的显著性检验

$$\text{Salary} = \beta_0 + \beta_1 \text{Sex} + \beta_2 \text{Rank} + \beta_3 \text{Year} + \varepsilon$$

$$H_0 : \beta_2 = \beta_3 = 0$$

```
> m2   = lm(Salary~Sex+Rank+Year, data=salary)
> m1   = lm(Salary~Sex, data=salary)
> anova(m1, m2)
Analysis of Variance Table

Model 1: Salary ~ Sex
Model 2: Salary ~ Sex + Rank + Year
  Res.Df        RSS   Df   Sum of Sq       F      Pr(>F)
1     50 1671623638
2     48  276016717    2  1395606921  121.35  < 2.2e-16 ***
```

$$\text{F} = \frac{n-p}{q} \times \frac{RSS_0 - RSS}{RSS} = \frac{52-4}{2} \times \frac{1671623638 - 276016717}{276016717} = 121.35$$