2-5

OLS的分布性质与区间估计

主讲人：范国斌

OLS估计的分布性质

基本思想：

- $\hat{\beta}$ 是随机变量，必须确定其分布性质才可能进行区间估计和假设检验
- u_i 是服从正态分布的随机变量，决定了 Y 也是服从正态分布的随机变量
- $\hat{\beta}$ 是 Y 的线性函数，决定了 $\hat{\beta}$ 也是服从正态分布的随机变量

$\hat{\beta}$ 的期望与方差

- $\hat{\beta}$ 的期望 $E(\hat{\beta})=\beta$ **(由无偏性)**
- $\hat{\beta}$ 的方差和标准误差： 可以证明 $\hat{\beta}$ 的方差—协方差矩阵为

$$\mathrm{Var\text{-}Cov}(\hat{\beta})=\sigma^2(X'X)^{-1}$$

$$Var(\hat{\beta}_j)=\sigma^2 c_{jj}$$

$$SE(\hat{\beta}_j)=\sigma\sqrt{c_{jj}}$$

这里的 $(X'X)^{-1}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kk} \end{bmatrix}$

(其中 c_{jj} 是矩阵 $(X'X)^{-1}$ 中第 j 行第 j 列的元素)

所以 $\hat{\beta}_j \sim N(\beta_j,\sigma^2 c_{jj})$ （j=1,2,---k）

对$\hat{\beta}_j$作标准化变换

分布函数

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

为什么要对 $\hat{\beta}_j$ 作标准化变换?

在u_i正态性假定下，由前面的分析已知 $\hat{\beta}_j \sim N[\beta_j, Var(\hat{\beta}_j)]$

但在对一般正态变量 $\hat{\beta}_j$ 作实际分析时，要具体确定 $\hat{\beta}_j$ 的取值及对应的概率，要通过正态分布密度函数或分布函数去计算是很麻烦的，为了便于直接利用“标准化正态分布的临界值”，需要对$\hat{\beta}_j$作标准化变换。

标准化的方式：$z_j=\dfrac{\hat{\beta}_j-E(\beta_j)}{SE(\hat{\beta}_j)}$

分布函数

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{x^2}{2}}dx$$

1.σ^2已知时，对$\hat{\beta}_j$作标准化变换

- 在σ^2已知时对$\hat{\beta}_j$作标准化变换，所得Z统计量为标准正态变量。

$$z_j = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{c_{jj}}} \sim N(0,1), j = 1,\ldots k$$

注意:这时 $SE(\hat{b}_j)$ 不是随机变量(**X**、**σ** 、***n*** 都是非随机的)

随机扰动项方差 σ^2 的估计

σ^2 一般未知，可证明多元回归中 σ^2 的无偏估计为：(证明见本科教材P93附录3.2)

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k} \quad \text{或表示为} \quad \hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k}$$

在**一元回归**的特例中，

$$se(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}} \qquad se(\hat{\beta}_1) = \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}\sigma$$

2. σ^2未知时，对$\hat{\beta}_j$作标准化变换

条件： 当 σ^2未知时，可用 $\hat{\sigma}^2$（随机变量）代替 σ^2 去估计参数的标准误差。这时参数估计的标准误差是个**随机变量**。

- 样本为**大样本**时,作标准化变换所得的统计量 z_j，也可以视为标准正态变量（根据中心极限定理）。
- 样本为**小样本**时，

用估计的参数标准误差对 $\hat{\beta}_j$ 作标准化变换，所得的统计量用t表示，这时t将不再服从正态分布，而是服从 t 分布：

（注意这时分母是随机变量）

$$t = \frac{\hat{\beta}_j - \beta_j}{\hat{SE}(\hat{\beta}_j)} \sim t(n-k)$$

回归系数的区间估计

由于 $t^* = \dfrac{\hat{\beta}_j - \beta_j}{\hat{SE}(\hat{\beta}_j)} = \dfrac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{c_{jj}}} \sim t(n-k)$

给定 α ，查 t 分布表的自由度为 n-k 的临界值 $t_{\alpha/2}(n-k)$

$$P[-t_{\alpha/2}(n-k) \le t^* = \frac{\hat{\beta}_j - \beta_j}{\hat{SE}(\hat{\beta}_j)} \le t_{\alpha/2}(n-k)] = 1-\alpha \quad (j = 1 \cdots k)$$

$$P[\hat{\beta}_j - t_{\alpha/2}\hat{SE}(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + t_{\alpha/2}\hat{SE}(\hat{\beta}_j)] = 1-\alpha$$

或 $$P[\hat{\beta}_j - t_{\alpha/2}\hat{\sigma}\sqrt{c_{jj}} \le \beta_j \le \hat{\beta}_j + t_{\alpha/2}\hat{\sigma}\sqrt{c_{jj}}] = 1-\alpha$$

或表示为 $\beta_j = (\hat{\beta}_j - t_{\alpha/2(n-k)}\hat{\sigma}\sqrt{c_{jj}}, \hat{\beta}_j + t_{\alpha/2(n-k)}\hat{\sigma}\sqrt{c_{jj}})$