第二讲　普通最小二乘估计量

一、基本概念：估计量与估计值

对总体参数的一种估计法则就是估计量。例如，为了估计总体均值为 u，我们可以抽取一个容量为 N 的样本，令 Y_i 为第 i 次观测值，则 u 的一个很自然的估计量就是 $\hat{u} = \bar{Y} = \frac{\sum Y_i}{N}$。A、B 两同学都利用了这种估计方法，但手中所掌握的样本分别是 $(y_1^A, y_2^A, \ldots, y_N^A)$ 与 $(y_1^B, y_2^B, \ldots, y_N^B)$。A、B 两同学分别计算出估计值

$\hat{u}^A = \frac{\sum y_i^A}{N}$ 与 $\hat{u}^B = \frac{\sum y_i^B}{N}$。因此，在上例中，估计量 $\hat{u}$ 是随机的，而 $\hat{u}^A, \hat{u}^B$ 是该随机变量可能的取值。估计量所服从的分布称为抽样分布。

如果真实模型是：$y = \beta_0 + \beta_1 x + \varepsilon$，其中 β_0, β_1 是待估计的参数，而相应的 OLS 估计量就是：

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}; \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

我们现在的任务就是，基于一些重要的假定，来考察上述 OLS 估计量所具有的一些性质。

二、高斯-马尔科夫假定

●**假定一**：真实模型是：$y=\beta_0+\beta_1 x+\varepsilon$。有三种情况属于对该假定的违背：⑴遗漏了相关的解释变量或者增加了无关的解释变量；⑵y 与 x 间的关系是非线性的；⑶β_0,β_1并不是常数。

●**假定二**：在重复抽样中，$(x_1,x_2,...,x_N)$被预先固定下来，即$(x_1,x_2,...,x_N)$是非随机的（进一步的阐释见附录），显然，如果解释变量含有随机的测量误差，那么该假定被违背。还存其他的违背该假定的情况。

笔记：

$(x_1,x_2,...,x_N)$是随机的情况更一般化，此时，高斯-马尔科夫假定二被更改为：对任意i,j，x_i与ε_j不相关，此即所谓的解释变量具有严格外生性。显然，当$(x_1,x_2,...,x_N)$非随机时，x_i与ε_j必定不相关，这是因为ε_j是随机的。

●**假定三**：误差项期望值为 0，即$E(\varepsilon_i)=0, i=1,2,...,N$。

笔记：

1、当$(x_1,x_2,...,x_N)$随机时，标准假定是：

$$E(\varepsilon_i|x_1,x_2,...,x_N)=0, i=1,2,...,N$$

根据迭代期望定律有：$E[E(\varepsilon_i|x_1,x_2,...,x_N)]=E(\varepsilon_i)$，因此，如果$E(\varepsilon_i|x_1,x_2,...,x_N)=0$成立，必定有：$E(\varepsilon_i)=0$。

另外，根据迭代期望定律也有：

$$E[E(\varepsilon_i x_j|x_1,x_2,\ldots,x_N)]=E(\varepsilon_i x_j)$$

而 $E(\varepsilon_i x_j|x_1,x_2,\ldots,x_N)=x_j E(\varepsilon_i|x_1,x_2,\ldots,x_N)$。故有：

$$E(\varepsilon_i|x_1,x_2,\ldots,x_N)=0\Rightarrow\begin{cases}E(\varepsilon_i)=0\\E(\varepsilon_i x_j)=0\end{cases}$$

$$\rightarrow Cov(\varepsilon_i,x_j)=E(\varepsilon_i x_j)-E(\varepsilon_i)E(x_j)=0$$

因此，在$(x_1,x_2,\ldots,x_N)$是随机的情况下，假定二、三可以修正为一个假定：$E(\varepsilon_i|x_1,x_2,\ldots,x_N)=0$。

2、所谓迭代期望定律是指：如果信息集 $\Theta\leq\Omega$，则有 $E[E(X|\Omega)|\Theta]=E(X|\Theta)$。为了理解上述等式， 考虑一个极端情况：$\Omega$ 包含了全部的信息，此时 X 丧失了随机性，故 $E(X|\Omega)=X$，因此必有$E[E(X|\Omega)|\Theta]=E(X|\Theta)$。无条件期望所对应的信息集是空集，因此$E[E(X|\Omega)]=E(X)$。

3、回忆第一讲，对模型$y=\beta_0+\beta_1 x+\varepsilon$，在 OLS 法下我们一定能保证：(1)残差均值为零；(2)残差与 x 样本不相关。残差是对误差的近似，如果假定二、三不成立，即误差项与解释变量相关，误差项期望值不为零，显然此时残差并不是对误差项的有效的近似，换句话说，此时 OLS 估计量是有严重问题的。因此，假定二、三非常重要。

●假定四： $\delta_{\varepsilon_i}^2=\delta^2$，即所谓的同方差假定。

笔记：

在$(x_1,x_2,...,x_N)$是随机的情况下，该假定修订为：

$$\mathrm{Var}(\varepsilon_i|x_1,x_2,...,x_N)=\delta^2$$

●**假定五：**$Cov(\varepsilon_i,\varepsilon_j)=0, i\neq j$，即所谓的序列不相关假定。

笔记：

在$(x_1,x_2,...,x_N)$是随机的情况下，该假定修订为：

$$\mathrm{Cov}(\varepsilon_i,\varepsilon_j|x_1,x_2,...,x_N)=0\quad, i\neq j$$

●**假定六：**$\sum(x_i-\bar{x})^2\neq 0$，在多元回归中，该假定演变为$XX'$的逆存在，即各解释变量不完全共线。

三、高斯-马尔科夫定理

当高斯-马尔科夫假定成立时，在所有线性无偏估计量中，OLS 估计量方差最小。或者说，OLS 估计量是最优线性无偏估计量（best linear unbiased estimator,BLUE）。这被称为高斯-马尔科夫定理。

（一）OLS 估计量是线性估计量

所谓 OLS 估计量是线性估计量，是指它能够被表示为y_i的线性函数。例如：

$$\hat{\beta}_1=\sum\left[\frac{x_i-\bar{x}}{\sum(x_i-\bar{x})^2}\right]y_i=\sum k_i y_i$$

注意，在假定二下，k_i 是非随机的。

练习：把$\hat{\beta_0}$表示成y_i的线性函数。

笔记：

线性意味着简单，简单意味着普通。因此有称谓“普通最小二乘法”。二乘即为平方，故 OLS 即为“简单的最小平方法”。

（二）**OLS 估计量具有无偏性**：$E(\hat{\beta_1})=\beta_1$；$E(\hat{\beta_0})=\beta_0$ 证明$E(\hat{\beta_1})=\beta_1$：

$$\hat{\beta}_1=\sum\left[\frac{x_i-\bar{x}}{\sum(x_i-\bar{x})^2}\right]y_i=\sum k_i y_i=\sum k_i(\beta_0+\beta_1 x_i+\varepsilon_i)$$

$$\Rightarrow E(\hat{\beta}_1)=\beta_0\sum k_i+\beta_1\sum k_i x_i+\sum k_i E(\varepsilon_i)$$

而$\sum k_i=0$；$\sum k_i x_i=1$。因此在重要假定三：$E(\varepsilon_i)=0$下，有：$E(\hat{\beta_1})=\beta_1$。

笔记：

在$(x_1,x_2,\dots,x_N)$是随机的情况下，我们需证：

$$E(\hat{\beta_1}\mid x_1,x_2,\dots,x_N)=\beta_1$$

练习：证明$E(\hat{\beta_0})=\beta_0$

（三）**在所有线性无偏估计量中，OLS 估计量方差最**

小

1、关于方差

$$\delta_{\beta} = Var[\sum k_i(\beta_0+\beta_1 x_i+\varepsilon_i)] = Var(\sum k_i\varepsilon_i)$$

在重要假定五： $Cov(\varepsilon_i,\varepsilon_j)=0, i\neq j$ 及其重要假定四：$\delta_{\varepsilon 2} = \delta 2$ 下，有：

i

$$\delta^2_{\hat{\beta}_1} = \delta^2 \sum k_i^2$$

注意到$\sum k_i^2 = \sum(\frac{x_i-\bar{x}}{\sum(x_i-\bar{x})^2})^2 = \frac{1}{\sum(x_i-\bar{x})^2}$

因此有：

$$\delta^2_{\hat{\beta}_1} = \frac{\delta^2}{\sum(x_i-\bar{x})^2}$$

笔记：

$\delta^2_{\hat{\beta}_1} = \frac{\delta^2}{\sum(x_i-\bar{x})^2} = \frac{\frac{1}{N}\delta^2}{\frac{1}{N}\sum(x_i-\bar{x})^2}$，当 N 趋于无穷大时，样本方差$\frac{1}{N}\sum(x_i-x)^2$收敛于总体方差，故当 N 趋于无穷大时，$\delta^2_{\hat{\beta}_1}$趋于 0。由于$E(\hat{\beta_1})=\beta_1$，因此，当 N 趋于无穷

大时，$\hat{\beta}_1$在概率上收敛于β_1，即$\hat{\beta}_1$是β_1的一致估计量。你能 够表明$\hat{\beta}_0$是β_0的一致估计量吗？应该注意到，一致性是估计

量应该满足的最低要求。想一想，如果把总体都告诉了你，但你的估计或者猜测却与真实参数不一致，你是不是应该检讨一下你的估计方法？

练习：

（1）证明在高斯-马尔科夫假定下：

$$\delta^2_{\hat{\beta}_0}=(\frac{1}{N}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2})\delta^2=\delta^2\frac{\sum x_i{}^2}{N\sum(x_i-\bar{x})^2}$$

（2）证明在高斯-马尔科夫假定下：

$$E(\varepsilon_i-\bar{\varepsilon})^2=\delta^2(1-\frac{1}{N})$$

（3）证明在高斯-马尔科夫假定下：

$$Cov(\varepsilon,\hat{\beta}_1)=0$$

（4）证明在高斯-马尔科夫假定下：

$$Cov(\hat{\beta}_0,\hat{\beta}_1)=-\frac{\bar{x}\delta^2}{\sum(x_i-\bar{x})^2}$$

2、证明方差最小

把任意一种线性估计量表示为 $\sum w_i y_i$，当 $w_i=k_i=\frac{x_i-\bar{x}}{\sum(x_i-\bar{x})^2}$ 时，该估计量即为β_1的 OLS 估计量。现在我们将证明： 在所有无偏的β_1的线性估计量中，OLS 估计量具有最小的方差。

“在所有无偏的β_1 的线性估计量中”是一个前提条件。我们的任务是， 在给定前提下（约束条件），证明 OLS 估计量所对应的权数使方差（目标函数）取最小值。

首先分析前提条件：线性估计量的表达是

$$\hat{\beta}_1=\sum w_i y_i=\sum w_i(\beta_0+\beta_1 x_i+\varepsilon_i)$$

为了保证$\hat{\beta}_1$的无偏性，那么应该保证：

$$E(\hat{\beta}_1)=\beta_0\sum w_i+\beta_1\sum w_i x_i+\sum w_i E(\varepsilon_i)=\beta_0\sum w_i+\beta_1\sum w_i x_i=\beta_1$$

因此，$\sum w_i=0;\sum w_i x_i=1$

其次分析方差表示：

$\delta^2_{\hat{\beta}_1}=Var(\sum w_i y_i)=Var[\sum w_i(\beta_0+\beta_1 x_i+\varepsilon_i)]=Var(\sum w_i\varepsilon_i)$，在假定四、五下，有：$\delta^2_{\hat{\beta}_1}=Var(\sum w_i\varepsilon_i)=\delta^2\sum w_i^2$。

最后，形成数学问题：

$$\underset{w_1,w_2,..w_N}{Min}\ \delta^2\sum_{i=1}^{N}w_i^2$$

$$s.t.\ \sum w_i=0 \quad (1)$$

$$\sum w_i x_i=1 \quad (2)$$

常数δ^2对于该最优化问题并不重要，因此上述问题简化为：

$$\underset{w_1,w_2,..w_N}{Min}\ \sum_{i=1}^{N}w_i^2$$

$$s.t.\ \sum w_i=0 \quad (1)$$

$$\sum w_i x_i=1 \quad (2)$$

对上述极值问题，其拉格朗日函数是：

$$L=\sum w_i^2-\lambda_1\sum w_i+\lambda_2(1-\sum w_i x_i)$$

相应的一阶条件是：

$$\frac{\partial l}{\partial w_1} = 2w_1 - \lambda_1 - \lambda_2 x_1 = 0$$

$$\frac{\partial l}{\partial w_i} = 2w_i - \lambda_1 - \lambda_2 x_i = 0$$

M

$$\frac{\partial l}{\partial w_N} = 2w_N - \lambda_1 - \lambda_2 x_N = 0 \qquad (3group)$$

$$\frac{\partial l}{\partial \lambda_1} = -\sum w_i = 0$$

$$\frac{\partial l}{\partial \lambda_2} = 1 - \sum w_i x_i = 0$$

(4)

–

(5)

应该注意到，把（3group）中各式相加并利用（4）有：$N\lambda_1 + \lambda_2 \sum x_i = 0$ ，即$\lambda_1 = -\lambda_2 x$；把（3group）中第 i 式两边同乘以x_i并各式相加，然后利用（5），有：

$2 - \lambda_1 \sum x_i - \lambda_2 \sum x_i^2 = 0$， 即$2 + \lambda_2 \bar{x} \sum x_i - \lambda_2 \sum x_i^2 = 0$

因此，$\lambda_2 = \frac{2}{\sum x_i (x_i - \bar{x})}$； $\lambda_1 = -\frac{2\bar{x}}{\sum x_i (x_i - \bar{x})}$

因此，

$$w_i = \frac{\lambda_1 + \lambda_2 x_i}{2} = \frac{-\frac{2\bar{x}}{\sum x_i (x_i - \bar{x})} + \frac{2x_i}{\sum x_i (x_i - \bar{x})}}{2} = \frac{x_i - \bar{x}}{\sum x_i (x_i - \bar{x})}$$

$$= \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$$

而在前面我们已知道这个权数正是β_1的 OLS 估计量所对应的权数！

练习：证明 OLS 估计量$\hat{\beta_0}$在所有β_0的线性无偏估计量中方差是最小的。