计量经济学(Econometrics)

胡华平

西北农林科技大学

经济管理学院数量经济教研室

huhuaping01@hotmail.com

2023-02-15

第3章：一元回归：参数估计

3.1 普通最小二乘法（OLS）

3.2 最小二乘估计的精度

3.3 经典线性回归模型(CLRM)

3.4 最小二乘估计的性质：BLUE

3.5 变异分解与拟合优度

3.6 一个数值例子

3.7 经典正态线性回归模型（N-CLRM）

3.8 极大似然估计法(ML)

3.1 普通最小二乘法(OLS)

引子：如何估计回归函数中的系数？

我们已经知道如何科学表达总体回归和样本回归的关系：

总体回归：

$$\begin{cases} E(Y|X_i) = \beta_1 + \beta_2 X_i & \text{(PRF)} \\ Y_i = \beta_1 + \beta_2 X_i + u_i & \text{(PRM)} \end{cases}$$

样本回归：

$$\begin{cases} \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i & \text{(SRF)} \\ Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i & \text{(SRM)} \end{cases}$$

首先需要回答的问题是，我们该如何估计得出样本回归函数中的系数？事实上，方法有多种多样：

- 图解法：比较粗糙，但提供了基本的视觉认知
- 最小二乘法(order lease squares, OLS)：最常用的方法
- 最大似然法(maximum likelihood, ML)
- 矩估计方法(Moment method, MM)

样本回归与总体回归的比较

总体回归函数PRF:

$$E(Y|X_i) = \beta_1 + \beta_2 X_i$$

总体回归模型PRM:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

样本回归函数SRF:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

样本回归模型SRM:

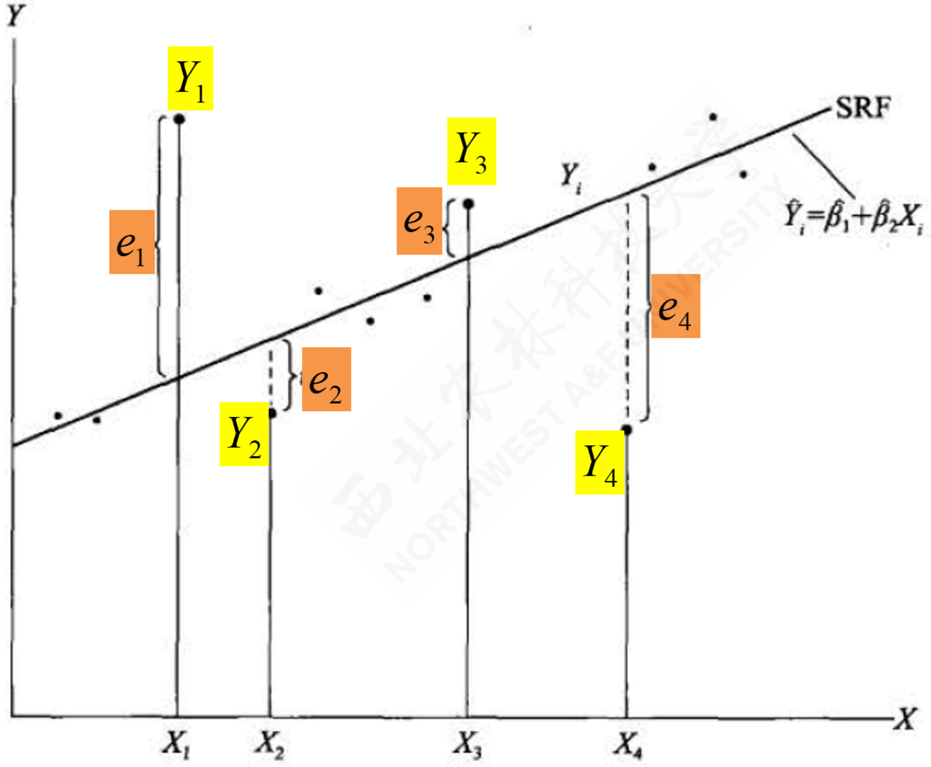
$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

思考：

- PRF无法直接观测，只能用SRF近似替代
- 估计值与观测值之间存在偏差
- SRF又是怎样决定的呢?

普通最小二乘法的原理

认识普通最小二乘法的原理：一个图示



最小二乘法的原理

普通最小二乘法的原理

OLS的基本原理：残差平方和最小化。

$$
\begin{aligned}
e_i &= Y_i - \hat{Y}_i \\
&= Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_i)
\end{aligned}
$$

$$
\begin{aligned}
Q &= \sum e_i^2 \\
&= \sum (Y_i - \hat{Y}_i)^2 \\
&= \sum \left(Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_i)\right)^2 \\
&\equiv f(\hat{\beta}_1, \hat{\beta}_2)
\end{aligned}
$$

$$
Min(Q) = Min\left(f(\hat{\beta}_1, \hat{\beta}_2)\right)
$$

普通最小二乘法的原理

认识普通最小二乘法的原理：一个数值试验。假设存在下面所示的4组观测值 (X_i, Y_i)：

X_i	Y_i		
(1)	(2)		
1	4		
4	5		
5	7		
6	12		
Sum:			

数值试验：数据

普通最小二乘法的原理

假设随便猜想了如下两个SRF，完成下表计算，并分析哪个SRF给出的 $(\hat{\beta}_1, \hat{\beta}_2)$要更好？

$$SRF1：\hat{Y}_{1i} = \hat{\beta}_1 + \hat{\beta}_2 X_i = 1.572 + 1.357 X_i$$
$$SRF2：\hat{Y}_{2i} = \hat{\beta}_1 + \hat{\beta}_2 X_i = 3.0 + 1.0 X_i$$

X_i	Y_i	$\hat{Y}_{1i}$	e_{1i}	e_{1i}^2	$\hat{Y}_{2i}$	e_{2i}	e_{2i}^2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	4	2.929	1.071	1.147	4	0	0
4	5	7.000	2.000	4.000	7	-2	4
5	7	8.357	1.357	1.841	8	-1	1
6	12	9.714	2.285	5.226	9	3	9
Sum:			0.000	12.214		0	14

回归参数的OLS点估计

最小化求解：

$$
\begin{aligned}
Min(Q) &= Min\left(f(\hat{\beta}_1, \hat{\beta}_2)\right) \\
&= Min\left(\sum(Y_i-(\hat{\beta}_1+\hat{\beta}_2X_i))^2\right) \\
&= Min\sum(Y_i-\hat{\beta}_1-\hat{\beta}_2X_i)^2
\end{aligned}
$$

方程组变形，得到**正规方程组**：

$$
\begin{cases}
\sum[\hat{\beta}_1-(Y_i-\hat{\beta}_2X_i)]=0 \\
\sum[X_i^2\hat{\beta}_2-(Y_i-\hat{\beta}_1)X_i]=0
\end{cases}
$$

$$
\begin{cases}
\sum Y_i-n\hat{\beta}_1-(\sum X_i)\hat{\beta}_2=0 \\
\sum X_iY_i-(\sum X_i)\hat{\beta}_1-(\sum X_i^2)\hat{\beta}_2=0
\end{cases}
$$

回归参数的OLS点估计

进而得到回归系数的计算公式1（Favorite Five，FF）：

$$\begin{cases} \hat{\beta}_2 = \dfrac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \\ \hat{\beta}_1 = \dfrac{n\sum X_i^2 Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \end{cases} \quad \text{(FF solution)}$$

此外我们也可以得到如下的离差公式(favorite five，ff)

$$\begin{cases} \hat{\beta}_2 = \dfrac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_1 = \bar{Y}_i - \hat{\beta}_2 \bar{X}_i \end{cases} \qquad \text{(ff solution)}$$

其中离差计算 $x_i = X_i - \bar{X}$; $y_i = Y_i - \bar{Y}$。

以下式子为什么是等价的？你能推导出来么？

$$
\begin{cases}
\sum x_i y_i = \sum[(X_i - \bar{X})(Y_i - \bar{Y})] & = \sum X_i Y_i - \dfrac{1}{n}\sum X_i \sum Y_i \\
\sum x_i^2 = \sum (X_i - \bar{X})^2 & = \sum X_i^2 - \dfrac{1}{n}(\sum X_i)^2
\end{cases}
$$

随机干扰项参数的OLS点估计

PRM公式变形：

$$\left.\begin{aligned} Y_i &= \beta_1 - \beta_2 X_i + u_i \ (\text{PRM}) \Rightarrow \\ \overline{Y} &= \beta_1 - \beta_2 \bar{X} + \bar{u} \end{aligned}\right\} \Rightarrow$$

$$y_i = \beta_2 x_i + (u_i - \bar{u})$$

残差公式变形：

$$\left.\begin{aligned} e_i &= y_i - \hat{\beta}_2 x_i \\ e_i &= \beta_2 x_i + (u_i - \bar{u}) - \hat{\beta}_2 x_i \end{aligned}\right\} \Rightarrow$$

$$e_i = -(\hat{\beta}_2 - \beta_2) x_i + (u_i - \hat{u})$$

求解残差平方和：

$$\sum e_i^2 \quad = (\hat{\beta}_2 - \beta_2)^2 \sum x_i^2 + \sum (u - \bar{u})^2 - 2(\hat{\beta}_2 - \beta_2) \sum x_i (u - \bar{u})$$

求残差平方和的期望：

$$\begin{aligned} E(\sum e_i^2) &= \sum x_i^2 E[(\hat{\beta}_2 - \beta_2)^2] + E[\sum (u - \bar{u})^2] \\ &\quad - 2E[(\hat{\beta}_2 - \beta_2) \sum x_i (u - \bar{u})] \\ &\equiv A + B + C \\ &= \sigma^2 + (n-1)\sigma^2 - 2\sigma^2 \\ &= (n-2)\sigma^2 \end{aligned}$$

随机干扰项参数的OLS点估计

回归误差方差（Deviation of Regression Error）：

- 采用OLS方法下，总体回归模型PRM中随机干扰项 u_i的总体方差的无偏估计量，记为 $E(\sigma^2) \equiv \hat{\sigma}^2$，简单地记为 $\hat{\sigma}^2$。

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$$

回归误差标准差（Standard Deviation of Regression Error）：有时候也记为**se**。

$$\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-2}}$$

附录：证明I

$$\begin{cases} Y_i & = \beta_1 + \beta_2 X_i + u_i \\ \bar{Y} & = \beta_1 + \beta_2 \bar{X} + \bar{u} \end{cases} \Rightarrow$$

$$y_i = \beta_2 x_i + (u_i - \bar{u}) \quad (\text{式}11)$$

$$y_i = \hat{\beta}_2 x_i + e_i \quad (\text{srf-de}) \Rightarrow$$

$$e_i = y_i - \hat{\beta}_2 x_i \quad (\text{式}12)$$

式(11)-式(12)

$$e_i = \beta_2 x_i + (u_i - \bar{u}) - \hat{\beta}_2 x_i \quad (\text{式}13)$$

两边平方再求和：

$$\sum e_i^2=(\hat{\beta}_2-\beta_2)^2\sum x_i^2+\sum(u_i-\bar{u})^2-2(\hat{\beta}_2-\beta_2)\sum x_i(u_i-\bar{u})$$

然后再两边取期望：

$$\begin{aligned}E(\sum e_i^2)&=\sum x_i^2 E(\hat{\beta}_2-\beta_2)^2+E[\sum(u_i-\bar{u})^2]-2E[(\hat{\beta}_2-\beta_2)\sum x_i(u_i-\bar{u})]\\&=A+B+C\\&=\sigma^2+(n-1)\sigma^2-2\sigma^2\\&=(n-2)\sigma^2\end{aligned}$$

$$\hat{\sigma}^2\equiv\frac{\sum e_i^2}{n-2}$$

$$E(\hat{\sigma}^2)=\frac{1}{n-2}E(\sum e_i^2)=\sigma^2$$

$$
\begin{aligned}
A &= \sum x_i^2 E[(\hat{\beta}_2 - \beta_2)^2] \\
&= \sum [x_i^2 \cdot var(\hat{\beta}_2)] \\
&= var(\hat{\beta}_2) \cdot \sum x_i^2 \\
&= \frac{\sigma^2}{\sum x_i^2} \cdot \sum x_i^2 \\
&= \sigma^2
\end{aligned}
$$

附录：B过程证明

$$
\begin{aligned}
B = E[\sum(u-\bar{u})^2] &= E(\sum u_i^2) - 2E[\sum(u_i\bar{u})] + nE(\bar{u}^2) \\
&= n \cdot Var(u_i) - 2E\left[\sum(u_i \cdot \frac{\sum u_i}{n})\right] + nE(\frac{\sum u_i}{n})^2 \\
&= n\sigma^2 - 2E\left[\frac{\sum u_i}{n}\sum u_i\right] + E\left[\frac{(\sum u_i)^2}{n}\right] \\
&= n\sigma^2 - E[(\sum u_i)^2/n] = n\sigma^2 - \frac{E(u_i^2) + E(u_2^2) + \cdots + E(u_n^2))}{n} \\
&= n\sigma^2 - \frac{nVaru_i}{n} = n\sigma^2 - \sigma^2 = (n-1)\sigma^2
\end{aligned}
$$

附录：C过程证明

$$
\begin{aligned}
C &= -2E[(\hat{\beta}_2 - \beta_2)\sum x_i(u_i - \bar{u})] \\
&= -2E[\frac{\sum x_i u_i}{\sum x_i^2}(\sum x_i u_i - \bar{u}\sum x_i)] \\
&= -2E[\frac{(\sum x_i u_i)^2}{\sum x_i^2}] \\
&= -2E[(\hat{\beta}_2 - \beta_2)^2] = -2\sigma^2
\end{aligned}
$$

- 其中：

$$
\hat{\beta}_2 = \sum k_i Y_i = \sum k_i(\beta_1 + \beta_2 X_i + u_i) = \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i = \beta_2 + \sum k_i u_i
$$

$$
\hat{\beta}_2 - \beta_2 = \sum k_i u_i = \frac{\sum x_i u_i}{\sum x_i^2}
$$

随机干扰项参数的OLS点估计

估计出总体参数 σ^2：

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{(n-2)}$$

$$\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{(n-2)}}$$

OLS方法讨论："估计值"与"估计量"

理解OLS方法下的"估计值"与"估计量"

回归系数的计算公式1（Favorite Five，FF）：

$$\begin{cases} \hat{\beta}_2 = \dfrac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \\ \hat{\beta}_1 = \dfrac{n\sum X_i^2Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \end{cases} \qquad \text{(FF solution)}$$

- 如果给出的参数估计结果是由一个具体样本资料计算出来的，它是一个"估计值"，或者"点估计"，是参数估计量的一个具体数值；
- 如果把上式看成参数估计的一个表达式，那么，则它是 (X_i, Y_i)的函数，而 Y_i是随机变量，所以参数估计也是随机变量，在这个角度上，称之为"估计量"。

OLS方法下SRF和SRM的特征

OLS估计量是纯粹由可观测的(即样本)量(指X和Y)表达的，因此它们很容易计算。

它们是点估计量(point estimators)，即对于给定样本，每个估计量仅提供有关总体参数的一个(点)值。[我们以后还将考虑区间估计量(interval Estimators)]

一旦从样本数据得到OLS估计值，便容易画出样本回归线。

OLS方法下SRF和SRM的特征

- 特征1：样本回归线一定会经过样本均值点 $(\bar{X}, \bar{Y})$：

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

- 特征2：Y_i的**估计值**$(\hat{Y}_i)$的均值$(\bar{\hat{Y}}_i)$等于Y的样本均值$(\bar{Y})$

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 \bar{X} \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i \\ &= \bar{Y} - \hat{\beta}_2 (X_i - \bar{X}) \end{aligned}$$

$$\Rightarrow 1/n \sum \hat{Y}_i = 1/n \sum \bar{Y} - \hat{\beta}_2 (X_i - \bar{X})$$

$$\Rightarrow \bar{\hat{Y}}_i = \bar{Y}$$

OLS方法下SRF和SRM的特征

- 特征3：残差的均值($\bar{e}_i$)为零：

$$
\begin{aligned}
\sum[\hat{\beta}_1 - (Y_i - \hat{\beta}_2 X_i)] &= 0 \\
\sum[Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)] &= 0 \\
\sum(Y_i - \hat{Y}_i) &= 0 \\
\sum e_i &= 0 \\
\bar{e}_i &= 0
\end{aligned}
$$

OLS方法下SRF和SRM的特征

- 特征4：SRM和SRF可以写成离差形式：

$$\left.\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \\ \bar{Y} &= \hat{\beta}_1 + \hat{\beta}_2 \bar{X} \end{aligned}\right\} \Rightarrow$$

$$Y_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X}) + e_i \Rightarrow$$

$$y_i = \hat{\beta}_2 x_i + e_i \qquad \text{(SRM-dev)}$$

$$\left.\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\ \bar{Y} &= \hat{\beta}_1 + \hat{\beta}_2 \bar{X} \end{aligned}\right\} \Rightarrow$$

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X}) \Rightarrow$$

$$\hat{y}_i = \hat{\beta}_2 x_i \qquad \text{(SRF-dev)}$$

OLS方法下SRF和SRM的特征

- 特征5：残差(e_i)和 Y_i的拟合值($\hat{Y}_i$)不相关

$$
\begin{aligned}
Cov(e_i, \hat{Y}_i) &= E\left[(e_i - E(e_i)) \cdot (\hat{Y}_i - E(\hat{Y}_i))\right] = E(e_i \cdot \hat{y}_i) \\
&= \sum (e_i \cdot \hat{\beta}_2 x_i) \\
&= \sum \left[(y_i - \hat{\beta}_2 x_i) \cdot \hat{\beta}_2 x_i\right] \\
&= \hat{\beta}_2 \sum \left[(y_i - \hat{\beta}_2 x_i) \cdot x_i\right] \\
&= \hat{\beta}_2 \sum \left[(y_i x_i - \hat{\beta}_2 x_i^2)\right] \\
&= \hat{\beta}_2 \sum x_i y_i - \hat{\beta}_2^2 \sum x_i^2 \qquad \Leftarrow \hat{\beta}_2 = \frac{\sum x_i y_i}{x_i^2} \\
&= \hat{\beta}_2^2 \sum x_i^2 - \hat{\beta}_2^2 \sum x_i^2 = 0
\end{aligned}
$$

- 特征6：残差(e_i)和自变量(X_i)不相关

OLS方法下的离差公式总结

- 离差定义与符号：

$$
\begin{aligned}
x_i &= X_i - \bar{X} \\
y_i &= Y_i - \bar{Y} \\
\hat{y}_i &= \hat{Y}_i - \bar{\hat{Y}}_i = \hat{Y}_i - \bar{Y}
\end{aligned}
$$

- PRM及其离差形式：

$$
\left.\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_i + u_i \\
\bar{Y} &= \beta_1 + \beta_2 \bar{X} + \bar{u}
\end{aligned}\right\} \Rightarrow
$$

$$
Y_i - \bar{Y} = \beta_2 x_i + (u_i - \bar{u}) \Rightarrow
$$

$$
y_i = \hat{\beta}_2 x_i + (u_i - \bar{u}) \qquad \text{(PRM-dev)}
$$

OLS方法下的离差公式总结

- SRM及其离差形式：

$$\left.\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \\ \bar{Y} &= \hat{\beta}_1 + \hat{\beta}_2 \bar{X} \end{aligned}\right\} \Rightarrow$$

$$Y_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X}) + e_i \Rightarrow$$

$$y_i = \hat{\beta}_2 x_i + e_i$$

- SRF及其离差形式：

$$\left.\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\ \bar{Y} &= \hat{\beta}_1 + \hat{\beta}_2 \bar{X} \end{aligned}\right\} \Rightarrow$$

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X}) \Rightarrow$$

$$\hat{y}_i = \hat{\beta}_2 x_i$$

- 残差的离差形式：

$$\begin{aligned} y_i &= \hat{\beta}_2 x_i + e_i && \text{(SRM-dev)} \Rightarrow \\ e_i &= y_i - \hat{\beta}_2 x_i && \text{(residual-dev)} \end{aligned}$$

本节**内容小结**：

- 普通最小二乘方法（OLS）采用“铅垂线距离平方和最小化”的思想，来拟合一条样本回归线，进而求解出模型参数估计量。
- 拟合样本回归线，并求解出模型参数估计量的方法有很多，还有**极大似让估计法**（MLE）、**据估计方法**（MM）等等。不同估计方法代表不同的思想理念，当然各种方法的优劣势差异都需要考虑到“模型环境”（或模型假设）。
- 大家需要很熟练地记住OLS参数估计量公式，以及它们的几大重要特征！

小结与讨论

思考与讨论：

- OLS采用的“铅垂线距离平方和最小化”这一方案，凭什么它被奉为计量分析的经典方法？你觉得还有其他可行替代方案么？可以是“垂线距离平方和最小化”么？如果是距离的3次方或4次方之和，又会怎样？距离的绝对值之和可以么？对于这些方案，你有什么想法？
- 回归标准误差 se的现实含义是什么？回归参数估计与随机干扰项的方差估计有什么内在联系么？
- OLS方法的几个特征，是不是使它“天生丽质”、“娘胎里生下来就含着金钥匙”？为什么能这么说？

3.2 OLS估计量的精度

如何知道OLS方法点估计量是否可靠？

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

我们已经使用OLS方法分别得到总体回归模型(PRM)的3个重要参数（实际不止3个）的点估计量：

$$\begin{cases} \hat{\beta}_2 = \dfrac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_1 = \bar{Y}_i - \hat{\beta}_2 \bar{X}_i \end{cases} \qquad \hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$$

> 问题是：
>
> - OLS方法的点估计量是否稳定？ OLS方法的点估计量是否可信？

因此，我们需要找到一种表达OLS方法估计稳定性或估计精度的指标！

> 点估计量的**方差**（variance）和**标准差**（standard deviation）就是衡量估计稳定性或估计精度的一类重要指标！

斜率系数的方差和样本方差

斜率系数（$\hat{\beta}_2$）的**总体方差**（$\sigma^2_{\hat{\beta}_2}$）和**总体标准差**（$\sigma_{\hat{\beta}_2}$）：

$$Var(\hat{\beta}_2) \equiv \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2}$$

$$\sigma_{\hat{\beta}_2} = \sqrt{\frac{\sigma^2}{\sum x_i^2}}$$

- 其中，$Var(u_i) \equiv \sigma^2$表示随机干扰项 u_i的总体方差。

斜率系数（$\hat{\beta}_2$）的**样本方差**（$S^2_{\hat{\beta}_2}$）和**样本标准差**（$S_{\hat{\beta}_2}$）：

$$S^2_{\hat{\beta}_2} = \frac{\hat{\sigma}^2}{\sum x_i^2}$$

$$S_{\hat{\beta}_2} = \sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}}$$

- 其中，$E(\sigma^2) = \hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$表示对随机干扰项（$u_i$）的总体方差的**无偏估计量**。

证明过程I

步骤1 $\hat{\beta}_2$**的变形：**

$$
\begin{aligned}
\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} &= \frac{\sum [x_i(Y_i - \bar{Y})]}{\sum x_i^2} \\
&= \frac{\sum x_i Y_i - \sum x_i \bar{Y}}{\sum x_i^2} \\
&= \frac{\sum x_i Y_i - \bar{Y} \sum x_i}{\sum x_i^2} && \leftarrow [\sum x_i = \sum (X_i - \bar{X}) = 0] \\
&= \sum (\frac{x_i}{\sum x_i^2} \cdot Y_i) && \leftarrow [k_i \equiv \frac{x_i}{\sum x_i^2}] \\
&= \sum k_i Y_i
\end{aligned}
$$

- 其中， $k_i \equiv \frac{x_i}{\sum x_i^2}$。

步骤2：计算 $\hat{\beta}_2$的总体方差（$\sigma^2_{\hat{\beta}_2}$）：

$$
\begin{aligned}
\sigma^2_{\hat{\beta}_2} &\equiv Var(\hat{\beta}_2) = Var(\sum k_i Y_i) \\
&= \sum (k_i^2 Var(Y_i)) \\
&= \sum (k_i^2 Var(\beta_1 + \beta_2 X_i + u_i)) \\
&= \sum (k_i^2 Var(u_i)) \qquad \leftarrow [k_i \equiv \frac{x_i}{\sum x_i^2}] \\
&= \sum \left((\frac{x_i}{\sum x_i^2})^2 \cdot \sigma^2 \right) \\
&= \frac{\sigma^2}{\sum x_i^2}
\end{aligned}
$$

> 其中，$Var(u_i) \equiv \sigma^2$表示随机干扰项 u_i的总体方差。

截距系数（$\hat{\beta}_1$）的**总体方差**（$\sigma^2_{\hat{\beta}_1}$）和**总体标准差**（$\sigma_{\hat{\beta}_1}$）：

$$Var(\hat{\beta}_1) \equiv \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n} \cdot \frac{\sigma^2}{\sum x_i^2}$$

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{\sum X_i^2}{n} \cdot \frac{\sigma^2}{\sum x_i^2}}$$

- 其中，$Var(u_i) \equiv \sigma^2$表示随机干扰项u_i的总体方差。

截距系数（$\hat{\beta}_1$）的**样本方差**（$S^2_{\hat{\beta}_1}$）和**样本标准差**（$S_{\hat{\beta}_1}$）：

$$S^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n} \cdot \frac{\hat{\sigma}^2}{\sum x_i^2}$$

$$S_{\hat{\beta}_1} = \sqrt{\frac{\sum X_i^2}{n} \cdot \frac{\hat{\sigma}^2}{\sum x_i^2}}$$

- 其中，$E(\sigma^2) = \hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$表示对随机干扰项（$u_i$）的总体方差的**无偏估计量**。

证明过程I

步骤1 $\hat{\beta}_1$的变形：

$$
\begin{aligned}
\hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2\bar{X} && \leftarrow [\hat{\beta}_2 = \sum k_iY_i] \\
&= \frac{1}{n}\sum Y_i - \sum (k_iY_i \cdot \bar{X}) \\
&= \sum \left((\frac{1}{n} - k_i\bar{X}) \cdot Y_i \right) && \leftarrow \left[w_i \equiv \frac{1}{n} - k_i\bar{X} \right] \\
&= \sum w_iY_i
\end{aligned}
$$

- 其中：令 $w_i \equiv \frac{1}{n} - k_i\bar{X}$

步骤2计算 $\hat{\beta}_1$ **的总体方差**（$\sigma^2_{\hat{\beta}_1}$）：

$$\sigma^2_{\hat{\beta}_1} \equiv Var(\hat{\beta}_1) = Var(\sum w_i Y_i)$$

$$= \sum (w_i^2 Var(\beta_1 + \beta_2 X_i + u_i)) \qquad \leftarrow \left[w_i \equiv \frac{1}{n} - k_i \bar{X} \right]$$

$$= \sum (\left(\frac{1}{n} - k_i \bar{X} \right)^2 Var(u_i))$$

$$= \sigma^2 \cdot \sum \left(\frac{1}{n^2} - \frac{2\bar{X} k_i}{n} + k_i^2 \bar{X}^2 \right) \qquad \leftarrow [\sum k_i = \sum (\frac{x_i}{\sum x_i^2}) = \frac{\sum x_i}{\sum x_i^2} = 0]$$

$$= \sigma^2 \cdot \left(\frac{1}{n} + \bar{X}^2 \sum k_i^2 \right) \qquad \leftarrow [k_i \equiv \frac{x_i}{\sum x_i^2}]$$

$$= \sigma^2 \cdot \left(\frac{1}{n} + \bar{X}^2 \sum (\frac{x_i}{\sum x_i^2})^2 \right)$$

步骤2计算 $\hat{\beta}_1$的**总体方差**（$\sigma^2_{\hat{\beta}_1}$）（续前）：

$$
\begin{aligned}
&= \sigma^2 \cdot \left(\frac{1}{n} + \bar{X}^2 \frac{\sum x_i^2}{(\sum x_i^2)^2}\right) \\
&= \sigma^2 \cdot \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right) \\
&= \frac{\sum x_i^2 + n\bar{X}^2}{n \sum x_i^2} \cdot \sigma^2 \qquad \leftarrow \left[\sum x_i^2 + n\bar{X}^2 = \sum (X_i - \bar{X})^2 + n\bar{X}^2 = \sum X_i^2\right] \\
&= \frac{\sum X_i^2}{n} \cdot \frac{\sigma^2}{\sum x_i^2}
\end{aligned}
$$

现在做一个**内容小结**：

- 为了衡量OLS方法的点估计量是否稳定或是否可信，我们一般采用方差和标准差指标来表达。
- 大家应熟记**斜率**和**截距**估计量的**总体方差**和**样本方差**最终公式。

请大家**思考**如下问题：

- 总体方差和样本方差都是确定的数么？
- 二者分别受那些因素的影响？二者又有什么联系？
- 证明过程中，约定的ki和wi，有什么特征？

$$\begin{cases} \sum k_i = 0 \\ \sum k_i X_i = 1 \end{cases}$$

$$\begin{cases} \sum w_i = 1 \\ \sum w_i X_i = 0 \end{cases}$$

3.3 经典线性回归模型假设(CLRM)

引子：仅仅利用OLS估计方法就足够了么？

我们已经知道OLS方法的原理和基本特征。

问题是：

- OLS方法凭什么能在其他众多拟合估计方法中“脱颖而出”？
- 要跨越“从样本推断总体”的巨大“鸿沟”，仅仅使用OLS方法就足够了么？

答案是：

OLS估计方法，还需要**经典线性回归模型（CLRM）假设**的加持，二者“双剑合璧”才能真正完成“从样本推断总体”的逻辑证明过程。

经典线性回归模型(CLRM)假设

经典线性回归模型(classical linear regression model, CLRM)：

- 又称为高斯或标准线性回归模型
- 成为计量经济学理论的基石，主要包括7个基本假设
- 本章以双变量回归模型为讨论基础。

关于模型的假设

CLRM假设1（模型是正确设置的）：这里大有学问，也是一切计量分析问题的根本来源。

> 思考：
>
> - 我们怎么知道自己设置的模型是“正确的”？
> - 我们有可能知道“正确的”模型么？

CLRM假设2（模型是参数线性的）：模型应该是参数线性的，具体而言模型中**参数**和**随机干扰项**必须线性，变量可以不是线性。

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

> 思考：为什么需要模型是“线性的”？

课堂讨论

以下模型都是**线性的**：

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i \qquad \text{(quadratic polynomial)}$$

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i \qquad \text{(cubic polynomial)}$$

$$Y_i = \beta_1 + \beta_2 ln(X_i) + u_i \qquad \text{(linear-log)}$$

$$ln(Y_i) = \beta_1 + \beta_2 X_i + u_i \qquad \text{(log-linear)}$$

$$Y_i = \beta_1 + \beta_2 \frac{1}{X_i} + u_i \qquad \text{(reciprocal)}$$

$$Q_t = A K_t^{\alpha} L_t^{\beta} u_t \qquad \text{(Cobb-douglas)}$$

关于自变量X的假设

CLRM假设3（自变量X是外生的）：X是固定的（给定的）或**独立于**误差项。也即自变量X**不是**随机变量。

$$
\begin{aligned}
Cov(X_i, u_i) &= 0 \\
E(u_i|X_i) &= 0
\end{aligned}
$$

自变量 X是固定的（给定的）是什么含义？

- 其方差 $Var(X)$是有限的正数。
  - 如X取值不能全部相同。如果全部X取值都一样，也即 $Var(X) = 0$，则会形成什么样的散点图？
  - 又例如回归系数估计值公式中分母为0，无法求解！
- X变量没有异常值(outlier)，即没有一个X值对于其他值过大或过小。

> **课堂思考：**
>
> - X值固定不变现实么？为什么要假设这种情形？

扁数据形态："非标准"数据形态（但很直观）

60个家庭的收入和支出情况：假设的总体

Mark	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
X	80	100	120	140	160	180	200	220	240	260
Y1	55	65	79	80	102	110	120	135	137	150
Y2	60	70	84	93	107	115	136	137	145	152
Y3	65	74	90	95	110	120	140	140	155	175
Y4	70	80	94	103	116	130	144	152	165	178
Y5	75	85	98	108	118	135	145	157	175	180
Y6		88		113	125	140		160	189	185
Y7				115				162		191

Showing 1 to 8 of 8 entries

Previous 1 Next

长数据形态：标准数据形态（但不直观）

60个家庭的收入和支出情况：假设的总体

group	X	Y
1	80	55
1	80	60
1	80	65
1	80	70
1	80	75
2	100	65
2	100	70

Showing 1 to 7 of 60 entries

Previous 1 2 3 4 5 … 9 Next

样本数据示例

- X为**固定值**情形下，两份随机样本：

var	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
样本1										
X1	80	100	120	140	160	180	200	220	240	260
Y1	65	88	90	93	107	140	140	157	165	185
样本2										
X2	80	100	120	140	160	180	200	220	240	260
Y2	55	70	90	108	125	135	136	135	155	178

样本数据示例

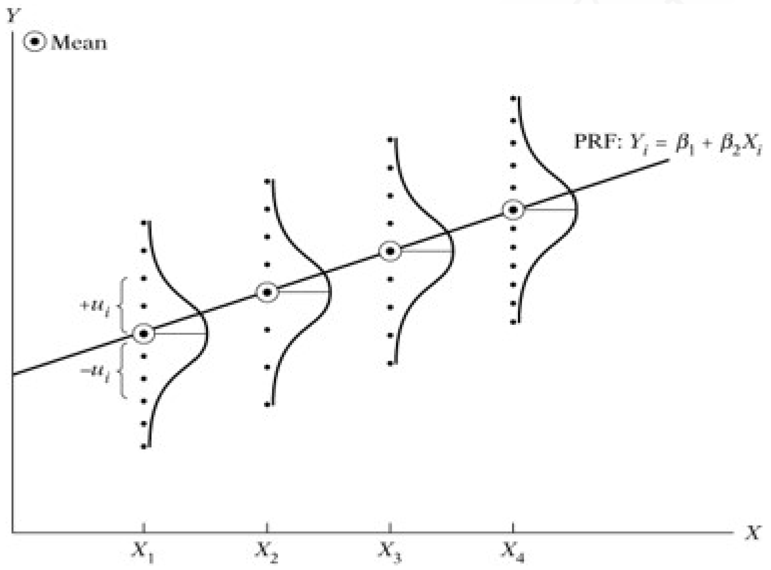
- X为**随机变量**时，两份随机样本：

var	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
样本3										
X3	100	100	140	160	180	180	220	220	240	260
Y3	74	80	103	102	110	140	137	157	165	191
样本4										
X4	120	160	160	160	180	220	220	240	260	260
Y4	79	116	110	125	135	137	160	189	152	185

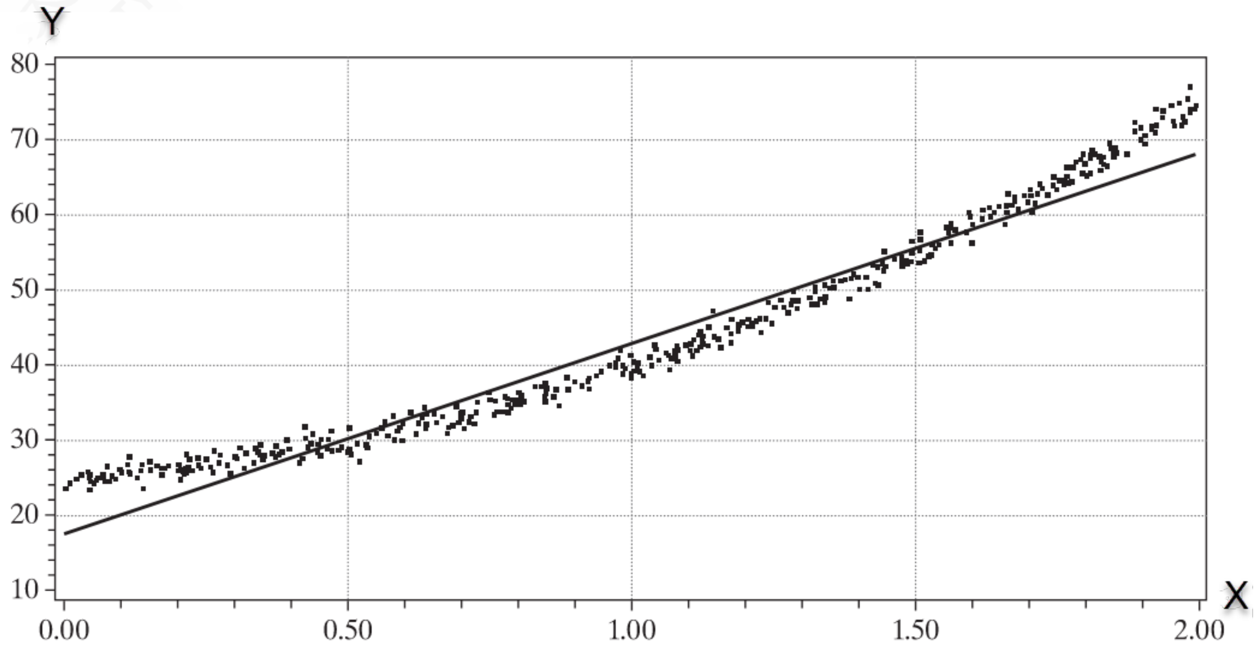
关于随机干扰项的假设I

CLRM假设4（随机干扰项条件期望值为零）：假设随机干扰项条件期望值为零。也即给定 X_i的情形下，假定随机干扰项 u_i的**条件期望**为零。

$$E(u|X_i) = 0$$



条件期望 $E(u|X_i)=0$ 比**无条件期望** $E(u)$的假设要更严格：



- 此时，在 X的特定范围内 $E(u|X_i)\neq 0$；但是对于全部的 X则有 $E(u)=0$。
- 实际的数据是由如下生成机制（DGP）模拟得到：
  $Y_i=25+5X_i(1+2X_i)+u_i$。

课堂讨论

- 讨论1：这条假设是多余的么？——因为前面已经假定 X_i与 u_i不相关（也即，$E(X_i, u_i) = 0$见**CLRM假设3**），是不是就必然 $E(u|X_i) = 0$？

> 答案：如果 X_i与 u_i**相关**（也即 $E(X_i, u_i) \neq 0$），则随机干扰项的条件期望往往不等于零（也即 $E(u|X_i) \neq 0$）；同时，如果 X_i与 u_i**不相关**（也即 $E(X_i, u_i) = 0$），则随机干扰项的条件期望也可能等于零（也即 $E(u|X_i) = 0$）。因此这一条假设是必须的。

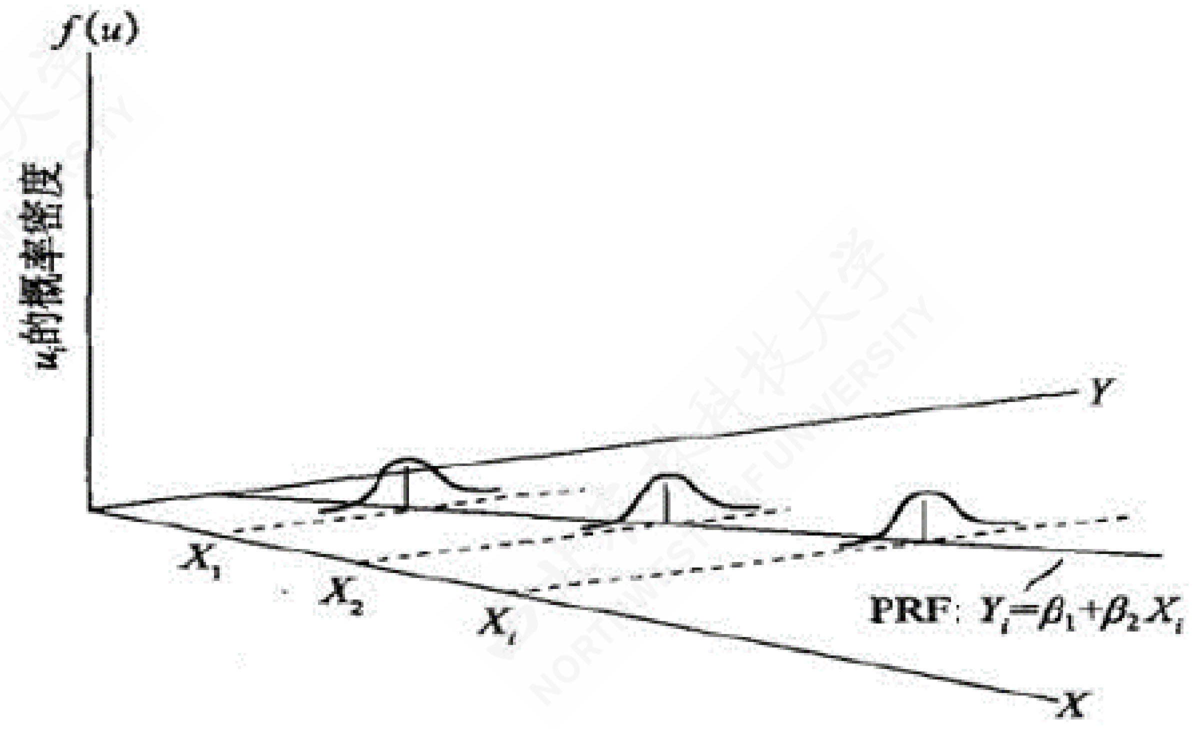
- 讨论2：若自变量 X_i为**随机变量**，我们还应当假定随机干扰项 u_i的**无条件期望**为零么？也即能否假定 $E(u_i) = 0$？

关于随机干扰项的假设2

CLRM假设5（随机干扰项的方差为同方差）：随机干扰项的方差为同方差。也即给定 X_i 的情形下，随机干扰项 u_i 的方差，处处都是相等的。记为：

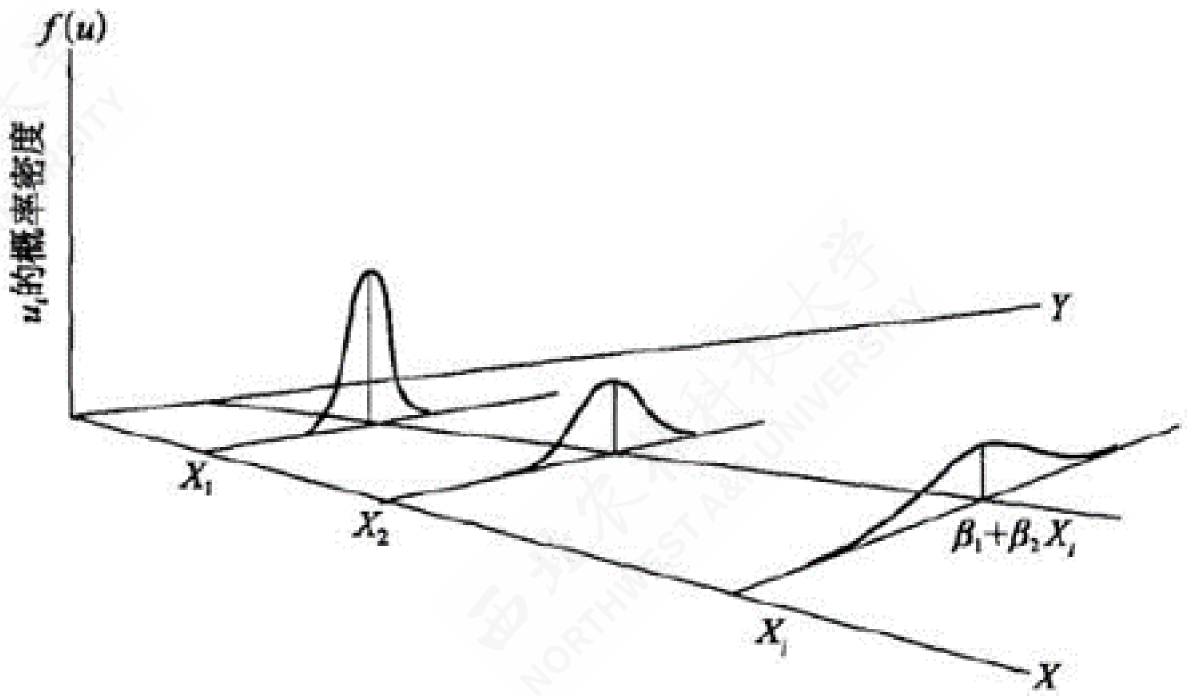
$$
\begin{aligned}
Var(u_i|X_i) &= E\left[(u_i - E(u_i))^2 |X_i\right] \\
&= E(u_i^2|X_i) \\
&= E(u_i^2) \\
&\equiv \sigma^2
\end{aligned}
$$

随机干扰项的方差为同方差



随机干扰项的方差处处相等

随机干扰项的方差为同方差



随机干扰项的方差随X取值不同而不同

随机干扰项的方差为同方差

- 同方差性(homoscedasticity)： $Var(u_i|X_i) \equiv \sigma^2$
- 异方差性(heteroscedasticity)： $Var(u_i|X_i) \equiv \sigma_i^2$

课堂讨论：

- 讨论1: 如果 $Var(u_i|X_1) < Var(u_2|X_2)$，是否意味着来自 $X = X_1$的总体，相比来自 $X = X_2$的总体，更靠近总体回归线PRL?
- 讨论2：如何看待**随机样本**的质量？或者，那些离均值较近的Y总体的随机样本，与远为分散的Y总体的随机样本，前者是不是质量更好?
- 讨论3：此时，Y_i的条件方差 $Var(Y_i|X_i)$是多少？Y_i的无条件方差 $Var(Y_i)$又是多少？
- 讨论4：如果出现异方差，会对OLS估计产生什么后果?

关于随机干扰项的假设3

CLRM假设6（随机干扰项之间无自相关）：各个随机干扰之间无自相关。也即给定两个不同的自变量取值（$X_i, X_j; i \neq j$）情形下，随机干扰项 u_i, u_j的相关系数为0。或者说 u_i, u_j最好是相互独立的。记为：

在 X_i为给定情形下，且 $i, j \in (1, 2, \cdots, n); i \neq j$，假定：

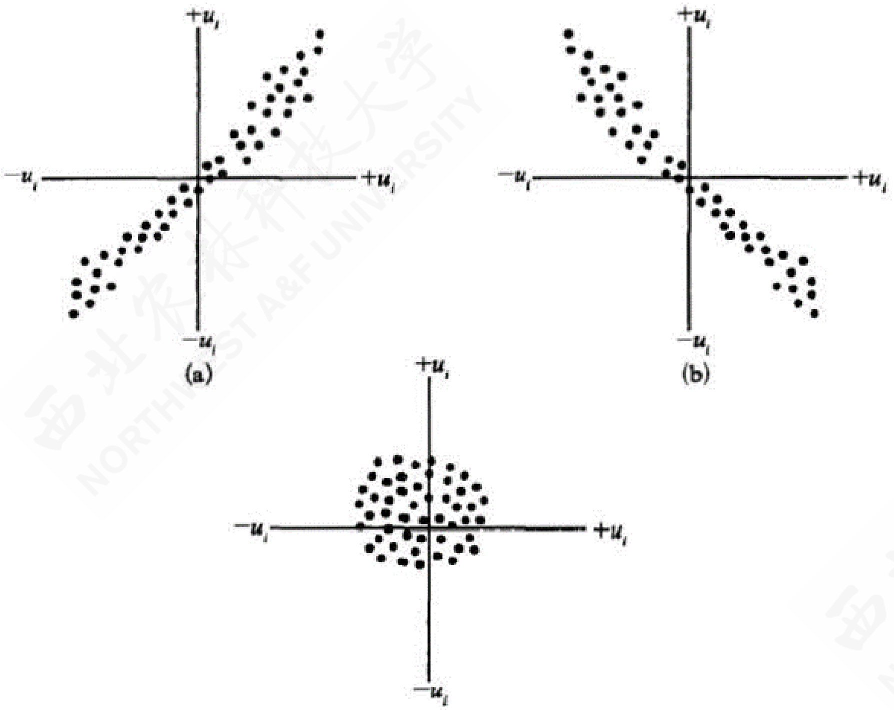
$$
\begin{aligned}
Cov(u_i, u_j|X_i, X_j) &= E\left[(u_i - E(u_i))\,(u_i - E(u_i))\right] \\
&= E(u_i u_j) \equiv 0
\end{aligned}
$$

重要概念区别：

- 无**序列相关**(no serial correlation)：
- 无**自相关**(no autocorrelation)：

$$
\begin{aligned}
Y_t &= \beta_1 + \beta_2 X_t + u_t \\
Y_t &= \beta_1 + \beta_2 X_{2t} + \beta_2 X_{3t} + u_t
\end{aligned}
$$

课堂讨论



随机干扰项之间的相关情形

课堂讨论：

- 讨论1：该假设的目的和用处是什么？
- 讨论2：如果出现自相关，会对OLS估计产生什么影响？

关于样本数的要求

CLRM假设7（观测样本数假设）：观测次数n，要大于待估计参数个数。否则方程无法解出，参数不能估计出来。

小结与讨论

下面我们对本小节内容做一个总结和讨论：

- 所有这些假设有多真实?

> “假定无关紧要论”——弗里德曼

- 上述所有假设都是针对PRF，而不是SRF！

> 例如：PRF中随机干扰项有无自相关的假设 $Cov(u_i, u_j) = 0$；但是在SRF中，可能就会出现 $Cov(e_i, e_j) \neq 0$

- 前面提到的OLS方法正是试图“复制”CLRM的假设！

> OLS方法中，$\sum e_i X_i = 0$，就类似于自变量X与随机干扰项不相关的假设（也即 $Cov(u_i, X_i) = 0$）。
> OLS方法中，$\sum e_i = 0, (\bar{e}_i = 0)$，就类似于随机干扰项期望值为0的假设（也即 $E(u|X_i) = 0$）

思考1：CLRM假设本质上是在讨论什么？

> **回答**：数据是依据什么机制产生的？(data-generating process, DGP)我们手头只有 n个样本数据对 (Y_i, X_i)但是我们希望能得到对总体**参数集** Φ的合理推断。因此，如果不对总体回归模型（PRM）作任何假设的话，我们就没有更多的信息，来对总体参数集进行任何有价值的推断。

思考2：CLRM假设既然有很多地方明显不符合现实，那么我们可以**放宽**这些假设么？如果现实根本就是**违背**了CLRM假设，OLS方法又将何去何从？对于参数估计量的性质会造成致命性的打击么？

> **回答**：**放宽**CLRM假设和**违背**CLRM假设的后果是不同的。如果只是**放宽**CLRM假设，则不会影响OLS方法的参数估计量的BLUE性质。但是如果是**违背**了CLRM假设，则OLS方法的参数估计量的BLUE性质很可能无法保持！

3.4 最小二乘估计量的性质

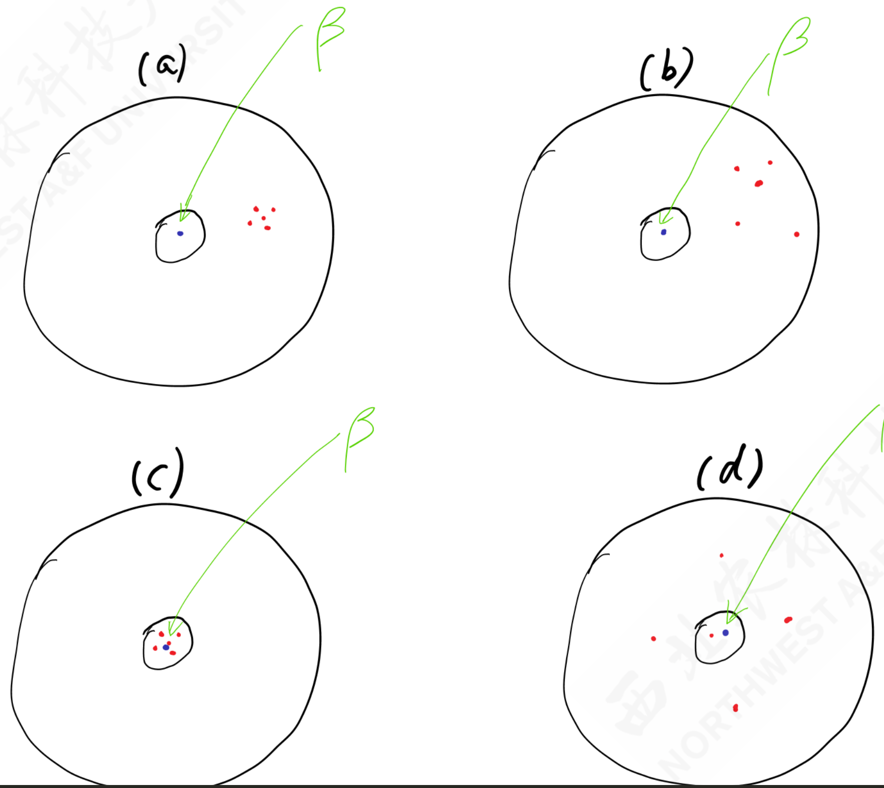
OLS方法怎么就“天生丽质”了？

我们已经知道了，OLS方法和CLRM假设“双剑合璧”下，参数估计量是最优线性无偏估计量（BLUE）。

问题是：

- OLS拟合估计方法很有“特点”，是不是意味着它就很“优秀”呢？
- 我们怎么知道，OLS方法和CLRM假设“双剑合璧”就是所向披靡呢？
- 同样在CLRM假设下，有没有一种不同于OLS的其他估计方法，也是同样那么优秀，甚至更好呢？

参数估计量的可能行为：



OLS方法怎么就“天生丽质”了？

某种参数**估计方法**（如OLS方法），得到的**估计量**（如 $\hat{\beta}_2, \hat{\beta}_1, \hat{\sigma}^2$）是总体**参数**（如 $\beta_2, \beta_1, \sigma^2$）的**最优线性无偏估计量**（**Best Linear Unbiased Estimate, BLUE**）需要满足如下三个条件：

- 线性的(Linear)：估计量是因变量 Y_i的线性函数。
- 无偏的(Unbiased)：**估计量**的均值或期望值（$E(\hat{\beta}_i)$）等于**参数**的真值（β_i）。
- 方差最小的（Best）：也即估计量是最有效的(Efficient)，是所有线性无偏估计量中有最小方差的那个估计量。

我们下面将证明：OLS方法在给定条件下就是那么“天生丽质”!

> - 用记号表达为：$\hat{\Phi}$ of $\underset{CLRM}{\overset{OLS}{\Longrightarrow}}$ Φ is BLUE
> - 以上表达读作：在经典详细回归模型假设下（CLRM），采用普通最小二乘法（OLS），得到参数 Φ的估计量 $\hat{\Phi}$，是最优线性无偏估计量（BLUE）。

高斯-马尔可夫定理(Gauss-Markov Theorem)：

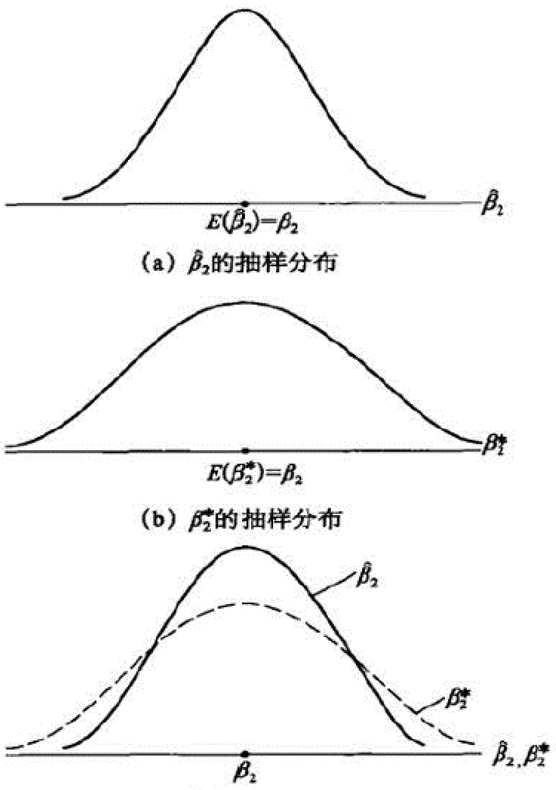
高斯-马尔可夫定理(Gauss-Markov Theorem)：在给定经典线性回归模型(CLRM)的假定下，最小二乘(OLS)**估计量**（如 $\hat{\beta}_2, \hat{\beta}_1, \hat{\sigma}^2$），在无偏线性估计量一类中，有最小方差，就是说它们是**总体参数**（如 $\beta_2, \beta_1, \sigma^2$）的**最优线性无偏估计量**(BLUE)。

课堂讨论：

- 讨论1：为什么最小二乘法（OLS）被计量学家奉为神明？还有其他选择吗？
- 讨论2：OLS得到的BLUE为到底有什么值得你称赞？
- 讨论3：OLS得到BLUE还需要CLRM假设以外的更多假设吗？(正态性？？)

OLS方法最优线性无偏估计性质的证明

不同估计方法下两个估计量的抽样分布



(a) $\hat{\beta}_2$的抽样分布

(b) β_2^*的抽样分布

(c) $\hat{\beta}_2$和β_2^*的抽样分布

- 图(a) **OLS方法**下估计量 $\hat{\beta}_2$是总体参数 β_2的一个**线性无偏估计量**
- 图(b) **其他某种方法**下估计量 $\hat{\beta}_2^*$也是总体参数 β_2的一个**线性无偏估计量**
- 图(c) 那么哪一个估计量（$\hat{\beta}_2$还是 $\hat{\beta}_2^*$）更能为我们所接受呢？

课程讨论：

- 讨论1：什么是抽样分布？
- 讨论2：怎样获得估计量分布？
- 讨论3：没有比OLS估计量更好的估计量吗？

性质1：线性性

线性性（Linearity）：是指 $\hat{\beta}_2$和 $\hat{\beta}_1$对 Y_i是线性的。

具体证明过程如下：

步骤1：证明斜率系数估计量 $\hat{\beta}_2$对 Y_i是线性的。

$$\hat{\beta}_2 = \sum k_i Y_i \qquad \leftarrow [k_i = \frac{x_i}{\sum x_i^2}]$$

又因为 $k_i = \frac{x_i}{\sum x_i^2}$是不全为0的（为什么？），所以斜率系数估计量 $\hat{\beta}_2$对 Y_i是线性的。

性质1：线性性

详细证明（反证法）：

- 假设 H_0： $k_i = \frac{x_i}{\sum x_i^2} = 0$，也即全为零。
- 则有： $x_i = X_i - \bar{X} = 0$,
- 则有： X_i处处等于 $\bar{X}$,
- 也就意味着： x_i是一个不变的量（只有一个取值）
- 因此，这是明显违背CLRM假设中关于自变量 X_i的设定（见前面）。
- 因此， H_0是显然不成立的，认为 k_i不能全为零。[证明完毕]

性质1：线性性

步骤2： 证明截距系数估计量 $\hat{\beta}_1$对 Y_i是线性的。

$$\hat{\beta}_1 = \sum w_i Y_i \qquad \leftarrow \left[w_i = \frac{1}{n} - k_i \bar{X} \right]$$

又因为 $w_i = \frac{1}{n} - k_i \bar{X}$是不全为0的（为什么？），所以斜率系数估计量 $\hat{\beta}_2$对 Y_i是线性的。

详细证明（反证法）：

- 假设 H_0： $w_i = \frac{1}{n} - k_i\bar{X} = 0$，也即全为零。
- 则有： $\sum w_i = \sum(\frac{1}{n} - k_i\bar{X}) = 0$，
- 则有： $1 - \bar{X}\sum k_i = 0$，
- 又因为： $\sum k_i = \sum \frac{x_i}{\sum x_i^2} = \frac{\sum x_i}{\sum x_i^2} = 0$，
- 因此有： $1 - 0 = 0$，也即 $1 = 0$
- 因此，这显然是错误的。
- 因此 H_0是显然不成立的，认为 w_i不能全为零。[证明完毕]

无偏性(Unbias)：**估计量**期望值（$E(\hat{\beta}_i)$）等于**参数**的真值（β_i）。

步骤1：证明斜率系数估计量$\hat{\beta}_2$是无偏的，也即$E(\hat{\beta}_2) = \beta_2$。

容易有：

$$
\begin{aligned}
\hat{\beta}_2 &= \sum k_i Y_i \qquad \leftarrow [k_i = \frac{x_i}{\sum x_i^2}] \\
&= \sum k_i(\beta_1 + \beta_2 X_i + u_i) \\
&= \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i \\
&= 0 + \beta_2 + \sum k_i u_i
\end{aligned}
$$

性质2：无偏性

（续前）因为有：

$$\sum k_i = \sum \frac{x_i}{\sum x_i^2} = \frac{\sum x_i}{\sum x_i^2} = 0$$

$$\begin{aligned}\sum k_i X_i &= \sum [\frac{x_i}{\sum x_i^2} \cdot X_i] = \frac{\sum x_i X_i}{\sum x_i^2} = \frac{\sum x_i(x_i + \bar{X})}{\sum x_i^2} \\ &= \frac{\sum x_i^2 + \sum x_i \bar{X}}{\sum x_i^2} = \frac{\sum x_i^2 + \bar{X} \sum x_i}{\sum x_i^2} = 1\end{aligned}$$

所以有：

$$\begin{aligned}\hat{\beta}_2 &= \beta_2 + \sum k_i u_i \\ E(\hat{\beta}_2) &= E(\beta_2 + \sum k_i u_i) = \beta_2 + E(\sum k_i u_i) \\ &= \beta_2 + \sum [k_i E(u_i)] = \beta_2\end{aligned}$$

性质2：无偏性

步骤2： 证明截距系数估计量 $\hat{\beta}_1$是无偏的，也即 $E(\hat{\beta}_1) = \beta_1$。

> 容易有：

$$
\begin{aligned}
\hat{\beta}_1 &= \sum w_i Y_i && \leftarrow \left[w_i = \frac{1}{n} - k_i \bar{X} \right] \\
&= \sum w_i (\beta_1 + \beta_2 X_i + u_i) \\
&= \beta_1 \sum w_i + \beta_2 \sum w_i X_i + \sum w_i u_i \\
&= \beta_1 + 0 + \sum w_i u_i
\end{aligned}
$$

性质2：无偏性

（续前）因为有：

$$\sum w_i = \sum\left[\frac{1}{n} - k_i\bar{X}\right] = 1 - \bar{X}\sum k_i = 1$$

$$\sum w_i X_i = \sum\left[\left(\frac{1}{n} - k_i\bar{X}\right)\cdot X_i\right] = \sum\left(\frac{X_i}{n} - \bar{X}k_iX_i\right) = \bar{X} - \bar{X}\sum(k_iX_i) = 0$$

所以有：

$$\begin{aligned}
\hat{\beta}_1 &= \beta_2 + \sum w_i u_i \\
E(\hat{\beta}_1) &= E(\beta_1 + \sum k_i u_i) \\
&= \beta_1 + E(\sum k_i u_i) \\
&= \beta_1 + \sum[k_i E(u_i)] \\
&= \beta_1
\end{aligned}$$

[证明完毕]!

性质3：方差最小性

方差最小性（Best）：也即估计量是最有效的(Efficient)，是所有线性无偏估计量中，方差为最小的那个估计量。

> 证明：

- 已知估计量 $\hat{\beta}_2$和 $\hat{\beta}_1$的**总体方差**分别是：

$$Var(\hat{\beta}_2) \equiv \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2}$$

$$Var(\hat{\beta}_1) \equiv \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n} \cdot \frac{\sigma^2}{\sum x_i^2}$$

性质3：方差最小性

假设存在用其他方法估计的线性无偏估计量 $\hat{\beta}_2^*$和 $\hat{\beta}_1^*$：

$$\hat{\beta}_2^* = \sum[(k_i + d_i)Y_i] = \sum c_i Y_i$$
$$\hat{\beta}_1^* = \sum[(w_i + g_i)Y_i] = \sum h_i Y_i$$

其中，d_i和 g_i为不全为零的常数（证明略），则可以证明（此处略）：

$$Var(\hat{\beta}_2^*) \geq Var(\hat{\beta}_2)$$
$$Var(\hat{\beta}_1^*) \geq Var(\hat{\beta}_1)$$

因此，方差最小性得以证明！

小结与讨论

小结：

- 评价不同估计方法的**参数估计量性质**，一般是从线性的(Linear)、无偏性(Unbiased)和有效性（方差最小性，Best）三个维度来共同测量。
- OLS估计方法，在CLRM假设下，其参数估计量很好地满足如上三个性质，因此我们称OLS方法估计的参数估计量是最优线性无偏估计量（BLUE）。

思考：

- 关于OLS估计量**有效性**（方差最小性，best）的证明过程你满意么？你能不能自己查阅资料证明一下？找到自己满意的证明过程！（参考答案：建议参阅Greene的《计量经济分析》，他采用的矩阵方法做了完美的证明！）
- 还有没有**其他维度**来评价不同估计方法的**参数估计量性质**？（参考答案：还可以从“一致性”（consistency）维度来评价，主要考察参数估计量的**渐进性质**，也即在样本不断接近总体时估计量的表现。）
- 使得参数估计量具备BLUE性质，仅有OLS方法么（独孤求败）？你能说出一个么？

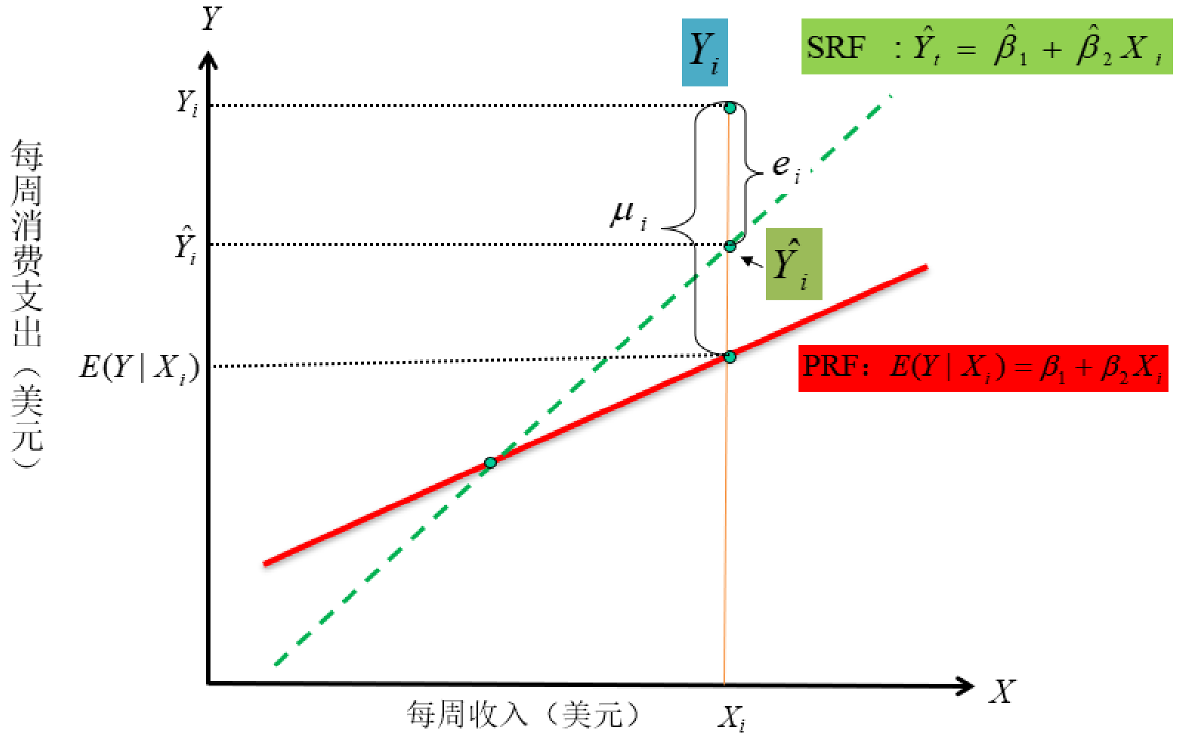
3.5 变异分解与拟合优度

引子：怎么来判定OLS方法对特定样本数据拟合的好坏？

请大家思考如下几个**问题**：

- 样本数据不完全落在拟合的直线（或曲线）上，是经常发生的么？
- 怎么来表达或测量这种对样本数据拟合的不完全性？
- 在OLS方法和CLRM假设“双剑合璧”下，对特定样本数据的拟合不是已经证明最好的么（BLUE）？为什么还要说“拟合”有“好坏之分”？

Y变异的分解



$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

$$y_i = \hat{y}_i \qquad + e_i$$

平方和分解

$$
\begin{aligned}
(Y_i - \bar{Y}) &= (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i) \\
y_i &= \hat{y}_i \qquad + e_i \\
\sum y_i^2 &= \sum \hat{y}_i^2 \qquad + \sum e_i^2 \\
TSS &= ESS \qquad + RSS
\end{aligned}
$$

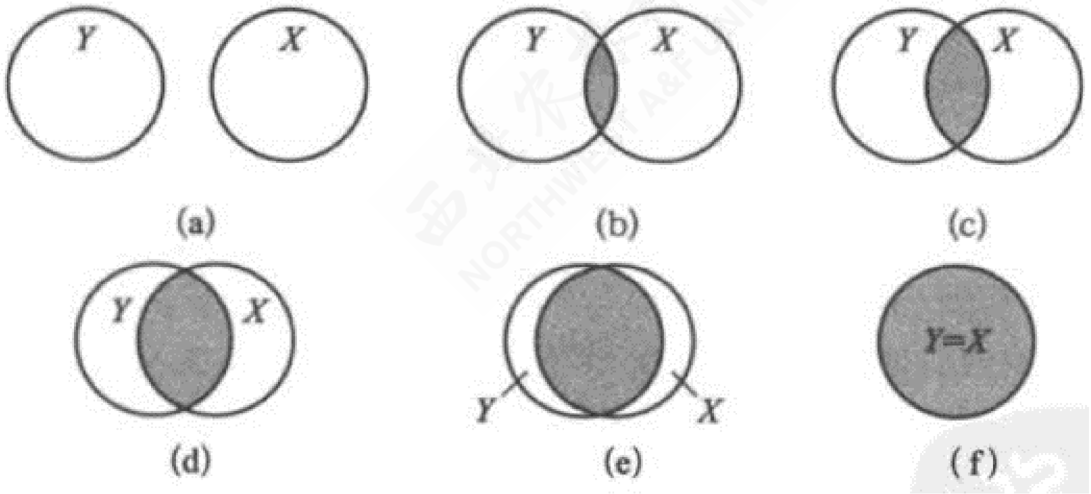
> TSS为总离差平方和; ESS为回归平方和; RSS为残差差平方和

$$
\begin{aligned}
\sum y_i^2 &= \sum (\hat{y}_i + e_i)^2 \\
&= \sum (\hat{y}_i^2 + 2\hat{y}_i e_i + e_i^2) \\
&= \sum \hat{y}_i^2 + 2\sum \hat{y}_i e_i + \sum e_i^2 \\
&= \sum \hat{y}_i^2 + 2\sum ((\hat{\beta}_2 x_i) e_i) + \sum e_i^2 \\
&= \sum \hat{y}_i^2 + 2\hat{\beta}_2 \sum (x_i e_i) + \sum e_i^2 && \leftarrow [\sum x_i e_i = 0] \\
&= \sum \hat{y}_i^2 + \sum e_i^2
\end{aligned}
$$

拟合优度的度量

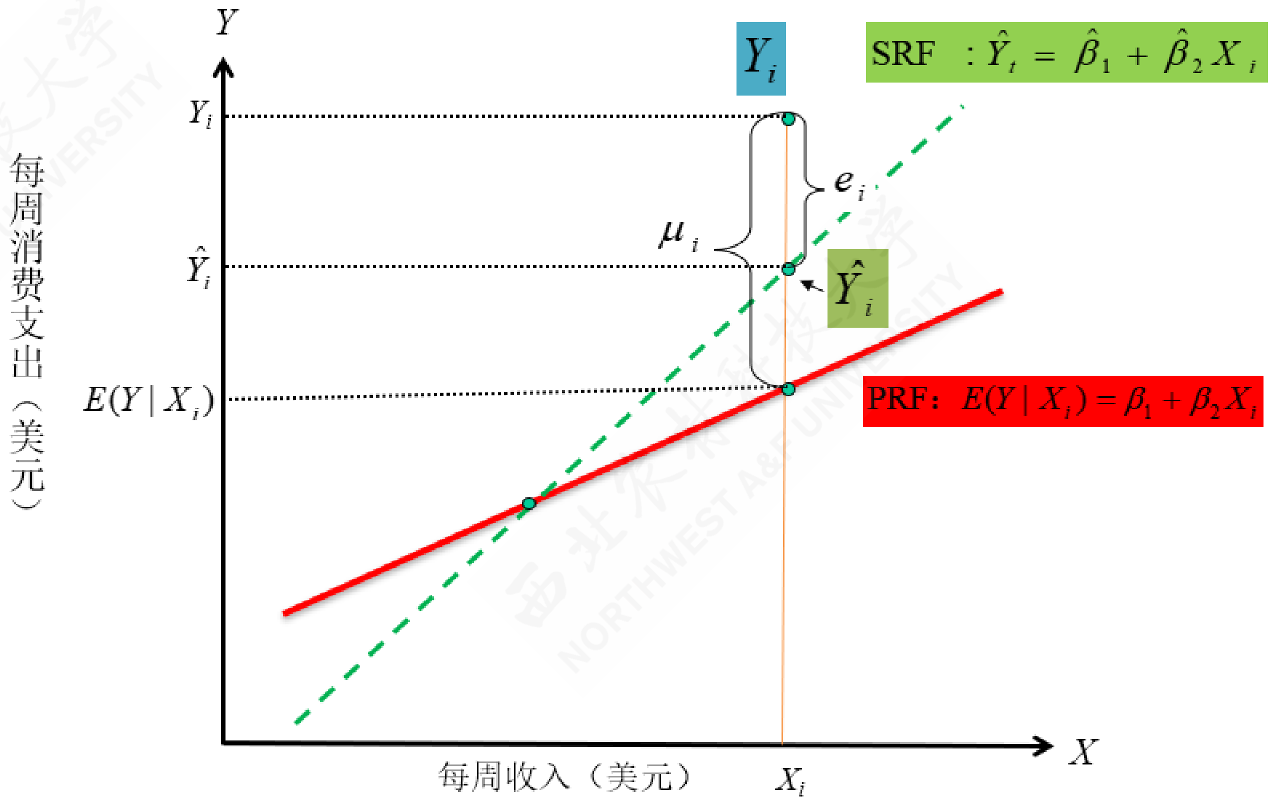
拟合优度（Goodness of fit）：判断样本回归线对一组数据拟合优劣水平的度量。

判定系数（coefficient of determination）：一种利用平方和分解，考察样本回归线对数据拟合效果的总度量。一元回归中，一般记为 r^2；多元回归中，一般记为 R^2。



维恩图看拟合优度

拟合优度的度量



平方和分解看拟合优度

判定系数 r^2 计算公式1：

$$r^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

判定系数 r^2 计算公式2：

$$r^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$$

拟合优度的度量

判定系数 r^2 计算公式3：

$$r^2 = \frac{ESS}{TSS} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\sum (\hat{\beta}_2 x_i)^2}{\sum y_i^2} = \hat{\beta}_2^2 \frac{\sum x_i^2}{\sum y_i^2} = \hat{\beta}_2^2 \frac{S_{X_i}^2}{S_{Y_i}^2}$$

判定系数 r^2 计算公式4：

$$r^2 = \hat{\beta}_2^2 \cdot \frac{\sum x_i^2}{\sum y_i^2} = \left(\frac{\sum x_i y_i}{\sum x_i^2}\right)^2 \cdot \left(\frac{\sum x_i^2}{\sum y_i^2}\right) = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2}$$

课堂讨论：

- 讨论1： r^2 是一个非负量。为什么？
- 讨论2： $0 \leq r^2 \leq 1$，两个端值分别意味什么？

判定系数和简单相关系数的区别与联系

总体相关系数： 是变量 X_i与变量 Y_i总体相关关系的参数，一般记为 ρ。

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X_i)Var(Y_i)}} = \frac{E(X_i - EX)(Y_i - EY)}{\sqrt{E(X_i - EX)^2 E(Y_i - EY)^2}}$$

样本相关系数： 是从总体中抽取随机样本，获得变量 X_i与变量 Y_i样本相关关系的统计量度量，一般记为 r。

$$r = \frac{S_{XY}^2}{S_X * S_Y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

判定系数和相关系数的区别与联系

判定系数和简单相关系数的联系：

- 在一元回归中，判定系数 r^2 等于样本相关系数 r 的平方。

判定系数和简单相关系数的区别：

- 判定系数 r^2 表明因变量变异由解释变量所解释的比例，而相关系数 r 只能表明变量间的线性关联强度。
- 在多元回归中，这种区别会更加凸显！因为那时的相关系数r出现了偏相关的情形(交互关联)！

本节内容小结：

- 即使采用OLS方法，它对样本数据的拟合也是不完全的。意味着实际数据点在样本回归线附近，而不是在样本回归线上。我们可以把样本点行为的“变异”，划分为“回归”能解释的部分和“随机”的部分。并进一步获得变异平方和的分解。
- 判定系数 R^2 是对OLS拟合程度的测量，它使用了变异平方和分解的思想。在一元线性回归（含截距）中，判定系数与相关系数存在如下关系 $R^2 = r^2_{(X_i,Y_i)}$。注意，在多元回归中则不存在这种关系。

小结与思考

问题与思考：

- OLS方法的参数估计量，在CLRM假设满足情况下，就是最优线性无偏估计量（BLUE），为什么还要用**判定系数**来判断“拟合好还是不好？”。对此，你的回答是什么？
- 还有没有其他指标，来反映估计方法对样本数据的拟合好坏程度？请说出一两个。

> 参考答案：还可以有**均方误差和**（MSE）$MSE = RSS/n = 1/n\sum(Y_i - \hat{Y}_i)^2$，以及**均方误差根**（RMSE）等。

3.6 一个数值例子

说一千道一万，重要的是我们自己能不能动手，利用统计软件对前述的计量分析理论进行计算和验证！

下面，我们利用样本数据对**教育和工资案例**进行一次完整的计算和验证吧！

> **教育和工资案例**的总体回归模型（PRM）如下：

$$
\begin{aligned}
Wage_i &= \beta_1 + \beta_2 Edu_i + u_i \\
Y_i &= \beta_1 + \beta_2 X_i + u_i
\end{aligned}
$$

计算表的和项

obs	`$X_i$`	`$Y_i$`	`$X_iY_i$`	`$X_i^2$`	`$Y_i^2$`	`$x_i$`	`$y_i$`	`$x_iy_i$`	`$x_i^2$`	`$y_i^2$`
1	6.00	4.46	26.74	36.00	19.86	-6.00	-4.22	25.31	36.00	17.79
2	7.00	5.77	40.39	49.00	33.29	-5.00	-2.90	14.52	25.00	8.44
3	8.00	5.98	47.83	64.00	35.74	-4.00	-2.70	10.78	16.00	7.27
4	9.00	7.33	65.99	81.00	53.75	-3.00	-1.34	4.03	9.00	1.80
5	10.00	7.32	73.18	100.00	53.56	-2.00	-1.36	2.71	4.00	1.84
6	11.00	6.58	72.43	121.00	43.35	-1.00	-2.09	2.09	1.00	4.37
7	12.00	7.82	93.82	144.00	61.12	0.00	-0.86	-0.00	0.00	0.73
8	13.00	7.84	101.86	169.00	61.39	1.00	-0.84	-0.84	1.00	0.70
9	14.00	11.02	154.31	196.00	121.49	2.00	2.35	4.70	4.00	5.51
10	15.00	10.67	160.11	225.00	113.93	3.00	2.00	6.00	9.00	4.00
11	16.00	10.84	173.38	256.00	117.42	4.00	2.16	8.65	16.00	4.67
12	17.00	13.62	231.46	289.00	185.37	5.00	4.94	24.70	25.00	24.41
13	18.00	13.53	243.56	324.00	183.09	6.00	4.86	29.14	36.00	23.58
sum	156.00	112.77	1485.04	2054.00	1083.38	0.00	0.00	131.79	182.00	105.12

计算回归系数

公式1：（Favorite Five，FF形式）

$$\hat{\beta}_2 = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$
$$= \frac{13 * 1485.04 - 156 * 112.771}{13 * 2054 - 156^2} = 0.7241$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X} = 8.6747 - 0.7241 * 12 = -0.0145$$

计算回归系数

公式2：（离差形式，favorite five，ff形式）

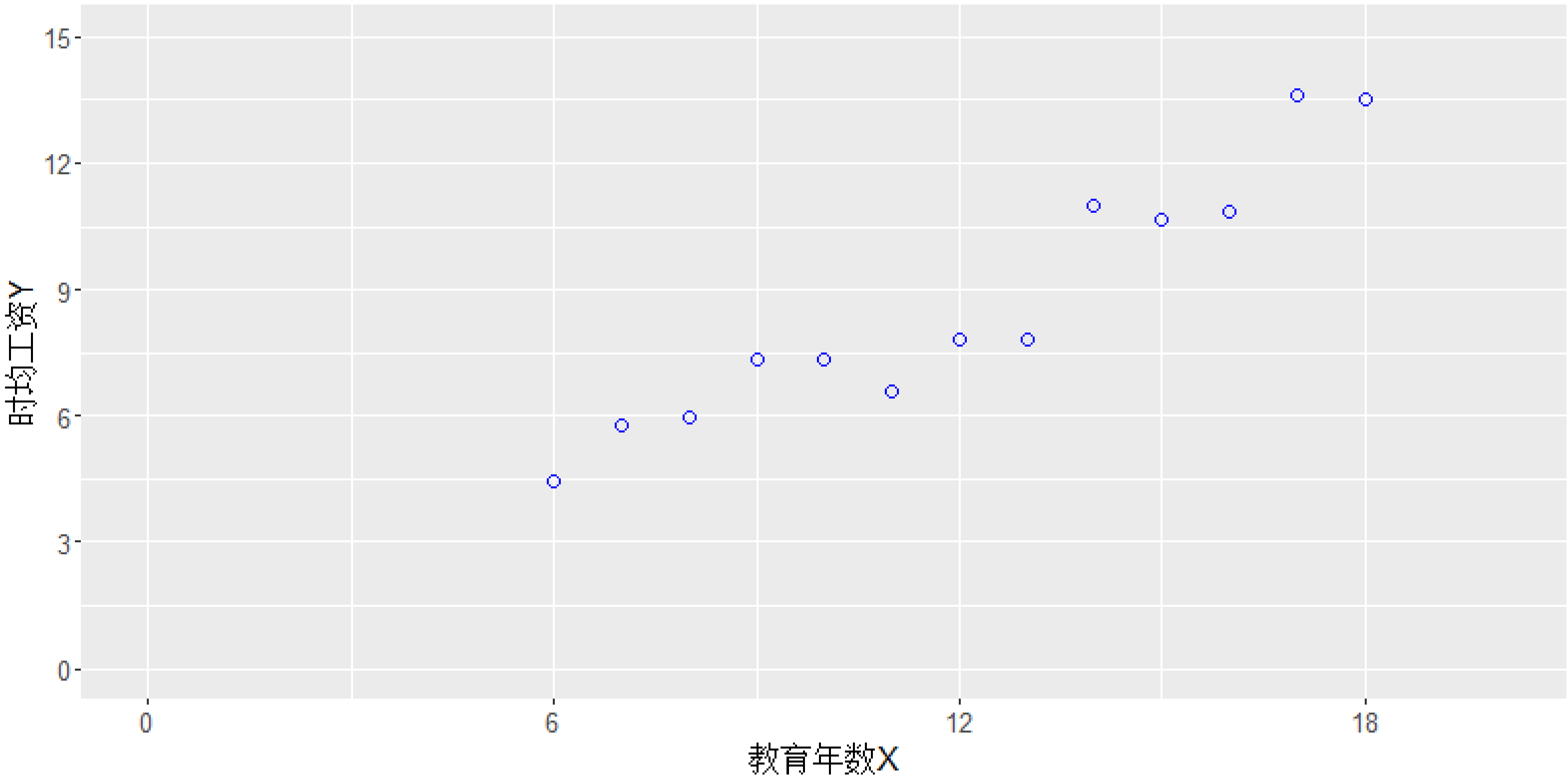
$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{131.786}{182} = 0.7241$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 8.6747 - 0.7241 * 12 = -0.0145$$

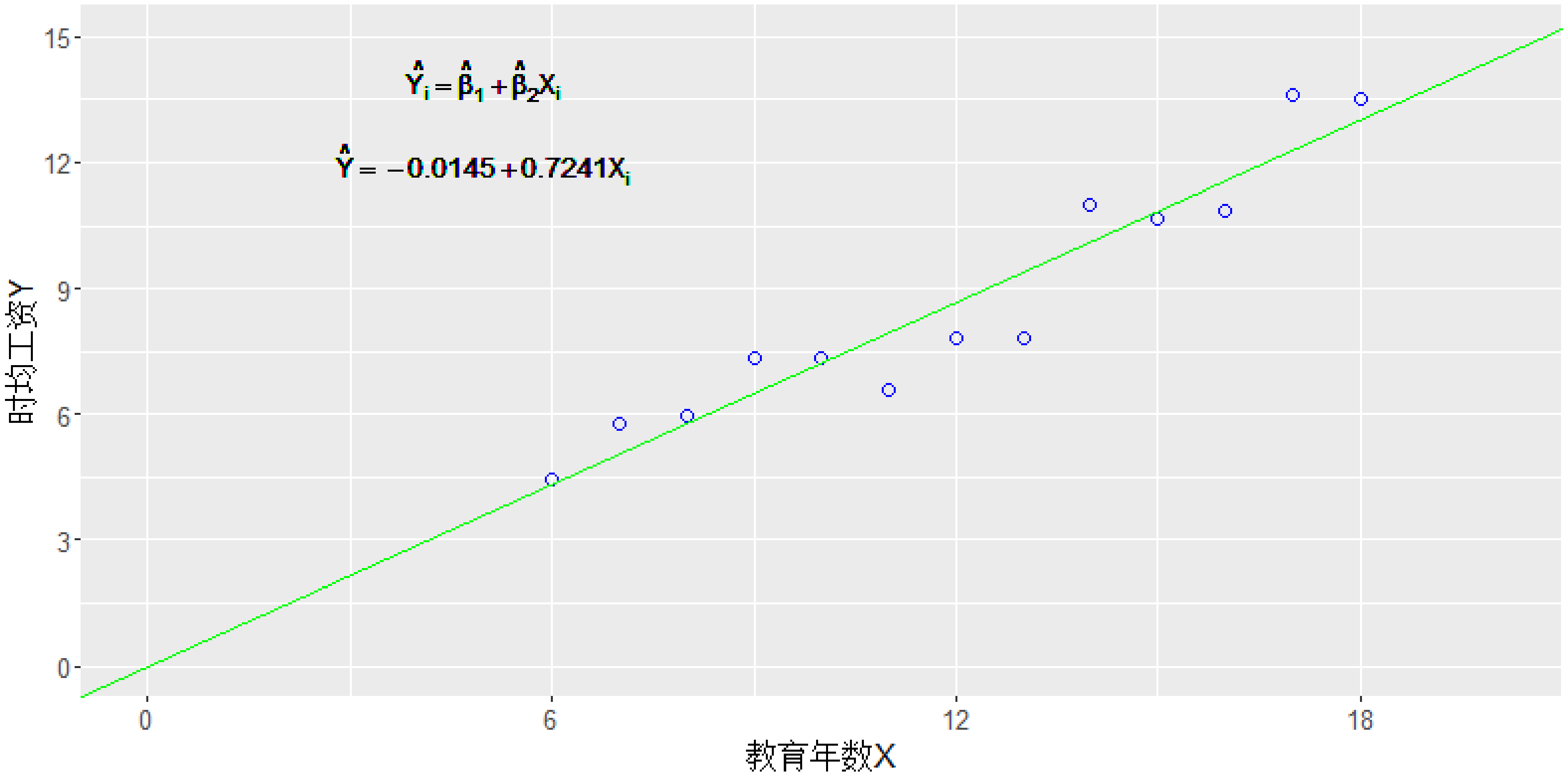
样本回归方程SRF

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i = -0.0145 + 0.7241 X_i$$

样本回归线SRL



样本回归线SRL



$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$
$\hat{Y} = -0.0145 + 0.7241X_i$
时均工资Y
教育年数X
15
12
9
6
3
0
0
6
12
18

计算得到拟合值和残差

obs	`$X_i$`	`$Y_i$`	`$\hat{Y}_i$`	`$e_i$`	`$e_i^2$`
1	6.0000	4.4567	4.3301	0.1266	0.0160
2	7.0000	5.7700	5.0542	0.7158	0.5123
3	8.0000	5.9787	5.7783	0.2004	0.0402
4	9.0000	7.3317	6.5024	0.8293	0.6877
5	10.0000	7.3182	7.2265	0.0917	0.0084
6	11.0000	6.5844	7.9506	-1.3662	1.8665
7	12.0000	7.8182	8.6747	-0.8565	0.7336
8	13.0000	7.8351	9.3988	-1.5637	2.4452
9	14.0000	11.0223	10.1229	0.8994	0.8089
10	15.0000	10.6738	10.8470	-0.1732	0.0300
11	16.0000	10.8361	11.5711	-0.7350	0.5402
12	17.0000	13.6150	12.2952	1.3198	1.7419
13	18.0000	13.5310	13.0193	0.5117	0.2618
sum	156.0000	112.7712	112.7712	0.0000	9.6928

根据以上样本回归方程，可以计算得到 Y_i 的回归拟合值 $\hat{Y}_i$，以及回归残差 e_i。

$$
\begin{aligned}
\hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\
e_i &= Y_i - \hat{Y}_i
\end{aligned}
$$

回归误差方差和标准差

回归误差方差 $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{(n-2)} = \frac{9.693}{11} = 0.8812$$

回归误差标准差 $\hat{\sigma}$:

$$\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{(n-2)}} = \sqrt{0.8812} = 0.9387$$

计算回归系数的样本方差

$$S_{\hat{\beta}_2}^2 = \frac{\hat{\sigma}^2}{\sum x_i^2} = \frac{0.8812}{182} = 0.0048$$

$$S_{\hat{\beta}_2} = \sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}} = \sqrt{0.0048} = 0.0696$$

$$S_{\hat{\beta}_1}^2 = \frac{\sum X_i^2}{n} \frac{\hat{\sigma}^2}{\sum x_i^2} = \frac{2054}{13} \frac{0.8812}{182} = 0.765$$

$$S_{\hat{\beta}_1} = \sqrt{\frac{\sum X_i^2}{n} \frac{\hat{\sigma}^2}{\sum x_i^2}} = \sqrt{0.765} = 0.8746$$

计算平方和分解

$$TSS = \sum (Y_i - \bar{Y})^2 = 105.1183$$

$$RSS = \sum (Y_i - \hat{Y}_i)^2 = 9.693$$

$$ESS = \sum (\hat{Y}_i - \bar{Y})^2 = 95.4253$$

相关系数和判定系数

样本相关系数 r:

$$r = \frac{S_{XY}^2}{S_X * S_Y} = \frac{10.9821}{3.8944 * 2.9597} = 0.9528$$

回归方程的判定系数 r^2:

$$r^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{9.693}{105.1183} = 0.9078$$

二者关系

下面对本节内容做一个**小结**：

- 通过样本数值例子，我们可以全部自己“手工计算”OLS方法参数估计、估计的精度（估计量样本方差）、变异的平方和分解，以及判定系数等过程环节。
- 实际上任何一款统计软件都会很快**帮我们**完成这些计算过程，实证分析中我们不需要亲自去计算它们。
- 送上一句忠告，统计软件就像一个“技术黑箱”，如果你不理解“黑箱”里面的运作原理，那么你就永远只是被它“牵着鼻子走”！所以，大家起码要自己手工计算**一遍**！
- 而且，随着开源软件（如`R`或`Python`等）的普遍流行，**“你”**的介入和作用可能越来越重要！因为你可以更加灵活、更加自由地进行改造、重塑和创新！

问题与思考：

- 请大家自己使用任何熟悉的统计软件，完成本节的所有环节的计算！
- 如果你和其他同学的随机样本数据，都来自同一个总体，你们的计算结果会是一样的么？全班同学全部计算结果，会给“从样本推断总体”带来什么启示？

3.7 经典正态线性回归模型（N-CLRM）

引子：我们对总体参数的认识已经足够了么？

我们已经证明了在经典详细回归模型假设下（CLRM），采用普通最小二乘法（OLS），得到参数 Φ的估计量 $\hat{\Phi}$，是最优线性无偏估计量（BLUE）。

而且也知道了，在随机抽样下，OLS的参数估计量是“随机变量”，其期望和方差分别是（以斜率参数的估计量为例）：

$$\begin{cases} E(\hat{\beta}_2) = \beta_2 \\ var(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum x_i^2} \end{cases}$$

我们马上想要提出的问题就是：既然这个参数估计量是**随机变量**，那么它服从什么分布呢？

随机干扰项的正态性假设

CLRM假设下：OLS方法得到的已经是BLUE了

CLRM假设下：随机干扰项 u_i的期望值为零（$Eu_i = 0$），u_i, u_j不相关（$E(u_i, u_j) = 0$），且有一个不变方差（$Var(u_i) \equiv \sigma^2$）。

总体回归模型PRM:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

样本回归模型SRM:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

讨论：

- 一个样本怎样才推断总体PRF？
- 多份样本而言，OLS估计量（$\hat{\beta}_i, \hat{\sigma}^2$）因为样本变化而变化(随机变量)

随机干扰项的正态性假设

随机干扰项 u_i对于估计SRF $\rightarrow$推断PRF的重要性：

$$\hat{\beta}_2=\sum\left(\frac{x_i}{\sum x_i^2}\cdot Y_i\right)=\sum k_iY_i=\sum\left[k_i(\beta_1+\beta_2X_i+u_i)\right]$$

$$\hat{\beta}_1=\sum\left(\left(\frac{1}{n}-k_i\bar{X}\right)\cdot Y_i\right)=\sum w_iY_i=\sum\left[w_i(\beta_1+\beta_2X_i+u_i)\right]$$

$$\hat{\sigma}^2=\frac{\sum e_i^2}{n-2}$$

经典正态线性回归模型假设（N-CLRM）

经典正态线性回归模型(classical normal linear regression model , N-CLRM)：在经典线性回归模型(CLRM)假设中再增加干扰项 u_i服从正态性的相关假设。

- 均值为0： $E(u|X_i) = 0$
- 同方差： $Var(u_i|X_i) \equiv \sigma^2$
- 无自相关： $E(u_i, u_j|X_i) = 0$
- 正态性分布： $u_i \sim N(0, \sigma^2)$

以上几条也可以统写为： $u_i \sim iid.\ N(0, \sigma^2)$

其中，iid表示独立同分布(Independent Identical Distribution, iid)。

N-CLRM假设下OLS估计量的性质

在N-CLRM假设下，OLS估计量有如下统计性质：

- 性质1：无偏性
- 性质2：有效性（方差最小）
- 性质3：一致性（收敛到它们的总体参数上）

- 性质4：估计量 $\hat{\beta}_2$是正态分布的：

$$\hat{\beta}_2 \sim N(\mu_{\hat{\beta}_2}, \sigma^2_{\hat{\beta}_2})$$

$$\mu_{\hat{\beta}_2} = E(\hat{\beta}_2) = \beta_2$$

$$\sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2}$$

随机变量 Z_2服从标准正态分布：

$$Z_2 = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \sim N(0,1)$$

$$\mu(Z_2) = E(\hat{\beta}_2 - \beta_2) = 0$$

$$\sigma^2_{Z_2} = Var\left(\frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}}\right) = \frac{Var(\hat{\beta}_2)}{\sigma^2_{\hat{\beta}_2}} = 1$$

N-CLRM假设下OLS估计量的性质

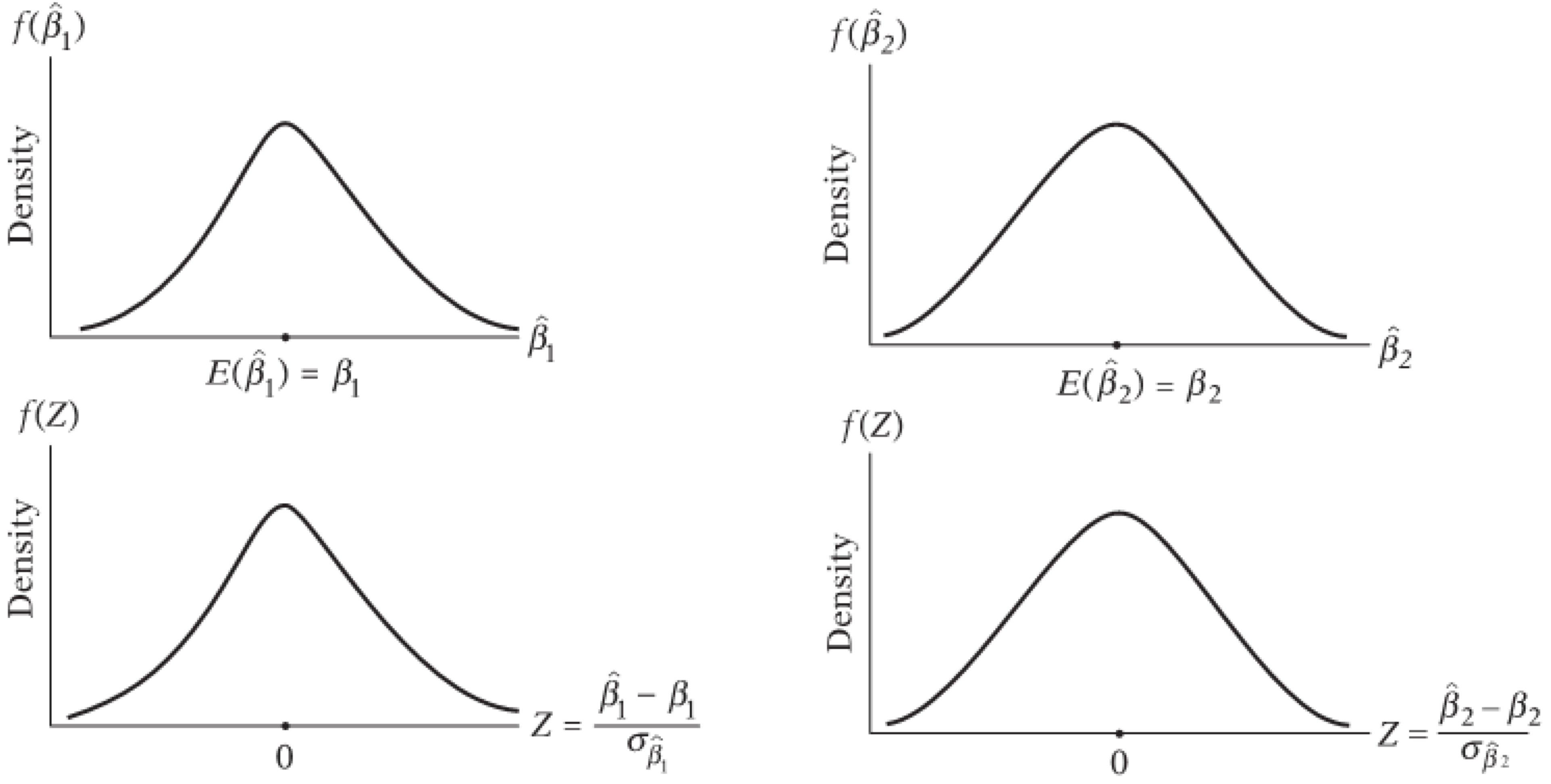
- 性质5：估计量 $\hat{\beta}_1$是正态分布的：

$$\hat{\beta}_1 \sim N(\mu_{\hat{\beta}_1}, \sigma^2_{\hat{\beta}_1})$$
$$\mu_{\hat{\beta}_1} = E(\hat{\beta}_1) = \beta_1$$
$$\sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n} \cdot \frac{\sigma^2}{\sum x_i^2}$$

随机变量 Z_1服从标准正态分布：

$$Z_1 = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0, 1)$$
$$\mu(Z_1) = E(\hat{\beta}_1 - \beta_1) = 0$$
$$\sigma^2_{Z_1} = Var\left(\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}}\right) = \frac{Var(\hat{\beta}_1)}{\sigma^2_{\hat{\beta}_1}} = 1$$

N-CLRM假设下OLS估计量的性质



N-CLRM假设下OLS估计量的性质

- 性质6： $X \equiv (n-2)\hat{\sigma}^2/\sigma^2$服从自由度为$(n-2)$的卡方分布。

$$X \equiv (n-2)\hat{\sigma}^2/\sigma^2$$
$$X \sim \chi^2(n-2)$$

- 性质7：随机变量$(\hat{\beta}_2, \hat{\beta}_1)$的分布独立于随机变量$\hat{\sigma}^2$
- 性质8：估计量$(\hat{\beta}_2, \hat{\beta}_1)$在所有无偏估计中，无论是线性还是非线性，都有最小的方差。也即，它们是最优无偏估计量（**B**est **U**nbiased **E**stimators, **BUE**）。

小结与讨论

本节**内容小结**：

- 除了关心参数估计量的性质（是否BLUE），我们还需要关注：参数估计量作为一个**随机变量**，会服从什么概率分布？这样才会为后面的**假设检验**提供基础！它将成为完成“由样本推断总体”逻辑分析的最后一个台阶，也是最重要的一个环节之一！
- N-CLRM假设体系，是在CLRM假设基础之上，额外再增加了一条关于随机干扰项服从**正态分布**的假设。基于此，我们可以推断得到回归系数估计量也将服从正态分布，从而进一步可以构造出很多有用的**样本统计量**，例如后面要学的**t统计量**、**F统计量**等。

思考与讨论：

- 你能说出现实中，随机干扰项 u_i服从其他概率分布的情形？

> 参考答案：显然，现实社会现象中有很多不服从正态分布的情形，比如t分布、二项分布等。

3.8 极大似然估计法(ML)

极大似然估计法(maximum likelihood, ML):

- 是由Fisher提出的一种参数估计方法基本思想：设总体分布的函数形式已知，但有未知参数 Θ， Θ可以取很多值，在 Θ的一切可能取值中选一个使样本观察值出现的概率为最大的 $\hat{\Theta}$值作为 Θ的估计值，并称估计值 $\hat{\Theta}$为参数 Θ的极大似然估计值。这种求估计量的方法称为极大似然估计法。

似然函数表达式：

- 设总体 X_i的概率密度函数 $f(X_i;\Theta)$为,其中 Θ为待估计参数。对于从总体中取得的样本观测值 $(X_1;X_2,\cdots,X_n)$， 其联合密度函数为 $\prod f(X_i;\Theta)$，它是参数 Θ的函数，称之为的似然函数，记为 $L(\Theta)$：

$$L(\Theta)=\prod f(X_i;\Theta)$$

ML估计法与OLS估计法的关系

极大似然估计法(ML)比较复杂，我们仅需知道。在随机干扰项正态性假设下（N-CLRM）：

- 回归系数 β_i的ML估计量和OLS估计量是相同的——无论是一元回归还是多元回归!
- 对于 σ^2的估计，其ML估计量为$\sum e_i^2/n$，是**有偏**的；其OLS估计量是$\sum e_i^2/(n-2)$，是**无偏**的。
- 关于 σ^2的两种估计量，随着样本容量n的增大，两者将趋于相等！

启示：OLS方法真好！

本章结束