计量经济学：作业一

刘泓尊　2018011446　计 84

liu-hz18@mails.tsinghua.edu.cn

2020.10

1.

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{SST_x} \\
&= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{SST_x} \\
&= \frac{0 + \beta_1 SST_x + \sum_{i=1}^{n} u_i(x_i - \bar{x})}{SST_x} \\
&= \beta_1 + \frac{\sum_{i=1}^{n} u_i(x_i - \bar{x})}{SST_x} \\
&= \beta_1 + \frac{\sum_{i=1}^{n} d_i u_i}{SST_x} \quad (d_i = x_i - \bar{x})
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
Var(\hat{\beta}_1) &= (\frac{1}{SST_x})^2 Var(\sum_{i=1}^{n} d_i u_i) \\
&= (\frac{1}{SST_x})^2 \sum_{i=1}^{n} {d_i}^2 Var(u_i) \\
&= (\frac{1}{SST_x})^2 \sum_{i=1}^{n} {d_i}^2 \sigma^2 \\
&= (\frac{\sigma}{SST_x})^2 \sum_{i=1}^{n} {d_i}^2 \\
&= (\frac{\sigma}{SST_x})^2 \\
&= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
E[(\hat{\beta}_1-\beta_1)\bar{u})] &= E[\bar{u}\frac{\sum_{i=1}^{n} d_i u_i}{SST_x}] \\
&= \frac{1}{SST_x}\sum_{i=1}^{n} d_i \bar{u} E(u_i) \\
&= \frac{1}{SST_x}\sum_{i=1}^{n} (\bar{u})^2 \\
&= \frac{\bar{u}^2}{SST_x}\sum_{i=1}^{n} d_i \\
&= 0
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
Cov(\hat{\beta}_1, \bar{u}) &= E[\hat{\beta}_1 \bar{u}] - E[\hat{\beta}_1]E[\bar{u}] \\
&= 0 - 0 \\
&= 0
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
Var(\hat{\beta}_0) &= Var[\beta_0 + \bar{u} + (\beta_1 - \hat{\beta}_1)\bar{x}] \\
&= Var[\bar{u} - \hat{\beta}_1 \bar{x}] \\
&= Var[\bar{u}] + Var[\hat{\beta}_1]\bar{x}^2 \quad (Cov(\hat{\beta}_1, \bar{u}) = 0) \\
&= Var[\bar{u}] + \bar{x}^2 \frac{\sigma^2}{SST_x} \\
&= \frac{\sigma^2}{n} + \frac{\bar{x}^2\sigma^2}{SST_x} \\
&= \frac{\sigma^2}{nSST_x}(SST_x + n\bar{x}^2) \\
&= \frac{\sigma^2 \sum_{i=1}^{n} {x_i}^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned} \tag{5}
$$

2. 不正确。

   对于给定抽样 x，虽然有 $E[u|x] = 0$，但是因为 u_i 仅有有限个样本，不能从统计意义上推断 $\sum_{i=1}^{n} u_i = 0$.

   此外，$E[\hat{\beta}_1] = \beta_1$ 的无偏性也不能说明 $\hat{\beta}_1 = \beta_1$.

   此题混淆了统计意义和抽样意义的区别。

3. 代码如下：

```
cd "D:\大三上\计量\作业\" //注意修改数据路径
use "EduIncome.dta", clear

summarize gender birthyear marriage empjob_twage schooling_yr

generate female = 1 if gender == 1
replace female = 0 if female==.
tab female

reg empjob_twage schooling_yr female
```

```
generate pred_twage = 1984.774*schooling_yr - 14048.56 * female + 46554.54

generate pred_u = empjob_twage - pred_twage

summarize empjob_twage pred_twage pred_u

drop pred_twage pred_u
```

(a) 除 id 之外所有变量的均值、标准差、最小值和最大值：

```
. summarize gender birthyear marriage empjob_twage schooling_yr
```

Variable	Obs	Mean	Std. Dev.	Min	Max
gender	2,539	1.688854	.4630536	1	2
birthyear	2,539	1974.824	11.30764	1914	1997
marriage	2,539	.6025994	.4894565	0	1
empjob_twage	2,539	57321.29	40688.22	2113.569	608707.9
schooling_yr	2,539	7.627019	2.942299	0	15

(b) 创建新变量 female 并统计女性比例 女性占比 31.11%

```
. generate female = 1 if gender == 1
(1,749 missing values generated)

. replace female = 0 if female==.
(1,749 real changes made)

. tab female
```

female	Freq.	Percent	Cum.
0	1,749	68.89	68.89
1	790	31.11	100.00
Total	2,539	100.00	

(c) 回归方程，给出结果

```
. reg empjob_twage schooling_yr female

      Source |       SS           df       MS      Number of obs   =     2,539
-------------+----------------------------------   F(2, 2536)      =     65.77
       Model |  2.0719e+11         2  1.0360e+11   Prob > F        =    0.0000
    Residual |  3.9945e+12     2,536  1.5751e+09   R-squared       =    0.0493
-------------+----------------------------------   Adj R-squared   =    0.0486
       Total |  4.2017e+12     2,538  1.6555e+09   Root MSE        =     39688

------------------------------------------------------------------------------
empjob_twage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
schooling_yr |   1984.774    268.381     7.40   0.000     1458.506    2511.042
      female |  -14048.56   1705.326    -8.24   0.000    -17392.54   -10704.59
       _cons |   46554.54   2289.326    20.34   0.000      42065.4    51043.68
------------------------------------------------------------------------------
```

(d) 受教育年限一样时，平均来讲男性年收入比女性年收入高 14048.56 元。

(e) 教育年限每增加一年，收入平均增加 1984.774 元

(f) 计算均值

```
. generate pred_twage = 1984.774*schooling_yr - 14048.56 * female + 46554.54

.
. generate pred_u = empjob_twage - pred_twage

.
. summarize empjob_twage pred_twage pred_u

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
empjob_twage |      2,539    57321.29    40688.22   2113.569   608707.9
  pred_twage |      2,539    57321.29    9035.289   32505.98   76326.15
      pred_u |      2,539   -.0007692    39672.35  -64690.55   544290.4
```

(g) $R^2 = 0.0493$，所以有 4.93% 的比例的收入波动被教育年限和性别解释。(修正的 $R^2 = 0.0486$)

4. (a) 不对。满足 $x_1 + x_2 = 1$, 并不是完全线性关系 (如 $x_1 = \alpha x_2$)。两个变量存在相关性，但并不是完全线性。所以依然满足 $MLR.3$，可以用 OLS。

(b)

$$E[ux_1] = E[E[(ux_1|x_1, x_2)]] = E[x_1 E[u|x_1, x_2]] = E[x_1 0] = 0 \tag{6}$$

$$E[ux_2] = E[E[(ux_2|x_1, x_2)]] = E[x_2 E[u|x_1, x_2]] = E[x_2 0] = 0 \tag{7}$$

(c) 由 $E[ux_1] = 0$ 和 $E[ux_2] = 0$ 得到

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} x_{1_i}(y_i - \hat{\beta}_1 x_{1_i} - \hat{\beta}_2 x_{2_i}) = 0 \\ \frac{1}{n}\sum_{i=1}^{n} x_{2_i}(y_i - \hat{\beta}_1 x_{1_i} - \hat{\beta}_2 x_{2_i}) = 0 \end{aligned} \tag{8}$$

(d) 目标函数

$$H = \sum_{i=1}^{n}(y_i - b_1 x_{1_i} - b_2 x_{2_i})^2 \tag{9}$$

$$(\hat{\beta}_1, \hat{\beta}_2) = \arg_{(b_1, b_2)} \min H \tag{10}$$

(e) 一阶条件：

$$
\begin{aligned}
\frac{\partial H}{\partial b_1} &= -2\sum_{i=1}^{n} x_{1_i}(y_i - b_1 x_{i_1} - b_2 x_{i_2}) = 0 \\
\frac{\partial H}{\partial b_2} &= -2\sum_{i=1}^{n} x_{2_i}(y_i - b_1 x_{i_1} - b_2 x_{i_2}) = 0
\end{aligned} \tag{11}
$$

(f) 参数估计是一样的。从数学上看，两者的约束方程是等价的。因此得到的解是一致的。

(g) 解得

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_{2_i}^2 \sum_{i=1}^{n} x_{1_i} y_i - \sum_{i=1}^{n} x_{2_i} y_i \sum_{i=1}^{n} x_{1_i} x_{2_i}}{\sum_{i=1}^{n} x_{1_i}^2 x_{2_i}^2 - (\sum_{i=1}^{n} x_{1_i} x_{2_i})^2} \\
\hat{\beta}_2 &= \frac{\sum_{i=1}^{n} x_{1_i}^2 \sum_{i=1}^{n} x_{2_i} y_i - \sum_{i=1}^{n} x_{1_i} y_i \sum_{i=1}^{n} x_{1_i} x_{2_i}}{\sum_{i=1}^{n} x_{1_i}^2 x_{2_i}^2 - (\sum_{i=1}^{n} x_{1_i} x_{2_i})^2}
\end{aligned} \tag{12}
$$

由数据计算的 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 为

$$
\begin{aligned}
\hat{\beta}_1 &= 5.0 \\
\hat{\beta}_2 &= 6.0
\end{aligned} \tag{13}
$$

附上述计算的 Python 程序：

```python
import numpy as np
y = np.array([5, 9, 7, 3, 1, 6, 4, 8, 2, 10])
x1 = np.array([1, 1, 1, 1, 1, 0, 0, 0, 0, 0])
x2 = np.array([0, 0, 0, 0, 0, 1, 1, 1, 1, 1])
x1_x2 = x1 * x2

beta1 = (x2.dot(x2) * x1.dot(y) - x2.dot(y) * x1.dot(x2)) / (x1.dot(x1)
    * x2.dot(x2) - (x1_x2.sum())**2)
beta2 = (x1.dot(x1) * x2.dot(y) - x1.dot(y) * x1.dot(x2)) / (x1.dot(x1)
    * x2.dot(x2) - (x1_x2.sum())**2)
print(beta1, beta2)
```