计量经济学：作业二

刘泓尊　2018011446　计 84

liu-hz18@mails.tsinghua.edu.cn

2020.10

1.

$$H = \sum {u_i}^2 = \sum (y_i - \beta x_i)^2 \tag{1}$$

令残差平方和对 β 的偏导为 0:

$$\frac{\partial H}{\partial \beta} = \sum -2(y_i - \beta x_i)(x_i) = -2\sum x_i y_i - \beta {x_i}^2 = 0$$

解得:

$$\beta = \frac{\sum x_i y_i}{\sum {x_i}^2}$$

2. 考虑一个通式:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

将 x_1 对 $x_2, x_3, \cdots, x_k$ 和常数做回归:

$$x_{i1} = \delta_0 + \delta_1 x_{i2} + \cdots \delta_k x_{ik} + r_{i1}$$

因为

$$\sum \hat{r_{i1}} = 0, \sum \hat{r_{i1}} x_{ik} = 0 (\forall k), \sum \hat{r_{i1}} \hat{x_{i1}} = 0, x_{i1} = x_{i1} \hat{+}\ \hat{r_{i1}}]$$

代入: $\sum x_{i1}\hat{u_i} = 0$ 得到

$$\begin{aligned}
\sum x_{i1}\hat{u_i} &= (\sum \delta_1 + \delta_2 x_{i2} + \cdots + \delta_k x_{ik} + \hat{r_{i1}})\hat{u_i} \\
&= \delta_1 \sum \hat{u_i} + \delta_2 \sum x_{i2}\hat{u_i} + \cdots + \delta_k \sum x_{ik} + \hat{u_i} \\
&= \sum \hat{r_{i1}}(y_i - \hat{\beta_0} - \beta_1 \hat{x}_{i1} - \cdots - \beta_k \hat{x}_{ik}) \\
&= \sum r_{i1} y_i - \beta_0 \sum r_{i1} - \beta_1 \sum r_{i1}(x_{i1} + r_{i1}) - \cdots \\
&= \sum r_{i1} \hat{y}_i - \beta_1 \sum \hat{r}_{i1}^2 = 0
\end{aligned} \tag{2}$$

解得:

$$\beta_1 = \frac{\sum r_{i1} y_i}{\sum r_{i1}^2}$$

因为

$$\sum r_{i1} = 0$$

所以上式中 β_0 是否存在是没有影响的，因为无论如何这一项都为 0.

3. 因为残差平方和对 β 的偏导为 0 导出条件:

$$\frac{\partial H}{\partial \beta} = \sum -2(y_i - \beta x_i)(x_i) = 0$$

根据上式并不能确定 $\sum u_i = \sum (y_i - \beta x_i) = 0$，因为还有 x_i 的影响。

实际上可以举例 $(x_1, y_1) = (1, 1), (x_2, y_2) = (2, 1.5)$

计算得出：

$$\beta = 0.8$$

但是

$$\sum u_i = u_1 + u_2 = (1 - 0.8) + (1.5 - 1.6) = -0.1$$

所以不包含截距回归的残差平方和不为 0

4.

$$\tilde{\beta_1} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum (x_i)^2 - (\sum x_i)^2}$$

$$\tilde{\beta_0} = \bar{y} - \tilde{\beta_1}\bar{x}$$

$$\hat{\beta_1} = \frac{\sum x_i y_i}{\sum (x_i)^2}$$

右侧:

$$\begin{aligned}
\tilde{\beta_1} + \tilde{\beta_0}\frac{\sum x_i}{\sum {x_i}^2} &= \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum (x_i)^2 - (\sum x_i)^2}(1 - \frac{(\sum x_i)^2}{\sum {x_i}^2}) + \frac{\sum y_i \sum x_i}{n\sum {x_i}^2} \\
&= \frac{n\sum x_i y_i - \sum x_i \sum y_i + \sum x_i \sum y_i}{n\sum {x_i}^2} \\
&= \frac{\sum x_i y_i}{\sum {x_i}^2} \\
&= \hat{\beta_1}
\end{aligned} \tag{3}$$

原式成立。

Summary for variables: cigs
by categories of: agegrp

agegrp	mean	N
0	7.674912	283
1	10.775	280
2	8.693467	199
3	2.022222	45
Total	8.686493	807

图 1:

5. (a) 年龄和抽烟数目不是线性关系，因为从图中可以看到，随着年龄上升，抽烟数目呈现先上升后下降的趋势。

```
. * Estimate: cigs ~ age, age^2, educ, restaurn. agesq: square of age
. reg cigs age agesq educ restaurn
```

Source	SS	df	MS		
Model	7739.48459	4	1934.87115	Number of obs	= 807
Residual	144014.198	802	179.568826	F(4, 802)	= 10.78
				Prob > F	= 0.0000
				R-squared	= 0.0510
				Adj R-squared	= 0.0463
Total	151753.683	806	188.280003	Root MSE	= 13.4

cigs	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.822327	.1541866	5.33	0.000	.51967	1.124984
agesq	-.0095886	.0016779	-5.71	0.000	-.0128822	-.006295
educ	-.4504	.1614857	-2.79	0.005	-.7673845	-.1334156
restaurn	-2.746372	1.09685	-2.50	0.012	-4.899408	-.5933367
_cons	.1521404	3.503322	0.04	0.965	-6.724623	7.028904

图 2:

(b) 回归方程:

$$cigs = 0.1521 + 0.822age - 0.0095age^2 - 0.4504educ - 2.746restaurn$$

在控制了教育水平和禁烟政策之后，平均来说年龄 42.880513 的人群吸烟最多。

(c) $H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$

$H_1 : H_0$ 不正确

进行 F 检验，通过 p 值与显著性水平 α 的大小，来确定是拒绝原假设还是接受原假设。P 值越

```
( 1)  age = 0
( 2)  agesq = 0
( 3)  educ = 0

      F(  3,    802) =   12.23
           Prob > F =    0.0000
```

图 3:

小，我们拒绝原假设的理由越充分。

在 %1 的显著性水平下，因为 $p = 0.000 < \alpha = 0.01$，所以拒绝原假设。

在 %5 的显著性水平下，因为 $p = 0.000 < \alpha = 0.05$，所以拒绝原假设。

(d) $H_0 : \beta_4 = 0$

$H_1 : H_0$ 不正确. 在 %1 的显著性水平下，因为 $p = 0.0125 > \alpha = 0.01$，所以接受原假设。

在 %5 的显著性水平下，因为 $p = 0.0125 < \alpha = 0.05$，所以拒绝原假设。

```
( 1)  restaurn = 0

      F(  1,    802) =    6.27
           Prob > F =    0.0125
```

图 4:

(e) 回归结果：

```
. reg cigs age agesq restaurn educ i.restaurn#c.edu

      Source |       SS           df       MS      Number of obs   =       807
-------------+----------------------------------   F(5, 801)       =      8.61
       Model |  7740.79264         5  1548.15853   Prob > F        =    0.0000
    Residual |  144012.89        801  179.791373   R-squared       =    0.0510
-------------+----------------------------------   Adj R-squared   =    0.0451
       Total |  151753.683       806  188.280003   Root MSE        =    13.409

------------------------------------------------------------------------------
        cigs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .8225172   .1542982     5.33   0.000     .5196405    1.125394
       agesq |   -.0095893    .001679    -5.71   0.000    -.0128849   -.0062936
    restaurn |   -2.357973    4.68395    -0.50   0.615    -11.55224    6.836294
        educ |   -.4426885   .1851587    -2.39   0.017    -.8061421   -.0792349
             |
restaurn#c.educ |
          1  |   -.0306016   .3587693    -0.09   0.932    -.7348406    .6736375
             |
       _cons |    .0502483   3.703441     0.01   0.989    -7.219348    7.319844
------------------------------------------------------------------------------
```

图 5:

restaurn 对 cigs 的偏效应为 $\beta_4 + \beta_5 \cdot educ = -2.3579 - 0.0306 educ$

β_5 的含义是在其他条件不变时，禁烟政策的实施对教育水平高的人起的作用更大，也就是吸烟数量更少。