# Homework_Econometrics_2

工物 70 向杰 2017011842 2020/10/31

(1) 最小二乘法 $H \equiv \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - bx_i)^2$，一阶条件，$\frac{\partial H}{\partial b} = 0$，$-\sum_{i=1}^{n} 2x_i(y_i - \hat{\beta}x_i) = 0$

$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} \hat{\beta} x_i^2$，$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ 即所求的估计量 $\hat{\beta}$

(2) 考虑二元线性回归方程 $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \hat{u}_i$，根据OLS代数性质有 $\sum_i \hat{u}_i = 0$，$\sum_i \hat{u}_i x_{i2} = 0$

将 $x_1$ 对 $x_2$ 和常数回归 $x_{i1} = \hat{\delta}_0 + \hat{\delta}_1 x_{i2} + \hat{r}_{i1}$，根据OLS同样有 $\sum_i \hat{r}_{i1} = 0$，$\sum_i \hat{r}_{i1} x_{i2} = \sum_i \hat{r}_{i1} \hat{x}_{i1} = 0$，

$\Rightarrow \sum_i x_{i1} \hat{u}_i = \sum_i (\hat{\delta}_0 + \hat{\delta}_1 x_{i2} + \hat{r}_{i1}) \hat{u}_i = \hat{\delta}_0 \sum_i \hat{u}_i + \hat{\delta}_1 \sum_i x_{i2} \hat{u}_i + \sum_i \hat{r}_{i1} \hat{u}_i$

$= \sum_i \hat{r}_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}) = \sum_i \hat{r}_{i1} y_i - \hat{\beta}_1 \sum_i \hat{r}_{i1}^2 = 0$

由此可以解出 $\beta_1$，可以看出第一步中得到的残差均值为0，所以是否包括常数不影响估计值

(3) 不是的，由一阶条件得到的是 $\sum_{i=1}^{n} x_i \hat{u}_i = 0$，不能直接推断 $\sum_{i=1}^{n} \hat{u}_i = 0$

比如 $x_1 = 1$，$x_2 = 2$，$\hat{u}_1 = -2$，$\hat{u}_2 = +1$，满足一阶~~对~~条件但不满足 $\sum_{i=1}^{n} \hat{u}_i = 0$

(4) $\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$，$\tilde{\beta}_0 = \bar{y} - \bar{x}\tilde{\beta}_1$，则 $\tilde{\beta}_1 + \tilde{\beta}_0 \frac{\sum x_i}{\sum x_i^2} = \tilde{\beta}_1 + (\bar{y} - \bar{x}\tilde{\beta}_1) \frac{\sum x_i}{\sum x_i^2}$

$\Rightarrow \frac{\bar{y} \sum x_i}{\sum x_i^2} + \left(1 - \frac{\bar{x} \sum x_i}{\sum x_i^2}\right) \cdot \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\bar{y} \sum x_i \sum (x_i - \bar{x})^2 + (\sum x_i^2 \cancel{\sum(x_i - \bar{x})^2} - \bar{x} \sum x_i) \sum (x_i - \bar{x})(y_i - \bar{y})}{\sum x_i^2 \sum (x_i - \bar{x})^2}$

$= \frac{\sum x_i y_i}{\sum x_i^2} = \hat{\beta}$，要使 $\tilde{\beta}_1 = \hat{\beta}$，则有 $\tilde{\beta}_0 \frac{\sum x_i}{\sum x_i^2} = 0$，$\sum x_i = 0$ 或 $\tilde{\beta}_0 = \bar{y} - \bar{x}\tilde{\beta}_1 = 0$ 均可

2. (1)用 Stata 计算得到各个年龄组的平均日吸烟量和样本数如下表：

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

从以上统计结果不能看出吸烟数和年龄有线性关系，或者推断得到线性关系的可信度很低。

(2) 回归结果和 Stata 截屏如下，

根据回归结果，在控制了教育水平和禁烟政策之后，44.2805 的人群吸烟数目最多

$cigs = 0.1521 + 0.8223age - 0.00959age^2 - 0.4504educ - 2.7464restaurn$

```
. reg cigs age agesq educ restaurn

      Source |       SS           df       MS      Number of obs   =        80
> 7
-------------+----------------------------------   F(4, 802)       =      10.7
> 8
       Model |  7739.48459         4  1934.87115   Prob > F        =     0.000
> 0
    Residual |  144014.198       802  179.568826   R-squared       =     0.051
> 0
-------------+----------------------------------   Adj R-squared   =     0.046
> 3
       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    1.12498
> 4
       agesq |  -.0095886   .0016779    -5.71   0.000    -.0128822    -.00629
> 5
        educ |     -.4504   .1614857    -2.79   0.005    -.7673845   -.133415
> 6
    restaurn |  -2.746372    1.09685    -2.50   0.012    -4.899408   -.593336
> 7
       _cons |   .1521404   3.503322     0.04   0.965    -6.724623    7.02890
> 4
-------------+----------------------------------------------------------------
```

(3)$H_0$: $\beta 1 = 0, \beta 2 = 0, \beta 3 = 0$;　$H_1$: $\beta 1, \beta 2, \beta 3$ 不全为零;

检验思路：确定一个显著性水平α，计算 p 值，若有 p < α拒绝原假设。

在 1%和 5%的显著性水平下均可以拒绝原假设，因为三个数值 P 均小于 1%

(4)在其他条件一样的情况下，是否实施禁烟政策对每天抽烟数的影响。

$H_0$: $\beta 4 = 0$;　$H_1$: $\beta 1 \neq 0$;

在 1%的显著性水平下不可以拒绝原假设，但在 5%的显著性水平下可以，因为 1%<P<5%

(5)回归结果如下所示：

```
. 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
------------------------------------------------------------------------------
```

偏效应表达式为 $\beta_4 + \beta_5 \cdot educ$，教育对抽样数的偏效应在是否禁烟样本的不同。