

## Homework2

司可经济系 15320171151903

2019 年 3 月 31 日

采用 CHIPS2013 年的数据，旨在考察收入与教育水平、家庭收入、父母职业及其交互项的影响，其中父母职业分为在公共部门和不在公共部门。

本文直接采用中国家庭收入分配研究院 2013 年的数据，来研究用该数据是否仍然满足假设。本文归纳整理了 CHIP13 城镇和农村收入和资产，CHIP13 城镇和农村户主工作经历，CHIP13 城镇和农村父母，CHIP13 城镇个人信息数据库的数据。包涵了全国 15 个省份 126 个城市 234 个县区抽选出的 18948 个住户样本和 64777 个个体样本，其中包括 7175 户城镇住户样本、11013 户农村住户样本中的有求职经历的个体。如下所示：

变量	计算方法
工资收入 inc（单位：元）	工资收入-（工资收入均值）
学历层次 edu	未上过学、小学、初中、高中、职高/技校、中专、大专、大学本科、研究生分别赋值 1~9
家庭收入 hou（单位：元）	家庭财产负债总额-（家庭财产负债总额均值）
父母职业 par	公共部门取值为 1，否则为 0

第一次回归：被解释变量为工资收入，解释变量为教育水平。

```
call:
lm(formula = log(inc) ~ log(edu), data = hw2)

Residuals:
    Min       1Q   Median       3Q      Max
-6.2357 -0.6793  0.0327  0.7334  3.5422

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.64881    0.06537  132.31  <2e-16 ***
log(edu)       0.62020    0.03891   15.94  <2e-16 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.019 on 3452 degrees of freedom
Multiple R-squared:  0.06855,    Adjusted R-squared:  0.06828
F-statistic:  254 on 1 and 3452 DF,  p-value: < 2.2e-16
```

我们发现教育水平和收入成正相关，教育水平越高，收入越高。

第二次回归：被解释变量为工资收入，解释变量为教育水平和家庭财产。

```
call:
lm(formula = log(inc) ~ log(edu) * log(hou), data = hw2)

Residuals:
    Min       1Q   Median       3Q      Max
-6.2734 -0.6789  0.0359  0.7143  3.4185

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.097098   0.121621  74.799 < 2e-16 ***
log(edu)        0.185581   0.079562   2.333  0.01973 *
log(hou)       -0.044752   0.011792  -3.795  0.00015 ***
log(edu):log(hou) 0.043429   0.007551   5.752  9.6e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.01 on 3450 degrees of freedom
Multiple R-squared:  0.08583,    Adjusted R-squared:  0.08503
F-statistic: 108 on 3 and 3450 DF,  p-value: < 2.2e-16
```

发现收入与家庭收入成负相关，但与交互项为正相关，说明教育水平高且家庭收入高的，工资收入更高。但家庭收入高的，工资收入反而较低，可能原因是家庭收入高的，更少的依赖工资收入。

第三次回归：被解释变量为工资收入，解释变量为教育水平、家庭财产和父母职业。

```
Call:
lm(formula = log(inc) ~ log(edu) * log(hou) * par, data = hw2)

Residuals:
    Min       1Q   Median       3Q      Max
-6.3953 -0.6589  0.0561  0.6932  3.5842

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.206694   0.136794  67.303 < 2e-16 ***
log(edu)        0.099442   0.095527   1.041 0.297956
log(hou)       -0.045394   0.013383  -3.392 0.000702 ***
par            -0.535083   0.348641  -1.535 0.124933
log(edu):log(hou) 0.040574   0.009148   4.435 9.49e-06 ***
log(edu):par      0.337797   0.205186   1.646 0.099795 .
log(hou):par      0.048100   0.032978   1.459 0.144780
log(edu):log(hou):par -0.019852  0.019239  -1.032 0.302190
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.006 on 3446 degrees of freedom
Multiple R-squared:  0.09376, Adjusted R-squared:  0.09192
F-statistic: 50.93 on 7 and 3446 DF, p-value: < 2.2e-16
```

发现收入与父母职业成负相关，但与交互项也为负相关，可能原因是父母在公共部门的，孩子可能也倾向于更稳定的工作，可能工资收入就比较低。

附录：

```
## regression 1
reg1=lm(log(inc)~log(edu),data=hw2)
summary(reg1)
## regression 2
reg2 = lm(log(inc) ~ log(edu)*log(hou), data=hw2)
```

```
summary(reg2)
## regression 3
reg3 = lm(log(inc) ~ log(edu)*log(hou)*par, data=hw2)
summary(reg3)
```