# 2018年11月27日：习题讲解

## 第六、七章

程振兴 2018年11月26日

## 习题6.5

**【题目】：**
使用数据集`grilic.dta`，以稳健标准误估计下面的回归方程：

$$lnw = \beta_1 + \beta_2 s + \beta_3 expr + \beta_4 tenure + \beta_5 smsa + \varepsilon \tag{6.43}$$

（1）使用全样本，估计方程（6.43）。
（2）使用美国南方的子样本，估计方程（6.43）。
（3）使用美国北方的子样本，估计方程（6.43）。
（4）与全样本相比，子样本估计量的标准误有何变化，为什么？

**【解答】：**

（1）使用全样本：

```
cuse grilic, clear web
// 因为我的数据里面的lnw的名字是lw，为了和书上的统一，重命名为lnw
ren lw lnw
reg lnw s expr tenure smsa, r
```

结果：

```
. reg lnw s expr tenure smsa, r

Linear regression                               Number of obs     =        758
                                                F(4, 753)         =      98.36
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3448
                                                Root MSE          =     .34813

------------------------------------------------------------------------------
             |               Robust
         lnw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           s |   .1035073   .0062235    16.63   0.000     .0912899    .1157247
        expr |   .0381933   .0066362     5.76   0.000     .0251656     .051221
      tenure |   .0363505   .0081018     4.49   0.000     .0204457    .0522554
        smsa |   .1523258   .0276534     5.51   0.000      .098039    .2066127
       _cons |   4.059067   .0861023    47.14   0.000     3.890038    4.228096
------------------------------------------------------------------------------
```

根据估计结果，拟合的模型为：

$$lnw = 4.059 + 0.104s + 0.036expr + 0.152smsa + \hat{\varepsilon}$$

（2）使用南方样本(rns == 1)估计：

```
reg lnw s expr tenure smsa if rns, r
```

结果：

```
. reg lnw s expr tenure smsa if rns, r

Linear regression                               Number of obs     =        204
                                                F(4, 199)         =      36.04
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4203
                                                Root MSE          =     .34929

------------------------------------------------------------------------------
             |               Robust
         lnw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           s |   .1198242   .0120804     9.92   0.000     .0960021    .1436463
        expr |   .0451903   .0143038     3.16   0.002     .0169839    .0733967
      tenure |   .0092643   .0177777     0.52   0.603    -.0257926    .0443211
        smsa |   .1746563   .0496961     3.51   0.001     .0766579    .2726548
       _cons |   3.806148   .1582838    24.05   0.000     3.494019    4.118276
------------------------------------------------------------------------------
```

根据估计结果，拟合的模型为：

$$lnw = 3.806 + 0.120s + 0.045expr + 0.175smsa + \hat{\varepsilon}$$

（3）使用北方样本(rns == 0)估计：

```
reg lnw s expr tenure smsa if !rns, r
```

结果：

```
. reg lnw s expr tenure smsa if !rns, r

Linear regression                               Number of obs     =        554
                                                F(4, 549)         =      59.45
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3127
                                                Root MSE          =     .34356

------------------------------------------------------------------------------
             |               Robust
         lnw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           s |   .0944787   .0072808    12.98   0.000      .080177    .1087804
        expr |   .0358675   .0073509     4.88   0.000     .0214281    .0503068
      tenure |   .0455117   .0088423     5.15   0.000     .0281429    .0628806
        smsa |   .1199364    .034029     3.52   0.000     .0530934    .1867794
       _cons |   4.214014   .103448    40.74   0.000     4.010812    4.417217
------------------------------------------------------------------------------
```

根据估计结果，拟合的模型为：

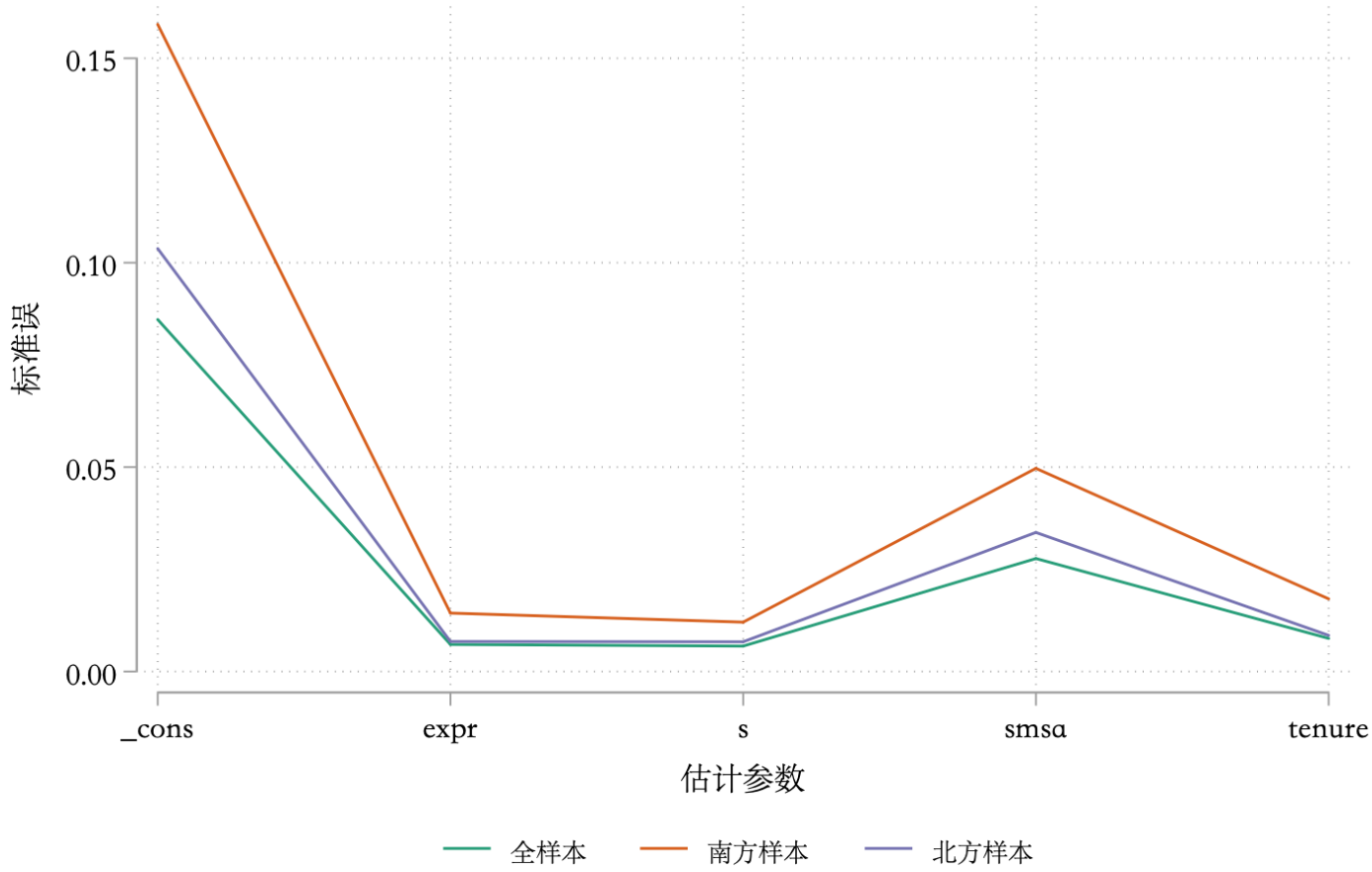
$$lnw = 4.021 + 0.094s + 0.036expr + 0.120smsa + \hat{\varepsilon}$$

（4）：比较

三个模型的估计量标准误如下：

| 估计参数 | 全样本 | 北方样本 | 南方样本 |
| --- | --- | --- | --- |
| s | 0.0062 | 0.0073 | 0.0120 |
| expr | 0.0066 | 0.0074 | 0.0143 |
| tenure | 0.0081 | 0.0088 | 0.0178 |
| smsa | 0.0277 | 0.0340 | 0.0497 |
| _cons | 0.0861 | 0.1034 | 0.1583 |

还可以把这个画出来：



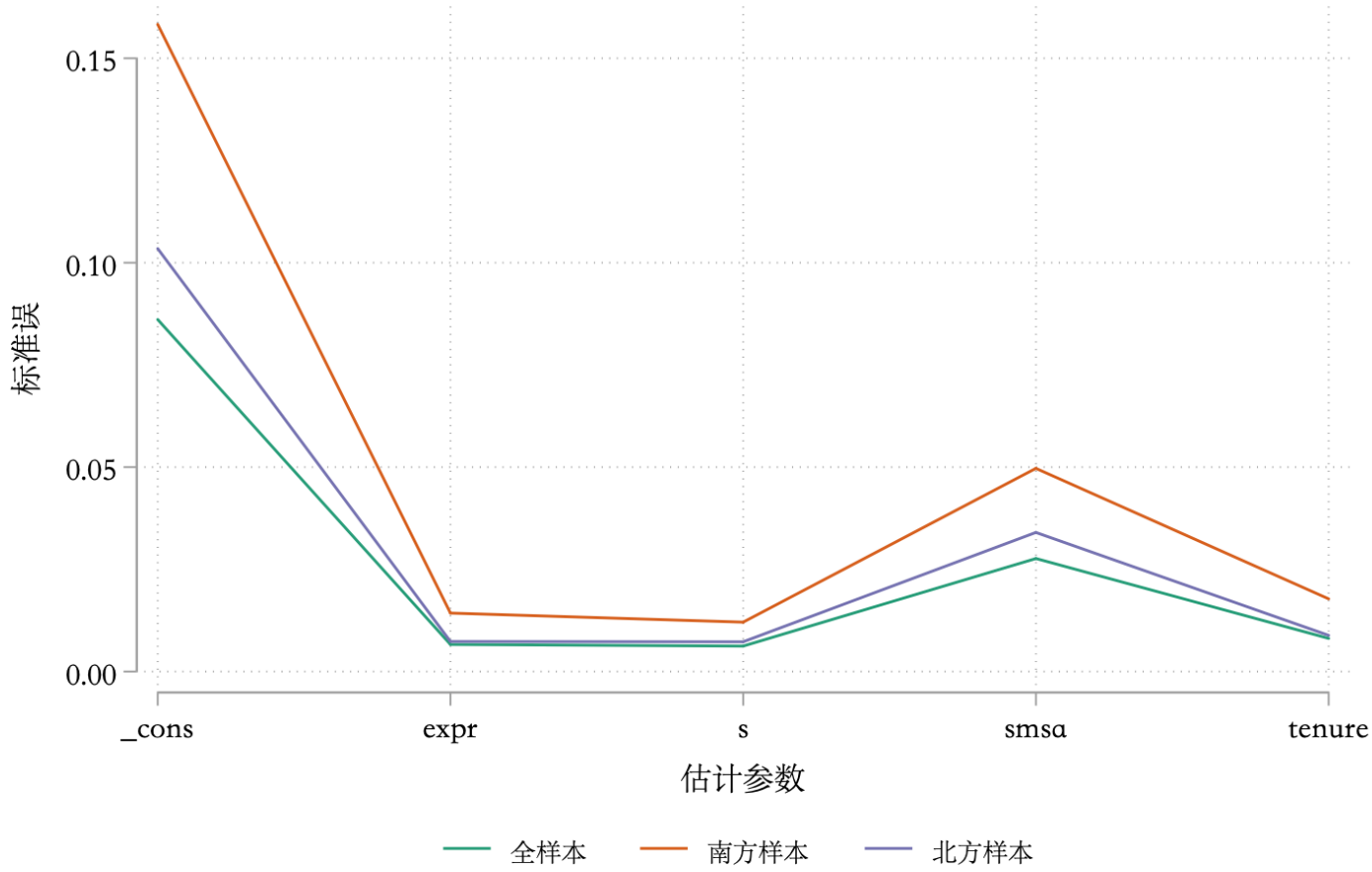

从图中可以很容易看出全样本估计的模型各个参数的标准误都较子样本的小。这是当样本量增加时，样本更加接近总体，对参数的估计自然更加准确，标准误会更小。

绘图代码：

```
clear
input str10 var sd1 sd2 sd3
"s" 0.0062  0.0073 0.0120
"expr" 0.0066 0.0074 0.0143
"tenure" 0.0081 0.0088 0.0178
"smsa" 0.0277 0.0340 0.0497
"_cons" 0.0861 0.1034 0.1583
end
reshape long sd, i(var) j(m)
encode var, gen(varlab) lab(var)
* colorscheme是一个选择配色的命令，安装方法：net install colorscheme.pkg, from("https://gith
* 详细介绍可以参考我的这篇博客：https://www.czxa.top/posts/16049/
colorscheme 3, palette(Dark2)
ret list
tw ///
line sd varlab if m == 1, lp(solid) lc("`r(color1)'") || ///
line sd varlab if m == 2, lp(solid) lc("`r(color2)'") || ///
line sd varlab if m == 3, lp(solid) lc("`r(color3)'") ||, ///
xlab(, val labsize(*1.2)) ylab(, format(%6.2f) labsize(*1.2)) ///
xti("估计参数") yti("标准误") ///
leg(order(1 "全样本" 2 "北方样本" 3 "南方样本") pos(6) row(1))
gr export "6_5模型比较.png", replace
```

另外一种办法是把标准误存储起来使用：

```
cuse grilic, clear
ren lw lnw
reg lnw s expr tenure smsa, r
ret list
mat list r(table)
mat m1 = r(table)
mat list m1

reg lnw s expr tenure smsa if rns, r
mat m2 = r(table)

reg lnw s expr tenure smsa if !rns, r
mat m3 = r(table)

mat list m1
mat list m2
mat list m3

clear
gen var = ""
gen sd1 = .
gen sd2 = .
gen sd3 = .
set obs 5

di m1[1,1]
forval i = 1/`=_N'{
    replace sd1 = m1[2, `i'] in `i'
    replace sd2 = m2[2, `i'] in `i'
    replace sd3 = m3[2, `i'] in `i'
}

local k = 1
foreach j in "s" "expr" "tenure" "smsa" "_cons"{
    replace var = "`j'" in `k'
    local ++k
}
reshape long sd, i(var) j(m)
encode var, gen(varlab) lab(var)
colorscheme 3, palette(Dark2)
ret list
tw ///
line sd varlab if m == 1, lp(solid) lc("`r(color1)'") || ///
line sd varlab if m == 2, lp(solid) lc("`r(color2)'") || ///
line sd varlab if m == 3, lp(solid) lc("`r(color3)'") ||, ///
xlab(, val labsize(*1.2)) ylab(, format(%6.2f) labsize(*1.2)) ///
xti("估计参数") yti("标准误") ///
leg(order(1 "全样本" 2 "南方样本" 3 "北方样本") pos(6) row(1))
gr export "6_5模型比较.png", replace
```




# 习题6.6

**【题目】：**

房屋的价格如何决定？一种理论认为，房价由房屋的性能决定，成为“特征价格法”。数据集`hprice2a.dta`包含了美国波士顿506个社区的房屋中位数价格的横截面数据。考虑以下特征价格回归：

$$lprice_i = \beta_1 + \beta_2 lnox_i + \beta_3 ldist_i + \beta_4 rooms_i + \beta_5 stratio_i + \varepsilon_i \tag{6.44}$$

其中，$lprice$为房价的对数，$lnox$为空气污染程度的对数，$ldist$为社区到就业中心距离的对数，$rooms$为房屋的平均房间数，$stratio$为社区学校的`学生 - 教师`比例，下标$i$表示社区$i$。

（1）使用普通标准误进行回归，并评论解释变量系数的符号、统计显著性及经济意义。
（2）使用稳健标准误进行回归，稳健标准误和普通标准误差别大么？
（3）使用稳健标准误，以5%的显著性水平，检验$H_0: \beta_3 = \beta_5$。
（4）使用稳健标准误，以5%的显著性水平，检验$H_0: \beta_4 = 0.31$与$H_0: \beta_4 = 0.30$。

**【解答】**

**（1）：普通标准误回归**

```
cuse hprice2a, clear
gen ldist = ln(dist)
reg lprice lnox ldist rooms stratio
```

结果：

```
. reg lprice lnox ldist rooms stratio

      Source |       SS           df       MS      Number of obs   =       506
-------------+----------------------------------   F(4, 501)       =    175.86
       Model |  49.3987735         4  12.3496934   Prob > F        =    0.0000
    Residual |  35.1834974       501  .070226542   R-squared       =    0.5840
-------------+----------------------------------   Adj R-squared   =    0.5807
       Total |  84.5822709       505  .167489645   Root MSE        =      .265

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnox |    -.95354   .1167418    -8.17   0.000    -1.182904   -.7241762
       ldist |  -.1343401   .0431032    -3.12   0.002    -.2190255   -.0496548
       rooms |   .2545271   .0185303    13.74   0.000     .2181203    .2909338
     stratio |  -.0524512   .0058971    -8.89   0.000    -.0640373   -.0408651
       _cons |   11.08387   .3181115    34.84   0.000     10.45887    11.70886
------------------------------------------------------------------------------
```

1. lnox：系数为负，在5%的显著性水平上显著，表示空气污染程度每加剧1%，房价平均下跌0.95%；
2. ldist：系数为负，在5%的显著性水平上显著，表示社区到就业中心的距离每增加1%，房价平均下跌0.13%；
3. rooms：系数为正，在5%的显著性水平上显著，表示房屋的平均房间数增加1，房价平均上涨25.45%；
4. stratio：系数为负，在5%的显著性水平上显著，表示社区学校的学生-教师比例每提高一个单位，房价平均下跌5.24%。

（2）：稳健标准误的回归

```
reg lprice lnox ldist rooms stratio, r
```

结果：

```
. reg lprice lnox ldist rooms stratio, r

Linear regression                               Number of obs     =        506
                                                F(4, 501)         =     146.27
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5840
                                                Root MSE          =       .265

------------------------------------------------------------------------------
             |               Robust
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnox |    -.95354   .1268006    -7.52   0.000    -1.202667   -.7044135
       ldist |  -.1343401   .0535287    -2.51   0.012    -.2395086   -.0291717
       rooms |   .2545271   .0247204    10.30   0.000     .2059586    .3030956
     stratio |  -.0524512   .0046082   -11.38   0.000     -.061505   -.0433974
       _cons |   11.08387   .3772952    29.38   0.000     10.34259    11.82514
------------------------------------------------------------------------------
```

显然差别不大。

（3）：5%显著性水平，检验$\beta_3 = \beta_5$

```
. test ldist = stratio

 ( 1)  ldist - stratio = 0

       F(  1,   501) =    2.27
            Prob > F =    0.1322
```

p值大于%5，不显著，因此无法拒绝原假设。

（4）：5%显著性水平，检验$\beta_4 = 0.31/0.30$

```
. test rooms = 0.31

 ( 1)  rooms = .31

       F(  1,   501) =    5.04
            Prob > F =    0.0253

. test rooms = 0.30

 ( 1)  rooms = .3

       F(  1,   501) =    3.38
            Prob > F =    0.0664
```

第一个检验的结果是拒绝原假设的，第二个结果是无法拒绝原假设。

# 习题7.2

**【题目】**

房价的回归是否存在异方差？继续考虑上题中的房价模型：

（1）以5%的置信度，使用BP检验，检验是否存在异方差（假设扰动项为iid，分别以拟合值$\hat{y}$以及所有解释变量进行检验）。

（2）以5%的置信度，使用怀特检验，检验是否存在异方差。

**【解答】**

（1）：BP检验

首先是以拟合值进行检验：

```
cuse hprice2a, clear
gen ldist = ln(dist)
qui reg lprice lnox ldist rooms stratio
estat hettest, iid
```

结果：

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lprice

         chi2(1)      =    37.57
         Prob > chi2  =   0.0000
```

然后再使用所有的解释变量进行检验：

```
estat hettest, iid rhs

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: lnox ldist rooms stratio

         chi2(4)      =    69.87
         Prob > chi2  =   0.0000
```

两次检验的结果都强烈拒绝拒绝同方差的原假设，因为认为存在异方差。

（2）：怀特检验

```
estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(14)     =    143.98
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |     143.98     14    0.0000
            Skewness |      16.99      4    0.0019
            Kurtosis |      11.30      1    0.0008
---------------------+-----------------------------
               Total |     172.26     19    0.0000
---------------------------------------------------
```

结果同样强烈拒绝同方差的原假设，认为存在异方差。

# 习题7.3

**【题目】**

恩格尔系数是否存在异方差？数据集`food.dta`包含有关每周食物开支(food_exp)和周收入(income)的40个观测值。

（1）将food_exp和income的散点图和线性拟合图画在一起。根据此图，是否可能存在异方差？此异方差和收入的关系是怎样的？

（2）将food_exp对income进行回归。

（3）以5%的置信度，使用BP检验，检验是否存在异方差（iid假设）。

（4）以5%的置信度，使用怀特检验，检验是否存在异方差。

（5）定义食物开支比例food_share = food_exp/income，将food_share与income的散点图与线性拟合图画在一起。从图上看，是否还存在异方差？

（6）将food_share对income进行回归。

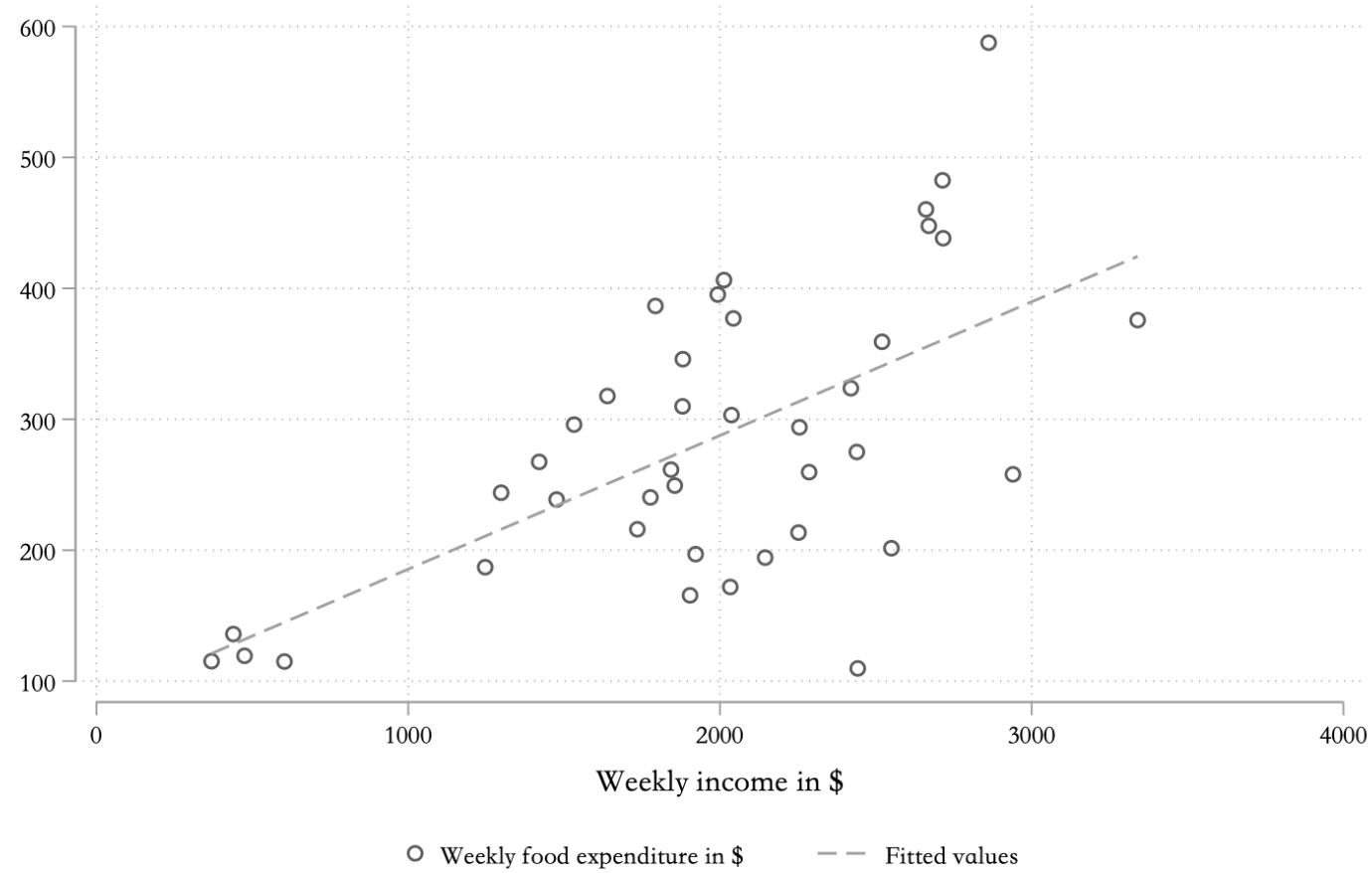
（7）5%、BP检验、iid。

（8）5%、怀特检验。

**【解答】**

（1）：图表

```
cuse food, clear
tw ///
sc food_exp income || ///
lfit food_exp income ||, ///
leg(pos(6) row(1))
gr export "7_3恩格尔系数1.png", replace
```




从图上可以看出，很可能存在异方差，并且收入越高方差越大。

（2）：回归

```
. reg food_exp income

      Source |       SS           df       MS      Number of obs   =        40
-------------+----------------------------------   F(1, 38)        =     23.79
       Model |  190626.976         1  190626.976   Prob > F        =    0.0000
    Residual |  304505.177        38  8013.29412   R-squared       =    0.3850
-------------+----------------------------------   Adj R-squared   =    0.3688
       Total |  495132.153        39  12695.6962   Root MSE        =    89.517

------------------------------------------------------------------------------
    food_exp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .1020964   .0209326     4.88   0.000     .0597205    .1444723
       _cons |   83.41601   43.41016     1.92   0.062    -4.46327    171.2953
------------------------------------------------------------------------------
```

拟合的模型为：

$$food\_exp = 83.416 + 0.102income + \varepsilon$$

（3）：BP检验

```
. estat hettest, iid

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of food_exp

         chi2(1)      =     7.38
         Prob > chi2  =   0.0066

. estat hettest, iid rhs

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: income

         chi2(1)      =     7.38
         Prob > chi2  =   0.0066
```

两个BP检验的结果都强烈拒绝同方差的原假设，因此认为存在异方差。

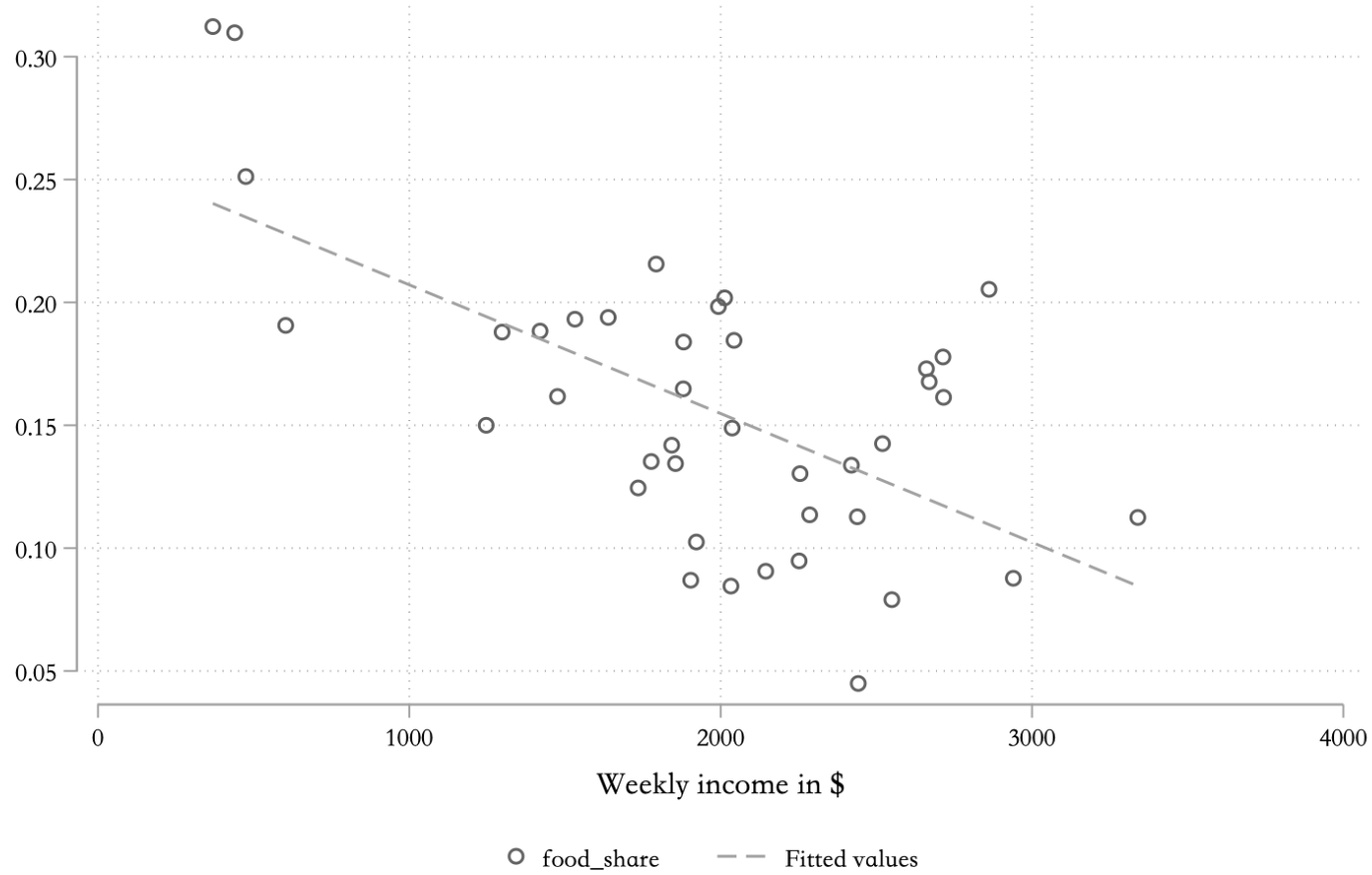
（4）：怀特检验

```
. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(2)      =     7.56
         Prob > chi2  =    0.0229

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |       7.56      2    0.0229
            Skewness |       0.13      1    0.7146
            Kurtosis |       0.00      1    0.9825
---------------------+-----------------------------
               Total |       7.69      4    0.1036
---------------------------------------------------
```

怀特检验的结果也强烈拒绝同方差的原假设，因此认为存在异方差。

（5）：food_share & income

```
gen food_share = food_exp / income
tw ///
sc food_share income || ///
lfit food_share income ||, ///
leg(pos(6) row(1)) ylab(, format(%6.2f))
gr export "7_3恩格尔系数2.png", replace
```




可以看出异方差现象不如刚刚那般明显了。

（6）：回归

```
. reg food_share income

      Source |       SS           df       MS      Number of obs   =        40
-------------+----------------------------------   F(1, 38)        =     24.39
       Model |  .050217947         1  .050217947   Prob > F        =    0.0000
    Residual |  .078224368        38  .002058536   R-squared       =    0.3910
-------------+----------------------------------   Adj R-squared   =    0.3749
       Total |  .128442315        39  .003293393   Root MSE        =    .04537

------------------------------------------------------------------------------
  food_share |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |  -.0000524   .0000106    -4.94   0.000    -.0000739   -.0000309
       _cons |   .2595986   .0220021    11.80   0.000     .2150576    .3041397
------------------------------------------------------------------------------
```

拟合的模型为：

$$food\_share = -5.24e^{-5}income + 0.26 + \varepsilon$$

（7）：BP检验

```
. estat hettest, iid

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of food_share

         chi2(1)      =     0.08
         Prob > chi2  =   0.7748

. estat hettest, iid rhs

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: income

         chi2(1)      =     0.08
         Prob > chi2  =   0.7748
```

两个BP检验的结果都无法拒绝同方差的原假设，因此不认为存在异方差。

（8）：怀特检验

```
. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(2)      =     2.60
         Prob > chi2  =    0.2722

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |       2.60      2    0.2722
            Skewness |       0.37      1    0.5421
            Kurtosis |       2.71      1    0.1000
---------------------+-----------------------------
               Total |       5.68      4    0.2244
---------------------------------------------------
```

怀特检验的结果无法拒绝同方差的原假设，因此不认为存在异方差。