Econométrie TP4

Hétéroscédasticité

Patrick Waelbroeck

Telecom Paris

February 27, 2020

Calcul de prix hédonique

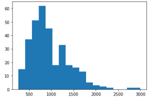
On utilise la base de données `hprice3.raw`.

```
df = pd.read_csv('hprice3.raw', delim_whitespace=True, header=None)
```

Exercice 1

Calculer la moyenne de $price/100$. Faire un histogramme de $price/100$. Faire ensuite la régression de $y = price/100$ en fonction de $const, age, nbh, inst, rooms, area, land, baths, dist, y81$. Interpéter le coefficient associé à $y81$ en le comparant à l'augmentation du prix moyen de 1978 à 1981.

```
price=df[7]/100
plt.hist(price,'auto')
```



Calcul de prix hédonique, suite

```
s=np.shape(price)
const=np.ones(s)
age=df[1]
nbh=df[3]
inst=df[5]
rooms=df[8]
area=df[9]
land=df[10]
baths=df[11]
dist=df[12]
y81=df[15]
y=price
X=np.column_stack((const,age,nbh,inst,rooms,area,land,baths,dist,y81))
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
s=y81==0
p0=np.mean(price[s])
s=y81==1
p1=np.mean(price[s])
print(p1-p0)
```

Prix hédoniques - résultats

```
==============================================================================
Dep. Variable:                      7   R-squared:                       0.715
Model:                            OLS   Adj. R-squared:                  0.707
Method:                 Least Squares   F-statistic:                     86.73
Date:                Thu, 27 Feb 2020   Prob (F-statistic):           2.14e-79
Time:                        00:52:37   Log-Likelihood:                -2201.6
No. Observations:                 321   AIC:                             4423.
Df Residuals:                     311   BIC:                             4461.
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       -211.1428    105.234     -2.006      0.046    -418.204      -4.081
x1            -2.2888      0.491     -4.659      0.000      -3.255      -1.322
x2           -19.8892      6.470     -3.074      0.002     -32.620      -7.158
x3            -0.0054      0.004     -1.331      0.184      -0.013       0.003
x4            41.3148     19.323      2.138      0.033       3.294      79.335
x5             0.2174      0.028      7.863      0.000       0.163       0.272
x6             0.0012      0.000      3.239      0.001       0.000       0.002
x7           130.0255     28.530      4.557      0.000      73.889     186.162
x8             0.0051      0.004      1.280      0.201      -0.003       0.013
x9           358.6543     27.666     12.964      0.000     304.218     413.091
==============================================================================
Omnibus:                       87.991   Durbin-Watson:                   1.596
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              581.142
Skew:                           0.942   Prob(JB):                    6.41e-127
Kurtosis:                       9.317   Cond. No.                     4.94e+05
==============================================================================
```

Le prix ajusté à la qualité augmente de 358.65 de 1978 à 1981. Le prix moyen augmente de 440.19.

Test d'hétéroscédasticité de forme linéaire

Exercice 2

Tester l'hypothèse d'homoscédasticité en utilisant la régression de u^2 en fonction des variables du modèles.

On doit tester la significativé globale de la régression suivante
$u^2 = \theta_0 + \theta_1 x_1 + ... + \theta_k x_k$
Ceci revient à tester
$H_0 : \theta_1 = ... = \theta_k = 0$

```
u=results.resid
u2=u**2
y=u2
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
```

F=13.53. On rejette l'hypothèse H_0.

Test d'hétéroscédasticité de forme linéaire

```
==============================================================================
Dep. Variable:                      y   R-squared:                       0.281
Model:                            OLS   Adj. R-squared:                  0.261
Method:                 Least Squares   F-statistic:                     13.53
Date:                Thu, 27 Feb 2020   Prob (F-statistic):           2.57e-18
Time:                        01:37:25   Log-Likelihood:                -4234.7
No. Observations:                 321   AIC:                             8489.
Df Residuals:                     311   BIC:                             8527.
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       2.307e+04   5.93e+04      0.389      0.697   -9.35e+04     1.4e+05
x1          -615.5935    276.676     -2.225      0.027   -1159.986     -71.201
x2         -3018.7279   3643.672     -0.828      0.408   -1.02e+04    4150.638
x3            -3.0761      2.285     -1.346      0.179      -7.571       1.419
x4         -7635.9441   1.09e+04     -0.702      0.483    -2.9e+04    1.38e+04
x5            89.8939     15.573      5.772      0.000      59.252     120.536
x6             1.5736      0.212      7.426      0.000       1.157       1.991
x7         -2.721e+04   1.61e+04     -1.694      0.091   -5.88e+04    4401.122
x8            -2.0748      2.225     -0.932      0.352      -6.454       2.304
x9          7940.4165   1.56e+04      0.510      0.611   -2.27e+04    3.86e+04
==============================================================================
Omnibus:                      366.671   Durbin-Watson:                   1.916
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            22235.419
Skew:                           4.972   Prob(JB):                         0.00
Kurtosis:                      42.542   Cond. No.                     4.94e+05
==============================================================================
```

Test d'hétéroscédasticité de forme linéaire

Exercice 3

Tester l'hypothèse $H_0 : \theta_{area} = \theta_{land} = 0$

`F=45.72`. On rejette l'hypothèse $H_0 (p < 0.01)$.

Transformation en variables binaires

Exercice 4

Donner la valeur minimale et maximale de *baths*. Transformer ensuite la variable *baths* en un ensemble de variables binaires. Refaire le test d'hétéroscédasticité de l'exercice 2.

Résultat pour la statistique de Fisher et la p-value :

```
F=10.87
3.93e-17
```

On rejette l'hypothèse $H_0 (p < 0.01)$.

y=price

```
==============================================================================
Dep. Variable:                      7   R-squared:                       0.720
Model:                            OLS   Adj. R-squared:                  0.710
Method:                 Least Squares   F-statistic:                     72.22
Date:                Thu, 27 Feb 2020   Prob (F-statistic):           1.33e-78
Time:                        02:24:55   Log-Likelihood:                -2198.8
No. Observations:                 321   AIC:                             4422.
Df Residuals:                     309   BIC:                             4467.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        -35.3626    113.493     -0.312      0.756    -258.679     187.953
x1            -2.1398      0.517     -4.142      0.000      -3.156      -1.123
x2           -18.7182      6.485     -2.887      0.004     -31.478      -5.959
x3            -0.0049      0.004     -1.213      0.226      -0.013       0.003
x4            40.2520     19.317      2.084      0.038       2.243      78.261
x5             0.2164      0.028      7.830      0.000       0.162       0.271
x6             0.0013      0.000      3.381      0.001       0.001       0.002
x7            56.5481     44.428      1.273      0.204     -30.871     143.967
x8           248.5475     58.807      4.226      0.000     132.834     364.261
x9           305.2797    141.172      2.162      0.031      27.500     583.059
x10            0.0042      0.004      1.057      0.291      -0.004       0.012
x11          351.9898     27.683     12.715      0.000     297.518     406.462
==============================================================================
Omnibus:                       84.992   Durbin-Watson:                   1.574
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              559.720
Skew:                           0.902   Prob(JB):                    2.87e-122
Kurtosis:                       9.212   Cond. No.                     6.97e+05
==============================================================================
```

$y = u^2$

```
==============================================================================
Dep. Variable:                      y   R-squared:                       0.281
Model:                            OLS   Adj. R-squared:                  0.255
Method:                 Least Squares   F-statistic:                     10.97
Date:                Thu, 27 Feb 2020   Prob (F-statistic):           3.93e-17
Time:                        02:28:51   Log-Likelihood:                -4227.2
No. Observations:                 321   AIC:                             8478.
Df Residuals:                     309   BIC:                             8524.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       7663.0126    6.3e+04      0.122      0.903   -1.16e+05    1.32e+05
x1          -436.3323    286.694     -1.522      0.129   -1000.452     127.787
x2         -3348.4394   3598.974     -0.930      0.353   -1.04e+04    3733.156
x3            -2.3966      2.248     -1.066      0.287      -6.821       2.027
x4         -8966.3938   1.07e+04     -0.836      0.404   -3.01e+04    1.21e+04
x5            86.6915     15.338      5.652      0.000      56.511     116.872
x6             1.5282      0.208      7.344      0.000       1.119       1.938
x7         -3.527e+04   2.47e+04     -1.430      0.154   -8.38e+04    1.32e+04
x8         -4.041e+04   3.26e+04     -1.238      0.217   -1.05e+05    2.38e+04
x9         -1.733e+05   7.84e+04     -2.212      0.028   -3.27e+05   -1.91e+04
x10           -2.6414      2.199     -1.201      0.231      -6.969       1.686
x11         7164.9064   1.54e+04      0.466      0.641   -2.31e+04    3.74e+04
==============================================================================
Omnibus:                      356.008   Durbin-Watson:                   1.891
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            18901.036
Skew:                           4.783   Prob(JB):                         0.00
Kurtosis:                      39.354   Cond. No.                     6.97e+05
==============================================================================
```

Exercice 5

En utilisant la spécification de l'exercice 4, refaire le test d'hétéroscédasticité en utilisant $log(area)$ et $log(land)$.

Résultats: `F=5.419` On rejette l'hypothèse $H_0 (p < 0.01)$.

Transformation log

Exercice 6

En utilisant la spécification de l'exercice 5, refaire le test d'hétéroscédasticité en utilisant $y = log(price/100)$.

Résultats: `F=1.905` On rejette l'hypothèse H_0 à 5% ($p = 0.0381$).

$y = u^2$

```
==============================================================================
Dep. Variable:                      y   R-squared:                       0.064
Model:                            OLS   Adj. R-squared:                  0.030
Method:                 Least Squares   F-statistic:                     1.905
Date:                Thu, 27 Feb 2020   Prob (F-statistic):             0.0381
Time:                        03:05:50   Log-Likelihood:                 268.51
No. Observations:                 321   AIC:                            -513.0
Df Residuals:                     309   BIC:                            -467.8
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.4437      0.209     -2.128      0.034      -0.854      -0.033
x1             0.0003      0.000      1.293      0.197      -0.000       0.001
x2            -0.0006      0.003     -0.205      0.838      -0.007       0.005
x3         -2.004e-06   2.11e-06     -0.949      0.343   -6.16e-06    2.15e-06
x4            -0.0089      0.009     -1.001      0.318      -0.026       0.009
x5             0.0347      0.027      1.282      0.201      -0.019       0.088
x6             0.0320      0.013      2.454      0.015       0.006       0.058
x7             0.0052      0.021      0.245      0.806      -0.036       0.047
x8             0.0162      0.028      0.573      0.567      -0.039       0.072
x9            -0.0028      0.064     -0.044      0.965      -0.129       0.123
x10        -1.173e-06   1.83e-06     -0.641      0.522   -4.77e-06    2.42e-06
x11           -0.0153      0.013     -1.212      0.226      -0.040       0.010
==============================================================================
Omnibus:                      552.886   Durbin-Watson:                   2.020
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           218514.258
Skew:                           9.627   Prob(JB):                         0.00
Kurtosis:                     129.360   Cond. No.                     1.03e+06
==============================================================================
```

Correction supplémentaire 1 : Weigthed Least Squares

Exercice 7

Utiliser la variable `lland` pour pondérer les observations dans la spécification de l'exercice 6. Refaire le test d'hétéroscédasticité.

On utilise la commande WLS

```
h=np.sqrt(lland)
y=np.log(price)
X=np.column_stack((const,age,nbh,inst,rooms,larea,lland,bath2,bath3,bath4,d
model=sm.WLS(y,X,weight=1/h)
results = model.fit()
print(results.summary())
```

WLS

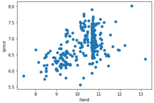
```
==============================================================================
Dep. Variable:                      7   R-squared:                       0.785
Model:                            WLS   Adj. R-squared:                  0.777
Method:                 Least Squares   F-statistic:                     102.6
Date:                Thu, 27 Feb 2020   Prob (F-statistic):           3.52e-96
Time:                        03:32:50   Log-Likelihood:                 56.624
No. Observations:                 321   AIC:                            -89.25
Df Residuals:                     309   BIC:                            -43.99
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.3297      0.403      5.774      0.000       1.536       3.124
x1            -0.0026      0.000     -5.635      0.000      -0.003      -0.002
x2            -0.0099      0.006     -1.645      0.101      -0.022       0.002
x3         -8.603e-06   4.08e-06     -2.106      0.036   -1.66e-05   -5.66e-07
x4             0.0575      0.017      3.338      0.001       0.024       0.091
x5             0.3465      0.052      6.608      0.000       0.243       0.450
x6             0.1163      0.025      4.608      0.000       0.067       0.166
x7             0.0779      0.041      1.903      0.058      -0.003       0.158
x8             0.2517      0.055      4.605      0.000       0.144       0.359
x9             0.4268      0.124      3.452      0.001       0.184       0.670
x10         6.041e-06   3.54e-06      1.707      0.089   -9.21e-07     1.3e-05
x11            0.3722      0.024     15.238      0.000       0.324       0.420
==============================================================================
Omnibus:                       70.271   Durbin-Watson:                   1.697
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              358.475
Skew:                          -0.787   Prob(JB):                     1.44e-78
Kurtosis:                       7.932   Cond. No.                     1.03e+06
==============================================================================
```

Correction supplémentaire 2 : régression par sous-groupes

Exercice 8

Faire le graphique en nuage de point entre $log(price/100)$ et *lland*. Diviser l'échantillon en deux groupes en fonction de lland et refaire le test d'hétéroscédasticité pour les deux sous-groupes.

```
lprice=np.log(price)
plt.scatter(lland,lprice)
plt.xlabel("lland")
plt.ylabel("lprice")
plt.show()
```



Construction de deux groupes : $lland \leq 10$ et $lland > 10$.
Pour le premier sous-groupe : `F=0.3985`, on ne rejette pas H_0.
Pour le deuxième sous-groupe : `F=1.373`, on ne rejette pas H_0 à 18.7%.