

Econométrie TP3

Test d'hypothèses multiples Python

Patrick Waelbroeck

Telecom Paris

February 11, 2020

On utilise la base de données `wage1.raw`. On explique $y = \log(\text{wage})$ en fonction de *const*, *educ*, *exper*, *tenure*.

Exercice 1

Tester l'hypothèse $H_0 : \beta_{educ} = 0, \beta_{exper} = 0$. Donner la p-value.

On doit calculer la somme des carrés des erreurs du modèle non contraint.

```
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
u=results.resid
SSR0=u.T@u
```

Résultat pour SSR0:

101.4555709121026

Exercice 1 - modèle non contraint

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	0.316			
Model:	OLS	Adj. R-squared:	0.312			
Method:	Least Squares	F-statistic:	80.39			
Date:	Sun, 09 Feb 2020	Prob (F-statistic):	9.13e-43			
Time:	20:23:15	Log-Likelihood:	-313.55			
No. Observations:	526	AIC:	635.1			
Df Residuals:	522	BIC:	652.2			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	0.2844	0.104	2.729	0.007	0.080	0.489
x1	0.0920	0.007	12.555	0.000	0.078	0.106
x2	0.0041	0.002	2.391	0.017	0.001	0.008
x3	0.0221	0.003	7.133	0.000	0.016	0.028
=====						
Omnibus:	11.534	Durbin-Watson:	1.769			
Prob(Omnibus):	0.003	Jarque-Bera (JB):	20.941			
Skew:	0.021	Prob(JB):	2.84e-05			
Kurtosis:	3.977	Cond. No.	135.			
=====						

Ensuit on doit calculer la somme des carrés des erreurs du modèle contraint. Le modèle contraint impose les contraintes de H_0 . Donc la matrice $X = [const, tenure]$

```
X0=X
X=np.column_stack((const, tenure))
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
u=results.resid
SSR1=u.T@u
```

Résultat pour SSR1:

132.6105094914973

Exercice 1 - modèle contraint

```

=====
Dep. Variable:                y    R-squared:                0.106
Model:                        OLS    Adj. R-squared:           0.104
Method:                        Least Squares    F-statistic:            62.11
Date:                Sun, 09 Feb 2020    Prob (F-statistic):    1.89e-14
Time:                        20:24:18    Log-Likelihood:       -383.98
No. Observations:            526    AIC:                    772.0
Df Residuals:                524    BIC:                    780.5
Df Model:                    1
Covariance Type:            nonrobust
=====
               coef    std err          t      P>|t|      [0.025    0.975]
-----
const                1.5010      0.027    55.870     0.000     1.448     1.554
x1                   0.0240      0.003     7.881     0.000     0.018     0.030
=====
Omnibus:                 8.882    Durbin-Watson:           1.776
Prob(Omnibus):           0.012    Jarque-Bera (JB):        11.058
Skew:                    0.185    Prob(JB):                0.00397
Kurtosis:                3.606    Cond. No.                10.9
=====

```

Calcul de la statistique de Fisher $F = \frac{(SSR1 - SSR0)/2}{SSR0/(n-k)}$. Pour la p-value, on doit importer f. La p-value est donnée par la fonction sf.

```
from scipy.stats import f
n,k=np.shape(X0)
F=((SSR1-SSR0)/2)/(SSR0/(n-k))
f.sf(F,2,n-k)
```

Résultat pour la p-value:

4.417396398302704e-31

Exercice 2

Tester l'hypothèse $H_0 : \beta_{exper} = 0$ en utilisant la méthode de Fisher et comparer au test de Student.

Test de student = 2.391. Résultat pour la statistique de Fisher et la p-value :

5.7189715598397175
0.01713562198649189

Exercice 3

Tester l'hypothèse $H_0 : \beta_{educ} = 0, \beta_{exper} = 0, \beta_{tenure} = 0$. Donner la p-value.

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

80.39091993522644

9.129958660791964e-43

Exercice 4

Tester l'hypothèse $H_0 : \beta_{educ} = 0.1, \beta_{exper} = 0.01$

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

5.819185550803654

0.0031660755577992738

Exercice 5

Construire une variable binaire *marrfem* correspondant aux femmes mariées. Faire la régression de $\log(\text{wage})$ sur une *constante*, *male*, *marrfem*, *educ*, *exper*, *tenure*. Tester et commenter l'hypothèse $H_0 : \beta_{\text{marrfem}} = 0$ à 5% :
Donner la p-value.

Attention : quelle est la catégorie de référence ?

```
female=df[5]
male=1-female
married=df[6]
marrfem=married*female
X=np.column_stack((const, male, marrfem, educ, exper, tenure))
n,k=np.shape(X)
y=np.log(wage)
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
```

Résultats OLS

=====						
Dep. Variable:	y	R-squared:	0.393			
Model:	OLS	Adj. R-squared:	0.387			
Method:	Least Squares	F-statistic:	67.32			
Date:	Mon, 10 Feb 2020	Prob (F-statistic):	3.44e-54			
Time:	19:55:49	Log-Likelihood:	-282.17			
No. Observations:	526	AIC:	576.3			
Df Residuals:	520	BIC:	601.9			
Df Model:	5					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	0.2107	0.100	2.110	0.035	0.015	0.407
x1	0.2793	0.047	5.939	0.000	0.187	0.372
x2	-0.0410	0.054	-0.761	0.447	-0.147	0.065
x3	0.0880	0.007	12.609	0.000	0.074	0.102
x4	0.0049	0.002	2.940	0.003	0.002	0.008
x5	0.0173	0.003	5.808	0.000	0.011	0.023
=====						
Omnibus:	12.433	Durbin-Watson:	1.772			
Prob(Omnibus):	0.002	Jarque-Bera (JB):	23.487			
Skew:	0.005	Prob(JB):	7.94e-06			
Kurtosis:	4.035	Cond. No.	138.			
=====						

On ne rejette pas l'hypothèse H_0 à 5%.

Exercice 6

Construire les 4 catégories suivantes : hommes mariés, femmes mariées, hommes célibataires, femmes célibataires. Tester l'hypothèse de discrimination salariale contre les femmes mariées par rapport aux hommes mariés. Donner la p-value.

Astuce : bien choisir la catégorie de référence.

Résultats OLS

=====						
Dep. Variable:	y	R-squared:	0.424			
Model:	OLS	Adj. R-squared:	0.417			
Method:	Least Squares	F-statistic:	63.63			
Date:	Mon, 10 Feb 2020	Prob (F-statistic):	4.53e-59			
Time:	20:19:22	Log-Likelihood:	-268.44			
No. Observations:	526	AIC:	550.9			
Df Residuals:	519	BIC:	580.7			
Df Model:	6					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	0.6799	0.107	6.370	0.000	0.470	0.890
x1	-0.2921	0.055	-5.273	0.000	-0.401	-0.183
x2	-0.3888	0.050	-7.723	0.000	-0.488	-0.290
x3	-0.4123	0.047	-8.737	0.000	-0.505	-0.320
x4	0.0835	0.007	12.175	0.000	0.070	0.097
x5	0.0032	0.002	1.912	0.056	-8.74e-05	0.006
x6	0.0157	0.003	5.370	0.000	0.010	0.021
=====						
Omnibus:	14.767	Durbin-Watson:	1.774			
Prob(Omnibus):	0.001	Jarque-Bera (JB):	29.522			
Skew:	0.071	Prob(JB):	3.88e-07			
Kurtosis:	4.152	Cond. No.	158.			
=====						

Exercice 7

Construire les variables binaires correspondants aux zones géographiques : *northcen*, *south*, *west*. Faire le test de Fisher d'absence d'effets régionaux dans le modèle $\log(wage)$ en fonction de *constante*, *northcen*, *south*, *west*, *educ*, *exper*, *tenure*. Donner la p-value.

On teste $H_0 : \beta_{northcen} = \beta_{south} = \beta_{west} = 0$.

Résultat pour la Fisher et la p-value:

1.9068882812088475

0.1274385871187363

On ne rejette pas l'hypothèse H_0 .

Exercice 8

Tester l'hypothèse que le rendement de l'éducation des hommes est le même que celui des femmes dans le modèle de l'exercice 1. Spécifier le modèle et donner la p-value

Résultats OLS

Dep. Variable:	y	R-squared:	0.391			
Model:	OLS	Adj. R-squared:	0.386			
Method:	Least Squares	F-statistic:	83.59			
Date:	Mon, 10 Feb 2020	Prob (F-statistic):	8.40e-55			
Time:	21:05:33	Log-Likelihood:	-283.05			
No. Observations:	526	AIC:	576.1			
Df Residuals:	521	BIC:	597.4			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.3767	0.099	3.802	0.000	0.182	0.571
x1	-0.0233	0.003	-8.003	0.000	-0.029	-0.018
x2	0.0966	0.007	13.910	0.000	0.083	0.110
x3	0.0047	0.002	2.856	0.004	0.001	0.008
x4	0.0177	0.003	5.969	0.000	0.012	0.024
Omnibus:	12.980	Durbin-Watson:	1.772			
Prob(Omnibus):	0.002	Jarque-Bera (JB):	24.934			
Skew:	0.027	Prob(JB):	3.85e-06			
Kurtosis:	4.065	Cond. No.	138.			
Omnibus:	14.767	Durbin-Watson:	1.774			
Prob(Omnibus):	0.001	Jarque-Bera (JB):	29.522			
Skew:	0.071	Prob(JB):	3.88e-07			
Kurtosis:	4.152	Cond. No.	158.			

Exercice 9

Tester l'hypothèse de changement de structure entre les hommes et les femmes dans le modèle de l'exercice 1:

Construire toutes les variables d'interaction pour *educ*, *exper*, *tenure* et les inclure dans le modèle :

$\log(\text{wage}) \sim \text{const}, \text{femeduc}, \text{femexper}, \text{femtenure}, \text{educ}, \text{exper}, \text{tenure}$

On teste ensuite l'hypothèse

$$H_0 : \beta_{\text{femeduc}} = \beta_{\text{femexper}} = \beta_{\text{femtenure}} = 0.$$

Résultat pour SSR0, F et p-value :

88.58752552493836

25.12963127469592

3.4039598845150574e-15

On rejette l'hypothèse H_0 à 5%.

Exercice 9

```

=====
Dep. Variable:            y    R-squared:            0.403
Model:                  OLS    Adj. R-squared:        0.396
Method:                 Least Squares    F-statistic:      58.33
Date:                   Mon, 10 Feb 2020    Prob (F-statistic): 4.54e-55
Time:                   21:26:51    Log-Likelihood:   -277.88
No. Observations:       526    AIC:              569.8
Df Residuals:           519    BIC:              599.6
Df Model:                6
Covariance Type:        nonrobust
=====
              coef    std err          t      P>|t|      [0.025      0.975]
-----
const          0.3375      0.099      3.402     0.001      0.143      0.532
x1             -0.0139      0.004     -3.339     0.001     -0.022     -0.006
x2             -0.0056      0.003     -1.950     0.052     -0.011     4.25e-05
x3             -0.0080      0.007     -1.233     0.218     -0.021      0.005
x4              0.0953      0.007     13.794     0.000      0.082      0.109
x5              0.0080      0.002      3.552     0.000      0.004      0.012
x6              0.0183      0.004      5.036     0.000      0.011      0.025
=====
Omnibus:                 16.668    Durbin-Watson:           1.781
Prob(Omnibus):            0.000    Jarque-Bera (JB):        36.874
Skew:                    0.009    Prob(JB):                9.84e-09
Kurtosis:                 4.297    Cond. No.                 153.
=====

```

Exercice 10

Refaire le test de l'exercice 9 en utilisant la méthode "simple".

Estimer le modèle de l'exercice 1 pour les femmes (SSR01), ensuite pour les hommes (SSR00). Calculer $SSR0 = SSR00 + SSR01$

Résultats pour SSR01 et SSR00:

39.03529730794003

49.547190393273375

Donc on retrouve $SSR0 = SSR00 + SSR01$