Econométrie TP2

> Patrick Waelbroeck

Le modèle linéaire

Test d'hypothès

# Econométrie TP2 Le modèle linéaire Python

Patrick Waelbroeck

Telecom Paris

February 6, 2020

broeck Le modèle

linéaire Test On utilise la base de données wage1.raw.

#### Exercice 1

Définir la variable dépendante y = wage. La matrice de variable explicative inclut une constante et les variables educ, exper, tenure.

La commande np.ones(shape) retourne une matrice de dimension shape. La commande np.column\_stack permet de définir une matrice à partir des vecteurs. On peut également définir la matrice X avec pandas ou utiliser la fonction reshape.

```
y=wage
s=np.shape(wage)
const=np.ones(s)
educ=df[1]
exper=df[2]
tenure=df[3]
X=np.column_stack((const, educ, exper, tenure))
```

# Le modèle

Test d'hypothès

#### Exercice 2

Calculer les estimateurs des moindres carrés ordinaires  $\hat{\beta} = (X'X)^{-1}X'y$ 

#### Plusieurs possibilités :

- utiliser la bibliothèque np.linalg. Solution retenue pour la suite.
- utiliser la bibliothèque statsmodels. Voir plus loin.

La fonction X.T retourne la transposée de X.

```
beta = np.linalg.inv(X.T @ X)@X.T@y
```

On obtient ainsi pour beta:

```
array([-2.87273489, 0.59896507, 0.02233952, 0.16926865])
```

Calculer la matrice de variance-covariance des estimateurs OLS  $Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$ . Puis, calculer les écart-types.

On doit d'abord calculer un estimateur de la variance des résidus. Ensuite, calculer les écart-types en utilisant les fonctions np.sqrt() et np.diag().

```
u=y-X@beta
n,k=np.shape(X)
sig2=u.T@u/(n-k)
Var=sig2*np.linalg.inv(X.T @ X)
std=np.sqrt(np.diag(Var))
```

Résultat pour std :

```
array([0.72896429, 0.05128355, 0.01205685, 0.02164461])
```

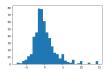
Patrick Waelbroeck

Le modèle linéaire

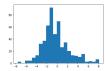
d'hypothès

#### Exercice 4

Faire l'histogramme des erreurs. Supprimer les observations pour lesquelles l'erreur est située à plus de trois écart-types de la moyenne. Refaire l'estimation.



Utiliser les commandes du TP 1 pour sélectionner les observations. Résultat pour  $\hat{\beta}$ : array([-1.70094791, 0.50063696, 0.01756833, 0.14558487])\



broeck

#### Exercice 5

Tester l'hypothèse de non significativité de exper, avec une hypothèse alternative des deux côtés à 5%:

 $H_0: \beta_{exper} = 0$ 

Calculer le seuil critique de rejet ainsi que la p-value.

Les distributions statistiques sont dans la bibliothèque scipy.stats. On a besoin de la distribution de student t.

from scipy.stats import t

On a également besoin de la fonction t.ppf(q,df) qui retourne le percentile d'ordre q avec un degré de liberté df. Les p-valeurs sont données par la fonction t.sf(x,df) qui calcule l'aire à droite de x sous la distribution de la loi de student avec df degrés de libertés

Résultats pour la student et la p-value

- 1.77887789559074
- 0.07585416697773316

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

Patrick Waelbroeck

Le modè linéaire

Test d'hypothèse

### Exercice 6

 $\label{lem:faire le même test en utilisant la bibliothèque statsmodels.}$ 

On importe la bibliothèque statsmodels. La commande OLS permet d'obtenir les estimateurs des moindres carrés ordinaires.

import statsmodels.api as sm
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())

#### OLS Regression Results

Dep. Variable:	у	R-squared:	0.314					
Model:	OLS	Adj. R-squared:	0.310					
Method:	Least Squares	F-statistic:	78.11					
Date:	Thu, 06 Feb 2020	Prob (F-statistic):	1.33e-41					
Time:	19:51:47	Log-Likelihood:	-1204.4					
No. Observations:	515	AIC:	2417.					
Df Residuals:	511	BIC:	2434.					
Df Model:	3							
C								

Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]		
const	-1.7009	0.604	-2.818	0.005	-2.887	-0.515		
x1	0.5006	0.043	11.736	0.000	0.417	0.584		
x2	0.0176	0.010	1.779	0.076	-0.002	0.037		
x3	0.1456	0.018	8.103	0.000	0.110	0.181		
Omnibus:		63.7	796 Durbin	-Watson:		1.832		
Prob(Omnil	bus):	0.0	000 Jarque	-Bera (JB):		90.134		

Le modèle linéaire

Test d'hypothèse Exercice 7

Refaire l'exercice 5 avec y=log(wage) (avec le même échantillon qu'en 2-5).

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

- 2.2412779619145424
- 0.025436638860709744

Cette fois, on rejette l'hypothèse  $H_0$ .

Tester l'hypothèse :

 $H_0: \beta_{educ} = 0.6$ 

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

-2.329221511737763

0.02023620756802269

Tester l'hypothèse :

$$H_0: \beta_{educ} = \beta_{exper}$$

Ecrire le modèle en fonction du paramètre  $\theta=\beta_{educ}-\beta_{exper}$ . Cela revient à faire une régression de y sur une  $constante,\ educ,\ educ+exper,\ tenure.$ 

Test d'hypothèse

OLS Regression Results   OLS Regression Results   OLS Regression Results   OLS Adj. R-squared: OLS Adj. R-squared: OLS Regression Results   OLS Adj. R-squared: OLS Regression Results   OLS Adj. R-squared: OLS Regression Results   OLS Regression   OLS Regression Results   OLS Regression   OLS Regression Results   OLS Regression   OLS Reg							
ep. Variable:	Exercice 9						
ep. Variable:			OLS Regr	ession R	esults		
odel:         OLS         Adj. R-squared:         0.3           ethod:         Least Squares         F-statistic:         78.1           ate:         Thu, 06 Feb 2020         Prob (F-statistic):         1.33e-4           ime:         20:48:22         Log-Likelihood:         -1204           o. Observations:         515         AIC:         2417           f Residuals:         511         BIC:         2436           f Model:         3         3         avariance Type:         nonrobust           ceef std err         t         P> t          [0.025         0.978           onst         -1.7009         0.604         -2.818         0.005         -2.887         -0.51           1         0.4831         0.041         11.884         0.000         0.403         0.62           2         0.0176         0.010         1.779         0.076         -0.002         0.03           3         0.1456         0.018         8.103         0.000         0.110         0.18           mnibus:         63.796         Durbin-Watson:         1.88							
odel:         OLS         Adj. R-squared:         0.3           ethod:         Least Squares         F-statistic:         78.1           ate:         Thu, 06 Feb 2020         Prob (F-statistic):         1.33e-4           ime:         20:48:22         Log-Likelihood:         -1204           o. Observations:         515         AIC:         2417           f Residuals:         511         BIC:         2436           f Model:         3         3         avariance Type:         nonrobust           ceef std err         t         P> t          [0.025         0.978           onst         -1.7009         0.604         -2.818         0.005         -2.887         -0.51           1         0.4831         0.041         11.884         0.000         0.403         0.62           2         0.0176         0.010         1.779         0.076         -0.002         0.03           3         0.1456         0.018         8.103         0.000         0.110         0.18           mnibus:         63.796         Durbin-Watson:         1.88	Dep. Variable:			y R-sq	uared:		0.314
ate: Thu, 06 Feb 2020 Prob (F-statistic): 1.33e-4 ime: 20:48:22 Log-Likelihood: -1204.  0. Observations: 515 AIC: 2417 f Residuals: 511 BIC: 2437 f Model: 3 ovariance Type: nonrobust	Model:		OL				0.310
ime: 20:48:22 Log-Likelihood: -1204. 0. Observations: 515 AIC: 2417 f Residuals: 511 BIC: 2434 f Model: 3 ovariance Type: nonrobust    Coef std err t P> t  [0.025 0.976   Const -1.7009 0.604 -2.818 0.005 -2.887 -0.51 1 0.4831 0.041 11.884 0.000 0.403 0.56 2 0.0176 0.010 1.779 0.076 -0.002 0.00 3 0.1456 0.018 8.103 0.000 0.110 0.16	Method:		Least Square	s F-st	atistic:		78.11
0. Observations: 515 AIC: 2417 f Residuals: 511 BIC: 2437 f Model: 3 ovariance Type: nonrobust    Coef std err	Date:	Th	u, 06 Feb 202	0 Prob	(F-statistic):		1.33e-41
f Residuals: 511 BIC: 2434 f Model: 3 ovariance Type: nonrobust  coef std err t P> t  [0.025 0.978 onst -1.7009 0.604 -2.818 0.005 -2.887 -0.51 1 0.4831 0.041 11.884 0.000 0.403 0.56 2 0.0176 0.010 1.779 0.076 -0.002 0.03 3 0.1456 0.018 8.103 0.000 0.110 0.18  mnibus: 63.796 Durbin-Watson: 1.88	Time:		20:48:2	2 Log-	Likelihood:		-1204.4
f Model: 3 ovariance Type: nonrobust    Coef   std err   t   P> t    [0.025   0.978]   Onst   -1.7009   0.604   -2.818   0.005   -2.887   -0.51	${\tt No.\ Observations:}$		51	5 AIC:			2417.
covariance Type:         nonrobust           coef         std err         t         P> t          [0.025]         0.978           const         -1.7009         0.604         -2.818         0.005         -2.887         -0.51           1         0.4831         0.041         11.884         0.000         0.403         0.65           2         0.0176         0.010         1.779         0.076         -0.002         0.03           3         0.1456         0.018         8.103         0.000         0.110         0.18           mnibus:         63.796         Durbin-Watson:         1.83	Df Residuals:						2434.
coef         std err         t         P> t          [0.025         0.978           onst         -1.7009         0.664         -2.818         0.005         -2.887         -0.51           1         0.4831         0.041         11.884         0.000         0.403         0.56           2         0.0176         0.010         1.779         0.076         -0.002         0.03           3         0.1456         0.018         8.103         0.000         0.110         0.18           mnibus:         63.796         Durbin-Watson:         1.83	Df Model:			-			
onst -1.7009 0.604 -2.818 0.005 -2.887 -0.51 1 0.4831 0.041 11.884 0.000 0.403 0.56 2 0.0176 0.010 1.779 0.076 -0.002 0.03 3 0.1456 0.018 8.103 0.000 0.110 0.18	Covariance Type:		nonrobus	t			
onst -1.7009 0.604 -2.818 0.005 -2.887 -0.51 1 0.4831 0.041 11.884 0.000 0.403 0.56 2 0.0176 0.010 1.779 0.076 -0.002 0.03 3 0.1456 0.018 8.103 0.000 0.110 0.18							
1 0.4831 0.041 11.884 0.000 0.403 0.56 2 0.0176 0.010 1.779 0.076 -0.002 0.03 3 0.1456 0.018 8.103 0.000 0.110 0.16 	•	coef	std err	t	P> t	[0.025	0.975]
2 0.0176 0.010 1.779 0.076 -0.002 0.03 3 0.1456 0.018 8.103 0.000 0.110 0.18 	const -1.7	7009	0.604	-2.818	0.005	-2.887	-0.515
3 0.1456 0.018 8.103 0.000 0.110 0.18 mnibus: 63.796 Durbin-Watson: 1.8	x1 0.4	4831	0.041	11.884	0.000	0.403	0.563
mnibus: 63.796 Durbin-Watson: 1.83	x2 0.0	0176	0.010	1.779	0.076	-0.002	0.037
	x3 0.:	1456	0.018	8.103	0.000	0.110	0.181
rob(Omnibus): 0.000 Jarque-Bera (JB): 90.13	Omnibus:					1.832	
	Prob(Omnibus):		0.00	0 Jarq	ue-Bera (JB):		90.134

On rejette l'hypothèse  $H_0$  à 5%.

Skew: Kurtosis: 0.866

4.097

Prob(JB):

Cond. No.

189.

2.68e-20

Le modèle

Test d'hypothèse Exercice 10

Tester l'hypothèse :

$$H_0: \beta_{educ} + \beta_{exper} = 1$$

Réécrire le modèle en fonction de  $\theta=1-\beta_{educ}-\beta_{exper}$ . Cela revient à redéfinir y'=wage-educ que l'on explique en fonction d'une constante, educ, educ-exper et tenure.

Le modè linéaire

Test d'hypothèse

#### Exercice 10

OLS	Regression	Results
-----	------------	---------

Dep. Variable:		y R-squared:				
Model:		OLS Adj. R-squared:			0.365	
Method:	Least S	quares F-	statistic:		99.36	
Date:	Thu, 06 Fe	b 2020 Pr	ob (F-statistic	:):	1.13e-50	
Time:	20	:57:15 Lo	g-Likelihood:		-1204.4	
No. Observations:		515 AIC:			2417.	
Df Residuals:		511 BIC:			2434.	
Df Model:		3				
Covariance Type:	non	robust				
	coef std er	r	t P> t	[0.025	0.975]	
	7009 0.604 4818 0.04			-2.887 -0.574	-0.515 -0.390	
	0176 0.01			-0.002	0.037	

x1	-0.4818	0.047	-10.	314	0.000	-0.574	-0.390
x2	0.0176	0.010	1.	779	0.076	-0.002	0.037
x3	0.1456	0.018	8.	103	0.000	0.110	0.181
Omnibus:		63.79	96	Durbin-Wa	atson:		1.832
Prob(Omnibus):		0.00	00	Jarque-Be	era (JB):		90.134
Skew:		0.86	66	Prob(JB):		:	2.68e-20
Kurtosis:		4.09	97	Cond. No.			93.1

On rejette l'hypothèse  $H_0$  à 5%.