Econométrie TP2

Le modèle linéaire Python

Patrick Waelbroeck

Telecom Paris

February 6, 2020

Le modèle linéaire - définition des variables explicatives

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))
```

Calcul des estimateurs OLS

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])
```

Estimation de la matrice de variance-covariance

Exercice 3

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))
```

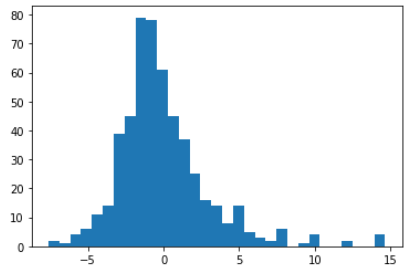
Résultat pour std :

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

Estimation de la matrice de variance-covariance

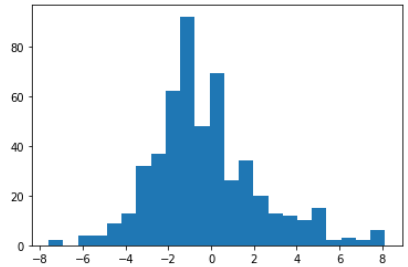
Exercice 4

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])\
```



Solution exercice 2-4

```
plt.hist(u,'auto')
s=(np.abs(u)<3*np.sqrt(sig2))
u1=u[s]
plt.hist(u1,'auto')
y=y[s]
X=X[s,:]
beta = np.linalg.inv(X.T @ X)@X.T@y
```

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
0.075854166977733 16
```

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

Solution exercice 2-5

```
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))
beta[2]/std[2]
t.sf(beta[2]/std[2],n-k)*2
```

Utilisation de statsmodels

Exercice 6

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())
```

Utilisation de statsmodels

Exercice 6

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   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
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:                       63.796   Durbin-Watson:                   1.832
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               90.134
Skew:                           0.866   Prob(JB):                     2.68e-20
Kurtosis:                       4.097   Cond. No.                         135.
==============================================================================
```

Effet de la forme fonctionnelle

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.

Solution exercice 2-7

```
y=np.log(y)
beta = np.linalg.inv(X.T @ X)@X.T@y
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))
beta[2]/std[2]
t.sf(beta[2]/std[2],n-k)*2
```

Test d'égalité à une valeur

Exercice 8

Tester l'hypothèse :
$H_0 : \beta_{educ} = 0.6$

Résultat pour la statistique de test et la p-value :
-2.329221511737763
0.020236207568022 69

Solution exercice 2-8

```
test=(beta[1]-0.6)/std[1]
2*(1-t.sf(test,n-k))
```

Test d'égalité des coefficients

Exercice 9

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'égalité des coefficients

Exercice 9

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   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:                        20:48:22   Log-Likelihood:                -1204.4
No. Observations:                 515   AIC:                             2417.
Df Residuals:                     511   BIC:                             2434.
Df Model:                           3
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             0.4831      0.041     11.884      0.000       0.403       0.563
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:                       63.796   Durbin-Watson:                   1.832
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               90.134
Skew:                           0.866   Prob(JB):                     2.68e-20
Kurtosis:                       4.097   Cond. No.                         189.
==============================================================================
```

On rejette l'hypothèse H_0 à 5%.

Solution exercice 2-9

```
toteduc=educ+exper
X=np.column_stack((const, educ, toteduc, tenure))
X=X[s,:]
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
```

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*.

Test de contrainte linéaire

Exercice 10

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.368
Model:                            OLS   Adj. R-squared:                  0.365
Method:                 Least Squares   F-statistic:                     99.36
Date:                Thu, 06 Feb 2020   Prob (F-statistic):           1.13e-50
Time:                        20:57:15   Log-Likelihood:                -1204.4
No. Observations:                 515   AIC:                             2417.
Df Residuals:                     511   BIC:                             2434.
Df Model:                           3
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            -0.4818      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:                       63.796   Durbin-Watson:                   1.832
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               90.134
Skew:                           0.866   Prob(JB):                     2.68e-20
Kurtosis:                       4.097   Cond. No.                         93.1
==============================================================================
```

On rejette l'hypothèse H_0 à 5%.

Solution exercice 2-10

```
y=wage-educ
diffeduc=exper-educ
X=np.column_stack((const, educ, diffeduc, tenure))

y=y[s]
X=X[s,:]
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
```