MODS202 - Econometrics

2021/11/21

YANG Yining

YANG Yuqing

Part I - Regression

Question 1-1

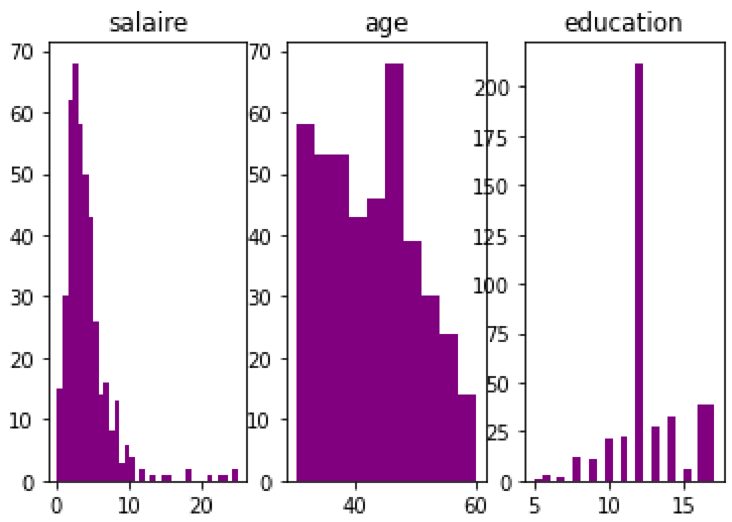
Lire le fichier mroz.txt. Ne sélectionner que les observations pour lesquelles la variable wage *est strictement positive.*

We used "pandas.read_fwf" function to read data from *MORZ.txt*, and then used "pandas.to_numeric" to transform the *wage* and *lwage* column into numbers. Finally, we threw lines where its *wage* is not strictly positive.

Question 1-2

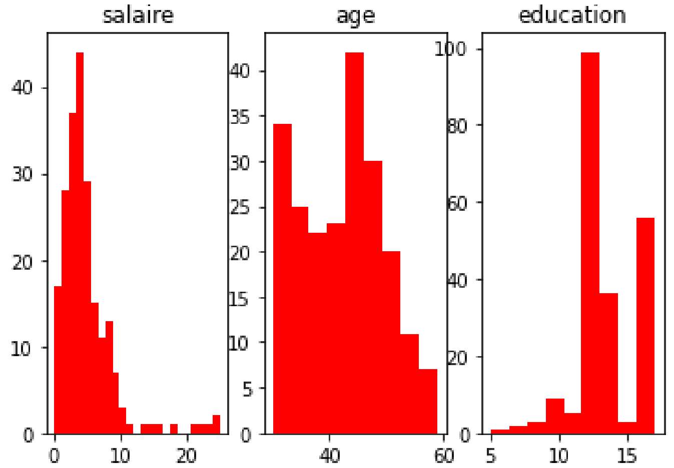
Faire les statistiques descriptives du salaire, de l'age et de l'éducation pour l'ensemble des femmes puis, pour les femmes dont le salaire du mari est supérieure à la médiane de l'échantillon, puis pour les femmes dont le salaire du mari est inférieur à la médiane de l'échantillon.

For all women, we have the distribution of their wage, age and education shown in the graph below:



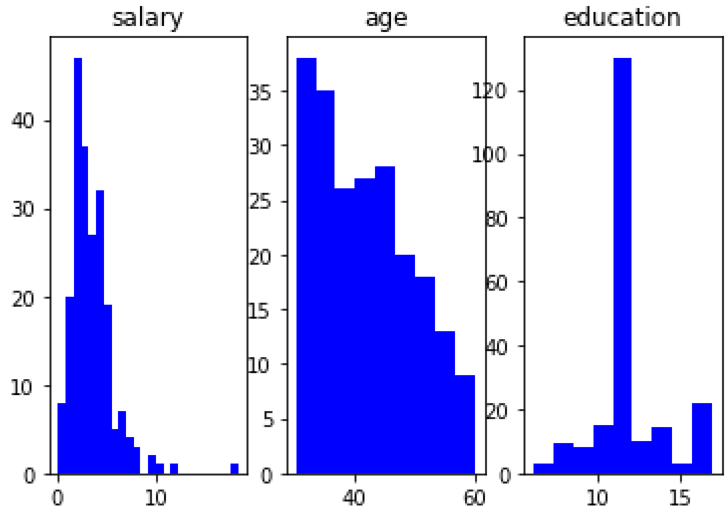
	wage	age	education
Mean	4.17	41.97	12.65
Median	3.48	42	12.0
Maximum	25.0	60	17
Minimum	0.12	30	5
Standard	3.30	7.71	2.28
Variance	10.93	59.47	5.21

For women whose husband's wage higher than the median, we have:



	wage	age	education
Mean	4.89	42.27	13.24
Median	3.84	43	12.0
Maximum	25.0	59	17
Minimum	0.16	30	5
Standard	4.03	7.37	2.35
Variance	16.25	54.33	5.53

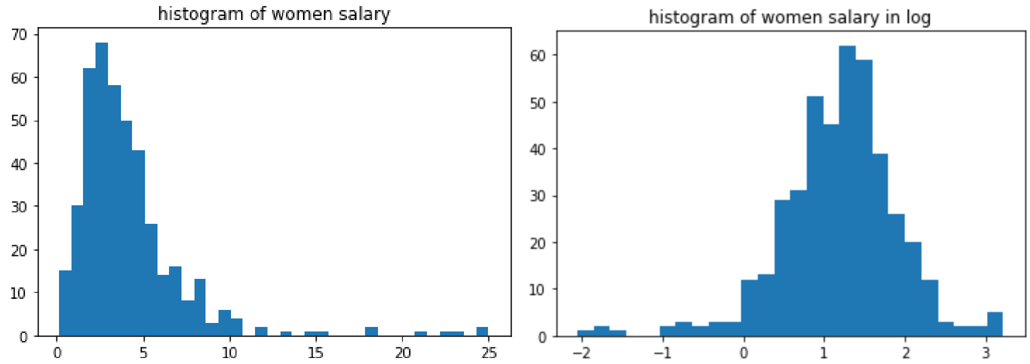
For women whose husband's wage lower than the median, we have:



	wage	age	education
Mean	3.45	41.66	12.07
Median	2.97	41	12.0
Maximum	18.26	60	17
Minimum	0.12	30	6
Standard	2.13	8.02	2.04
Variance	4.57	64.42	4.20

Question 1-3

Faire l'histogramme de la variable wage. *Calculer le log de* wage *et faire l'histogramme. Comparez les deux histogrammes et commentez*



By comparing these two figures, we see that the log-figure is denser, with a much smaller range of data. And the existence of some marginal data is fading in log-figure. The most important thigh for logarithm is that this operation does not change the relationship between variables. Besides, the log-figure resembles more to a normal distribution. Besides, in econometrics, the coefficient of variable in logarithmic form is its elasticity. Meanwhile, once the variable needs to be differentiated, the actual meaning of the logarithmic variable difference is the approximate growth rate.

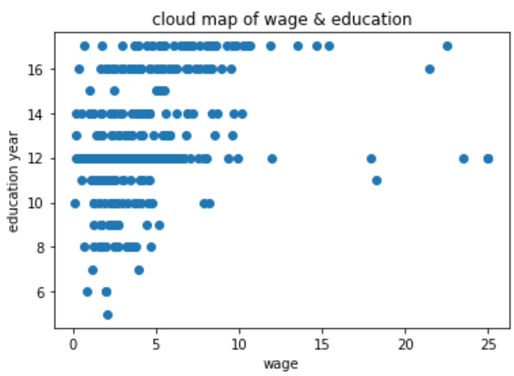
Question 1-4

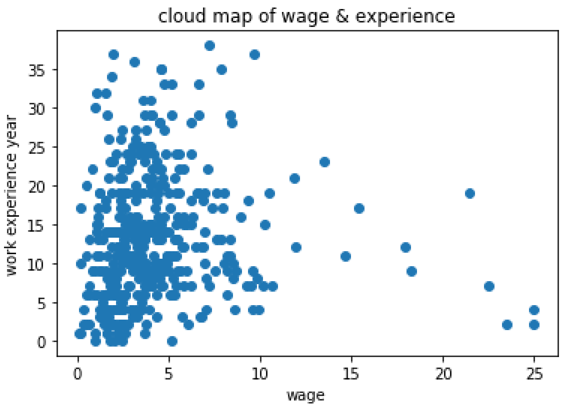
Calculer les corrélations motheduc *et* fatheduc*. Commentez. Il y a-t-il un problème de multicollinéarité si l'on utilise ces variables comme variables explicatives ?*

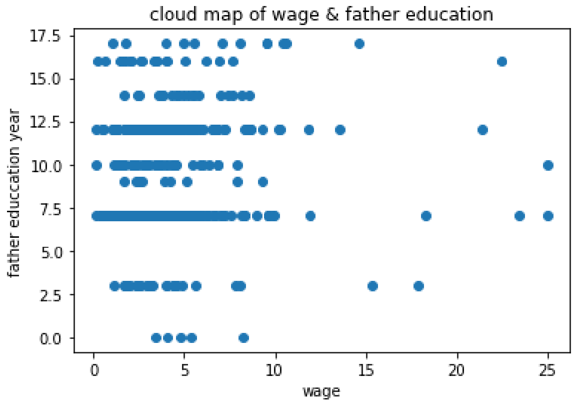
The correlation between mother's education and father's education is 0.554, which shows that the educational background of one's father and mother is comparatively closed, which can be considered moderately correlated. Thus, when we use these two variables as explanatory variables, we will not meet a serious problem of multicollinearity, at least far away from the perfect multicollinearity.

Question 1-5

Faites un graphique en nuage de point entre wage *et* educ*,* wage *et* exper*,* wage *et* fatheduc*. Commentez. S'agit-il d'un effet "toute chose étant égale par ailleurs ?"*







From the cloud maps, we can see that the wage is not closely connected to only one variable, neither education, experience, nor father education. We cannot simply predict one woman's salary by only looking at her education year/working experience/ father education year. This is not an effect of "everything else being equal", because apparently everything else is not equal. We didn't control the data to be only differed from only one variable listed above. Thus, although the cloud map shows that the relation between wage and other factors are not so close, we cannot conclude that these factors are not important, because we didn't control other variables to be equal.

Question 1-6

Quelle est l'hypothèse fondamentale qui garantit des estimateurs non biaisés ? Expliquer le biais de variable omise.

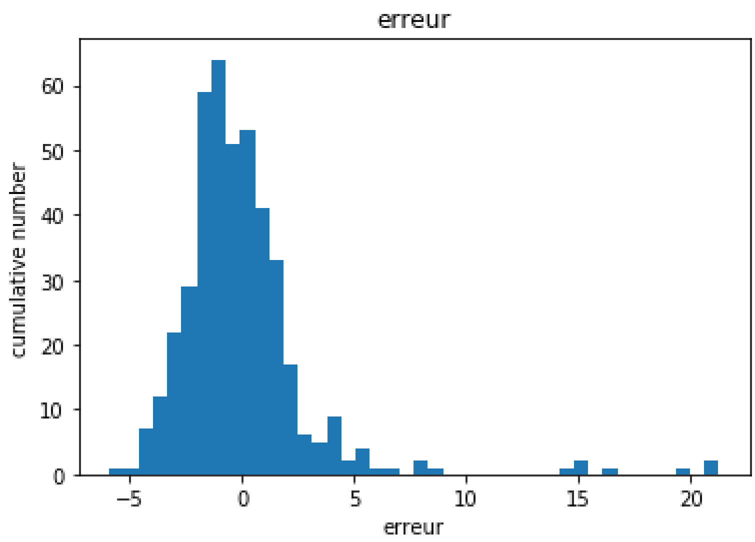
The fundamental assumption guaranteeing unbiased estimators is that the residuals $\epsilon\epsilon$ of the model are independent of the explanatory variables, which is to say:

$$E(\epsilon|X) = E(\epsilon) = 0$$

The omitted variable bias is to say that some variables that should be treated as explanatory variables, but are considered to be residuals in the model. It occurs when one of the explanatory variables is correlated both with the explained variable and with the residual that is not taken into account in the model. In this case, the fundamental assumption is no longer satisfied.

Question 1-7

Faire la régression de wage *en utilisant les variables explicatives un constante,* city, educ, exper, nwifeinc, kidslt6, kidsgt6. *Commentez l'histogramme des résidus.*

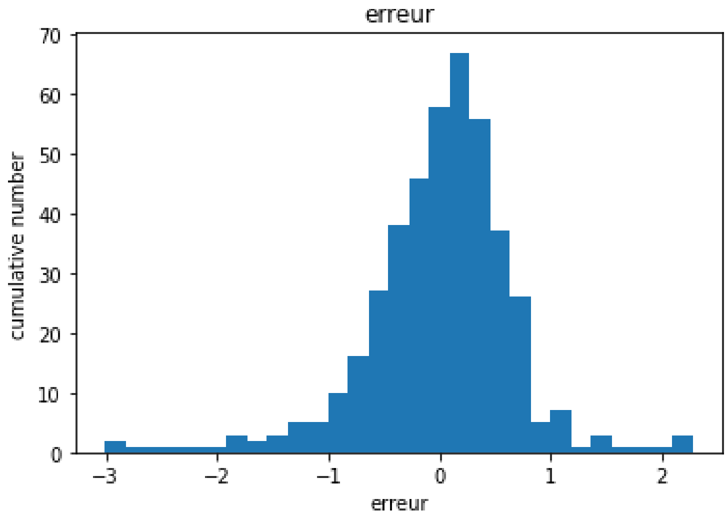



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      6   R-squared:                       0.127
Model:                            OLS   Adj. R-squared:                  0.115
Method:                 Least Squares   F-statistic:                     10.23
Date:                Sat, 20 Nov 2021   Prob (F-statistic):           1.41e-10
Time:                        23:55:52   Log-Likelihood:                -1090.0
No. Observations:                 428   AIC:                             2194.
Df Residuals:                     421   BIC:                             2222.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -2.4034      0.963     -2.495      0.013      -4.297      -0.510
x1             0.3697      0.327      1.132      0.258      -0.272       1.012
x2             0.4600      0.070      6.546      0.000       0.322       0.598
x3             0.0238      0.021      1.141      0.255      -0.017       0.065
x4             0.0152      0.015      0.984      0.326      -0.015       0.046
x5             0.0362      0.397      0.091      0.927      -0.744       0.816
x6            -0.0619      0.125     -0.494      0.622      -0.308       0.185
==============================================================================
Omnibus:                      345.825   Durbin-Watson:                   2.056
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6499.393
Skew:                           3.389   Prob(JB):                         0.00
Kurtosis:                      20.847   Cond. No.                         178.
==============================================================================
```

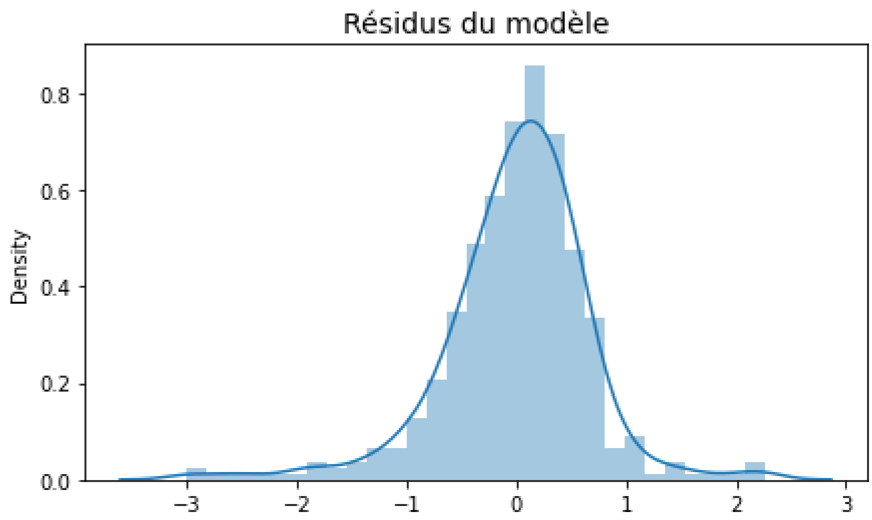
The residuals are centered around zero, but not in gaussian distribution. This is because that we didn't filter the data which are far from others. In other words, the cases which are considered abnormal should be thrown away to make the model more suitable for the common situation.

Question 1-8

Faire la régrssion de lwage sur une constante, city, educ, exper, nwifeinc, kidslt6, kidsgt6. Comparer l'histogramme obtenu à celui de la question 7.




```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     20   R-squared:                       0.156
Model:                            OLS   Adj. R-squared:                  0.144
Method:                 Least Squares   F-statistic:                     12.92
Date:                Sat, 20 Nov 2021   Prob (F-statistic):           2.00e-13
Time:                        21:15:07   Log-Likelihood:                -431.92
No. Observations:                 428   AIC:                             877.8
Df Residuals:                     421   BIC:                             906.3
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.3990      0.207     -1.927      0.055      -0.806       0.008
x1             0.0353      0.070      0.503      0.616      -0.103       0.173
x2             0.1022      0.015      6.770      0.000       0.073       0.132
x3             0.0155      0.004      3.452      0.001       0.007       0.024
x4             0.0049      0.003      1.466      0.143      -0.002       0.011
x5            -0.0453      0.085     -0.531      0.596      -0.213       0.122
x6            -0.0117      0.027     -0.434      0.664      -0.065       0.041
==============================================================================
Omnibus:                       79.542   Durbin-Watson:                   1.979
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              287.192
Skew:                          -0.795   Prob(JB):                     4.34e-63
Kurtosis:                       6.685   Cond. No.                         178.
==============================================================================
```



The first graph is obtained by our self-designed algorithm, and the second is drawn by the package “statsmodels.api”. The two methods give us the same results. By comparing the figure of log-wage and wage, we can see that the influence of extremities, the abnormal situation, is reduced by doing the “log” calculation of wage. By reading from the graph, we deduce that the education, living city and experience are significant to a woman’s salary.

Question 1-9

Tester l'hypothèse de non significativité de nwifeinc *avec un seuil de significativité de 1%, 5% et 10% (test alternatif des deux côtés). Commentez les p-values.*

By interpreting this question, we use the null hypothesis that the factor “nwifeinc”, (faminc - wage*hours)/1000, has a coefficient of zero, which means that it has no influence to women’s salary.

Significance threshold 1%: [-0.0037351537917680623, 0.01350231203119639]
Significance threshold 5%: [-0.0016635636668245763, 0.011430721906252904]
Significance threshold 10%: [-0.000607242838783161, 0.010374401078211488]
We can see that the coefficient of “nwifeinc” is 0.0049, which is in the interval of which of all these three threshold above. Thus, we cannot reject the null hypothesis.

Question 1-10

Tester l’hypothèse que le coefficient associé à nwifeinc est égal à 0.01 avec un seuil de significativité de 5% (test à alternatif des deux côtés)

We see that the p-value of the hypothesis that “nwifeinc = 0.01” is 0.125, which is bigger than 0.05. Thus, we cannot reject the hypothesis that “nwifeinc = 0.01” with the significance threshold of 5%.

Question 1-11

Tester l'hypothèse jointe que le coefficient de nwifeinc est égal à 0.01 et que celui de city est égal à 0.05

We removed the factor "nwifeinc" and "city" from the model, and we get the result table as shown below:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      6   R-squared:                       0.116
Model:                            OLS   Adj. R-squared:                  0.108
Method:                 Least Squares   F-statistic:                     13.88
Date:                Sat, 20 Nov 2021   Prob (F-statistic):           1.19e-10
Time:                        21:15:08   Log-Likelihood:                -1090.7
No. Observations:                 428   AIC:                             2191.
Df Residuals:                     423   BIC:                             2212.
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -2.2865      0.957     -2.390      0.017      -4.167      -0.406
x1             0.4774      0.067      7.141      0.000       0.346       0.609
x2             0.0224      0.021      1.089      0.277      -0.018       0.063
x3             0.0194      0.395      0.049      0.961      -0.756       0.795
x4            -0.0710      0.125     -0.568      0.570      -0.317       0.175
==============================================================================
Omnibus:                      348.571   Durbin-Watson:                   2.057
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6652.941
Skew:                           3.424   Prob(JB):                         0.00
Kurtosis:                      21.060   Cond. No.                         123.
==============================================================================
```

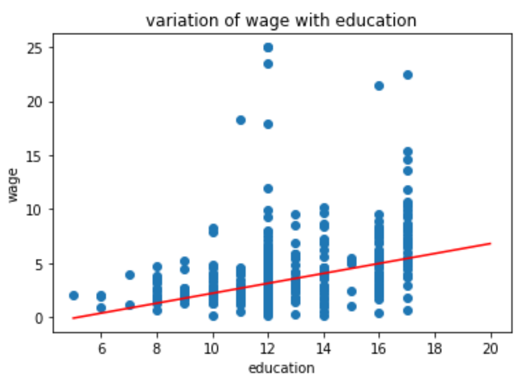
By calculating with the formula:

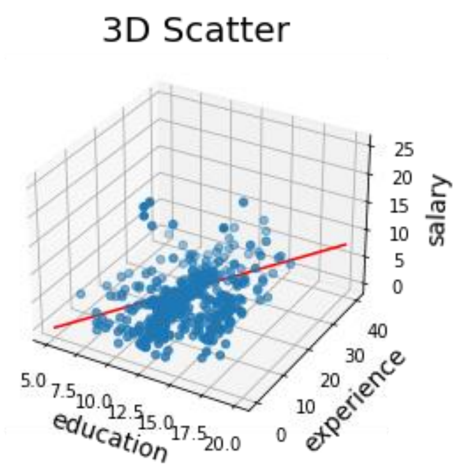
$$F \equiv \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n-k-1)},$$

We obtained the p-value of the conjoint hypothesis, 0.528, which is bigger than both 0.01 and 0.05. Thus, we cannot reject this hypothesis.

Question 1-12

Faites une représentation graphique de la manière dont le salaire augmente avec l'éducation et l'expérience professionnelle. Commentez





We take the theory value from question 1-7 and draw it in red line. Reading these three figures, we can say that the augmentation of duration of education and experience has a positive influence on the augmentation of women's salary. The slope of education-salary and experience-salary are quite similar to the real-world scatters' trend. Although the red line in the second graph is a little bit away from the scatters, the augmentation trend is quite similar. And this difference is due to the fact that we take many factors into consideration in the model of question 1-7. From the third graph, we can see that the sacrifice of the constant value in "experience-wage" fitting is favorable for the "education&experience-wage" fitting.

Question 1-13

Tester l'égalité des coefficients associés aux variables kidsgt6 *et* kidslt6*. Interprétez.*

We make the hypothesis that the coefficient of "kidsgt6" and "kidslt6" are the same, which can be transformed into that the coefficient of the factor "kidsgt6- kidslt6" is 0. We also used the all the rest factors in model 1-7, and we obtain:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      6   R-squared:                       0.127
Model:                            OLS   Adj. R-squared:                  0.115
Method:                 Least Squares   F-statistic:                     10.23
Date:                Sat, 20 Nov 2021   Prob (F-statistic):           1.41e-10
Time:                        23:01:19   Log-Likelihood:                -1090.0
No. Observations:                 428   AIC:                             2194.
Df Residuals:                     421   BIC:                             2222.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -2.4034      0.963     -2.495      0.013      -4.297      -0.510
x1             0.3697      0.327      1.132      0.258      -0.272       1.012
x2             0.4600      0.070      6.546      0.000       0.322       0.598
x3             0.0238      0.021      1.141      0.255      -0.017       0.065
x4             0.0152      0.015      0.984      0.326      -0.015       0.046
x5            -0.0257      0.412     -0.062      0.950      -0.835       0.784
x6            -0.0619      0.125     -0.494      0.622      -0.308       0.185
==============================================================================
Omnibus:                      345.825   Durbin-Watson:                   2.056
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             6499.393
Skew:                           3.389   Prob(JB):                         0.00
Kurtosis:                      20.847   Cond. No.                         178.
==============================================================================
```

And we have the p-value of this hypothesis 0.77.

Question 1-14

En utilisant le modèle de la question 7, faire le test d'hétéroscédasticité de forme linéaire en donnant la p-valeur. Déterminer la ou les sources d'hétéroscédasticité et corriger avec les méthodes vues en cours. Comparer les écarts-types des coefficients estimés avec ceux obtenus à la question 7. Commenter.

By the heteroskedasticity test, we obtain the model shown in the following page:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.022
Model:                            OLS   Adj. R-squared:                  0.008
Method:                 Least Squares   F-statistic:                     1.593
Date:                Sat, 20 Nov 2021   Prob (F-statistic):              0.148
Time:                        21:15:08   Log-Likelihood:                -2207.4
No. Observations:                 428   AIC:                             4429.
Df Residuals:                     421   BIC:                             4457.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.4856     13.111      0.113      0.910     -24.285      27.256
x1             5.9645      4.444      1.342      0.180      -2.770      14.699
x2             0.8077      0.956      0.845      0.399      -1.072       2.687
x3            -0.5341      0.284     -1.880      0.061      -1.093       0.024
x4             0.0435      0.211      0.206      0.837      -0.371       0.458
x5             4.9573      5.402      0.918      0.359      -5.661      15.575
x6            -0.4018      1.706     -0.236      0.814      -3.756       2.952
==============================================================================
Omnibus:                      638.793   Durbin-Watson:                   2.029
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            96122.153
Skew:                           8.127   Prob(JB):                         0.00
Kurtosis:                      74.595   Cond. No.                         178.
==============================================================================
```

Thus, we have the p-value 0.148. And we observe that the absolute value of variable1 and variable3 is big, which refers to the city and experience. However, due to the property of factor "city", which is either 0 or 1, we cannot separate it into different groups just like in class, so we turn to look at x2 and x3, education and experience. Instead of making logarithm, we see that making exponential with indices of 4 is more suitable, and then we have:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.013
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.9567
Date:                Sat, 20 Nov 2021   Prob (F-statistic):              0.454
Time:                        23:47:48   Log-Likelihood:                -2207.2
No. Observations:                 428   AIC:                             4428.
Df Residuals:                     421   BIC:                             4457.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.4521      5.822      0.936      0.350      -5.992      16.897
x1            -5.7903      4.432     -1.306      0.192     -14.502       2.922
x2          6.735e-05   9.74e-05      0.692      0.490      -0.000       0.000
x3         -1.205e-12   4.91e-12     -0.245      0.806   -1.09e-11    8.45e-12
x4             0.1126      0.210      0.537      0.591      -0.299       0.525
x5             6.3850      5.332      1.197      0.232      -4.096      16.866
x6             0.7677      1.622      0.473      0.636      -2.421       3.956
==============================================================================
Omnibus:                      647.495   Durbin-Watson:                   2.035
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           103735.063
Skew:                           8.315   Prob(JB):                         0.00
Kurtosis:                      77.434   Cond. No.                     1.37e+12
==============================================================================
```

With p-value raised up to 0.454, and the standard deviation is bigger.

Question 1-15

Tester le changement de structure de la question 8 entre les femmes qui ont plus de 43 ans et les autres : test sur l'ensemble des coefficients. Commentez et donnez les p-valeurs

We have the regression of women over 43 years old here:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     20   R-squared:                       0.145
Model:                            OLS   Adj. R-squared:                  0.120
Method:                 Least Squares   F-statistic:                     5.750
Date:                Sat, 20 Nov 2021   Prob (F-statistic):           1.51e-05
Time:                        21:15:09   Log-Likelihood:                -197.61
No. Observations:                 211   AIC:                             409.2
Df Residuals:                     204   BIC:                             432.7
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.1774      0.253     -0.701      0.484      -0.677       0.322
x1            -0.0216      0.101     -0.214      0.830      -0.220       0.177
x2             0.0780      0.018      4.262      0.000       0.042       0.114
x3             0.0155      0.005      2.865      0.005       0.005       0.026
x4             0.0084      0.004      1.886      0.061      -0.000       0.017
x5            -0.1228      0.285     -0.430      0.667      -0.685       0.440
x6             0.0112      0.043      0.259      0.796      -0.074       0.097
==============================================================================
Omnibus:                       47.345   Durbin-Watson:                   2.286
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              154.469
Skew:                          -0.886   Prob(JB):                     2.87e-34
Kurtosis:                       6.799   Cond. No.                         198.
==============================================================================
```

And women under 43 years old:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     20   R-squared:                       0.190
Model:                            OLS   Adj. R-squared:                  0.167
Method:                 Least Squares   F-statistic:                     8.209
Date:                Sat, 20 Nov 2021   Prob (F-statistic):           5.47e-08
Time:                        21:15:09   Log-Likelihood:                -228.61
No. Observations:                 217   AIC:                             471.2
Df Residuals:                     210   BIC:                             494.9
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.7940      0.353     -2.251      0.025      -1.489      -0.099
x1             0.0856      0.100      0.853      0.394      -0.112       0.283
x2             0.1326      0.026      5.139      0.000       0.082       0.184
x3             0.0258      0.009      2.878      0.004       0.008       0.044
x4             0.0017      0.005      0.332      0.740      -0.008       0.012
x5            -0.0860      0.098     -0.877      0.382      -0.279       0.107
x6            -0.0208      0.039     -0.535      0.593      -0.097       0.056
==============================================================================
Omnibus:                       37.393   Durbin-Watson:                   1.825
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              124.647
Skew:                          -0.654   Prob(JB):                     8.57e-28
Kurtosis:                       6.475   Cond. No.                         189.
==============================================================================
```

Here, the null hypothesis is that there is no change for the age of women. We use the test of Chow to check this. And we have the information of Fisher in this form:

$$\frac{(S_C - (S_1 + S_2))/k}{(S_1 + S_2)/(N_1 + N_2 - 2k)}.$$

And then we have the p-value 0.30.

Question 1-16

Refaire la question 15 en supposant que seuls les rendements de l'éducation et de l'expérience professionnelle changent selon l'âge de la femme. Formuler l'hypothèse H0 et tester-la. Donnez la p-valeur.

Since only the education and experience will change with age, we only need to change these two columns of data. After re-modeling, we then have:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     20   R-squared:                       0.163
Model:                            OLS   Adj. R-squared:                  0.147
Method:                 Least Squares   F-statistic:                     10.19
Date:                Sat, 20 Nov 2021   Prob (F-statistic):           4.96e-13
Time:                        21:15:09   Log-Likelihood:                -430.07
No. Observations:                 428   AIC:                             878.1
Df Residuals:                     419   BIC:                             914.7
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.4200      0.207     -2.029      0.043      -0.827      -0.013
x1            -0.0028      0.010     -0.273      0.785      -0.023       0.018
x2            -0.0081      0.009     -0.869      0.385      -0.026       0.010
x3             0.0514      0.071      0.729      0.467      -0.087       0.190
x4             0.1018      0.016      6.335      0.000       0.070       0.133
x5             0.0237      0.008      2.899      0.004       0.008       0.040
x6             0.0055      0.003      1.644      0.101      -0.001       0.012
x7            -0.0750      0.089     -0.846      0.398      -0.249       0.099
x8            -0.0232      0.028     -0.837      0.403      -0.078       0.031
==============================================================================
Omnibus:                       78.791   Durbin-Watson:                   1.982
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              290.946
Skew:                          -0.779   Prob(JB):                     6.63e-64
Kurtosis:                       6.726   Cond. No.                         193.
==============================================================================
```

The null hypothesis is that the wage of women will not be influenced by their education and experience, and then we got the p-value 0.163.

Part II - Séries temporelles

Question 2-1

Importer les données du fichier quarterly.xls (corriger le problème éventuel d'observations manquantes).

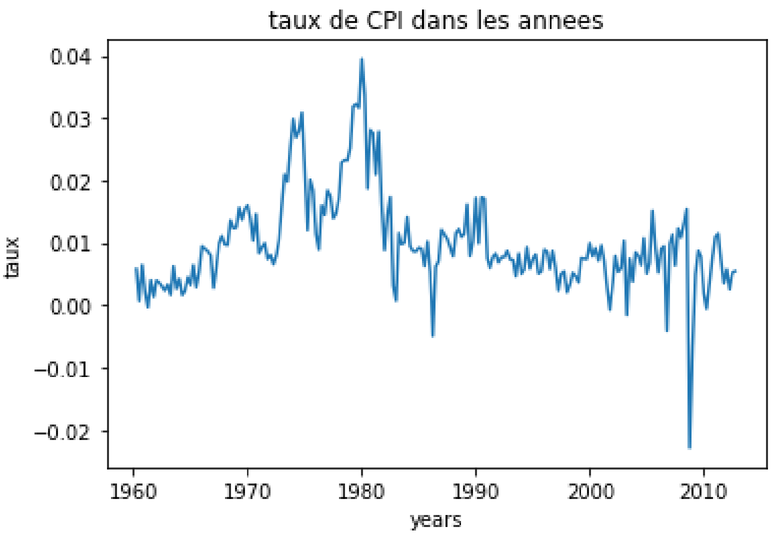
We used "pandas.read_excel" function to read data from quarterly.xls.

Question 2-2

Calculer inf, le taux d'inflation à partir de la variable CPI. Faire un graphique dans le temps de inf. Commentez.

We used this formula to calculate the inflation rate, and then we got:

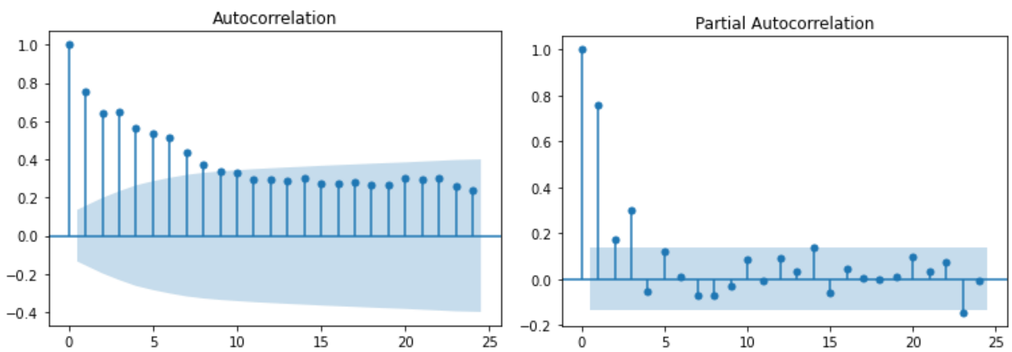
$$df_{inf} = (df_{CPI_t} - df_{CPI_{t-1}})/(df_{CPI_t})$$



We see that the inflation rate is usually positive, except for the significant year 2008, which is the time for the global financial crisis. The inflation rate held comparatively high during 1970s and 1980s, and is comparatively stable in 1960s and 1990s.

Question 2-3

Interpréter l'autocorrélogramme et l'autocorrélogrammes partiels de inf. Quelle est la différence entre ces deux graphiques ?



The autocorrelation graph shows that the autocorrelation dominus with time(delay), while the partial autocorrelation oscillates around 0. And we can deduct from these two figures that the past data pose an influence on the future data in terms of inflation rate.

The autocorrelation presents the influence of a series at time t-x in the past on the value of the series at time t, independent of the rest of the observations. The partial autocorrelation presents the influence of all of all the value in the past until time t-x. Thus, we identify the joint effects of the different years.

Question 2-4

Quelle est la différence entre la stationnarité et l'ergodicité ? Pourquoi a-t-on besoin de ces deux conditions

The stationarity means that the distribution of data is periodic. The ergodicity means that the limit of a dataset is independent with the starting point of analysis.

These two conditions are suffisante for applying the ergodicity theory, which proposes that the time average is equal to the spatial average of one dataset. Thus, a single trajectory makes it possible to analyze the stochastic process.

Question 2-5

Proposer une modélisation AR(p) de inf, en utilisant tous les outils vus au cours.

By using “statsmodels.tsa.api”, we get the parameters for different delays:

[0.00149584 0.60377471 -0.02700434 0.33513437 -0.06041087]

Akaike information criterion: -10.662277978959965

Bayesian Information Criterions: -10.565677434227636

Question 2-6

Estimer le modèle de la courbe de Philips qui explique le taux de chômage (Unemp) en fonction du taux d’inflation courant et une constante.

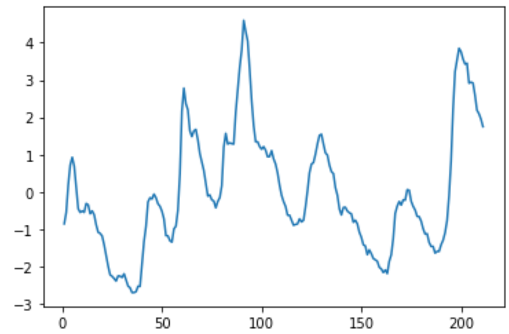
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Unemp   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.005
Method:                 Least Squares   F-statistic:                   0.01214
Date:                Sun, 21 Nov 2021   Prob (F-statistic):              0.912
Time:                        01:27:00   Log-Likelihood:                -400.28
No. Observations:                 211   AIC:                             804.6
Df Residuals:                     209   BIC:                             811.3
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          6.0708      0.181     33.576      0.000       5.714       6.427
x1             0.0159      0.144      0.110      0.912      -0.269       0.301
==============================================================================
Omnibus:                       13.872   Durbin-Watson:                   0.044
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               15.356
Skew:                           0.660   Prob(JB):                     0.000463
Kurtosis:                       2.937   Cond. No.                         2.99
==============================================================================
```

We can see that, with the inflation rate increase by one unit, the Unemp will increase by 1.59, suppose by common sense that its unit is percentage.

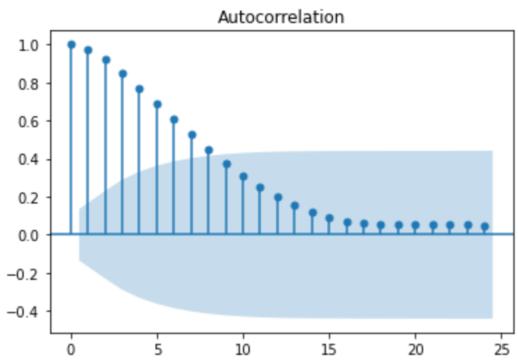
Question 2-7

Tester l'autocorrélation des erreurs.

We have the distribution of residues like this:



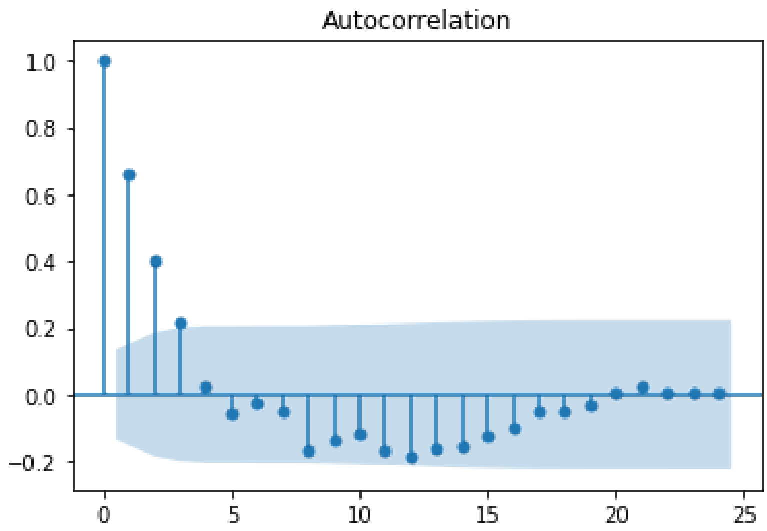
By doing the Durbin-Watson test, we have 0.04 which is close to 0. Thus, the autocorrelation of residues should be positive, and that's exactly what we saw.



Question 2-8

Corriger l'autocorrélation des erreurs par la méthode vue en cours.

To correct the autocorrelation of residues, we need to calculate the pho, which is 0.98 here. By calculating u' = u(t) – pho*u(t-1), we have then:



Question 2-9

Tester la stabilité de la relation chômage-inflation sur deux sous-périodes de taille identique (test de changement de structure avant et après la moitié de la période d'observation)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Unemp   R-squared:                       0.007
Model:                            OLS   Adj. R-squared:                 -0.003
Method:                 Least Squares   F-statistic:                    0.7071
Date:                Sun, 21 Nov 2021   Prob (F-statistic):              0.402
Time:                        01:27:01   Log-Likelihood:                -204.20
No. Observations:                 105   AIC:                             412.4
Df Residuals:                     103   BIC:                             417.7
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.9377      0.287     20.662      0.000       5.368       6.508
x1            15.6225     18.579      0.841      0.402     -21.224      52.469
==============================================================================
Omnibus:                        3.884   Durbin-Watson:                   0.059
Prob(Omnibus):                  0.143   Jarque-Bera (JB):                3.782
Skew:                           0.462   Prob(JB):                        0.151
Kurtosis:                       2.889   Cond. No.                         111.
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Unemp   R-squared:                       0.031
Model:                            OLS   Adj. R-squared:                  0.022
Method:                 Least Squares   F-statistic:                     3.315
Date:                Sun, 21 Nov 2021   Prob (F-statistic):             0.0715
Time:                        01:27:01   Log-Likelihood:                -193.35
No. Observations:                 106   AIC:                             390.7
Df Residuals:                     104   BIC:                             396.0
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          6.4269      0.259     24.822      0.000       5.913       6.940
x1           -54.4051     29.882     -1.821      0.072    -113.663       4.853
==============================================================================
Omnibus:                       12.639   Durbin-Watson:                   0.082
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               14.065
Skew:                           0.890   Prob(JB):                     0.000883
Kurtosis:                       3.121   Cond. No.                         203.
==============================================================================
```

The difference of the coefficient before and after the middle point of time is quite different, which changed from 15.6 to-54.4. We can then deduct that there is no stability in this relation.

Question 2-10

Estimer la courbe de Philips en supprimant l'inflation courante des variables explicatives mais en ajoutant les délais d'ordre 1, 2, 3 et 4 de l'inflation et du chômage. Faire le test de Granger de non causalité de l'inflation sur le chômage. Donnez la p-valeur.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Unemp   R-squared:                       0.979
Model:                            OLS   Adj. R-squared:                  0.978
Method:                 Least Squares   F-statistic:                     1145.
Date:                Sun, 21 Nov 2021   Prob (F-statistic):         2.80e-161
Time:                        02:13:03   Log-Likelihood:                 4.6497
No. Observations:                 207   AIC:                             8.701
Df Residuals:                     198   BIC:                             38.70
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.1457      0.072      2.014      0.045       0.003       0.288
x1             3.1105      3.759      0.827      0.409      -4.303      10.524
x2            -2.3578      4.089     -0.577      0.565     -10.421       5.706
x3             6.8949      3.989      1.729      0.085      -0.971      14.760
x4             1.6329      3.755      0.435      0.664      -5.773       9.039
x5             1.5937      0.071     22.383      0.000       1.453       1.734
x6            -0.6472      0.134     -4.832      0.000      -0.911      -0.383
x7             0.0222      0.135      0.164      0.870      -0.245       0.289
x8            -0.0080      0.070     -0.114      0.910      -0.146       0.130
==============================================================================
Omnibus:                       29.127   Durbin-Watson:                   1.997
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               68.886
Skew:                           0.625   Prob(JB):                     1.10e-15
Kurtosis:                       5.534   Cond. No.                     4.06e+03
==============================================================================
```

We have the first four variables delay of inflation, and the last four variables delay of unemployment.

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=1.7447  , p=0.1880  , df_denom=207, df_num=1
ssr based chi2 test:   chi2=1.7700  , p=0.1834  , df=1
likelihood ratio test: chi2=1.7626  , p=0.1843  , df=1
parameter F test:         F=1.7447  , p=0.1880  , df_denom=207, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=4.0000  , p=0.0198  , df_denom=204, df_num=2
ssr based chi2 test:   chi2=8.1960  , p=0.0166  , df=2
likelihood ratio test: chi2=8.0394  , p=0.0180  , df=2
parameter F test:         F=4.0000  , p=0.0198  , df_denom=204, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:         F=2.8197  , p=0.0401  , df_denom=201, df_num=3
ssr based chi2 test:   chi2=8.7537  , p=0.0328  , df=3
likelihood ratio test: chi2=8.5745  , p=0.0355  , df=3
parameter F test:         F=2.8197  , p=0.0401  , df_denom=201, df_num=3

Granger Causality
number of lags (no zero) 4
ssr based F test:         F=4.0061  , p=0.0038  , df_denom=198, df_num=4
ssr based chi2 test:   chi2=16.7526 , p=0.0022  , df=4
likelihood ratio test: chi2=16.1092 , p=0.0029  , df=4
parameter F test:         F=4.0061  , p=0.0038  , df_denom=198, df_num=4

Granger Causality
number of lags (no zero) 5
ssr based F test:         F=3.1204  , p=0.0098  , df_denom=195, df_num=5
ssr based chi2 test:   chi2=16.4820 , p=0.0056  , df=5
likelihood ratio test: chi2=15.8558 , p=0.0073  , df=5
parameter F test:         F=3.1204  , p=0.0098  , df_denom=195, df_num=5
```

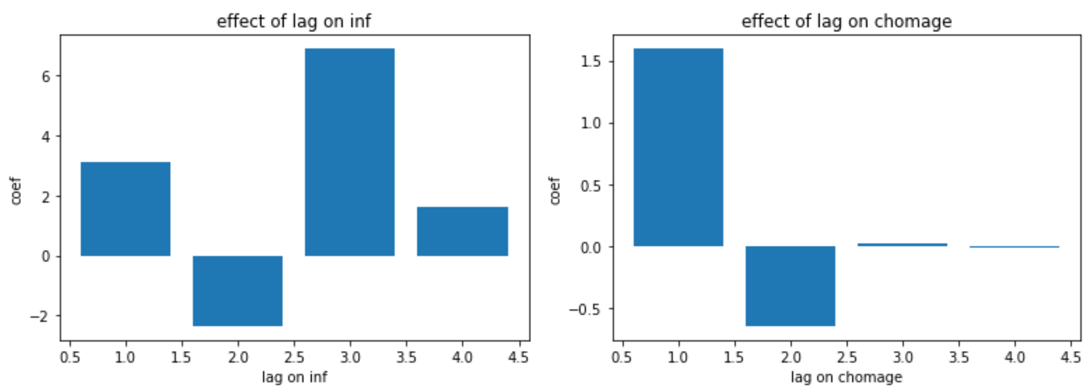
Except for the first one, lag=1, the p-values for the rest of them are comparatively small. Therefore, we deduct that the inflation can effectively influence the unemployment rate in the first year, but not for all the rest years.

Question 2-11

Représentez graphiquement les délais distribués et commentez. Calculer l'impact à long de terme de l'inflation sur le chômage.

The formula of delays can be expressed as:

$$Y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 x_{t-2} + \ldots + u_t$$



We can see that, the inflation rate will cause a raise in the next year, and s small drop in the second year, and then still rising-up. Which is to say that, the trend of inflation in inevitable.
For the effect of unemployment, it's similar to the inflation, but its influence is reduced significantly in two years. Although the overall trend of unemployment rate is going up, its velocity will be much slower than the inflation rate, in response to the obvious decrease of impact. However, in long term, the inflation and unemployment will continue to be a central problem for human society.
The effect of long term can be roughly calculated as the sum of coefficients, and that of inflation here is 9.28.

Question1-1

```
#import numpy as py
import pandas as pd

mroz = pd.read_fwf('../data/MROZ.txt', sep="/t", header=None)
mroz = mroz[mroz[6] != '.']
mroz[6] = pd.to_numeric(mroz[6])
mroz[20] = pd.to_numeric(mroz[20])
mroz = mroz[mroz[6] > 0]

inlf = mroz[0]          # =1 if in labor force, 1975
hours = mroz[1]    # hours worked, 1975
kidslt6 = mroz[2]                              # kids < 6 years
kidsge6 = mroz[3]                             # kids 6-18
age = mroz[4]                                        # woman's age in yrs
educ = mroz[5]    # years of schooling
wage = mroz[6]                                              # estimated wage from earns., hours
repwage = mroz[7]                                # reported wage at interview in 1976
hushrs = mroz[8]    # hours worked by husband, 1975
husage = mroz[9]    # husband's age
huseduc = mroz[10]    # husband's years of schooling
huswage = mroz[11]    # husband's hourly wage, 1975
faminc = mroz[12]    # family income, 1975
mtr = mroz[13]    # fed. marginal tax rate facing woman
motheduc = mroz[14]    # mother's years of schooling
fatheduc = mroz[15]    # father's years of schooling
unem = mroz[16]    # unem. rate in county of resid.
city = mroz[17]                                        # =1 if live in SMSA
exper = mroz[18]    # actual labor mkt exper
nwifeinc = mroz[19]    # (faminc - wage*hours)/1000
lwage = mroz[20]    # log(wage)
expersq = mroz[21]    # exper^2
```

Question 1-2

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
import dataloader_p1 as dt_p1
plt.subplot(131)
plt.hist(dt_p1.wage,'auto',color = 'purple')
plt.title('salaire')

plt.subplot(132)
```

```
plt.hist(dt_p1.age,'auto',color = 'purple')
plt.title('age')

plt.subplot(133)
plt.hist(dt_p1.educ,'auto',color = 'purple')
plt.title('education')
plt.show()

print('The mean is : ', np.mean(dt_p1.wage))
print('The median is : ', np.median(dt_p1.wage))
print('The maximum is : ', np.max(dt_p1.wage))
print('The minimum is : ', np.min(dt_p1.wage))
print('The standard deviation is : ', np.std(dt_p1.wage))
print('The variance is : ', np.var(dt_p1.wage))

print('The mean is : ', np.mean(dt_p1.age))
print('The median is : ', np.median(dt_p1.age))
print('The maximum is : ', np.max(dt_p1.age))
print('The minimum is : ', np.min(dt_p1.age))
print('The standard deviation is : ', np.std(dt_p1.age))
print('The variance is : ', np.var(dt_p1.age))

print('The mean is : ', np.mean(dt_p1.educ))
print('The median is : ', np.median(dt_p1.educ))
print('The maximum is : ', np.max(dt_p1.educ))
print('The minimum is : ', np.min(dt_p1.educ))
print('The standard deviation is : ', np.std(dt_p1.educ))
print('The variance is : ', np.var(dt_p1.educ))

mediane_e = np.median(dt_p1.huswage)

plt.subplot(131)
plt.hist(dt_p1.wage[dt_p1.huswage>=mediane_e],'auto',color = 'r')
plt.title('salaire')

plt.subplot(132)
plt.hist(dt_p1.age[dt_p1.huswage>=mediane_e],'auto',color = 'r')
plt.title('age')

plt.subplot(133)
plt.hist(dt_p1.educ[dt_p1.huswage>=mediane_e],'auto',color = 'r')
plt.title('education')
plt.show()
```

```
print('The mean is : ', np.mean(dt_p1.wage[dt_p1.huswage>=mediane_e]))
print('The median is : ', np.median(dt_p1.wage[dt_p1.huswage>=mediane_e]))
print('The maximum is : ', np.max(dt_p1.wage[dt_p1.huswage>=mediane_e]))
print('The minimum is : ', np.min(dt_p1.wage[dt_p1.huswage>=mediane_e]))
print('The standard deviation is : ', np.std(dt_p1.wage[dt_p1.huswage>=mediane_e]))
print('The variance is : ', np.var(dt_p1.wage[dt_p1.huswage>=mediane_e]))

print('The mean is : ', np.mean(dt_p1.age[dt_p1.huswage>=mediane_e]))
print('The median is : ', np.median(dt_p1.age[dt_p1.huswage>=mediane_e]))
print('The maximum is : ', np.max(dt_p1.age[dt_p1.huswage>=mediane_e]))
print('The minimum is : ', np.min(dt_p1.age[dt_p1.huswage>=mediane_e]))
print('The standard deviation is : ', np.std(dt_p1.age[dt_p1.huswage>=mediane_e]))
print('The variance is : ', np.var(dt_p1.age[dt_p1.huswage>=mediane_e]))

print('The mean is : ', np.mean(dt_p1.educ[dt_p1.huswage>=mediane_e]))
print('The median is : ', np.median(dt_p1.educ[dt_p1.huswage>=mediane_e]))
print('The maximum is : ', np.max(dt_p1.educ[dt_p1.huswage>=mediane_e]))
print('The minimum is : ', np.min(dt_p1.educ[dt_p1.huswage>=mediane_e]))
print('The standard deviation is : ', np.std(dt_p1.educ[dt_p1.huswage>=mediane_e]))
print('The variance is : ', np.var(dt_p1.educ[dt_p1.huswage>=mediane_e]))

plt.subplot(131)
plt.hist(dt_p1.wage[dt_p1.huswage<mediane_e],'auto',color = 'b')
plt.title('salary')

plt.subplot(132)
plt.hist(dt_p1.age[dt_p1.huswage<mediane_e],'auto',color = 'b')
plt.title('age')

plt.subplot(133)
plt.hist(dt_p1.educ[dt_p1.huswage<mediane_e],'auto',color = 'b')
plt.title('education')
plt.show()

print('The mean is : ', np.mean(dt_p1.wage[dt_p1.huswage<mediane_e]))
print('The median is : ', np.median(dt_p1.wage[dt_p1.huswage<mediane_e]))
print('The maximum is : ', np.max(dt_p1.wage[dt_p1.huswage<mediane_e]))
print('The minimum is : ', np.min(dt_p1.wage[dt_p1.huswage<mediane_e]))
print('The standard deviation is : ', np.std(dt_p1.wage[dt_p1.huswage<mediane_e]))
print('The variance is : ', np.var(dt_p1.wage[dt_p1.huswage<mediane_e]))

print('The mean is : ', np.mean(dt_p1.age[dt_p1.huswage<mediane_e]))
print('The median is : ', np.median(dt_p1.age[dt_p1.huswage<mediane_e]))
print('The maximum is : ', np.max(dt_p1.age[dt_p1.huswage<mediane_e]))
```

```
print('The minimum is : ', np.min(dt_p1.age[dt_p1.huswage<mediane_e]))
print('The standard deviation is : ', np.std(dt_p1.age[dt_p1.huswage<mediane_e]))
print('The variance is : ', np.var(dt_p1.age[dt_p1.huswage<mediane_e]))

print('The mean is : ', np.mean(dt_p1.educ[dt_p1.huswage<mediane_e]))
print('The median is : ', np.median(dt_p1.educ[dt_p1.huswage<mediane_e]))
print('The maximum is : ', np.max(dt_p1.educ[dt_p1.huswage<mediane_e]))
print('The minimum is : ', np.min(dt_p1.educ[dt_p1.huswage<mediane_e]))
print('The standard deviation is : ', np.std(dt_p1.educ[dt_p1.huswage<mediane_e]))
print('The variance is : ', np.var(dt_p1.educ[dt_p1.huswage<mediane_e]))
```

Question 1-3

```
plt.hist(dt_p1.wage,'auto')
plt.title('histogram of women salary')
plt.show()

log_wage = np.log(dt_p1.wage)
plt.hist(log_wage,'auto')
plt.title('histogram of women salary in log')
plt.show()
```

Question 1-4

```
corr = np.corrcoef(dt_p1.motheduc, dt_p1.fatheduc)
print(corr)
```

Question 1-5

```
plt.scatter(dt_p1.wage, dt_p1.educ)
plt.xlabel('wage')
plt.ylabel('education year')
plt.title('cloud map of wage & education')
plt.show()

plt.scatter(dt_p1.wage, dt_p1.exper)
plt.xlabel('wage')
plt.ylabel('work experience year')
plt.title('cloud map of wage & experience')
plt.show()

plt.scatter(dt_p1.wage, dt_p1.fatheduc)
plt.xlabel('wage')
plt.ylabel('father educcation year')
plt.title('cloud map of wage & father education')
plt.show()
```

Question 1-7

```
def linear_reg(X,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))
    return beta, u , sig2, Var, std

y = dt_p1.wage
const = np.ones(np.shape(y))
X=np.column_stack((const, dt_p1.city, dt_p1.educ,dt_p1.exper, dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6))

beta_7, u_7, sig2_7, Var_7, std_7 = linear_reg(X,y)

plt.hist(u_7,'auto')
plt.ylabel('cumulative number')
plt.xlabel('erreur')
plt.title('erreur')
plt.show()
```

Question 1-8

```
logy = dt_p1.lwage
const = np.ones(np.shape(logy))

X=np.column_stack((const, dt_p1.city, dt_p1.educ,dt_p1.exper, dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6))

beta_8, u_8, sig2_8, Var_8, std_8 = linear_reg(X,logy)

u_8.T@u_8

plt.hist(u_8,'auto')
plt.ylabel('cumulative number')
plt.xlabel('erreur')
plt.title('erreur')
plt.show()

std_8

import statsmodels.api as sm
model=sm.OLS(logy,X)
```

```
results = model.fit()
print(results.summary())

plt.figure(figsize=(7, 4))
sns.distplot(results.resid)
plt.title("Résidus du modèle", fontsize=14)
plt.show()
```

Question 1-9

```
n,k = np.shape(X)
test_9 = beta_8[4]/std_8[4]
2*t.sf(test_9,n-k)
test_9
seuil_signifi = [0.01, 0.05,0.1]

for i   in seuil_signifi:
    print(i,'left side:',t.ppf(i/2,n-k,loc = beta_8[4], scale = std_8[4]),'right side:',t.ppf(1-i/2,n-k,loc = beta_8[4], scale = std_8[4]))
```

Question 1-10

```
test_10 = (beta_8[4]-0.01)/std_8[4]
2*(1-t.sf(test_10,n-k))
```

Question 1-11

```
y = dt_p1.wage
const = np.ones(np.shape(y))
X0=np.column_stack((const, dt_p1.city, dt_p1.educ,dt_p1.exper, dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6))
y = y - 0.01*X0[:,4]-0.05*X0[:,1]
X = np.column_stack((const, dt_p1.educ,dt_p1.exper,dt_p1.kidslt6,dt_p1.kidsgt6))
beta_11, u_11, sig2_11, Var_11, std_11 = linear_reg(X,y)
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())

SSR_0 = u_7.T@u_7
SSR_11 = u_11.T@u_11
n,k=np.shape(X0)
F=((SSR_11-SSR_0)/2)/(SSR_0/(n-k))
f.sf(F,2,n-k)
```

Question 1-12

```
xxx_educ = np.linspace(5,20,100)
yyy_wage = xxx_educ*beta_7[2]+beta_7[0]
```

```
plt.scatter(dt_p1.educ,dt_p1.wage)
plt.plot(xxx_educ,yyy_wage, color = 'r')
plt.xlabel('education')
plt.ylabel('wage')
plt.title('variation of wage with education')

xxx_exper = np.linspace(0,40,100)
yyy_wage = xxx_exper*beta_7[3]+beta_7[0]
plt.scatter(dt_p1.exper,dt_p1.wage)
plt.plot(xxx_exper,yyy_wage, color = 'r')
plt.xlabel('experience professionnelle')
plt.ylabel('wage')
plt.title('variation of wage with experience professionnelle')

import numpy as np
import matplotlib.pyplot as mp
from mpl_toolkits.mplot3d import axes3d
mp.figure("3D Scatter", facecolor="lightgray")
ax3d = mp.gca(projection="3d")

mp.title('3D Scatter', fontsize=20)
ax3d.set_xlabel('education', fontsize=14)
ax3d.set_ylabel('experience', fontsize=14)
ax3d.set_zlabel('salary', fontsize=14)
mp.tick_params(labelsize=10)

zzz_wage = xxx_exper*beta_7[3]+beta_7[0]+xxx_educ*beta_7[2]
ax3d.scatter(dt_p1.educ, dt_p1.exper,dt_p1.wage, marker="o")
ax3d.plot(xxx_educ, xxx_exper, zzz_wage, color = 'r')

mp.show()
```

Question 1-13

```
y = dt_p1.wage
const = np.ones(np.shape(y))
#X0=np.column_stack((const, dt_p1.city, dt_p1.educ,dt_p1.exper, dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6))
X = np.column_stack((const, dt_p1.city, dt_p1.educ,dt_p1.exper,dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6-dt_p1.kidslt6))

model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
```

```
beta_13, u_13, sig2_13, Var_13, std_13 = linear_reg(X,y)
test_13 = beta_13[6]/std_13[6]
2*(1-t.sf(test_13,n -k))
```

Question 1-14

```
y = dt_p1.wage
const = np.ones(np.shape(y))
X=np.column_stack((const, dt_p1.city, dt_p1.educ,dt_p1.exper,
dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6))
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())

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

city2 = (dt_p1.city == 0)
y = dt_p1.wage
const = np.ones(np.shape(y))
X=np.column_stack((const, city2, (dt_p1.educ)**4, (dt_p1.expersq)**4,
dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6))
model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
u = results.resid
u2= u**2
y = u2
model = sm.OLS(y,X)
results  = model.fit()
print(results.summary())
```

Question 1-15

```
X0=np.column_stack((const, dt_p1.city, dt_p1.educ,dt_p1.exper,
dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6))
s_g = dt_p1.age>=43
y = dt_p1.lwage
X = X0[s_g,:]
y = y[s_g]
n,k = np.shape(X)
model_g = sm.OLS(y,X)
```

```
results_g = model_g.fit()
u_g = results_g.resid
SSR15_g = u_g.T@u_g


s_l = dt_p1.age<43
y = dt_p1.lwage
X = X0[s_l,:]
y = y[s_l]
n,k = np.shape(X)
model_l = sm.OLS(y,X)
results_l = model_l.fit()
u_l = results_l.resid
SSR15_l = u_l.T@u_l
print(results_g.summary())
print(results_l.summary())

F = (SSR_8- (SSR_G+SSR_15))/k/((SSR_G+SSR_15)/(n-2*k))
f.sf(F,2,n-k)
```

Question 1-16

```
X0=np.column_stack((const, dt_p1.city, dt_p1.educ,dt_p1.exper, dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6))
y = dt_p1.lwage
femegreateduc = (dt_p1.age>=43)*dt_p1.educ
femegreatexper = (dt_p1.age>=43)*dt_p1.exper
X = np.column_stack((const, femegreateduc, femegreatexper, dt_p1.city, dt_p1.educ,dt_p1.exper, dt_p1.nwifeinc,dt_p1.kidslt6,dt_p1.kidsgt6))
model = sm.OLS(y,X)
results   = model.fit()
print(results.summary())
SSR16_1 = results.resid.T@results.resid
SSR16_0 = u_8.T@u_8
n_16,k_16 = np.shape(X)
F = ((SSR16_0-SSR16_1)/2)/(SSR16_1/(n_16-k_16))
f.sf(F,2,n_16-k_16)
```

Question 2-1

```
import numpy as py
import pandas as pd

quarterly = pd.read_excel('../data/quarterly.xls')
```

Question 2-2

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from    scipy.stats import f
from scipy.stats import t
import statsmodels.api as sm

import dataloader_p2 as dt_p2

CPI = dt_p2.quarterly.CPI
year = pd.to_datetime(dt_p2.quarterly.DATE)
n = len(CPI)

inf = []
for i in range(1,n):
    inter = (CPI[i] - CPI[i-1])/CPI[i-1]
    inf.append(inter)

plt.plot(year[1:n],inf)
plt.xlabel('years')
plt.ylabel('taux')
plt.title('taux de CPI dans les annees')
plt.show()
```

Question 2-3

```
from statsmodels.tsa.stattools import acf
from statsmodels.tsa.stattools import pacf
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.graphics.tsaplots import plot_pacf
plot_acf(np.array(inf))
plt.show()
plot_pacf(np.array(inf))
plt.show()
```

Question 2-5

```
import statsmodels.tsa.api as smt
minf = smt.AR(inf).fit(maxlag = 4)
```

```
print(minf.params)
print(minf.aic)
print(minf.bic)
```

Question 2-6

```
y = dt_p2.quarterly.Unemp[1:]
const = np.ones(np.shape(y))
X = np.column_stack((const,np.array(inf)*100))

model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
```

Question 2-7

```
u = results.resid
plt.plot(u)
import statsmodels
statsmodels.stats.stattools.durbin_watson(u, axis=0)
acf(u)
plot_acf(np.array(u))
plt.show()
```

Question 2-8

```
rho = results1.params[0]
transform_u = np.array(u[1:n])-np.array(rho*u_1)
```

Question 2-9

```
n = len(dt_p2.quarterly.Unemp)
y = dt_p2.quarterly.Unemp[1:int((n/2))]
const = np.ones(np.shape(y))
X = np.column_stack((const,np.array(inf[:int((n/2))-1])))

model=sm.OLS(y,X)
results = model.fit()
print(results.summary())

n = len(dt_p2.quarterly.Unemp)
y = dt_p2.quarterly.Unemp[int(n/2):]
const = np.ones(np.shape(y))
X = np.column_stack((const,np.array(inf[int(n/2)-1:])))

model=sm.OLS(y,X)
results = model.fit()
print(results.summary())
```

Question 2-10

```
n =len(dt_p2.quarterly.Unemp)
y = dt_p2.quarterly.Unemp[5:n]
const = np.ones(np.shape(y))
inf_1 = inf[3:n-2]
inf_2 = inf[2:n-3]
inf_3 = inf[1:n-4]
inf_4 = inf[0:n-5]
cho_1 = dt_p2.quarterly.Unemp[4:n-1]
cho_2 = dt_p2.quarterly.Unemp[3:n-2]
cho_3 = dt_p2.quarterly.Unemp[2:n-3]
cho_4 = dt_p2.quarterly.Unemp[1:n-4]

X = np.column_stack((const,np.array(inf_1),np.array(inf_2),np.array(inf_3),np.array(inf_4),cho_1,cho_2,cho_3,cho_4))
X = np.column_stack((const,np.array(inf_1)*100,np.array(inf_2)*100,np.array(inf_3)*100,np.array(inf_4)*100,cho_1,cho_2,cho_3,cho_4))

model=sm.OLS(y,X)
results = model.fit()
print(results.summary())

X = np.column_stack((dt_p2.quarterly.Unemp[0:n-1],np.array(inf)*100))
statsmodels.tsa.stattools.grangercausalitytests(X,5)
```

Question 2-11

```
d_inf = (results.params[1],results.params[2],results.params[3],results.params[4])
x = (1,2,3,4)
plt.bar(x,d_inf)
plt.title('effect of lag on inf')
plt.xlabel('lag on inf')
plt.ylabel('coef')
plt.show()

d_cho = (results.params[5],results.params[6],results.params[7],results.params[8])
x = (1,2,3,4)
plt.bar(x,d_cho)
plt.title('effect of lag on chomage')
plt.xlabel('lag on chomage')
plt.ylabel('coef')
plt.show()
```

```
results.params[1]+results.params[2]+results.params[3]+results.params[4]
```