```
*CORRIGE DE L'EXERCICE CHAPITRE 5 - EXERCICE 4
. *Dalila Chenaf-Nicet  Université de Bordeaux.
.
. * Tout d'abord s'assurer d'avoir ouvert le fichier de données stata C5EX4 avant d'ouvrir
le do.fileC5EX4 qui est fichier programme.
. * Une fois le fichier de données ouvert ainsi que le do-file apppuyer sur Run (Execute en
haut à droite de la barre de menu du dofile)
. * pour démarrer
.
.
. * Toutefois il est possible en sélectionnant les parties du programme de l'exécuter pas à
pas afin de voir apparaître pas à pas les
. * différents résultats.
.
.
. *         CHAPITRE 5 EXERCICE 4
. *Test de détection de l'hétéroscedasticité
.
. * Il faut tout d'abord définir les 6 groupes
. egen m=group(x)


.
. * Comme les données sont constituées pour m=6 groupes d'observations il faut calculer la
variance empirique pour chacun groupe (i=1,..6)
.
. gsort -m

. bysort m: egen sdi=sd(y)

. bysort m : gen vari=sd^2


.
. * On peut faire un tableau noté V avec les 6 variances de chaque groupe
. mkmat vari, mat(variancei)

. matlist variancei

             |      vari
-------------+-----------
          r1 |     127.3
          r2 |     127.3
          r3 |     127.3
          r4 |     127.3
          r5 |     127.3
          r6 |  78.69999
          r7 |  78.69999
          r8 |  78.69999
          r9 |  78.69999
         r10 |  78.69999
         r11 |      63.7
         r12 |      63.7
         r13 |      63.7
         r14 |      63.7
         r15 |      63.7
         r16 |      18.8
         r17 |      18.8
         r18 |      18.8
         r19 |      18.8
         r20 |      18.8
         r21 |      17.8
         r22 |      17.8
         r23 |      17.8
         r24 |      17.8
         r25 |      17.8
         r26 |       3.2
         r27 |       3.2
         r28 |       3.2
         r29 |       3.2
         r30 |       3.2

.
```

```
. matrix var1=variancei[1,1]

. matrix var6=variancei[6,1]

. matrix var11=variancei[11,1]

. matrix var16=variancei[16,1]

. matrix var21=variancei[21,1]

. matrix var26=variancei[26,1]

.
. matrix V=[var1\var6\var11\var16\var21\var26]

. matlist V

             |        c1
-------------+-----------
          r1 |     127.3
          r1 |  78.69999
          r1 |      63.7
          r1 |      18.8
          r1 |      17.8
          r1 |       3.2

.
. * Calculer la variance totale à partir de la matrice V qui comporte les 6 variances des
groupes à partir de la
. *formule de l'ouvrage (en décomposant le numérateur "num" du dénominateur "den")
.
. svmat V, names(V)

. gen num=(4*V1)
(24 missing values generated)

. egen variancet=sum(num)

. gen den=(24)

. gen variancetotale =variancet/den

. display variancetotale
51.583328

.
. ** Les étapes du calcul du khi2
.
. *vi= ni-1 = 4 et la somme des vi= m*(ni-1) = 24 =den
.
. * la quantité à calculer selon la formule de l'ouvrage nécessite de calculer logvariance
totale et des 6 variances par groupe
. gen vlnvarT=den*log(variancetotale)

. gen vlnvari=4*log(V1)
(24 missing values generated)

. egen sommevlnvari=sum(vlnvari)

.
. * Le calcul de  Qprime
. gen Qprime=vlnvarT-sommevlnvari

. display Qprime
13.266449

.
. * Pour affiner le calcul on peut calculer la grandeur C pour Q avec Q=Qprime/C
. gen C=1 +[1/(3*(6-1))]*[6*(1/4)-(1/4)]

. display C
```

```
1.0833334

.
. gen Q=Qprime/C

. display Q
12.245953

.
. * Q est supérieur à 11,07 lu sur la table pour un risque de 1ere espéce de 5% et 6ddl. Le
modèle est donc hétéroscedastique
.
. **************************** On peut aussi faire le test de Golgfeld-Quandt
.
. * Etape 1 : il faut ordonner les observations en fonction des variables croissantes ou
décroissante de la variable à expliquer y.
. * Or elles sont déjà classées par ordre croissant
. * On choisit D observations situées au centre de l'échantillon en règle générale  D= _N/4.
En prenant la partie entière de la division.
.
. gen D=_N/4

. display D
7.5

.
. * On choisit donc D=8. On exlut donc les 8 valeurs centrales de la série des observations
. * On conserve donc deux échantillons le premier de la 1er à la 11ème observations et le
deuxième de la 20ème à la 30ème
.
.
. * Faire les régressions pour les deux échantillons et récupérer les résidus
. regress y x in 1/11

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =      0.02
       Model |        1.98         1        1.98   Prob > F        =    0.8895
    Residual |      872.02         9  96.8911111   R-squared       =    0.0023
-------------+----------------------------------   Adj R-squared   =   -0.1086
       Total |         874        10        87.4   Root MSE        =    9.8433

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |      -1.32    9.23386    -0.14   0.889    -22.20844    19.56844
       _cons |      23.08   8.117016     2.84   0.019     4.718033    41.44197
------------------------------------------------------------------------------

. gen rss1=e(rss)

. display rss1
872.02002

.
. regress y x in 20/30

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =      0.85
       Model |  15.5151515         1  15.5151515   Prob > F        =    0.3811
    Residual |  164.666667         9  18.2962963   R-squared       =    0.0861
-------------+----------------------------------   Adj R-squared   =   -0.0154
       Total |  180.181818        10  18.0181818   Root MSE        =    4.2774

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  -2.133333    2.31666    -0.92   0.381    -7.373982    3.107316
       _cons |   16.93333   8.418293     2.01   0.075    -2.110168    35.97683
------------------------------------------------------------------------------

. gen rss2=e(rss)
```

```
. display rss2
164.66667

.
. * Calculer le F avec ddl =n-2=11_2=9
. gen F=(rss1/9)/(rss2/9)

. display F
5.2956681

.
.  * F est supérieur à 3.18 lu sur la table pour un risque de 1ere espéce de 5% et (9,9)ddl.
Le modèle est donc hétéroscedastique
.
. * ******************************On peut faire le Test de Gleisjer
.
. * Etape 1: on régresse le modèle complet et on récupére le vecteur des résidus
. regress y x

      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 28)        =     17.64
       Model |  826.095708         1  826.095708   Prob > F        =    0.0002
    Residual |  1311.40429        28  46.8358676   R-squared       =    0.3865
-------------+----------------------------------   Adj R-squared   =    0.3646
       Total |      2137.5        29  73.7068966   Root MSE        =    6.8437

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  -4.125322   .9822724    -4.20   0.000    -6.137416   -2.113228
       _cons |   24.09442   2.397697    10.05   0.000     19.18296    29.00588
------------------------------------------------------------------------------

.  predict ei, re

.
.  * On régresse ensuite la valeur absolue des résidus par rapport à X
.
.  gen absei=abs(ei)

.  regress absei x

      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 28)        =      6.55
       Model |  104.501976         1  104.501976   Prob > F        =    0.0162
    Residual |  446.860339        28  15.9592978   R-squared       =    0.1895
-------------+----------------------------------   Adj R-squared   =    0.1606
       Total |  551.362315        29  19.0124936   Root MSE        =    3.9949

------------------------------------------------------------------------------
       absei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  -1.467253   .5733893    -2.56   0.016    -2.641787    -.292718
       _cons |   8.090138   1.399626     5.78   0.000     5.223135    10.95714
------------------------------------------------------------------------------

.
.  * On régresse ensuite la valeur absolue des résidus par rapport à X2
.  gen x2=sqrt(x)

.  regress absei x2

      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 28)        =      6.76
       Model |  107.281344         1  107.281344   Prob > F        =    0.0147
    Residual |   444.08097        28  15.8600347   R-squared       =    0.1946
-------------+----------------------------------   Adj R-squared   =    0.1658
       Total |  551.362315        29  19.0124936   Root MSE        =    3.9825

------------------------------------------------------------------------------
```

```
       absei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |  -4.147867   1.594832    -2.60   0.015    -7.414732   -.881003
       _cons |   10.71379   2.301941     4.65   0.000      5.99848   15.42911
------------------------------------------------------------------------------


.
.  * Les t-students empiriques des deux relations indiquent que les coefficients sont
significativement différentes de zéro.
.  * L'hétéroscedasticité est détectée.
.
.  ************************************* On peut faire le Test de White
.
.  * Il est fondé sur une relation significative entre le carré des résidus de la régression
y/ x et les x au carré
.  * Ici nous n'avons qu'une seule variable explicative. on pourra recourir à la statistique
LM distribuée comme un Khi2 ou F.
.
.  regress y x

      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 28)        =     17.64
       Model |  826.095708         1  826.095708   Prob > F        =    0.0002
    Residual |  1311.40429        28  46.8358676   R-squared       =    0.3865
-------------+----------------------------------   Adj R-squared   =    0.3646
       Total |      2137.5        29  73.7068966   Root MSE        =    6.8437

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  -4.125322   .9822724    -4.20   0.000    -6.137416   -2.113228
       _cons |   24.09442   2.397697    10.05   0.000     19.18296    29.00588
------------------------------------------------------------------------------

.  predict re, re

.  gen re2=re^2

.
.  * Soit ici le modèle testé
.  regress re2 x x2

      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(2, 27)        =      4.04
       Model |  31983.0303         2  15991.5151   Prob > F        =    0.0292
    Residual |  106848.104        27   3957.3372   R-squared       =    0.2304
-------------+----------------------------------   Adj R-squared   =    0.1734
       Total |  138831.135        29   4787.2805   Root MSE        =    62.907

------------------------------------------------------------------------------
         re2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   59.06827    67.6917     0.87   0.391    -79.82363    197.9602
          x2 |  -231.4947   188.8668    -1.23   0.231    -619.0173    156.0279
       _cons |   237.6825   120.9112     1.97   0.060    -10.40673    485.7717
------------------------------------------------------------------------------

.   gen R2=e(r2)

.   display R2
.23037362

.
.   gen LM=_N*R2

.   display LM
6.9112086

.
.  *La valeur de LM calculée est supérieur au khi2 de la table (5.99) pour un risque de 1ere
espece de 6 % et p=2*k ddl
```

```
.  *L'hétéroscedasticité est détectée
.
.  ****************************** Correction de l'hétéroscédasticité
.  * On crée un variable Z=(Y/sqrt(X)) , X1=1/(sqrt(X)) et  X2=X/(sqrt(X))
.
.  gen Z=y/sqrt(x)

.  gen  X1=1/sqrt(x)

.  gen X2=x/sqrt(x)

.
.  * On réalise la régression de ces variables
.  regress Z X1 X2, noconstant

      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(2, 28)        =     79.92
       Model |  9127.90514         2  4563.95257   Prob > F        =    0.0000
    Residual |  1599.01168        28  57.1075601   R-squared       =    0.8509
-------------+----------------------------------   Adj R-squared   =    0.8403
       Total |  10726.9168        30  357.563894   Root MSE        =     7.557

------------------------------------------------------------------------------
           Z |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          X1 |   24.94147   2.503469     9.96   0.000     19.81335    30.06959
          X2 |  -4.531905   1.535487    -2.95   0.006    -7.677207   -1.386603
------------------------------------------------------------------------------

.
.  * Le modèle estimé est donc Y= 24,96 - 4.53 *X +e
.
.  * Il existe bien une influence significative du temps de vérification sur le nombre de
défauts constatés.
.  * Chaque heure de vérification permet de supprimer en moyenne 4,5 défauts
.
end of do-file

. exit, clear
```