I EXERCICE-1

1. Une des hypothèses de la méthode MCO est homoscédasticité du terme d'erreur La variance $V(\varepsilon_i) = E\left((\varepsilon_i - E(\varepsilon_i))^2\right)$ est constante pour tout i, soit $V(\varepsilon_i) = \sigma_\varepsilon^2$. Cette hypothèse de variance constante est l'hypothèse d'homoscédasticité ; on parle alors de série homoscédastique; par opposition si le terme d'erreur n'a pas une cvariance constante, on parle de terme d'erreur hétéroscédastique.
2. On donne l'écriture d'un modèle sous la forme : $Y = \widehat{a_0} + \widehat{a_1}X_1 + \widehat{a_2}X_2 + \widehat{a_3}X_3 + \varepsilon$, Y étant la variable expliquée et les X_i trois variables explicatives.
   a. $\widehat{a_0}$ est la valeur estimée de Y pour $X_i = 0$. $\widehat{a_1}$, $\widehat{a_2}$ et $\widehat{a_3}$ donnent respectivement une estimation de la variation de Y quand une des variables explicatives X_1, X_2 ou X_3 augmente d'une unité, les autres restant constantes.
   b. Rappel de cours :

      ε étant le **terme d'erreur** regroupe en fait trois types d'erreur :

      Une erreur de spécification :

      On ne peut expliquer entièrement la variable Y par les variables explicatives retenues ; de nombreuses autres variables non prises en compte ont une influence Y ; elles ont été omises du fait de leur faible influence sur Y. Donc une part de l'erreur est dûe aux variables omises.

      Une erreur de mesure :

      les données ne représentent pas pleinement le phénomène,

      Une erreur d'échantillonnage :

      les observations variant d'un échantillon à l'autre engendrant les fluctuations de Y autour de sa moyenne,
3. $F = \dfrac{SCE/k}{SCR/(n-k-1)} = \dfrac{R^2/k}{(1-R^2)/(n-k-1)}$. Si les résidus sont normalement distribués, sous l'hypothèse nulle $H_0 : a_1 = a_2 = ... = a_k = 0$, la variable aléatoire F suit la loi de Fisher avec pour degrés de liberté k et $n-k-1$.

   Le ratio F fournit un test d'hypothèse permettant de tester l'hypothèse nulle H_0.

   Régle de décision :

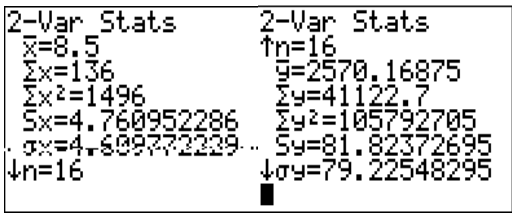
   On choisit un seuil de signification, puis :

   Si $F \leq F_{(k;n-k-1)}$, on ne rejette pas H_0 ; aucune variable n'est significative.

   Si $F > F_{(k;n-k-1)}$, on rejette H_0, on accepte l'hypothèse que les paramètre s ne sont pas tous nuls et que R^2 significativement de 0, il existe au moins une variable significative.
4. Estimateur BLUE : sans biais et de variance minimale et linéaire. Exemple : modèle de régression simple $\widehat{a}_1 = \dfrac{\sum_{i=1}^{i=n} x_i y_i}{\sum_{i=1}^{i=n} x_i^2}$; en posant $w_i = \dfrac{x_i}{\sum_{i=1}^{i=n} x_i^2}$, $\widehat{a}_1 = \sum_{i=1}^{i=n} w_i Y_i$.
5. Théorème de Gauss-Markov.Si les hypothèses de la MCO sont vérifiées, les estimateurs $\widehat{a}_1$ et $\widehat{a}_0$ **sont** $BLUE$ (Best Linear Unbiased Estimator).

II EXERCICE-2

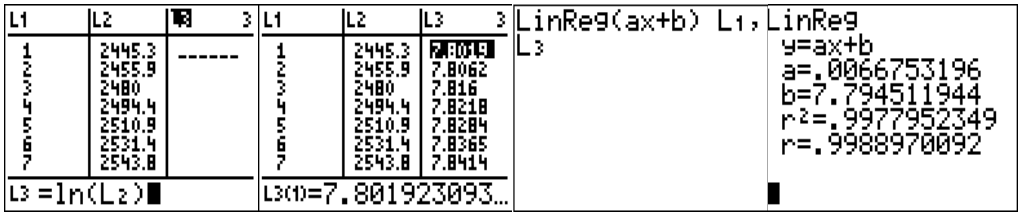
1. $Y_t = Y_0(1+r)^t \Longleftrightarrow \ln Y_t = \ln Y_0(1+r)^t = \ln Y_0 + t\ln(1+r)$; si on pose : $Z_t = \ln Y_t$, $B = \ln Y_0$ et $A = \ln(1+r)$, on a : $Z_t = At + B$, soit une relation linéaire entre t et $\ln Y_t$.
2. La calculatrice :
   a.



   On obtient : $\overline{t} = 8.5$ et $\overline{Y} = 2570.17$.

CORRIGE EXAMEN ECONOMETRIE

b.



On effectue la regression linéaire de $Ln\left(Y_t\right)$ en t, ce qui donne : $\widehat{Ln\left(Y_t\right)} = \widehat{a_0} + \widehat{a_1}t$, avec : $\widehat{a_0} \simeq 7.7945$ et $\widehat{a_1} \simeq 0.0067$.

c. $\widehat{a_0}$ donne l'estimation de la valeur de $\widehat{Ln\left(Y_t\right)}$ pour $t = 0$, donc au dernier trimestre 1992, soit : $\widehat{Y_0} \simeq e^{7.7945} \simeq 2427.22$.

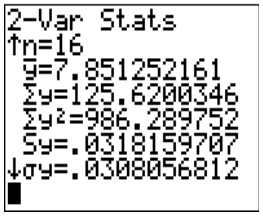
d. Pour interpréter $\widehat{a_1}$ on dérive : $\widehat{Ln\left(Y_t\right)} = \widehat{a_0} + \widehat{a_1}t$ donne $\dfrac{Y_t'}{Y_t} = \widehat{a_1}$ soit $\dfrac{dY_t}{dt} * \dfrac{1}{Y_t} = \widehat{a_1}$ soit $\widehat{a_1} = \left(\dfrac{dY_t}{Y_t}\right)/dt$; c'est le quotient de la variation relative de Y sur la variation absolue de t ; si $dt = 1$, $\widehat{a_1} = \dfrac{dY_t}{Y_t}$ donne une estimation de la variation relative de Y_t, soit de Y_t à Y_{t+1} une variation estimée à une augmentation de 0.67%.

e.

Source de variation	ddl	Somme des carrés	Moyenne des carrés (variances)
Régression	$\mathbf{1}$	$\mathbf{SCE} = \sum\left(\widehat{Z}_t - \overline{Z}\right)^2$	$\mathbf{SCE/1}$
Résiduelle	$\mathbf{n-2}$	$\mathbf{SCR} = \sum\left(\widehat{Z}_t - \mathbf{Z_t}\right)^2$	$\mathbf{SCR}/\left(n-2\right)$
Totale	$\mathbf{n-1}$	$\mathbf{SCT} = \sum\left(Z_t - \overline{Z}\right)^2$	

	Degré de liberté	*Somme des carrés*	*Moyenne des carrés*
Régression	**1**	**0.01515**	**0.01515**
Résidus	**14**	**0.00003**	**0.00000**
Total	**15**	**0.01518**	

La calculatrice donne les paramètres de $Z = \ln Y$:



et on a : $SCT = \sum\left(Z_t - \overline{Z}\right)^2 = nV\left(Z_t\right) = 16 * 0.030806^2 \simeq 0.015\,18$; par ailleurs : $\widehat{Z}_t - \overline{Z} = \widehat{a_1}\left(t - \overline{t}\right)$ et $\sum\left(\widehat{Z}_t - \overline{Z}\right)^2 = \widehat{a_1}^2\sum\left(t - \overline{t}\right)^2 = nV\left(t\right)\widehat{a_1}^2$ soit $SCE = 16 * 0.00667^2 * 4.60977^2 \simeq 0.015\,13$ et par différence : $SCR = SCT - SCE \simeq 0.015\,18 - 0.015\,13 = 0.000\,05$

Les degrés de liberté :

Pour SCT, nous devons estimer la moyenne $\overline{Y}$, donc le degré de liberté est $n-1$, soit ici 15 ; pour SCR, il faut estimer $\widehat{a}_0$ et $\widehat{a}_1$, ce qui constitue une perte de deux degrés de liberté donc un degré de liberté de $n-2$, soit 14 et enfin pour SCE, il suffit d'estimer $\widehat{a}_1$ car $V\left(\widehat{Y} - \overline{Y}\right) = \widehat{a_1}^2 V\left(X\right)$, le degré de liberté est 1.

f. L'erreur type, est donnée par : $S_\varepsilon^2 = \dfrac{\sum\limits_{i=1}^{i=n} e_i^2}{n-2} \simeq \dfrac{0.00005}{14} = 3.57 \times 10^{-6}$ et $S_\varepsilon \simeq \sqrt{\dfrac{0.00005}{14}} \simeq 0.0019$ (Excel donne 0.0015).

g. $I = \left[\widehat{a}_1 - t_{\alpha/2;\nu}S_{\widehat{a}_1}\ ;\ \widehat{a}_1 + t_{\alpha/2;\nu}S_{\widehat{a}_1}\right]$, avec : $S_{\widehat{a}_1}^2 = \dfrac{S_\varepsilon^2}{\sum\limits_{i=1}^{i=n} x_i^2} = \dfrac{S_\varepsilon^2}{nV\left(t\right)} \simeq \dfrac{3.57 \times 10^{-6}}{340} = 0.000\,000\,010\,5$ soit

$S_{\widehat{a}_1} \simeq \sqrt{\dfrac{3.57 \times 10^{-6}}{340}} = 1.024\,695 \times 10^{-4}$ soit $I = \left[\widehat{a}_1 - t_{\alpha/2;\nu}S_{\widehat{a}_1}\ ;\ \widehat{a}_1 + t_{\alpha/2;\nu}S_{\widehat{a}_1}\right]$

soit $I = \left[0.006675 - t_{\alpha/2;\nu}S_{\widehat{a}_1}\ ;\ 0.006675 + t_{\alpha/2;\nu}S_{\widehat{a}_1}\right]$.

$t_{0.025;14} = 2.145$; on obtient : $0.006675 - 2.145 * 1.024\,695 \times 10^{-4} = 0.006\,455$ et

$0.006675 + 2.145 * 1.024\,695 \times 10^{-4} = 0.006895$, soit $\boxed{I = [0.006\,455\ ;\ 0.006895]}$

CORRIGE EXAMEN ECONOMETRIE

h. Le premier trimestre 97 correspond à $t = 17$; on a pour estimation ponctuelle : $\widehat{Ln\,(Y_t)} = 7.7945 + 0.0067t$, soit $\widehat{Ln\,(Y_{17})} = 7.7945 + 0.0067 * 17 = 7.908\,4$ et $\widehat{Y_{17}} \simeq e^{7.908\,4} = 2720.03$, de plus $S_f^2 = S_\epsilon^2 \left(1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{i=n} x_i^2} \right)$ soit $S_f^2 = 3.57 \times 10^{-6} \left(1 + \frac{1}{16} + \frac{(17-8.5)^2}{16 * 4.60977^2} \right) \simeq 4.55 \times 10^{-6}$, soit $S_f \simeq \sqrt{4.55 \times 10^{-6}} = 2.1331 \times 10^{-3}$.
$\frac{f}{S_f}$ suit une loi de Student avec un ddl de $(n-2)$, soit 14, ce qui donne pour Z_t, l'intervalle : $7.908\,4 - 2.145 * 2.1331 \times 10^{-3} = 7.\,903\,8$ et $7.908\,4 + 2.145 * 2.1331 \times 10^{-3} = 7.912\,98$ soit pour Y_t : $e^{7.\,903\,8} = 2707.55$ et $e^{7.912\,98} = 2732.52$ soit $\boxed{I = [2707.55\ ;\ 2732.52]}$

III EXERCICE-3

Un chercheur a mené une étude sur le salaire horaire Y d'ouvriers, à partir d'un échantillon comportant 250 hommes et 280 femmes.
On considère la variable indicatrice définie par : $M = \begin{cases} 1 \text{ si l'ouvrier est un homme} \\ 0 \text{ s'il s'agit d'une femme.} \end{cases}$

1. a. Il a obtenu pour la régression :
$\begin{cases} \widehat{a_1} = 2.12 & S_{\widehat{a_1}} = 0.36 \\ \widehat{a_0} = 12.52 & S_{\widehat{a_0}} = 0.23 \end{cases}$ que l'on note : $\boxed{\widehat{Y} = 2.12\ M + 12.52}$ avec $S_\varepsilon \simeq 4.2$.

2. $M = 0$, donne une estimation du salaire horaire moyen des femmes, soit $\widehat{a_0} = 12.52$ et $M = 1$, donne une estimation du salaire horaire moyen des hommes, soit $\widehat{a_0} + \widehat{a_1} = 2.12 + 12.52 = 14.64$, $\widehat{a_1}$ représentant donc la différence entre le salaire moyen des hommes et celui des femmes.

3. La différence de salaire dans les deux sous échantillons (hommes et femmes) est $\widehat{a_1}$. Il nous faut faire un test sur $\widehat{a_1}$; $\begin{cases} H_0 : a_1 = 0 \\ H_1 : a_1 \neq 0 \end{cases}$
. Sous l'hypothèse H_0, $t_{\widehat{a}_1} = \dfrac{\widehat{a}_1}{S_{\widehat{a}_1}}$, le ratio de Student, suit une distribution de Student avec $(n-2)$ degrés de liberté, soit ici 528 ; ici $\dfrac{\widehat{a}_1}{S_{\widehat{a}_1}} = \dfrac{2.12}{0.36} = 5.89$; il reste à comparer ce quotient avec la valeur lue dans la table de Student, soit ici 2, car $ddl > 20$. La valeur du quotient $\dfrac{\widehat{a}_1}{S_{\widehat{a}_1}}$ est supérieure à 2, on en déduit donc que l'on rejette l'hypothèse H_0 et donc $\widehat{a}_1$ suffisamment différent de zéro pour affirmer que a_1 est significativement différent de zéro. La différence de salaire dans les deux sous échantillons est différente de zéro au niveau de 5%.

4. L'autre chercheur : $\boxed{\widehat{Y} = \widehat{\beta}_1\ M + \widehat{\beta}_0}$. On a ici $\widehat{\beta}_0$ qui est une estimation du salaire des hommes ($M = 0$, donc $\widehat{\beta}_0 = 14.64$ et $\widehat{\beta}_1 + \widehat{\beta}_0$ une estimation du salaire des femmes donc $\widehat{\beta}_1 = 12.52 - 14.64 = -2.12$.