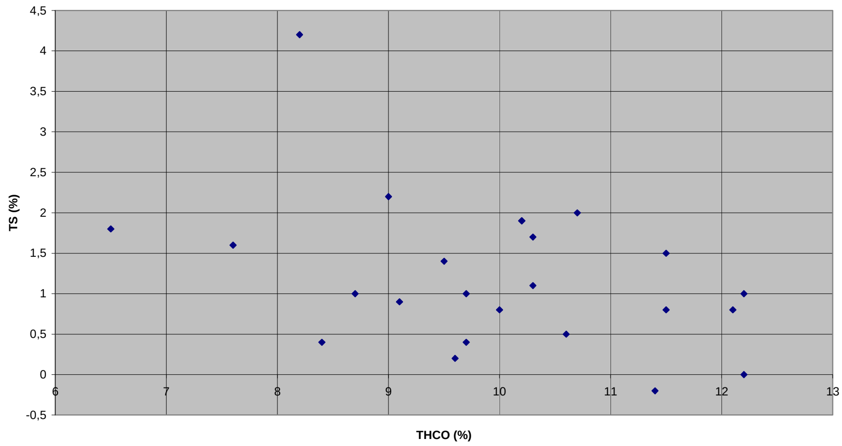
UNIVERSITE DE PARIS X **Année universitaire 2008-2009**
UFR SEGMI L2 Economie & Gestion ***Travaux dirigés***
Statistiques Economiques ***Fascicule 3***
N. CHEZE et D. ABECASSIS **Exercices repris ou adaptés de G. NEUBERG**

RÉGRESSION

Exercice 1

Graphique 1: nuages de points (THCO, TS)



Année	THCO	TS	Année	THCO	TS
1980	6,5	1,8	1992	10,3	1,7
1981	7,6	1,6	1993	11,4	-0,2
1982	8,2	4,2	1994	12,2	0
1983	8,4	0,4	1995	11,5	0,8
1984	9,6	0,2	1996	12,1	0,8
1985	10,3	1,1	1997	12,2	1
1986	10,2	1,9	1998	11,5	1,5
1987	10,6	0,5	1999	10,7	2
1988	10,2	1,9	2000	9,7	1
1989	9,7	0,4	2001	8,7	1
1990	9	2,2	2002	9,1	0,9
1991	9,5	1,4	2003	10	0,8

Le tableau ci-dessus présente le *taux de chômage* (en %) en France, noté *TCHO*, et le *taux de croissance annuel réel* (en % ; salaire horaire du secteur marchand), noté *TS*, sur la période 1971-1994. Sur le graphique 1, on a reporté les points de coordonnées $(TCHO_t, TS_t)$. Bien que le nuage de points ait une grande dispersion, on observe en première approximation une baisse du taux de croissance du salaire réel lorsque le taux de chômage augmente. On envisage de résumer le nuage de points à l'aide du modèle : $TS = a\,TCHO + b$.

1.- Quelle est la variable expliquée? Quelle est la variable explicative ? En termes économiques, que représentent les coefficients a et b et pourquoi sont-ils respectivement négatif et positif ?

2.- Dans la suite, on désigne par y le taux *TS* et par x le taux *TCHO*. Les calculs donnent

$$\sum_{t=1}^{24} x_t = 239{,}2\ ;\ \sum_{t=1}^{24} y_t = 28{,}9\ ;\ \sum_{t=1}^{24} x_t^2 = 2433{,}66\ ;\ \sum_{t=1}^{24} x_t y_t = 275{,}5\ ;\ \sum_{t=1}^{24} y_t^2 = 54{,}18.$$

Calculer $\bar{x}$, $\bar{y}$, $V(y)$ et $Cov(x,y)$.

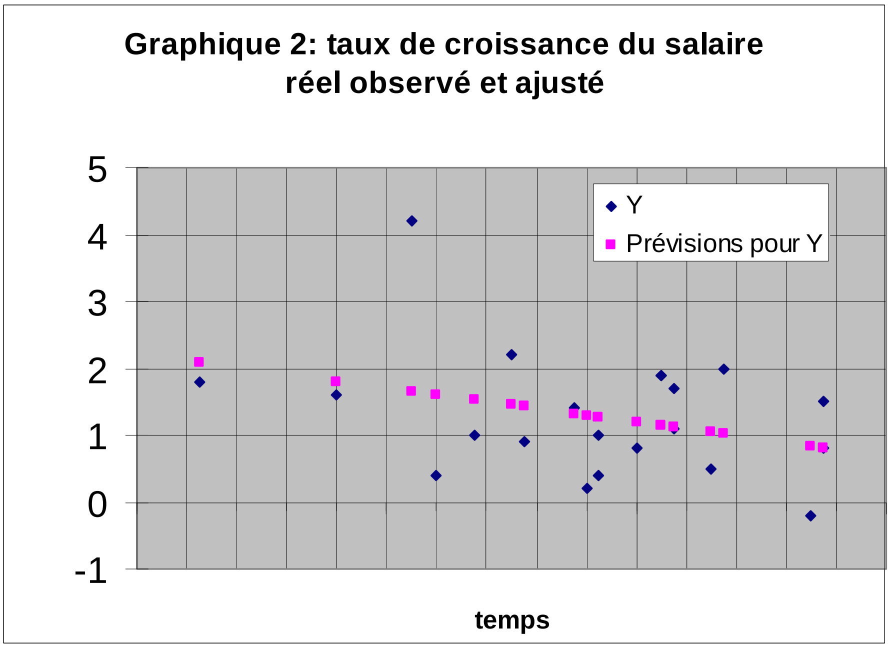
3.- Calculer les estimations $\widehat{a}$ et $\widehat{b}$ des coefficients a et b telles qu'obtenues par la méthode des moindres carrés. Interpréter les valeurs numériques trouvées.

4.- Si le taux de chômage et le taux de croissance annuel du salaire réel étaient mesurés en "pour mille" (donc, multipliés par 10) quels seraient les estimations des moindres carrés $\widehat{a}'$ et $\widehat{b}'$ des coefficients a et b ?

5.- Calculer le coefficient de détermination du modèle. Rappeler son interprétation.

6.- Calculer l'indicateur de précision moyenne $s(\widehat{\varepsilon})$ et interpréter ce résultat.

7.- Calculer le taux du salaire réel ajusté en 1992, 1993 et 1994 ainsi que les résidus de l'ajustement pour ces trois points.



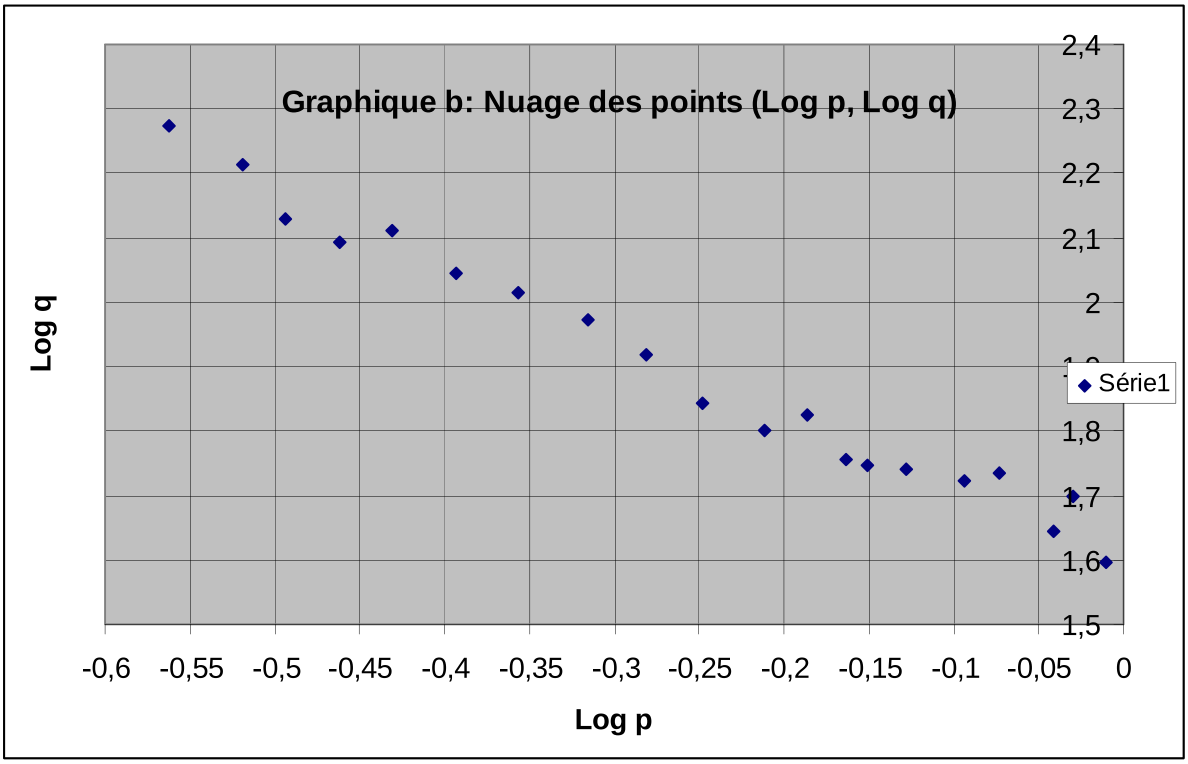
Sur le graphique 2 ci-dessus, on a représenté le taux de salaire réel observé et ajusté en fonction du temps. Quel commentaire peut-on faire sur la capacité du modèle à retracer l'évolution du taux de croissance du salaire réel sur la période considérée ?

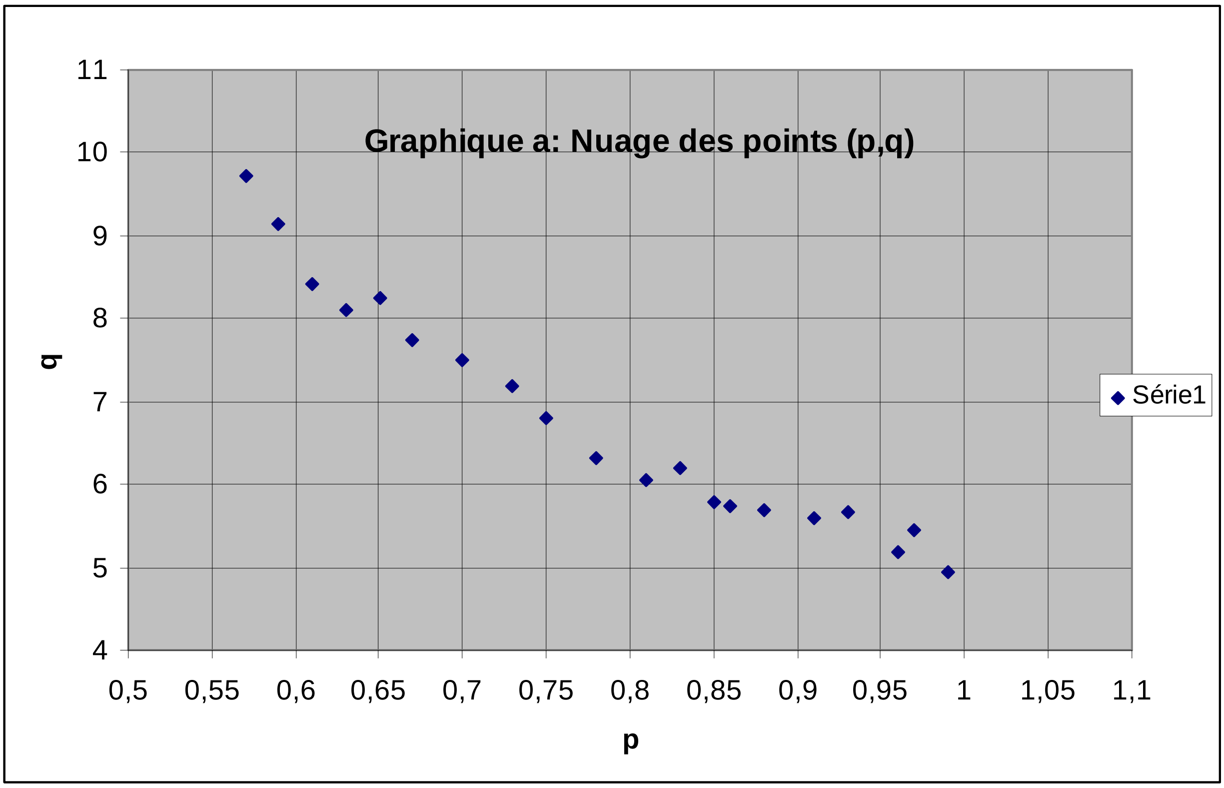
8.- Avec le modèle estimé, quel taux de chômage annulerait le taux de croissance du salaire réel ? Quel serait le taux de croissance du salaire réel pour un taux de chômage de 10%, de 12%, de 14% ? Quel taux de croissance du salaire réel peut-on prévoir pour 2004 si le taux de chômage atteint 11%. Dans ce cas, si le taux d'inflation en 2004 s'établit à 1,5%, quel sera le taux de croissance du salaire nominal ?

Exercice 2

Le tableau ci-après présente, en colonne 3, l'évolution du volume q_t (en milliers) des ventes de micro-ordinateurs d'un constructeur informatique du 3ème trimestre 2002 au 2$^{\text{ème}}$ trimestre 2007.

t	p	q	Log p	Log q
1	0,99	4,93	-0,01	1,5963
2	0,97	5,46	-0,03	1,699
3	0,96	5,18	-0,041	1,645
4	0,93	5,67	-0,073	1,736
5	0,91	5,6	-0,094	1,723
6	0,88	5,69	-0,128	1,74
7	0,86	5,74	-0,151	1,748
8	0,85	5,79	-0,163	1,757
9	0,83	6,2	-0,186	1,825
10	0,81	6,05	-0,211	1,801
11	0,78	6,31	-0,248	1,842
12	0,75	6,81	-0,281	1,918
13	0,73	7,19	-0,315	1,974
14	0,7	7,5	-0,357	2,015
15	0,67	7,73	-0,393	2,046
16	0,65	8,26	-0,431	2,111
17	0,63	8,11	-0,462	2,094
18	0,61	8,41	-0,494	2,13
19	0,59	9,15	-0,519	2,214
20	0,57	9,72	-0,562	2,274





Les ventes augmentent car le prix des micro-ordinateurs a considérablement baissé sur les cinq dernières années. On a porté dans la deuxième colonne du tableau, l'indice du prix relatif p_t (base 1 au au 2ème semestre 2002) des micro-ordinateurs vendus au trimestre t (t = 1, …, 20). Le prix relatif est le rapport de l'indice des prix des micro-ordinateurs et de l'indice des prix à la consommation. Il vaut 1 au 92 (t = 0). Sur les graphiques a et b sont représentés les points de coordonnées (p_t, q_t) d'une part et de coordonnées $(Log\ p_t, Log\ q_t)$ d'autre part.

On s'intéresse aux modèles : [M1] $Log\ q = a\ Log\ p + b$

et [M2] $q = \alpha\ p + \beta$

Le calcul donne :

$$\sum_{t=1}^{20} Log\ p_t = -5{,}149\ ;\ \sum_{t=1}^{20} Log\ q_t = 37{,}888\ ;\ \sum_{t=1}^{20} (Log\ p_t)(Log\ q_t) = -10{,}41315\ ;$$

$$\sum_{t=1}^{20} Log^2\ p_t = 1{,}91175\ ;\ \sum_{t=1}^{20} Log^2 q_t = 72{,}53228.$$

Première partie

1.- En comparant les deux graphiques a et b, expliquer pourquoi il semble plus raisonnable d'utiliser l'équation [M1] que l'équation [M2] pour décrire la loi de la demande. Quelle est l'interprétation économique du coefficient *a* ? A quel signe s'attend-on pour *a* ?

Le paramètre *b* a-t-il une interprétation économique intéressante. Ecrire la loi de comportement obtenue sous la forme $q = kp^a$.

2.- Calculer les estimations des moindres carrés, $\hat{a}$ et $\hat{b}$, des coefficients de [M1] sur la période. On pose $Log\ \hat{q}_t = \hat{a}\ Log\ \hat{q}_t + \hat{b}$.

Ecrire cette expression sous la forme $\hat{q}_t = \hat{k}\,\hat{p}_t^{\,\hat{a}}$ et calculer $\hat{k}$.

3.- Calculer la précision moyenne $s(\hat{\varepsilon})$ de l'ajustement de l'équation [M1] et interpréter le résultat. Calculer au 2ème trimestre 2007 (t = 20)

- le résidu de l'ajustement de [M1] : $\hat{\varepsilon}_{20} = Log(q_{20}) - Log(\hat{q}_{20})$
- la valeur ajustée $\hat{q}_{20}$ puis l'erreur relative $\dfrac{q_{20} - \hat{q}_{20}}{q_{20}}$; la comparer à $\hat{\varepsilon}_{20}$.

Interpréter la valeur numérique $\sigma(\hat{\varepsilon})$ trouvée ci-dessus en supposant que $\hat{q}_t / q_t$ est suffisamment proche de l'unité pour qu'on puisse approximer $Log(\hat{q}_t / q_t)$ par $(\hat{q}_t / q_t) - 1$ (t = 1, …, 20) .

4.-Calculer le coefficient de détermination de la régression effectuée à la question 2.

5.- Calculer le taux trimestriel moyen de variation du prix des micro-ordinateurs entre le 2ème trimestre 2002 et le 2ème trimestre 2007 (ce prix vaut 1 en t = 0).

En supposant que la baisse des prix se poursuit à ce même rythme trimestriel jusqu'à la fin de l'année 2007 (t = 21 et t = 22), prévoir les ventes de ces deux trimestres avec l'équation [M1] trouvée à la question 2.

6.- On désigne par V le chiffre d'affaire trimestriel ($V = pq$)

On désigne par [M3] l'équation : $Log\,V = Log\;pq = c\;Log\;p + d$ selon laquelle le logarithme du chiffre d'affaires est une fonction linéaire du logarithme de l'indice des prix.

Montrer que cette équation se déduit de [M1].

On désigne par $\hat{c}$ et $\hat{d}$ les estimations des coefficients c et d obtenues par la méthode des moindres carrés. Montrer que $\hat{c} = \hat{a} + 1$ et que $\hat{d} = \hat{b}$. En déduire que les équations [M1] et [M3] ont les mêmes résidus. Réécrire l'ajustement de l'équation [M3] sous la forme $\hat{V}_t = \hat{k}'\,p_t^{\hat{c}}$.

7.-On pose : $y = Log\,q$ et $x = Log\;p$. V (.) désignant ici la variance, montrer que

$$V(x + y) = V(y) + (1 + 2\hat{a})\,V(x)$$

En déduire que le coefficient de détermination de la régression qui fait l'objet de la question 6 est inférieur ou supérieur à celui trouvé à la question 3 selon que $\hat{a}$ est supérieur ou inférieur à -0,5. Calculer le coefficient de détermination de la régression du modèle [M3].

Déduire de ces calculs et de la question 6 pourquoi on ne doit pas apprécier la pertinence d'une modélisation en se basant seulement sur la valeur du coefficient de détermination d'une régression.

<u>Deuxième partie</u>

Les calculs donnent

$$\sum_{t=1}^{20} p_t = 15{,}685\ ;\ \sum_{t=1}^{20} q_t = 135{,}575\ ;\ \sum_{t=1}^{20} p_t q_t = 102{,}862\ ;\ \sum_{t=1}^{20} p_t^2 = 12{,}6445\ ;\ \sum_{t=1}^{20} q_t^2 = 956{,}179.$$

8.-Déterminer les estimations des moindres carrés, $\hat{\alpha}$ et $\hat{\beta}$, des coefficients de [M2].

Avec ce modèle, donner l'expression algébrique de l'élasticité prix et la calculer en t =20.

9.- Soit $\hat{e}$ la variable résidu du modèle [M2]. Calculer la précision moyenne $s(\hat{e})$ de l'ajustement de l'équation [M2] et interpréter le résultat. Expliquer pourquoi on ne peut pas comparer $s(\hat{e})$ et $s(\hat{\varepsilon})$ trouvée à la question 3. Préciser comment on calculerait un indicateur de précision moyenne pour le modèle [M2] qui serait comparable à $s(\hat{\varepsilon})$ du modèle [M1].

Troisième partie

10.- On pose $z_t = Log(\frac{q_t}{q_{t-1}}) / Log(\frac{p_t}{p_{t-1}})$. Montrer que l'équation [M1] implique [M4] : $z = a$.

Sous quelles conditions peut-on approximer $Log(\frac{q_t}{q_{t-1}})$ par $\frac{q_t}{q_{t-1}} - 1$ et $Log(\frac{p_t}{p_{t-1}})$ par $\frac{p_t}{p_{t-1}} - 1$?

Interpréter z_t sous ces conditions.

Estimer le paramètre a du modèle [M4] par la méthode des moindres carrés.

11.- On considère le modèle [M5] : $Log\, q = \gamma t + \delta$ où t prend les valeurs 1, …, 20. Comment s'interprète le paramètre γ ? Calculer les estimations $\hat{\gamma}$ et $\hat{\delta}$ obtenues par la méthode des moindres carrés et le coefficient de détermination de la régression.

Exercice 3

On considère les loyers y_j trimestriels d'un échantillon de 30 appartements de 3 pièces (j = 1 , …, 30) mis en location dans un même quartier de Paris, classés selon leurs surfaces x_j (en m^2). On pose : $z_j = y_j / x_j \, (j = 1, ... 30)$. On obtient

$\bar{y} = 4025\, euros$; $\bar{x} = 59{,}20\, m^2$; $\bar{z} = 68{,}31\, euros / m^2$; $\sum_{j=1}^{30} x_j y_j = 7370247$ eurosm^2 ;

$\sum_{j=1}^{30} x_j^2 = 108890\, m^4$; $V(y) = \frac{1}{30}\sum_{j=1}^{30}(y_j - \bar{y})^2 = 569802\, euros^2$; $\sum_{j=1}^{j=30} z_j^2 = 141322$ euros$^2 / m^4$.

__1ère partie__ : Pour tenter de caractériser simplement la relation entre le loyer et la surface d'un appartement, on s'intéresse aux deux modèles équivalents suivants :

[Ml] : $y = m\, x$ et [M2] : $z = y / x = m$.

1.- Calculer z_1 et z_{30}. Comment le paramètre m s'interprète-t-il économiquement ?

2.- On se propose d'appliquer la méthode des moindres carrés pour estimer m à l'aide de l'échantillon des 30 appartements. On désigne respectivement par $\hat{m}$ et par m^* son estimation dans le modèle [M1] et dans le modèle [M2].

2.1- Quelle est la variable expliquée dans le modèle [M2] et quelle est la variable explicative ? Rappeler l'expression de l'estimation m^* du paramètre m du modèle [M2] en fonction des valeurs z_j (j = 1, …, 30). Comparer m^* au rapport $\bar{y} / \bar{x}$. Calculer la précision absolue de la régression. Que vaut le coefficient de détermination de la régression ?

2.2- On désigne par $\hat{y}_j$ la valeur ajustée du loyer de l'appartement n° j obtenue à l'aide du modèle [M1], et par $\hat{\varepsilon}_j$ le résidu d'ajustement correspondant.

Calculer $\hat{m}$. Montrer que la moyenne arithmétique des résidus d'ajustement $\bar{\hat{\varepsilon}}$ est positive ou négative selon que $\hat{m}$ est inférieur ou supérieur au rapport $\bar{y} / \bar{x}$. Calculer $\bar{\hat{\varepsilon}}$. Montrer que l'estimation m peut s'interpréter comme une moyenne arithmétique pondérée des valeurs z_j (j de 1 à 30) ; préciser l'expression des coefficients de pondération. Calculer et comparer les pondérations des valeurs z_1 et z_{30} (1er et 30ème appartement) dans cette moyenne.

3.- Montrer que $V(y) \neq V(\hat{y}) + V(\hat{\varepsilon})$. Comment s'exprime $V(\hat{\varepsilon})$ en fonction de $V(y)$, $V(x)$ $Cov(x,y)$ et $\hat{m}$. Calculer $V(\hat{\varepsilon})$. En déduire la moyenne quadratique des résidus d'ajustement : $\sqrt{\frac{1}{30}\sum_{j=1}^{30}\hat{\varepsilon}_j^2}$.

<u>2ème partie</u> : Pour tenter de mieux caractériser la relation entre le loyer et la surface d'un appartement, on s'intéresse aux modèles : [M3] $y = ax + b$ et [M4] $z = y/x = a + b\frac{1}{x}$.

4.1- Dans chacun des deux modèles, dire quelle est la variable expliquée et quelle est la variable explicative. Le modèle [M4] est-il linéaire ?
Expliquer pourquoi le signe de *b* permet de préciser si le loyer du m^2 marginal est inférieur ou non au loyer du m^2 moyen.

4.2- La méthode des moindres carrés appliquée au modèle [M4] conduit à la régression suivante :
$\hat{z} = \frac{y}{x} = a + b\frac{1}{x}$ ($R^2 =$ 0,065) où R^2 est le coefficient de détermination.

On se propose d'estimer les coefficients *a* et *b* du modèle [M3] par la méthode des moindres carrés. On désigne par $\hat{a}$ l'estimation de *a* obtenue à l'aide du modèle [M3], par y_j * la valeur ajustée du loyer de l'appartement n° *j* et par ε_j * le résidu d'ajustement correspondant.

Écrire l'équation de la droite de régression. Calculer la précision absolue de la régression. Calculer le coefficient de détermination et l'interpréter.
Peut-on comparer les coefficients de détermination des deux régressions [M3] et [M4] ?

4.3-Montrer que $\hat{a}$ peut s'écrire comme une moyenne arithmétique pondérée des nombres $p_j = \frac{y_j - \bar{y}}{x_j - \bar{x}}$, soit $\hat{a} = \sum_{j=1}^{30}\alpha_j p_j$ où $\alpha_j = \frac{(x_j - \bar{x})^2}{\sum_{j=1}^{30}(x_j - \bar{x})^2}$

Calculer $\alpha_1, \alpha_{16}, \alpha_{30}$ et z_{16}. En déduire les contributions des 1ère, 16ème et 30ème observations à $\hat{a}$.

4.4- Comparer les formules d'estimation de m^* et de $\hat{a}$. Quelles sont les contributions des 1ère, 16ème et 30ème observations à m^* ?

<u>*Exercice 4*</u>

On trouve dans les Comptes de la Nation de l'année 1997, pour chaque l'année *t* (*t* = 1970, …, 1997), la valeur de la consommation des ménages ainsi que l'évaluation des quantités produits consommés (q_t^h) aux prix de l'année 1980 (p_{80}^h), où $h = a, b, \ldots$ désigne la nomenclature des produits consommés.
On désigne par : $C_t = \sum_h p_{80}^h\, q_t^h$, la consommation des ménages de l'année *t* aux prix de l'année 1980 et par R_t le revenu disponible des ménages de l'année *t*, en francs 1980.

1.

1.1- Sachant que la consommation finale des ménages de l'année 1997 valait

$4714\,\overline{\overline{MF}}$ et que $C_{97} = 2253\,\overline{\overline{MF}}$, calculer l'indice des prix à la consommation $I_{97/80}(p)$ issu des Comptes Nationaux. Préciser le nom de l'indice des prix ainsi calculé.
De quelle autre information a-t-on besoin pour calculer l'indice $I_{97/96}(p)$?

1.2- Calculer le revenu disponible de l'année 1997 en francs 1980 : R_{97} sachant que le revenu disponible des ménages en francs courants de l'année 1997 valait $\tilde{R}_{97}$ =5685 Mds F.

2.- On envisage le modèle [M] : $Log\,C_t = \alpha\ Log\,C_{t-1} + (1-\alpha)\ LogR_t + \beta + \varepsilon_t$, où α et β désignent deux paramètres à estimer, selon lequel la consommation des ménages C_t dépend de la consommation de l'année précédente C_{t-1} (du fait des habitudes) et du revenu disponible R_t de l'année *t*.
D'après ce modèle, calculer l'élasticité de C_t par rapport à R_t, puis de C_{t+1} par rapport à C_t. En déduire quel serait l'impact (en %) d'une augmentation du revenu disponible des ménages de l'année *t* de 1 %, sur la consommation de l'année, puis sur celle de l'année *t*+1 ?
Application numérique : $\alpha = 0{,}8$.

3.-

3.1- Comment doit-on choisir y_t et x_t pour que le modèle se présente sous la forme habituelle de la régression simple : [M'] $y_t = \alpha\, x_t + \beta + \varepsilon_t$? Interpréter économiquement la variable expliquée de ce modèle.

3.2- On pose $c_t = Log\,C_t$ et $r_t = Log\,R_t$, et on donne les résultats numériques ci-dessous :

n = 27 (t = 0 en 1970 et t = 27 en 1997)

$\sum_{t=1}^{27} c_{t-1} = 387{,}14188$; $c_0 = 13{,}93321$; $c_{27} = 14{,}62798$; $\sum_{t=1}^{27} r_t = 393{,}19406$; $\sum_{t=1}^{27} c_{t-1}^2 = 5552{,}15543$
$\sum_{t=1}^{27} r_t^2 = 5726{,}71466$; $\sum_{t=1}^{27} c_{t-1} r_t = 5638{,}73306$; $\sum_{t=1}^{27} c_{t-1} c_t = 5562{,}06554$; $\sum_{t=1}^{27} c_t r_t = 5648{,}80886$.

Estimer les coefficients α et β par la méthode des moindres carrés.

3.3- Calculer le coefficient de détermination de la régression de *y* sur *x* ainsi que la précision absolue $\sigma(\hat{\varepsilon})$. Comment peut-on interpréter cette valeur ?

Exercice 5

On suppose que la valeur ajoutée *Y* d'une entreprise est liée à la valeur du capital utilisé *X* et au nombre de salariés employés *Z* par la relation suivante : [M] $Y = kX^{\alpha} Z^{1-\alpha}$ où α et *k* désignent deux constantes.
Disposant d'observations des grandeurs *X*, *Y* et *Z* d'un échantillon d'entreprises (c'est-à-dire de triplets $(X_j, Y_j, Z_j),\ j = 1, \ldots, n\rfloor$), on désire estimer le coefficient α.

1.- Quelle est la variable expliquée ? Quelles sont les variables explicatives ? Quelle est l'interprétation des coefficients α et (1-α) ?

Montrer que la fonction $f : Y = f(X, Z)$ est homogène de degré 1.

2.- Pourquoi [M] n'est-il pas un modèle de régression linéaire ?

On pose $y = Log\ Y$, $x = Log\ X$, $z = Log\ Z$ et $c = Log\ k$. Montrer que le modèle [M] implique [M'] : $y - z = \alpha\ (x - z) + c$.

Pourquoi est-il possible d'estimer les coefficients α et c en appliquant la méthode des moindres carrés au modèle [M']?

3.- Exprimer $Cov\ (x - z,\ y - z)$, $Var(x - z)$ et $Var(y - z)$ en fonction de $Cov(x, y)$, $Cov(x, z)$, $Cov(y, z)$ $Var(x)$, $Var(z)$ et $Var(y)$.

4.- On donne: $\bar{x} = 4{,}76$; $\bar{y} = 9{,}97$; $\bar{z} = 4{,}71$; $V(x) = 0{,}1887$; $V(y) = 0{,}1804$; $V(z) = 0{,}1720$; $Cov\ (x, y) = 0{,}1836$; $Cov\ (x, z) = 0{,}1791$; $Cov\ (y, z) = 0{,}1760$.

Calculer les estimations $\hat{\alpha}$ et $\hat{c}$ des coefficients α et c obtenues par la méthode des moindres carrés. Avec ce modèle, de combien augmentera la valeur ajoutée de l'entreprise si elle augmente le nombre de salariés de 1%, en laissant la valeur du capital utilisé inchangé ?

Écrire l'équation d'analyse de la variance de la régression, calculer le coefficient de détermination R^2. Calculer l'indicateur de précision moyenne de l'ajustement : $\sqrt{\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2}$.

Exercice 6

Age i du véhicule	Effectif n_i	Moyenne $\bar{y}_i$	Dispersion D_i
1 an	26	0,7961	0,05581
2 ans	26	0,6410	0502749
3 ans	23	0,5422	0,01825
4 ans	21	0,4468	0,01247
5 ans	17	0,3502	0,00743
Ensemble	113	0,5767	2,73384

On s'intéresse ici au facteur de dépréciation $y = \dfrac{prix\ du\ véhicule\ d'occasion}{prix\ du\ véhicule\ neuf}$ des automobiles âgées d'un, deux, trois, quatre ou cinq ans (i = 1, …, 5). On dispose d'un fichier de données qui contient les couples de valeurs $(âge_j, y_j)$ de 113 cas (j = 1, …, 113) publiés par une revue spécialisée. On désigne par E_i les véhicules d'âge i dans le fichier (i = 1, …, 5). Le tableau présente les effectifs, les moyennes $\bar{y}_i$, et les dispersions $D_i = \sum_{j \in E_i}(y_j - \bar{y}_i)^2$ dans chacun des échantillons E_i.

1.- On se place d'abord dans l'échantillon El des 26 véhicules âgés d'un an. Quel est l'écart-type $\sigma_1(y)$ de y dans cet échantillon ?

On considère le modèle [M1] : $y = m_1$ où m_1 désigne une constante. Quelle est l'estimation des moindres carrés $\hat{m}_1$ du paramètre m_1 ?

2.- En se plaçant maintenant dans l'échantillon entier des 113 véhicules, on s'intéresse au modèle [M2] : $y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5$ (appelé modèle d'analyse de la variance à un facteur), où $m_1, m_2, \dots m_5$ désignent 5 constantes, et $x_1, x_2, \dots x_5$ cinq variables *indicatrices*, (c'est-à-dire ?) des sous-échantillons El, E2, E3, E4 et E5 des automobiles d'un, deux, ..., cinq ans d'âge.

2.1.- Montrer comment on obtient les valeurs indiquées dans le tableau pour la rubrique ensemble.

2.2.- Quelles sont les estimations $\hat{m}_i$ des moindres carrés des paramètres m_i ?

2.3.- Montrer que $\bar{\hat{y}} = \bar{y}$. En déduire la valeur de la somme des résidus d'estimation de l'ajustement. Ecrire l'équation d'analyse de la variance. Calculer le coefficient de détermination et la précision de la régression (c'est-à-dire la moyenne quadratique des résidus d'ajustement).

3.- En se plaçant encore dans l'échantillon entier des 113 véhicules, on considère le modèle :
[M3] : $Log\ y = a\,i$, où i désigne l'âge du véhicule $(i = 1, \dots, 5)$ et a une constante.

3.1.- Interpréter le coefficient $e^a - 1$.

3.2.- Sachant que l'estimation des moindres carrés $\hat{a}$ du coefficient a vaut -0,208227, calculer l'estimation de $e^a - 1$.

On désigne par $\hat{y}* = e^{\hat{a}\,i}$ l'estimation du facteur de dépréciation y d'un véhicule d'âge i obtenue à l'aide de l'ajustement du modèle [M3] et on donne $\sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j^*)^2} = 0{,}0353$.

Comparer les deux ajustements obtenus pour [M1] et pour [M3].

3.3.- Selon le modèle [M3], après combien d'années la décote d'une automobile d'occasion est-elle égale à la moitié du prix du véhicule neuf ?