Université A/MIRA de Béjaia
Faculté des sciences de Gestions, commerciales et économiques

Département de sciences économiques

Laboratoire de Mathématiques Appliquées

Polycopié pédagogique

Dossier numéro (à remplir par l'administration) : SE / NIHR / 2021.

Titre

Techniques quantitatives

Régression Linéaire simple et multiple

Cours destiné aux étudiants de

Licence (spécialité et niveau) : $3^{ème}$ année EMB

Année : 2021.

Introduction

Ce cours est destiné en particulier aux étudiants de troisième année License en Economie Monétaire et Bancaire, notamment aux étudiants en sciences économiques et sciences de gestions.

L'objectif de ce présent cours est de présenter la théorie de l'analyse des données par régression au sens statistique du terme. Pour cela nous allons proposer des détails sur la régression simple et multiple, où on va expliquer la démarche à suivre pour ces deux modèles en donnant quelques exemples.

L'analyse par régression linéaire est l'une des solutions qui existe pour définir les liens entre une variable quantitative dépendante (endogène) avec une variable ou plusieurs variables quantitatives indépendantes (exogènes), en acceptant des hypothèses fondatrices de la statistique paramétrique et la notion d'ajustement par les moindres carrés.

Ce type de modèles, nous permet de décrire et de comprendre la relation entre une variable dépendante Y et une ou plusieurs variables indépendantes X_i. Ces modèles, nous permet également prédire les valeurs que peut prendre la variable de réponse Y à partir des valeurs prisent par les variables prédictives X_i.

La régression linéaire, nous permet non seleument de faire le diagnostic de la régression à l'aide de l'analyse des résidus en utilisant des tests statistiques, mais aussi permet l'amélioration du modèle en sélectionnant les variables et de détecter l'existence de la colinéarité entre les variables exogènes.

Les c de régression linéaire sont utilisés dans divers domaines, à savoir : en économie, en science politique, en sociologie, en psychologie, en géographie, en physique, en mécanique, en biologie,. . .. Comme nous pouvons les utilisés dans certains cas non linéaire en effectuant un changement de variable (puissance, logarithme).

Ces Les modèles sont excellents car ils sont faciles à utiliser et à interpréter. Cependant, leur simplicité inhérente présente également quelques inconvénients et dans de nombreux cas, ils ne constituent pas vraiment le meilleur choix du modèle de régression. Il existe en fait plusieurs types de régressions, chacune avec ses avantages et ses inconvénients.

Dans ce polycopié, on va faire appel à quelques notions à savoir : notions statistiques (le calcul de la moyenne, variance, intervalle de confiance et les tests) et notions algébrique, l'algèbre linéaire (le calcul des déterminants et l'inverse d'une matrice).

Chapitre 1

Modèle de régression linéaire simple

L'objectif de la régression linéaire simple est d'établir un lien entre une variable dépendante Y et une variable indépendante X, ce qui va nous permettre de faire ensuite des prévisions de Y lorsque X est mesurée.

1.1 *Présentation du modèle*

Un modèle de régression linéaire simple est défini par :

$$Y_t = B_0 + B_1 X_t + \varepsilon_t, \quad t = 1...n \tag{1.1}$$

où

- Y_t est la variable dépendante ou endogène (une variable à expliquer) au temps t.
- B_0, B_1 sont les coefficients de la régression.
- X_t est la variable indépendante ou exogène (variable explicative).
- ε_t est une erreur aléatoire.
- n est le nombre d'observations.

Ce modèle doit satisfaire certaines conditions :
H_1 : Le modéle est linéaire en X_t.
H_2 : Les valeurs de X_t sont des grandeurs numériques mesurées sans erreur.
H_3 : Les ε_t sont i.i.d. (indépendants et identiquement distribués).
H_4 : La moyenne des erreurs s'annulent, le modèle est bien spécifié, $E(\varepsilon_t) = 0$.
H_5 : La variance de l'erreur est constante et ne dépend pas de l'observation : homoscédasticité, $Var(\varepsilon_t) = \sigma_\varepsilon^2$.
H_6 : L'erreur est indépendante de la variable exogène, $COV(\varepsilon_t, X_t) = 0$.

H_7 : Indépendance des erreurs, les erreurs relatives à 2 observations sont indépendantes (on dit aussi que les erreurs sont non corrélées), $COV(\varepsilon_t, \varepsilon_s) = 0$.
H_8 : ε_t suit une loi normale, $\varepsilon_t \rightsquigarrow N(0, \sigma_\varepsilon^2)$

1.2 *Estimation des paramètres par la méthode Moindres carrés ordinaires (MCO)*

Dans la réalité les valeurs de B_0 et B_1 sont inconnus, on doit les estimer. Leurs estimateurs sont donnés par $\widehat{B_0}$ et $\widehat{B_1}$ qui sont obtenus par la méthode Moindres carrés ordinaires (MCO). Cette méthode consiste à minimiser la somme des carrés des erreurs.

$$\min \sum_{t=1}^{n} \varepsilon_s^2 = \min \sum_{t=1}^{n} (Y_t - B_0 - B_1 X_t)^2 = \min S(B_0, B_1) \tag{1.2}$$

Pour minimiser $S(B_0$ et $B_1)$, il suffit de trouver $\widehat{B_0}$ et $\widehat{B_1}$.

$$\begin{cases} \frac{\partial S(B_0,B_1)}{\partial B_0} = 0 \\ \\ \frac{\partial S(B_0,B_1)}{\partial B_1} = 0 \end{cases}$$

Par conséquent, on obtient

$$\begin{cases} -2 \sum\limits_{t=1}^{n} \left(Y_t - \widehat{B_0} - \widehat{B_1} X_t \right) = 0 \\ \\ -2 \sum\limits_{t=1}^{n} X_t \left(Y_t - \widehat{B_0} - \widehat{B_1} X_t \right) = 0 \end{cases}$$

Et après calculs que,

$$\begin{cases} \widehat{B_1} = \frac{\sum\limits_{t=1}^{n} X_t Y_t - n\ \overline{X}\ \overline{Y}}{\sum\limits_{t=1}^{n} X_t^2 - n\overline{X}^2} = \frac{\sum\limits_{t=1}^{n} \left(X_t - \overline{X}\right)\left(Y_t - \overline{Y}\right)}{\sum\limits_{t=1}^{n} \left(X_t - \overline{X}\right)^2} \\ \\ \widehat{B_0} = \overline{Y} - \widehat{B_1}\overline{X} \end{cases} \tag{1.3}$$

Remarque

1. Si B_0 est constant, alors

$$\widehat{B_1} = \frac{\sum\limits_{t=1}^{n} X_t Y_t}{\sum\limits_{t=1}^{n} X_t^2} \tag{1.4}$$

2. Les estimateurs peuvent s'écrire :

$$\begin{cases} \widehat{B_1} = B_1 + \dfrac{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)\,\varepsilon_t}{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2} \\ \\ \widehat{B_0} = B_0 - \left(\widehat{B_1} - B_1\right)\overline{X} + \overline{\varepsilon} \end{cases} \tag{1.5}$$

1.3 *Propriétés des estimateurs* $\widehat{B_0}$ *et* $\widehat{B_1}$

D'après la remarque précédente, nous allons montrer que ces estimateurs sont sans biais et convergents :

Les estimateurs sont sans biais

Nous savons que les bruits ε_t sont aléatoires, et puisqu'ils sont centrés, et D'après la formule précédente de $\widehat{B_1}$, on déduit bien que $E[\widehat{B_1}] = B_1$ ($\widehat{B_1}$ est sans biais) .
Pour $\widehat{B_0}$, on a : $E[\widehat{B_0}] = B_0 - E[\left(\widehat{B_1} - B_1\right)]\overline{X}$, et comme $\widehat{B_1}$ est sans biais, on déduit alors :$E[\widehat{B_0}] = B_0$, donc $\widehat{B_0}$ est sans biais.

Les estimateurs sont convergents

$$Var(\widehat{B_1}) = E\left(\widehat{B_1} - B_1\right)^2 = E\left[\frac{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)\,\varepsilon_t}{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2}\right]^2$$

$$Var(\widehat{B_1}) = \frac{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2}{\left(\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2\right)^2} E\left[\varepsilon_t\right]^2 + \sum_{t\leq s}\left[\frac{\left(X_t-\overline{X}\right)}{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2}\right]\left[\frac{\left(X_s-\overline{X}\right)}{\sum_{s=1}^{n}\left(X_s-\overline{X}\right)^2}\right] E\left[\varepsilon_t\varepsilon_s\right]$$

$Var(\widehat{B_1}) = \frac{E[\varepsilon_t]^2}{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2}$ car les erreurs sont non corrélées.

$$Var(\widehat{B_1}) = \frac{\sigma_\varepsilon^2}{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2} \tag{1.6}$$

$Var(\widehat{B_0}) = Var(B_0) + Var(\widehat{B_1})\overline{X}^2 + Var(B_1)\overline{X}^2 + Var(\overline{\varepsilon}) = Var(\widehat{B_1})\overline{X}^2 + Var(\overline{\varepsilon})$,
car $Var(B_0) = Var(B_1) = 0$

$$Var(\widehat{B_0}) = \frac{1}{n}\sigma_\varepsilon^2 + \frac{\overline{X}^2\sigma_\varepsilon^2}{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2} = \sigma_\varepsilon^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2}\right] \tag{1.7}$$

Les deux estimateurs des moindres carrés sont convergents lorsque le nombre d'observations est important ou lorsque les valeurs de variable explicative sont très disperées autour de la moyenne.

Remarque

$Cov(\widehat{B_1}, \widehat{B_0}) = -\overline{X} Var(\widehat{B_1}) = -\frac{\sigma_\varepsilon^2 \ \overline{X}}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2}$.

1.4 *Analyse de la variance*

Les estimateurs $\widehat{B_0}$ et $\widehat{B_1}$ des coefficients B_0 et B_1 prmettent de calculer pour chaque observation la réponse estimée. Celle-ci sera comparée à la valeur observée de Y_t par l'intermédiaire des résidus e_t,
où $e_t = Y_t - \widehat{Y_t}$, avec

- $E[e_t] = 0$
  car $\sum_{t=1}^{n} e_t = 0$
- $Var(e_t) = \sum_{t=1}^{n} e_t^2$
  $Var(e_t) = \sum_{t=1}^{n} (Y_t - \widehat{Y_t})^2 = (n-2)\sigma_\varepsilon^2$.
  Ce qui nous permet de déduire que :

$$\widehat{\sigma_\varepsilon^2} = \frac{\sum_{t=1}^{n} e_t^2}{(n-2)} = \frac{SCR}{(n-2)} \tag{1.8}$$

Alors les estimateurs des variances des deux paramètres $\widehat{B_1}$ et $\widehat{B_0}$ sont donnés par

$$\begin{cases} Var(\widehat{B_1}) = \frac{\widehat{\sigma_\varepsilon^2}}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2}, \\ \\ Var(\widehat{B_0}) = \widehat{\sigma_\varepsilon^2} \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2} \right], \end{cases} \tag{1.9}$$

1.4.1 *Décomposition de la variance*

L'objectif de la régression est de minimiser $\sum_{t=1}^{n} e_t^2$ et ce qui nous permet de juger la qualité de l'ajustement de ce modèle.

$$\sum_{t=1}^{n}\left(Y_t-\overline{Y}\right)^2=\sum_{t=1}^{n}\left(Y_t-\widehat{Y}_t+\widehat{Y}_t-\overline{Y}\right)^2=\sum_{t=1}^{n}\left(Y_t-\widehat{Y}_t\right)^2+\sum_{t=1}^{n}\left(\widehat{Y}_t-\overline{Y}\right)^2+2\underbrace{\sum_{t=1}^{n}\left(Y_t-\widehat{Y}_t\right)\left(\widehat{Y}_t-\overline{Y}\right)}_{=0}$$

$$\sum_{t=1}^{n}\underbrace{\left(Y_t-\overline{Y}\right)^2}_{SCT}=\sum_{t=1}^{n}\underbrace{\left(Y_t-\widehat{Y}_t\right)^2}_{SCR}+\sum_{t=1}^{n}\underbrace{\left(\widehat{Y}_t-\overline{Y}\right)^2}_{SCE} \tag{1.10}$$

SCT : somme des carrés totaux (variabilité totale)

SCE : somme des carrés expliqués par le modèle (variabilité expliquée)

SCR : somme des carrés résiduels, non expliqués par le modèle (variabilité résiduelle)
Les estimateurs sont d'autant plus précis lorsque La variance de l'erreur est faible et quand la dispersion des X est forte.

Tableau d'analyse de la variance (ANOVA)

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression (expliquée)	$SCE=\sum_{t=1}^{n}\left(\widehat{Y}_t-\overline{Y}\right)^2$	1	SCE
Résiduelle	$SCR=\sum_{t=1}^{n}\left(Y_t-\widehat{Y}_t\right)^2$	$n-2$	$\frac{SCR}{n-2}$
Totale	$SCT=\sum_{t=1}^{n}\left(Y_t-\overline{Y}\right)^2$	$n-1$	

Le coefficient de détermination R^2

Le coefficient de détermination R^2, nous donne une idée sur la qualitée de l'ajustement et il est meilleur lorsque la variabilitée expliquée est très proche de la variabilitée totale ($c-\grave{a}-d$ R^2 est très proche de 1) qui est donné par la formule suivante :

$$R^2=\frac{\sum_{t=1}^{n}\left(\widehat{Y}_t-\overline{Y}\right)^2}{\sum_{t=1}^{n}\left(Y_t-\overline{Y}\right)^2}=1-\frac{\sum_{t=1}^{n}\left(Y_t-\widehat{Y}_t\right)^2}{\sum_{t=1}^{n}\left(Y_t-\overline{Y}\right)^2} \tag{1.11}$$

ou

$$R^2=\frac{SCE}{SCT}=1-\frac{SCR}{SCT}=1-\frac{\sum_{t=1}^{n}e_t^2}{SCT} \tag{1.12}$$

Remarque

Le coefficient decrrélation linéaire simple est défini par

$$r = \rho = R = \frac{\sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)\left(X_t - \overline{X}\right)}{\sqrt{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2}\sqrt{\sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)^2}} \tag{1.13}$$

1.5 *Tests de significativités du modèle*

1.5.1 *Test de significativité globale du modèle (Fisher)*

Le test de Fisher nous permet de tester la significativité de la régression, ce test est défini par les deux hypothèses suivantes :
$H_0 : B_1 = 0$, Le modèle n'explique pas la variable Y
$H_1 : B_1 \neq 0$, Le modèle est pertinent (globalement significatif)
Pour cela, nous devons calculer la statistique de Fisher empirique F_c afin de la comparer à celle lu sur la table $F_{\frac{\alpha}{2}}(1, n-2)$, où $(1, n-2)$ sont les degrés de liberté la statistique de Fisher au seuil $\frac{\alpha}{2}$.

$$F_c = \frac{SCE/1}{SCR/(n-2)} = \frac{R^2/1}{(1-R^2)/(n-2)} \tag{1.14}$$

Aolrs :

- si $F_c > F_{\frac{\alpha}{2}}(1, n-2)$, alors le modèle est globalement significatif.
- si $F_c < F_{\frac{\alpha}{2}}(1, n-2)$, alors le modèle n'est pas globalement significatif, et que la variable X n'est pas explicative (nous acceptons l'hypothèse H_0).

1.5.2 *Test de Student*

D'après l'hypothèse H_8, les erreurs suivent une loi normale, ce qui nous permet de vérifier que :

$$\begin{cases} \frac{\widehat{B_0} - B_0}{\sigma^2_{\widehat{B_0}}} \rightsquigarrow N(0,1), \\ \frac{\widehat{B_1} - B_1}{\sigma^2_{\widehat{B_1}}} \rightsquigarrow N(0,1), \\ \frac{\sum_{t=1}^{n} e_t^2}{\sigma^2_\varepsilon} = \frac{(n-2)\widehat{\sigma}^2_\varepsilon}{\sigma^2_\varepsilon} \rightsquigarrow \chi^2_{(n-2)}, \end{cases}$$

où $\chi^2_{(n-2)}$ est le khi-deux à $(n-2)$ degrés de liberté.

$$\text{ET} \begin{cases} \frac{\widehat{B_0}-B_0}{\widehat{\sigma}^2_{\widehat{B_0}}} \rightsquigarrow t_{\frac{\alpha}{2}}(n-2), \\ \\ \frac{\widehat{B_1}-B_1}{\widehat{\sigma}^2_{\widehat{B_1}}} \rightsquigarrow t_{\frac{\alpha}{2}}(n-2), \end{cases}$$

où $t_{\frac{\alpha}{2}}(n-2)$ est la Student à $(n-2)$ degrés de liberté.
Par conséquent, nous pouvons mettre en place des tests statistiques pour voir si les paramètres sont significatifs ?

Test bilatéral

$H_0 : B_1 = 0$ contre $H_1 : B_1 \neq 0$. Dans un premier, nous douvons calculer la statistique de Student empirique t_c afin de la comparer à celle lu sur la table $t_{\alpha/2}(n-2)$, où $(n-2)$ est le degré de liberté la statistique de Student au seuil $\alpha/2$, avec

$$t_c = \left| \frac{\widehat{B_1} - B_1}{\widehat{\sigma}^2_{\widehat{B_1}}} \right| \tag{1.15}$$

Aolrs :

- si $t_c > t_{\alpha/2}(n-2)$, alors B_1 est significatif.
- si $F_c < t_{\alpha/2}(n-2)$, alors B_1 n'est pas significatif, et que la variable X n'est pas explicative (nous acceptons l'hypothèse H_0).

De la même maniere, nous pouvons vérifier si B_0 est significatif ou non ?

Intervalle de confiance

Nous pouvons aussi donner un intervalle de confiance pour chaque paramètre. Dans la plus part des cas σ_ε est inconnu, alors
L'intervalle de confiance pour B_0 (I_{cB_0})

$$B_0 = \widehat{B_0} \pm t_{\alpha/2}(n-2)\widehat{\sigma}_{\widehat{B_0}} \tag{1.16}$$

où $t_{\alpha/2}(n-2)$ est la valeur de la statistique de Student au seuil $\alpha/2$ à $(n-2)$ est le degré de liberté lu sur la table.

Remarque

$H_0 : B_0 = 0$ contre $H_1 : B_0 \neq 0$.

- si $0 \notin I_{cB_0}$, alors B_0 est significatif.
- si $0 \in I_{cB_0}$, alors B_0 n'est pas significatif, et que ce modèle ne contient de constante.

La même chose pour l'intervalle de confiance pour B_1, si $0 \in I_{cB_1}$, alors B_1 n'est pas significatif, et que la variable X n'est pas explicative (nous acceptons l'hypothèse H_0). Sinon, nous acceptons l'hypothèse H_1 et que la variable X est une variable explicative pour la variable endogène Y.

1.6 *Prévision*

L'un des buts de la régression est de proposer des prévisions pour la variable endogène Y. Une fois le modèle est validé, nous pouvons faire des prévisions pour une valeur non observée X_{n+h}. Dans un premier temps, nous douvons calculer Y_{n+h} qui est donné par

$$Y_{n+h} = B_0 + B_1 X_{n+h} + \varepsilon_{n+h} \tag{1.17}$$

avec $E[\varepsilon_{n+h}] = 0$ et $Cov(\varepsilon_{n+h}, \varepsilon_t) = 0$ et $Var(\varepsilon_{n+h}^2) = \sigma_\varepsilon^2$.

- La variance de l'erreur de prévision est donnée par

$$Var(\widehat{\varepsilon}_{n+h}) = Var(B_0 - \widehat{B_0}) + Var(B_1 - \widehat{B_1})X_{n+h}^2 + 2X_{n+h}Cov(\widehat{B_0}, \widehat{B_1}) + Var(\varepsilon_{n+h})$$

$$Var(\widehat{\varepsilon}_{n+h}) = Var(\widehat{B_0}) + Var(\widehat{B_1})X_{n+h}^2 + Var(\varepsilon_{n+h})$$

$$Var(\widehat{\varepsilon}_{n+h}) = \frac{\sigma_\varepsilon^2}{n} + (X_{n+h} - \overline{X})^2 \frac{\sigma_\varepsilon^2}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2} + \sigma_\varepsilon^2$$

$$Var(\widehat{\varepsilon}_{n+h}) = \sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{(X_{n+h} - \overline{X})^2}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2} + 1 \right] \tag{1.18}$$

- La variance de l'erreur de prévision est fonction de l'ecart quadratique entre la variable exogène prévue et la moyenne de la même variable. Donc plus la valeur prévue s'éloigne de cette moyenne, plus le risque d'erreur est important.

Remarque

D'après l'hypothèse H_8, nous déduisons que :

$$\widehat{\varepsilon}_{n+h} \rightsquigarrow N\left(0, \sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{(X_{n+h} - \overline{X})^2}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2} + 1 \right]\right).$$

- La valeur de prévision correspondante au modèle estimé est défini par

$$\widehat{Y}_{n+h} = \widehat{B_0} + \widehat{B_1} X_{n+h} \tag{1.19}$$

Et sa variance est donnée par

$$Var(\widehat{Y}_{n+h}) = \widehat{\sigma}_{\varepsilon}^2 \left[\frac{1}{n} + \frac{(X_{n+h} - \overline{X})^2}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2} \right] \tag{1.20}$$

• L'intervalle de prédiction est défini par

$$Y_{n+h} = \widehat{Y}_{n+h} \pm t_{\frac{\alpha}{2}}(n-2)\ \widehat{\sigma}_{\varepsilon} \sqrt{\left[\frac{1}{n} + \frac{(X_{n+h} - \overline{X})^2}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2} + 1 \right]}. \tag{1.21}$$

1.7 *Quelques exemples*

Quelques exemples

1.8 Exercices

exemple 1.
Soit le modèle : $Y_i = B_0 + B_1 X_i + \varepsilon_i, \quad i = \overline{1,8}$
Avec les données suivantes :

X_i	11.2	13.2	16	16	16.8	17.2	18.6	19
Y_i	28	33	30.4	34	32.2	38	40	36.4

1. Représenter le nuage de points.

2. Déterminer les estimateurs B_0, B_1 et étudier la significativité de chaque paramètre.

3. Calculer le SCE, SCT, SCR et le coefficient de détermination. Conclusion ?

4. Donner la prévision à la date 9 sachant que $X_9 = 19.5$

5. Donner le tableau de l'analyse de la variance.

Corrigé 1.

1. Représentation du nuage de points D'après la figure, nous constatons que le nuage de points ne donne pas une idée claire sur le type de relation qu'il existe entre les deux variables, et que la droite de régression ne passe par l'origine.

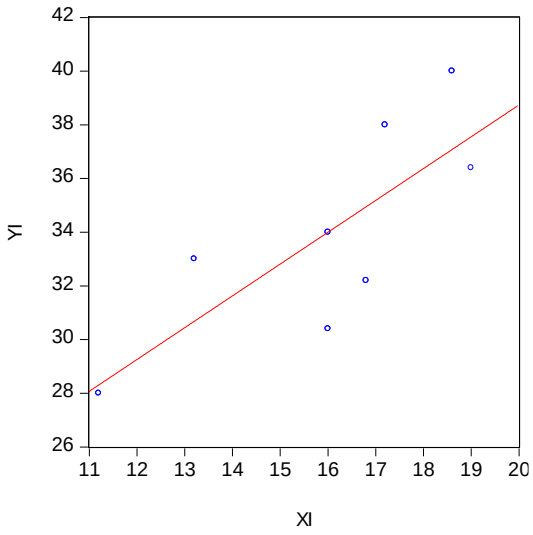

Figure 1 - Nuage de points Xt et Yt

$\widehat{B_0}$ et $\widehat{B_1}$ sont les quantités qui minimisent les distances verticales entre les observations Y_t et la droite de régression théorique $Y_t = B_0 + B_1 X_t + \varepsilon_t$, et la droite de régression estimée est donnée par :
$\widehat{Y_t} = \widehat{B_0} + \widehat{B_1} X_t$.

Afin de caculer les estimateurs des coefficients de régression, nous réalisons le tableau suivant :

n	X_i	Y_i	X_iY_i	X_i^2	Y_i^2	$\widehat{Y}_i$	$e_i = Y_i - \widehat{Y}_i$	e_i^2
1	11.2	28	313.6000	125.44	784.0000	28.309359	-0.309359	0.095703
2	13.2	33	435.6000	174.24	1089.0000	30.680460	2.319540	5.380265
3	16	30.4	486.4000	256.0000	924.1600	33.99999	-3.599999	12.960002
4	16	34	544.0000	256.0000	1156.0000	34.00000	0.000000	0.000000
5	16.8	32.2	540.9600	282.2400	1036.8400	34.00000	-2.748440	7.553924
6	17.2	38	653.6000	295.8400	1444.0000	34.948440	2.577339	6.6426795
7	18.6	40	744.0000	345.9600	1600.0000	35.422660	2.917569	8.512211
8	19	36.4	691.6000	361.0000	1324.9600	37.556650	-1.156650	1.337840
$\sum$	128	272	4409.7600	2096.7200	9358.9600		0	42.48263

$\overline{X} = \frac{128}{8} = 16$ et $\overline{Y} = \frac{272}{8} = 34$

$\widehat{B_1} = \frac{\sum_{t=1}^{n} X_tY_t - n\ \overline{X}\ \overline{Y}}{\sum_{t=1}^{n} X_t^2 - n\overline{X}^2} = \frac{4409.76 - 8(16)(34)}{2096.72 - 8(16)^2} = 1.18555008$

$\widehat{B_0} = \overline{Y} - \widehat{B_1}\overline{X} = 34 - 1.18555008(16) = 15.031199$

On écrit

$$\widehat{Y}_i = 15.031199 + 1.18555008X_i \tag{1.22}$$

2. ♣ On a $\sum_{t=1}^{n} e_i^2 = 42.48263$. Et $\widehat{\sigma_\varepsilon^2} = \frac{\sum_{t=1}^{n} e_t^2}{(n-2)} = \frac{42.48263}{8-2} = 7.080438$

♣ $Var(\widehat{B_1}) = \frac{\widehat{\sigma}_\varepsilon^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2} = \frac{7.080438}{2096.7200 - 8(16)^2} = 0.145329$

d'ou

$\widehat{\sigma}_{\widehat{B}_1} = 0.381220$

♣ $Var(\widehat{B_0}) = \widehat{\sigma_\varepsilon^2}\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum_{t=1}^{n}(X_t - \overline{X})^2}\right]$

$$Var(\widehat{B_0}) = 7.08044\left[\frac{1}{8} + \frac{(16)^2}{2096.7200 - 8(16)^2}\right] = 38.089324$$

$$\widehat{\sigma}_{\widehat{B}_0} = 6.171655$$

♣ On utilise le test de Student pour la significativité de B_0, qui est définie par :

$H_0 : \ B_0 = 0$ contre $H_1 : \ B_0 \neq 0$
Pour cela, on doit calculer $t_{c\ B_0}$

$$t_{c_{B_0}} = \left| \frac{\widehat{B_0} - B_0}{\widehat{\sigma}_{\widehat{B_0}}} \right| = \frac{15.03120 - 0}{6.171655} = 2.435522 \tag{1.23}$$

D'après la valeur de la Statistique de Student calculer pour B_0 ($t_{c_{B_0}} = 2.435522$) est infieure à $t_T(6, 0.025) = 2.447$ ($t_T(6, 0.025)$ est la valeur de la statistique de Student lu sur la table avec 6 est degré de liberté et 0.025 est le seuil de confiance). on conclut donc que le paramètre n'est pas significatif.

♣ De même pour étudier la significativité de B_1. Le test de Student est défini par :

$H_0 : \ B_1 = 0$ contre $H_1 : \ B_1 \neq 0$
Pour cela, on doit calculer $t_{c\ B_1}$

$$t_{c_{B_1}} = \left| \frac{\widehat{B_1} - B_1}{\widehat{\sigma}_{\widehat{B_1}}} \right| = \frac{1.185550 - 0}{0.381221} = 3.109879 \tag{1.24}$$

D'après la valeur de la Statistique de Student calculer pour B_0 ($t_{c_{B_1}} = 3.109879$) est superieure à $t_T(6, 0.025) = 2.447$ ($t_T(6, 0.025)$ est la valeur de la statistique de Student lu sur la table avec 6 est degré de liberté et 0.025 est le seuil de confiance). on conclut donc que le paramètre est significatif.

3. Le calcul SCT, SCE et R^2

♣ On sait que : $SCR = \sum_{t=1}^{n} e_t^2 = 42.48263$

♣

$$SCT = \sum_{t=1}^{n} \left[Y_t - \overline{Y} \right]^2 = \sum_{t=1}^{n} Y_t^2 - n\overline{Y}^2 = 9358.96000 - 8(34)^2 = 110.96000 \tag{1.25}$$

♣ On a $SCT = SCE + SCR$, donc

$$SCE = SCT - SCR = 110.960000 - 42.48263 = 68.47737 \tag{1.26}$$

Remarque

On peut également calculer SCE en utilisant la formule suivante :

$$SCE = \sum_{t=1}^{n} \left[\widehat{Y}_t - \overline{Y}\right]^2 = \widehat{\beta}_1^2 \left[\sum_{t=1}^{n} X_t^2 - n\overline{X}^2\right] = 1.185550^2(48.72) = 68.477363 \tag{1.27}$$

♣

$$R^2 = \frac{SCE}{SCT} = \frac{68.47737}{110.960000} = 0.617136. \tag{1.28}$$

D'aprsè la valeur de R^2, le modèle est moyennement significatif

4. Le calcul de la prévision à la date 9 sachant que $X_9 = 19.5$

   D'après la Statistique de Student calculer pour chaque paramètre calculer précidament est superieure à $t_T(6, 0.05) = 1.943$ ($t_T(6, 0.05)$ est la valeur de la statistique de Student lu sur la table avec 6 est degré de liberté et 0.05 est le seuil de confiance). on conclut donc que les deux paramètres sont significatifs. Ce qui nous permis de dire que ce modèle est validé et de calculer la prévision à la date 9 .

   ♣ La valeur de prévision correspondante au modèle estimé est défini par

   $$\widehat{Y}_{n+1} = \widehat{B_0} + \widehat{B_1} X_{n+1} \tag{1.29}$$

   Donc

   $$\widehat{Y}_9 = 15.03120 + 1.185550 X_9 = 15.03120 + 1.185550(19.5) = 38.149425$$

♣ La variance de l'erreur de prévision $Var(\widehat{\varepsilon}_{n+h})$ est donnée par :

$$Var(\widehat{\varepsilon}_{n+h}) = \widehat{\sigma}_{\varepsilon}^2 \left[\frac{1}{n} + \frac{(X_{n+1} - \overline{X})^2}{\sum_{t=1}^{n} (X_t - \overline{X})^2} + 1 \right] \tag{1.30}$$

$$Var(\widehat{\varepsilon}_9) = \widehat{\sigma}_{\varepsilon}^2 \left[\frac{1}{8} + \frac{(X_9 - \overline{X})^2}{\sum_{t=1}^{n} (X_t - \overline{X})^2} + 1 \right] = 2.660909 \left[\frac{1}{8} + \frac{(19.5 - 16)^2}{48.72} + 1 \right] = 3.662569$$

Alors $\sigma_{\widehat{\varepsilon}_9} = 1.913784$

♣ L'intervalle de prédiction est défini par :

$$Y_{n+1} = \widehat{Y}_{n+1} \pm t_{\frac{\alpha}{2}}(n-2) \ \sqrt{Var(\widehat{\varepsilon}_{n+1})}. \tag{1.31}$$

$$Y_9 = \widehat{Y}_9 \pm t_{0.05}(6) \ \sigma_{\widehat{\varepsilon}_9} = 38.149425 \pm 1.943(1.913784) = [34.43094 \ ; \ 41.86791]$$

5. Le tableau de l'analyse de la variance

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression X_1	$SCE = 68.47737$	1	$\frac{SCE}{1} = 68.47737$
Résiduelle	$SCR = 42.48263$	$n-2 = 8-2 = 6$	$\frac{SCR}{n-2} = \frac{42.48263}{6} = 7.08044$
Totale	$SCT = 110.96$	$n-1 = 8-1 = 7$	

exemple 2.
Nous modélisons la relation existante entre la variable Y et la variable X par un modèle de régression linéaire simple.
Pour tout $i = \overline{1, 15}$, nous avons : $Y_i = B_0 + B_1 X_i + \varepsilon_i$
O

Le résultat de l'estimation du modèle de régression linéaire multiple est donné par le tableau suivant :

Dependent Variable: Y
Method: Least Squares
Date: 05/31/15 Time: 06:46
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-23.74045	1?	-11.58787	0.0000
X	0.765572	0.031140	2?	0.0000
R-squared	0.978944	Mean dependent var		24.13333
Adjusted R-squared	0.977324	S.D. dependent var		16.37449
S.E. of regression	3?	Akaike info criterion		4.766422
Sum squared resid	79.03797	Schwarz criterion		4.860829
Log likelihood	-33.74816	F-statistic		4?
Durbin-Watson stat	2.659911	Prob(F-statistic)		0.000000

1) Ecrire l'équation du modèle de régression.

2) Compléter les valeurs manquantes dans le tableau.

3) Ce modèle est-il globalement significatif au seuil de 5% ?

4) Testez au seuil de 5% la signification de B_0 et B_1. Conclusion.

5) Donner le tableau de l'analyse de la variance.

Corrigé 2.

1. L'équation du modèle de régression est donnée par :

$$\widehat{Y}_t = \widehat{B}_0 + \widehat{B}_1 X_t \tag{1.32}$$

Alors, d'après le tableau, les valeurs des paramètres sont données par $\ll$ $coefficient$ $\gg$, et on écrit

$$\widehat{Y}_t = -23.74045 + 0.765572\ X_t \tag{1.33}$$

2. ♣ $1 \rightarrow \widehat{\sigma}_{\widehat{B}_0}$,
On sait que :

$$t_{c_{B_0}} = \frac{\widehat{B}_0 - 0}{\widehat{\sigma}_{\widehat{B}_0}} \tag{1.34}$$

Donc :

$$\widehat{\sigma}_{\widehat{B}_0} = \frac{\widehat{B}_0 - 0}{t_{c_{B_0}}} = \frac{-23.74045}{-11.58787} = 2.048733 \tag{1.35}$$

♣ $2 \rightarrow t_{c_{B_1}}$,

$$t_{c_{B_1}} = \frac{\widehat{B}_1 - 0}{\widehat{\sigma}_{\widehat{B}_1}} = \frac{0.765572}{0.031140} = 24.58467 \tag{1.36}$$

♣ $3 \rightarrow \widehat{\sigma}_\varepsilon$, on sait que :

$$\widehat{\sigma}^2_\varepsilon = \frac{SCR}{n-2} \tag{1.37}$$

où $SCR = Sum\ squared\ resid = \sum_{t=1}^{n} e_t^2 = 79.03797$, d'ou

$\widehat{\sigma}^2_\varepsilon = \frac{79.03797}{15-2} = 6.079844$, et par conséquent :

$\widehat{\sigma}_\varepsilon = \sqrt{6.07984} = 2.465734$

♣ $4 \rightarrow F - statistic = F_c$, sa valeur est donnée par la formule suivante :

$$F_c = \frac{R^2/1}{(1-R^2)/(n-2)} = \frac{0.978944}{(1-0.978944)/(15-2)} = 604.40121 \tag{1.38}$$

3. Comme $F_c = 604.40121 > F_{\frac{\alpha}{2}}(1, n-2) = F_{0.05}(1, 15-2) = 4.67$ et $R^2 = 0.978944$ est très proche de 1, alors ce modèle est globalement significatif au seuil 10%.

4. les deux paramètres sont significatif au seuil 5% car la probabilité de significativité de chaque paramètre $Prod = 0.0000 < 0.025$.
   Donc, la variable X explique bien le modèle et que ce modèle contient une constante.

5.
$$B_1 = \widehat{B_1} \pm t_{0.025}(15-2)\widehat{\sigma}_{\widehat{B_1}} = 0.765572 \pm 2.160(0.031140) \tag{1.39}$$

$$I_{C\ B_1} = [0.69831, 0.83283] \tag{1.40}$$

   Puisque $0 \notin I_{C\ B_1}$, alors B_1 est significatif et la variable X explique bien le modèle.

6. Avant de donner le tableau d'analyse de la variance (ANOVA) , on doit d'abord calculer SCT et SCE

   ♣ On sait que : $R^2 = 0.978944$, $SCR = \sum_{t=1}^{n} e_t^2 = 79.03797$ et que
   $R^2 = 1 - \frac{SCR}{SCT}$.
   Alors

$$SCT = \frac{SCR}{(1-R^2)} = \frac{79.03797}{(1-0.978944)} = 3753.70298 \tag{1.41}$$

   ♣ On a $SCT = SCE + SCR$, donc

$$SCE = SCT - SCR = 3753.70298 - 79.03797 = 3674.66501 \tag{1.42}$$

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression X_1	$SCE = 3674.66501$	1	$\frac{SCE}{1} = 3674.66501$
Résiduelle	$SCR = 79.03797$	$n-2 = 15-2 = 13$	$\frac{SCR}{n-2} = \frac{79.03797}{13} = 6.07984$
Totale	$SCT = 3753.70298$	$n-1 = 15-1 = 14$	

exemple 3.
Nous souhaitons exprimer la hauteur Y d'un arbre en fonction de son diamètre X. Pour cela, nous avons obtenu les résultats suivants :

$$\overline{X} = 34.9;\ \frac{1}{20}\sum_{1}^{20}(X_i - \overline{X})^2 = 28.29;\ \overline{Y} = 18.34;$$

$$\frac{1}{20}\sum_{1}^{20}(Y_i - \overline{Y})^2 = 2.85;\ \frac{1}{20}\sum_{1}^{20}(X_i - \overline{X})(Y_i - \overline{Y}) = 6.26$$

1. On note $\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta}X_i,\ \ i = \overline{1,20}$, alors donner la valeur de chaque paramètre ($\widehat{\alpha}$ et $\widehat{\beta}$).

2. Quelle est la qualité de cet ajustement ? commenter.

3. Les paramètres de ce modèle sont-ils significativement non nuls ?

Corrigé 3.

$$\overline{X} = 34.9;\ \frac{1}{20}\sum_{1}^{20}(X_i - \overline{X})^2 = 28.29;\ \overline{Y} = 18.34;$$

$$\frac{1}{20}\sum_{1}^{20}(Y_i - \overline{Y})^2 = 2.85;\ \frac{1}{20}\sum_{1}^{20}(X_i - \overline{X})(Y_i - \overline{Y}) = 6.26.$$

1. On sait que : $\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta}X_i,\ \ i = \overline{1,20}$

   $\widehat{\beta} = \frac{\sum_{1}^{20}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{1}^{20}(X_i - \overline{X})^2} = \frac{20*6.26}{20(28.29)} = 0.22$

   Et $\widehat{\alpha} = \overline{Y} - \widehat{\beta}\ \overline{X} = 18.34 - 0.22(34.9) = 10.66$

   Donc le modèle est :

$$\widehat{Y}_i = 10.66 + 0.22\ X_i,\ \ i = \overline{1,20} \tag{1.43}$$

2. $R^2 = \frac{SCE}{SCT}$
On sait que :

$$\begin{cases} SCE = \widehat{\beta}^2 \sum_{1}^{20}(X_i - \overline{X})^2, \\ SCT = \sum_{1}^{20}(Y_i - \overline{Y})^2, \end{cases} \tag{1.44}$$

Donc

$$R^2 = \frac{\widehat{\beta}^2 \sum_{1}^{20}(X_i - \overline{X})^2}{\sum_{1}^{20}(Y_i - \overline{Y})^2} = \frac{(0.22)^2(28.29)20}{20(2.85)} = 0.48 \tag{1.45}$$

D'après la valeur du R^2, on déduit que la qualité de l'ajustement est moyenne.

3. $SCR = SCT - SCE = 20\,(2.85 - (0.22)^2(28.29)) = 29.62$

$$\widehat{\sigma}_{\varepsilon}^2 = \frac{SCR}{n-2} \tag{1.46}$$

$\widehat{\sigma}_{\varepsilon}^2 = \frac{29.62}{20-2} = 1.65$

$$\widehat{\sigma}_{\widehat{\beta}}^2 = \frac{\widehat{\sigma}_{\varepsilon}^2}{\sum_{1}^{20}(X_i - \overline{X})^2} \tag{1.47}$$

$\widehat{\sigma}_{\widehat{\beta}}^2 = \frac{1.65}{20(28.29)} = 0.0029$
D'où $\widehat{\sigma}_{\widehat{\beta}} = 0.054$
Donc, le test de Student pour β est définie par :
$H_0 :\ \beta = 0$ contre $H_1 :\ \beta \neq 0$
Pour cela, on doit calculer $t_{c\ \beta}$

$$t_{c_\beta} = \left| \frac{\widehat{\beta} - \beta}{\widehat{\sigma}_{\widehat{\beta}}} \right| \tag{1.48}$$

Sous H_0, on obtient :
$t_{c_\beta} = \left|\frac{0.22}{0.054}\right| = 4.07 > t_T(18, 0.025) = 2.101$, $t_T(18, 0.025)$ est la valeur de la statistique de Student lu sur la table avec 18 est degré de liberté et 0.025 est le seuil de confiance.
Donc, la variable X_i est significative.
De la même manière, on va étudier la significativité de α

Le test de Student pour α est définie par :

$H_0 : \ \alpha = 0$ contre $H_1 : \ \alpha \neq 0$
Pour cela, on doit calculer $t_{c\ \alpha}$

$$t_{c_\alpha} = \left| \frac{\widehat{\alpha} - \alpha}{\widehat{\sigma}_{\widehat{\alpha}}} \right| \tag{1.49}$$

Dans un premier temps, on doit calculer $\widehat{\sigma}_{\widehat{\alpha}}$

$$\widehat{\sigma}^2_{\widehat{\alpha}} = \widehat{\sigma}^2_{\varepsilon} \left[\frac{\overline{X}^2}{\sum_1^{20} (X_i - \overline{X})^2} + 1 \right] \tag{1.50}$$

$\widehat{\sigma}^2_{\widehat{\alpha}} = 1.65 \left[\frac{(34.9)^2}{20(28.29)} + 1 \right] = 5.20$

$\widehat{\sigma}_{\widehat{\alpha}} = 2.28$.

Alors, sous H_0, on obtient :
$t_{c_\alpha} = \left| \frac{10.66}{2.28} \right| = 4.68 > t_T(18, 0.025) = 2.101$, $t_T(18, 0.025)$ est la valeur de la statistique de Student lu sur la table avec 18 est degré de liberté et 0.025 est le seuil de confiance.
D'après ce résultat, on conclut que α est significatif. Ce qui ne permis de dire que ce modèle est validé, mais il faut ajouter d'autre variables pour améliorer le R^2.

1.9 *Énoncés des exercices*

exercice 1.
Un père a deux garçons, et s'inquiète de la croissance de son cadet qu'il trouve petit. Il décide de faire un modèle familial à partir des mesures de taille en fonction de l'âge de l'aîné :

$Y_t = taille$	96	104.8	112.3	115.3	124.9	127.4	132.8	136	136.7	144.5
$X_t = âge$	3	4	5	6	7	8	9	10	11	12

1. les données sur un graphique et justifier l'utilisation d'un modèle de régression linéaire simple.

2. Déterminer les estimateurs B_0, B_1 et étudier la significativité de chaque paramètre.

3. Calculer le SCE, SCT, SCR et le coefficient de détermination. Conclusion ?

4. Donner la prévision à la date 11 sachant que $X_{11} = 13.5$

5. Donner le tableau de l'analyse de la variance.

exercice 2.
Douze personnes sont inscrites à une formation. Au début de la formation, ces stagiaires subissent une épreuve A notée sur 20. A la fin de la formation, elles subissent une épreuve B de niveau identique. Les résultats sont donnés dans le tableau suivant :

Epreuve A	3	4	6	7	9	10	9	11	12	13	15	4
Epreuve B	8	9	10	13	15	14	13	16	13	19	6	19

1. Représenter le nuage de points. Déterminer la droite de régression. Calculer le coefficient de détermination. Commenter.

2. Deux stagiaires semblent se distinguer des autres. Les supprimer et déterminer la droite de régression sur les dix points restants. Calculer le coefficient de détermination. Commenter.

exercice 3.
Le tableau ci-dessous donne l'évolution du nombre de personnes agées en milliers (Y) de plus de 85 *ans*, en France métropolitaine, de 1950 à 2000.

Année	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000
X	0	5	10	15	20	25	30	35	40	45	50
Y	201	231	290	361	423	498	567	684	874	1079	1267

1. Ajuster un modèle exponentiel de la forme $Y = \alpha e^{B\ X}$, cet ajustement est-il correct ?

2. Déterminer, au niveau 95%, l'intervalle de prévison du nombre de personnes agées de plus de 85 *ans* en 2010.

Chapitre 2

Modèle de régression linéaire multiple

Le modèle de régression linéaire multiple est une extension du modèle de régression simple lorsque les variables explicatives sont en nombre quelconque (c-à-d la relation entre la variable dépendante Y avec plusieurs variables indépendantes X_i, ce qui va nous permettre de faire ensuite des prévisions de Y lorsque $X_i$est mesurée.

2.1 *Présentation du modèle*

Nous supposons donc que les données collectées suivent le modèle défini par :

$$Y_t = B_0 + B_1X_{1t} + B_2X_{2t} + B_3X_{3t} + ... + B_kX_{kt} + \varepsilon_t, \quad t = \overline{1,n} \tag{2.1}$$

où

- Y_t est la variable dépendante ou endogène (une variable à expliquer) au temps t.
- B_0, ..., B_k sont les coefficients de la régression.
- X_{it} est la variable indépendante ou exogène i (variable explicative) au temps t.
- ε_t est une erreur aléatoire.
- k est le nombre de variables explicatives non aléatoires.
- n est le nombre d'observations.

2.2 *la forme matricielle*

En utilisant l'écriture précédente et en écrivant le modèle observation par observation, nous obtenons :

$$Y_1 = B_0 + B_1X_{11} + B_2X_{21} + B_3X_{31} + ... + B_kX_{k1} + \varepsilon_1$$
$$Y_2 = B_0 + B_1X_{12} + B_2X_{22} + B_3X_{32} + ... + B_kX_{k2} + \varepsilon_2$$
$$...$$
$$...$$
$$...$$
$$Y_n = B_0 + B_1X_{1n} + B_2X_{2n} + B_3X_{3n} + ... + B_kX_{kn} + \varepsilon_n$$

Ce qui nous permis d'obtenir la forme matricielle suivante :

$$\underset{(n,1)}{Y} = \underset{(n,k+1)}{X}\ \underset{(k+1,1)}{B} + \underset{(n,1)}{\varepsilon} \tag{2.2}$$

où

- Y est un vecteur aléatoire de dimension n,
- X est une matrice de taille $n \times (k+1)$ connue, appelée matrice du plan d'expérience,
- B est le vecteur des paramètres inconnus du modèle,
- ε est le vecteur des erreurs.

Avec

$$\underset{(n,1)}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} ; \underset{(n,k+1)}{X} = \begin{pmatrix} 1 & X_{11} & X_{21} & . & . & . & X_{k1} \\ 1 & X_{12} & X_{22} & . & . & . & X_{k2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & X_{1n} & X_{2n} & . & . & . & X_{kn} \end{pmatrix} ; \underset{(k+1,1)}{B} = \begin{pmatrix} B_0 \\ B_1 \\ . \\ . \\ . \\ B_k \end{pmatrix} ; \underset{(n,1)}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}.$$

Ce modèle doit satisfaire les hypothèses suivantes :

H_1 : les valeurs X_{it} sont observées sans erreurs.
H_2 : La moyenne des erreurs s'annulent, le modèle est bien spécifié, $E(\varepsilon_t) = 0$.
H_3 : La variance de l'erreur est constante et ne dépend pas des observations : homoscédasticité, $Var(\varepsilon_t) = \sigma_\varepsilon^2$.
H_4 : L'erreur est indépendante des variables exogènes, $COV(\varepsilon_t, X_{it}) = 0$.
H_5 : Indépendance des erreurs, les erreurs relatives à 2 observations sont indépendantes (on dit aussi que les erreurs sont non corrélées), $COV(\varepsilon_t, \varepsilon_s) = 0$.
H_6 : $(X^{'}X)$ est régulière et inversible ($(X^{'}X)^{-1}$ existe).
H_7 : $\frac{(X^{'}X)}{n}$ tend vers une matrice finie non singulière
H_8 : $n > k + 1$ Nombre d'observations est supérieur aux nombre des séries explicatives.

2.3 *Estimation des paramètres par la méthode Moindres carrés ordinaires (MCO)*

Comme pour la régression linéaire simple, nous allons estimer le vecteur des paramètres B par la méthode Moindres carrés ordinaires (MCO), en supposant que : $Y = X\ B\ +\ \varepsilon$. Pour cela nous minimisons la somme des carrés des erreurs.

$$\min \sum_{t=1}^{n} \varepsilon^2 = \min \left(\varepsilon^{'} \varepsilon\right) = \min \left(Y - X\ \widehat{B}\right)^{'} \left(Y - X\ \widehat{B}\right) = \min S \tag{2.3}$$

Et

$$S = \left(Y - X\ \widehat{B}\right)^{'} \left(Y - X\ \widehat{B}\right) = \left(Y'Y - Y'X\widehat{B} - \widehat{B}'X'Y + \widehat{B}'X'X\widehat{B}\right) = \left(Y'Y - 2\widehat{B}'X'Y + \widehat{B}'X'X\widehat{B}\right)$$

avec $\varepsilon^{'}$ est le transposé du vecteur ε. Pour minimiser S, il suffit de différencier S par rapport au vecteur B (résoudre cette équation $\frac{\partial S}{\partial B} = 0$) et on obtient :

$$\frac{\partial S}{\partial B} = -2X'Y + 2X'X\ \widehat{B} = 0.$$

Par conséquent, on obtient :

$$\widehat{B} = (X'X)^{-1}X'Y. \tag{2.4}$$

Avec

$$(X^{'}X) = \begin{pmatrix} n & \sum X_{1t} & \sum X_{2t} & ... & \sum X_{kt} \\ \sum X_{1t} & \sum X_{1t}^2 & \sum X_{2t}X_{1t} & ... & \sum X_{kt}X_{1t} \\ \sum X_{2t} & \sum X_{2t}X_{1t} & \sum X_{1t}^2 & ... & \sum X_{kt}X_{2t} \\ . & . & . & . & \\ . & . & . & . & \\ . & . & . & . & \\ \sum X_{kt} & \sum X_{1t}X_{kt} & \sum X_{2t}X_{kt} & ... & \sum X_{kt}^2 \end{pmatrix}$$

Et

$$(X^{'}Y) = \begin{pmatrix} \sum Y_t \\ \sum X_{1t}Y_t \\ . \\ . \\ . \\ \sum X_{kt}Y_t \end{pmatrix}$$

Cette solution existe car $(X^{'}X)^{-1}$ existe par hypothèse.

Remarque

1. Si les variables sont centrées, alors $\frac{(X^{'}X)}{n} =$ *matrice de variance covariance*

2. Si les variables sont centrées et réduites $\frac{(X^{'}X)}{n}$ = matrice de corrélation 1. Si B_0 est constant, alors

$$\widehat{B_1} = \frac{\sum_{t=1}^{n} X_t Y_t}{\sum_{t=1}^{n} X_t^2} \tag{2.5}$$

2. Les estimateurs peuvent s'écrire :

$$\begin{cases} \widehat{B_1} = B_1 + \frac{\sum_{t=1}^{n} \left(X_t - \overline{X}\right) \varepsilon_t}{\sum_{t=1}^{n} \left(X_t - \overline{X}\right)^2} \\ \widehat{B_0} = B_0 - \left(\widehat{B_1} - B_1\right) \overline{X} + \overline{\varepsilon} \end{cases} \tag{2.6}$$

2.4 *Propriétés des estimateurs*

L'estimateur est Sans biais

Comme en regression simple, l'estimateur obtenu est sans biais. Car :

$$\widehat{B} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(XB + \varepsilon) = B + (X'X)^{-1}X'\varepsilon$$

D'après l'hypothèse H_2, on déduit que $E[\widehat{B}] = B$

Matrice de variance covariance de $\widehat{B}$

On appelle la matrice de variance covariance du vecteur aleatoire la matrice de dispersion, qui est donnée par :

$$\Omega_{\widehat{B}} = E\left[\left(\widehat{B} - B\right)\left(\widehat{B} - B\right)'\right]$$

$$\Omega_{\widehat{B}} = \begin{pmatrix} V(\widehat{B_0}) & Cov(\widehat{B_0}, \widehat{B_1}) & \dots \\ & V(\widehat{B_1}) & \\ & & V(\widehat{B_k}) \end{pmatrix}$$

la variance de l'estimateur de chaque coefficient, se trouve sur la diagonale de $\Omega_{\widehat{B}}$.

On sait que :

$\widehat{B} = B + (X'X)^{-1}X'\varepsilon$, ce qui implique que

$$\Omega_{\widehat{B}} = E\left[(X'X)^{-1}X'\varepsilon\left((X'X)^{-1}X'\varepsilon\right)'\right] = E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right]$$
$$= (X'X)^{-1}X'E\left[\varepsilon\varepsilon'\right]X(X'X)^{-1}$$

Et

$$E\left[\varepsilon\varepsilon'\right] = \begin{pmatrix} E[\varepsilon_1^2] & E[\varepsilon_1\varepsilon_2] & \dots & E[\varepsilon_1\varepsilon_n] \\ & E[\varepsilon_2^2] & & \vdots \\ & & & \\ & & & E[\varepsilon_n^2] \end{pmatrix}$$

Et d'après les hypothèses H_2 et H_3, on déduit que :

$$E\left[\varepsilon\varepsilon'\right] = \sigma_\varepsilon^2\mathbf{I}_n$$

D'où

$$\Omega_{\widehat{B}} = \sigma_\varepsilon^2(X'X)^{-1} \tag{2.7}$$

Cette matrice tend vers la matrice nulle (toutes les cellules à 0) lorsque $n \to +\infty$ (hypothèses H_7), ce qui signifie que les estimateurs des moindres carrés sont convergents et à variance minimale .

Remarque

Dans la plupart des cas la variance des erreurs (σ_ε^2) est inconnue.

Détermination d'un estimateur Sans biais de la variance de l'erreur σ_ε^2

L'estimateur Sans biais de σ_ε^2 est donnée par :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{t=1}^{n} e_t^2}{(n-k-1)} \tag{2.8}$$

Preuve

Notons que $\sum_{t=1}^{n} e_t^2 = e^{'} e$

Avec $e = Y - \widehat{Y}$, et $\widehat{Y} = X\widehat{B}$

Alors $e = \varepsilon - X\left(\widehat{B} - B\right) = \left[\mathbf{I}_n - X(X^{'}X)^{-1}X^{'}\right]\varepsilon = \Gamma\varepsilon$

On peut facilement montrer que la matrice Γ est symétrique ($\Gamma^{'} = \Gamma$, car $(X^{'}X)$ est symétrique) et elle est idempotente d'ordre 2 ($\Gamma^2 = \Gamma$), de taille (n, n).

On déduit, alors :

$\sum_{t=1}^{n} e_t^2 = \varepsilon^{'}\Gamma\varepsilon$, et

$E\left[\sum_{t=1}^{n} e_t^2\right] = \sigma_\varepsilon^2 \mathbf{I}_n Tr(\Gamma)$ ($Tr(\Gamma)$ est la trace de la matrice Γ).

$E\left[\sum_{t=1}^{n} e_t^2\right] = \sigma_\varepsilon^2 \mathbf{I}_n (n-k-1)$. Par conséquent, on peut considérer l'estimateur sans biais de la variance de l'erreur suivant :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{t=1}^{n} e_t^2}{(n-k-1)} \tag{2.9}$$

Remarque

D'après le résultat précédent, on déduit que l'estimateur de la matrice variance covariane est donné par :

$$\widehat{\Omega}\varepsilon = \widehat{\sigma}_\varepsilon^2 (X^{'}X)^{-1} \tag{2.10}$$

Où, les estimateurs des variances des paramètres du modèle sont sur la diagonale de la matrice $\widehat{\Omega}\varepsilon$

2.5 *Equation d'analyse de la variance et la qualité de l'ajustement*

2.5.1 *Equation d'analyse de la variance*

L'equation d'analyse de la variance est donnée par :

$$\sum_{t=1}^{n} \underbrace{\left(Y_t - \overline{Y}\right)^2}_{SCT} = \sum_{t=1}^{n} \underbrace{\left(Y_t - \widehat{Y}_t\right)^2}_{SCR} + \sum_{t=1}^{n} \underbrace{\left(\widehat{Y}_t - \overline{Y}\right)^2}_{SCE} \tag{2.11}$$

SCT : variabilité totale

SCE : Variabilité expliquée par le modèle

SCR : Variabilité non-expliquée (Variabilité résiduelle)

Les estimateurs sont d'autant plus précis lorsque La variance de l'erreur est faible et quand la dispersion des X est forte.

Tableau d'analyse de la variance (ANOVA)

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression $X_1, \ , X_2, ..., \ X_k$	$SCE = \sum_{t=1}^{n} \left(\widehat{Y}_t - \overline{Y}\right)^2$	k	$\frac{SCE}{k}$
Résiduelle	$SCR = \sum_{t=1}^{n} \left(Y_t - \widehat{Y}_t\right)^2$	$n-k-1$	$\frac{SCR}{n-k-1}$
Totale	$SCT = \sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)^2$	$n-1$	

2.5.2 *La qualité de l'ajustement*

Le coefficient de détermination R^2

Les valeurs données par l'equation d'analyse de la variance dépendent des unités de mesure, c'est pourquoi on préfère utiliser le nombre sans dimension R^2.

$$R^2 = \frac{\sum_{t=1}^{n} \left(\widehat{Y}_t - \overline{Y}\right)^2}{\sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)^2} = 1 - \frac{\sum_{t=1}^{n} \left(Y_t - \widehat{Y}_t\right)^2}{\sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)^2} \tag{2.12}$$

ou

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{t=1}^{n} e_t^2}{SCT} \tag{2.13}$$

Le coefficient de détermination R^2, mesure la proportion de la variance de Y expliquée par la régression de Y sur X. La qualitée de l'ajustement est meilleur quand R^2 est très proche de 1.

Remarque

Dans le cas de données centrées le coefficient de détermination R^2 est défini comme suit :

$$R^2 = \frac{\widehat{Y}'\widehat{Y}}{Y'Y} = 1 - \frac{\widehat{e}'\widehat{e}}{Y'Y} = 1 - \frac{\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n} Y_t^2} \tag{2.14}$$

Le coefficient de détermination corrigé $\overline{R}^2$

D'aprés la formule du coefficient de détermination R^2, nous constatons qu'il ne tient pas compte ni du nombre d'observations n, ni du nombre variables explicative k. Donc, il faut condidérer un autre coefficient afin de tenir compte n et k. Ce coefficient est appelé le coefficient de détermination corrigé noté par $\overline{R}^2$ et qui est défini par :

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2) \tag{2.15}$$

Alors, la qualitée de l'ajustement est meilleur lorsque $\overline{R}^2$ est très proche de 1.

Remarque

1. Si $k = 0 \Longrightarrow \overline{R}^2 = R^2$
2. Si $k \geq 1 \Longrightarrow \overline{R}^2 < R^2$
3. Quand $n \to +\infty \Longrightarrow \overline{R}^2 \simeq R^2$

2.6 *Tests Statistiques*

Afin de faire des tests Statistiques, on doit introduire une Hypothèse supplémentaire qui est celle de la normalité des erreurs : $H_9 : \varepsilon_t \rightsquigarrow N(0, \sigma_\varepsilon^2)$

2.6.1 *Test de significativité globale du modèle (Fisher)*

Le test de Fisher nous permet de tester s'il existe au moins une variable X_i qui explique la variable Y, ce test est défini par les deux hypothèses suivantes :
$H_0 : B_1 = B_2 = ... = B_k = 0$, aucune variable n'explique la variable Y
$H_1 : \exists i = \overline{1,k} / B_i \neq 0$, Le modèle est globalement significatif (il y a au moins une variable explicative)
Pour cela, nous devons calculer la statistique de Fisher empirique F_c afin de la comparer à celle lu sur la table $F_{\frac{\alpha}{2}}(k, n-k-1)$, où $(k, n-k-1)$ sont les degrés de liberté la statistique de Fisher au seuil $\frac{\alpha}{2}$.

$$F_c = \frac{SCE/k}{SCR/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{2.16}$$

Aolrs :

- si $F_c > F_{\frac{\alpha}{2}}(k, n-k-1)$, alors le modèle est globalement significatif.
- si $F_c < F_{\frac{\alpha}{2}}(k, n-k-1)$, alors le modèle n'est pas globalement significatif, et que toutes les variables X_i ne sont pas explicatives (nous acceptons l'hypothèse H_0).

2.6.2 *Test de Student*

D'après l'hypothèse H_9, les erreurs suivent une loi normale, ce qui nous permet de vérifier que :

$$\begin{cases} \frac{\widehat{B_i} - B_i}{\sigma^2_{\widehat{B_i}}} \rightsquigarrow N(0,1), \\ \\ \frac{\sum_{t=1}^{n} e_t^2}{\sigma_\varepsilon^2} = \frac{(n-k-1)\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \rightsquigarrow \chi^2_{(n-k-1)}, \end{cases}$$

où $\chi^2_{(n-k-1)}$ est le khi-deux à $(n-k-1)$ degrés de liberté.

Et $\frac{\widehat{B_i}-B_i}{\widehat{\sigma}^2_{\widehat{B_i}}} \rightsquigarrow t_{\frac{\alpha}{2}}(n-k-1)$.
avec $t_{\frac{\alpha}{2}}(n-k-1)$ est la Student à $(n-k-1)$ degrés de liberté.

Par conséquent, nous pouvons mettre en place des tests statistiques (tests bilatéraux) pour vérifier la significativité des B_i.

Test de Student

$H_0 : B_i = 0$ contre $H_1 : B_i \neq 0$. Dans un premier, nous douvons calculer la statistique de Student empirique t_c afin de la comparer à celle lu sur la table $t_{\alpha/2}(n-k-1)$, où $(n-k-1))$ est le degré de liberté la statistique de Student au seuil $\alpha/2$, avec

$$t_c = \left| \frac{\widehat{B_i} - B_i}{\widehat{\sigma}^2_{\widehat{B_i}}} \right| \tag{2.17}$$

Aolrs :

- si $t_c > t_{\alpha/2}(n-k-1)$, alors B_i est significatif.
- si $F_c < t_{\alpha/2}(n-k-1)$, alors B_i n'est pas significatif, et que la variable X_i n'est pas explicative (nous acceptons l'hypothèse H_0).

Intervalle de confiance

Nous pouvons confirmer les résultats obtenus par le test bilatéral de Student à partir de l'intervalle de confiance de chaque paramètre. Dans la plus part des cas σ_ε est inconnu, alors

L'intervalle de confiance pour B_i (I_{CB_i})

$$B_i = \widehat{B_i} \pm t_{\alpha/2}(n-k-1)\widehat{\sigma}_{\widehat{B_i}} \tag{2.18}$$

où $t_{\alpha/2}(n-k-1)$ est la valeur de la statistique de Student au seuil $\alpha/2$ à $(n-k-1)$ est le degré de liberté lu sur la table.

Remarque

$H_0 : B_i = 0$ contre $H_1 : B_i \neq 0$.

- si $0 \notin I_{C\ B_i}$, alors B_i est significatif.
- si $0 \in I_{CB_i}$, alors B_i n'est pas significatif, et que la variable X_i n'est pas explicative (nous acceptons l'hypothèse H_0).

2.7 *Prévision*

Une fois le modèle est validé (toutes les variables sont significatives), nous pouvons faire des prévisions pour Y_{n+h} lorsque nous connaisons les valeurs $X_{i\ n+h}$.

$$Y_{n+h} = \widehat{B}_0 + \widehat{B}_1 X_{1\ n+h} + \widehat{B}_2 X_{2\ n+h} + ...\widehat{B}_k X_{k\ n+h} + \varepsilon_{n+h} \tag{2.19}$$

Sachant que : $E[\varepsilon_{n+h}] = 0$, et $Cov(\varepsilon_{n+h}, \varepsilon_t) = 0$ et $Var(\varepsilon^2_{n+h}) = \sigma^2_\varepsilon$
Dans un premier temps, nous douvons calculer $\widehat{Y}_{n+h}$ (la valeur ponctuelle ajustée de la prévision) qui est donné par :

$$\widehat{Y}_{n+h} = \widehat{B}_0 + \widehat{B}_1 X_{1\ n+h} + \widehat{B}_2 X_{2\ n+h} + ...\widehat{B}_k X_{k\ n+h} \tag{2.20}$$

Et

$$Y_{n+h} = \widehat{Y}_{n+h} + e_{t+h} \tag{2.21}$$

avec $E[\varepsilon_{n+h}] = 0$, et $Cov(\varepsilon_{n+h}, \varepsilon_t) = 0$ et $Var(\varepsilon^2_{n+h}) = \sigma^2_\varepsilon$ L'erreur de prévision est donnée par :

$$e_{t+h} = Y_{n+h} - \widehat{Y}_{n+h}$$

Remarque

$E[e_{n+h}] = 0$ et la valeur ajustée de la prévision $\widehat{Y}_{n+h}$ est sans biais. On peut facilement montrer que $E\left[\widehat{Y}_{n+h}\right] = Y_{n+h}$.

La variance de l'erreur de prévision

On sait que :
$e_{n+h} = \widehat{Y}_{n+h} - Y_{n+h} = X'_{n+h}\left(B - \widehat{B}\right) + \varepsilon_{n+h}$
Où :

$$X'_{n+h} = (1, X_{1\ n+h}, X_{2\ n+h}, .., X_{k\ n+h})$$

Alors
$Var(e_{n+h}) = Var\left[X'_{n+h}\left(B - \widehat{B}\right) + \varepsilon_{n+h}\right] = Var\left[X'_{n+h}\left(B - \widehat{B}\right)\right] + Var\left[\varepsilon_{n+h}\right]$

$Var(e_{n+h}) = X'_{n+h}Var\left[\left(B - \widehat{B}\right)\right]X_{n+h} + Var\left[\varepsilon_{n+h}\right] = X'_{n+h}Var\left[\widehat{B}\right]X_{n+h} + Var\left[\varepsilon_{n+h}\right]$

D'après les résultats précédents, on déduit que :

$$Var(e_{n+h}) = \sigma_\varepsilon^2\left[X'_{n+h}(X'X)^{-1}X_{n+h} + 1\right] \tag{2.22}$$

Et comme la variance de l'erreur σ_ε^2 est inconnue, alors la variance de l'erreur de prévision est donnée par :

$$Var(e_{n+h}) = \widehat{\sigma}_\varepsilon^2\left[X'_{n+h}(X'X)^{-1}X_{n+h} + 1\right] \tag{2.23}$$

Remarque

Nous constatons, comme pour le modèle de régression simple que la valeur de la variance de prévision est faible lorsque les valeurs prévues des variables exogènes se rapprochent de leurs moyennes.

L'intervalle de prévision

D'après l'hypothèse H_9, nous déduisons que :

$$\varepsilon_{n+h} \rightsquigarrow N\left(0, \sigma^2_{e_{n+h}}\right).$$

L'intervalle de prédiction est défini par :

$$Y_{n+h} = \widehat{Y}_{n+h} \pm t_{\frac{\alpha}{2}}(n-k-1)\ \widehat{\sigma}_\varepsilon\sqrt{\left[X'_{n+h}(X'X)^{-1}X_{n+h} + 1\right]}. \tag{2.24}$$

Où $t_{\frac{\alpha}{2}}(n-k-1)$ est la valeur de la loi de Student à $(n-k-1)$ degrés de liberté au seuil de signification $\frac{\alpha}{2}$.

2.8 *Exemple*

exemple 4. Pendant dix ans, de 1995 à 2004, une ferme a expérimenté le rendement du maïs Y associé à l'emploi de quantités croissantes d'un fertilisant X_1 et d'un insecticide X_2. Les données sont :

X_{1t}	6	10	12	14	16	18	22	24	26	32
X_{2t}	4	4	5	7	9	12	14	20	21	24
Y_t	40	44	46	48	52	58	60	68	74	80

Soit le modèle : $Y_t = B_0 + B_1 X_{1t} + B_2 X_{2t} + \varepsilon_t,\ t = \overline{1,n}$

1. Mettre le modèle sous forme matricielle en spécifiant les dimensions de chacune des matrices.

2. Estimer par la méthode des moindres carrés ordinaires les paramètres du modèle.

3. Calculer la variance résiduelle ainsi que les écarts-types de chacun des paramètres.

4. Calculer le coefficient de détermination et le coefficient de détermination corrigé. Conclusion ?

5. Le modèle est-il globalement significatif au seuil 5% ?

6. Les variables explicatives sont-elles significatives au seuil 5% ?

7. Donner la valeur de Y à la date 11 sachant que :$X_{1\ 11} = 36$, $X_{2\ 11} = 27$

8. Donner le tableau de l'analyse de la variance.

Corrigé 4.

1. La forme matricielle est par l'equation suivante :

$$\underset{(n=10,1)}{Y} = \underset{(n=10,k+1=3)}{X}\ \underset{(k+1=3,1)}{B} + \underset{(n=10,1)}{\varepsilon} \tag{2.25}$$

où

- Y est un vecteur aléatoire de dimension $n = 10$,
- X est une matrice de taille $n \times (k+1) = 10(3) = 30$, connue, appelée matrice du plan d'expérience,
- B est le vecteur des paramètres inconnus du modèle,
- ε est le vecteur des erreurs.

Avec

$$\underset{(n=10,1)}{Y} = \begin{pmatrix} Y_1 = 40 \\ Y_2 = 44 \\ Y_3 = 46 \\ Y_4 = 48 \\ Y_5 = 52 \\ Y_6 = 58 \\ Y_7 = 60 \\ Y_8 = 68 \\ Y_9 = 74 \\ Y_{10} = 80 \end{pmatrix}; \quad \underset{(n=10,k+1=3)}{X} = \begin{pmatrix} 1 & X_{1\ 1} = 6 & X_{2\ 1} = 4 \\ 1 & X_{1\ 2} = 10 & X_{2\ 2} = 4 \\ 1 & X_{1\ 3} = 12 & X_{2\ 3} = 5 \\ 1 & X_{1\ 4} = 14 & X_{2\ 4} = 7 \\ 1 & X_{1\ 5} = 16 & X_{2\ 5} = 9 \\ 1 & X_{1\ 6} = 18 & X_{2\ 6} = 12 \\ 1 & X_{1\ 7} = 22 & X_{2\ 7} = 14 \\ 1 & X_{1\ 8} = 24 & X_{2\ 8} = 20 \\ 1 & X_{1\ 9} = 26 & X_{2\ 9} = 21 \\ 1 & X_{1\ 10} = 32 & X_{2\ 10} = 24 \end{pmatrix};$$

$$\underset{(k+1=3,1)}{B} = \begin{pmatrix} B_0 \\ B_1 \\ B_2 \end{pmatrix}; \quad \underset{(n=10,1)}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_{n=10} \end{pmatrix}.$$

2. Nous allons estimer le vecteur des paramètres B par la méthode Moindres carrés ordinaires (MCO), en supposant que : $Y = X\ B\ +\ \varepsilon$.
   L'estimer du vecteur des paramètres B est donné par :

$$\widehat{B} = (X'X)^{-1}X'Y. \tag{2.26}$$

Afin de caculer les estimateurs des coefficients de régression, nous devons calculer dans un premier temps $(X'X)^{-1}$.

$$(X'X) = \begin{pmatrix} n & \sum X_{1t} & \sum X_{2t} \\ 10 & 180 & 120 \\ \sum X_{1t} & \sum X_{1t}^2 & \sum X_{1t}X_{2t} \\ 180 & 3816 & 2684 \\ \sum X_{2t} & \sum X_{2t}X_{1t} & \sum X_{2\ t}^2 \\ 120 & 2684 & 1944 \end{pmatrix}$$

Où, les résultats utilisés sont donnés dans le tableau suivant :

n	Y_t	$X_{1\ t}$	$X_{2\ t}$	$X_{1\ t}X_{2\ t}$	$X_{1\ t}^2$	$X_{2\ t}^2$	$X_{1\ t}Y_t$	$X_{2\ t}Y_t$	Y_t^2
1	40	6	4	24	36	16	240	160	1600
2	44	10	4	40	100	16	440	176	1936
3	46	12	5	60	144	25	552	230	2116
4	48	14	7	98	196	49	672	336	2304
5	52	16	9	144	256	81	832	468	2704
6	58	18	12	216	324	144	1044	696	3364
7	60	22	14	308	484	196	1320	840	3600
8	68	24	20	480	576	400	1632	1630	4624
9	74	26	21	546	676	441	1924	1554	5476
10	80	32	24	786	1024	676	2560	1920	6400
$\sum$	570	180	120	2684	3816	1944	11216	7740	34124

$$\det(X'X) = \begin{vmatrix} 10 & 180 & 120 \\ 180 & 3816 & 2684 \\ 120 & 2684 & 1944 \end{vmatrix}$$

$$\det(X^{'}X) = 10 \begin{vmatrix} 3816 & 2684 \\ 2684 & 1944 \end{vmatrix} - 180 \begin{vmatrix} 180 & 120 \\ 2684 & 1944 \end{vmatrix} + 120 \begin{vmatrix} 180 & 120 \\ 3816 & 2684 \end{vmatrix} = 157280$$

La comatrice de $(X^{'}X)$

$$C_{(X'X)} = \begin{pmatrix} + \begin{vmatrix} 3816 & 2684 \\ 2684 & 1944 \end{vmatrix} & - \begin{vmatrix} 180 & 120 \\ 2684 & 1944 \end{vmatrix} & + \begin{vmatrix} 180 & 120 \\ 3816 & 2684 \end{vmatrix} \\ - \begin{vmatrix} 180 & 120 \\ 2684 & 1944 \end{vmatrix} & + \begin{vmatrix} 10 & 120 \\ 120 & 1944 \end{vmatrix} & - \begin{vmatrix} 10 & 180 \\ 120 & 2624 \end{vmatrix} \\ + \begin{vmatrix} 180 & 120 \\ 3816 & 2684 \end{vmatrix} & - \begin{vmatrix} 10 & 120 \\ 180 & 2684 \end{vmatrix} & + \begin{vmatrix} 10 & 180 \\ 180 & 3816 \end{vmatrix} \end{pmatrix}$$

$$C_{(X'X)} = \begin{pmatrix} 214448 & -27840 & 25200 \\ -27840 & 5040 & -5240 \\ 25200 & -5240 & 5760 \end{pmatrix}$$

$$(X^{'}X)^{-1} = \frac{C^{'}_{(X'X)}}{\det(X'X)} = \frac{1}{157280} \begin{pmatrix} 214448 & -27840 & 25200 \\ -27840 & 5040 & -5240 \\ 25200 & -5240 & 5760 \end{pmatrix}$$

Et par conséquent

$$\widehat{B} = \frac{C'_{(X'X)}}{\det(X'X)}(X'Y) = \frac{1}{157280}\begin{pmatrix} 214448 & -27840 & 25200 \\ -27840 & 5040 & -5240 \\ 25200 & -5240 & 5760 \end{pmatrix}\begin{pmatrix} 570 \\ 11216 \\ 7740 \end{pmatrix}$$

Avec

$$(X'Y) = \begin{pmatrix} \sum Y_t = 570 \\ \sum X_{1\ t} Y_t = 11216 \\ \sum X_{2\ t} Y_t = 7740 \end{pmatrix}$$

$$\widehat{B} = \frac{1}{157280}\begin{pmatrix} 5029920 \\ 102240 \\ 174560 \end{pmatrix} = \begin{pmatrix} \frac{5029920}{157280} \\ \frac{102240}{157280} \\ \frac{174560}{157280} \end{pmatrix}$$

$$\widehat{B} = \begin{pmatrix} 31.98 \\ 0.65 \\ 1.11 \end{pmatrix} \tag{2.27}$$

Donc, le modèle s'écrit comme suit :

$$\widehat{Y_t} = 31.98 + 0.65X_{1\ t} + 1.11X_{2\ t}, \quad t = \overline{1,10} \tag{2.28}$$

3. Pour le calcul de la variance des paramètres du modèle, on doit calculer la matrice variance covariance $\widehat{\Omega}_{\widehat{B}}$

$$\widehat{\Omega}_{\widehat{B}} = \widehat{\sigma}_{\varepsilon}^{2}(X^{'}X)^{-1} \tag{2.29}$$

Où

$$\widehat{\sigma}_{\varepsilon}^{2} = \frac{\sum_{t=1}^{n} e_i^2}{n-(k+1)}$$

Avec

$$\sum_{t=1}^{n} e_i^2 = \sum_{t=1}^{n} Y_t^2 - \widehat{B}^{'}\,(X^{'}Y) = 34124 - (31.98, 0.65, 1.11)\begin{pmatrix} 570 \\ 11216 \\ 7740 \end{pmatrix} = 34124 - 34110.4 = 13.6$$

Et on déduit que :

$$\widehat{\sigma}_{\varepsilon}^{2} = \frac{13.6}{10-3} = 1.94 \Rightarrow \widehat{\sigma}_{\varepsilon} = 1.39$$

La variance des paramètres du modèle se trouve sur la diagonale de la matrice variance covariance $\widehat{\Omega}_{\widehat{B}}$, c'est à dire :

$$\widehat{\sigma}_{\widehat{B}_0}^{2} = 1.94 \times 1.363 = 2.64 \Rightarrow \widehat{\sigma}_{\widehat{B}_0} = 1.63$$

$$\widehat{\sigma}_{\widehat{B}_1}^{2} = 1.94 \times 0.032 = 0.062 \Rightarrow \widehat{\sigma}_{\widehat{B}_1} = 0.25$$

$$\widehat{\sigma}_{\widehat{B}_2}^{2} = 1.94 \times 0.036 = 0.07 \Rightarrow \widehat{\sigma}_{\widehat{B}_2} = 0.26$$

4. ♣ Le calcul du coefficient de détermination R^2

$$R^2 = 1 - \frac{SCR}{SCT}. \tag{2.30}$$

♣ Le calcul du SCT

$$SCT = \sum_{t=1}^{n} Y_t^2 - n\overline{Y}^2 = 34124 - 10(57)^2 = 1634$$

Alors

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{13.6}{1634} = 0.99$$

♣ Le calcul du coefficient de détermination corrigé $\overline{R}^2$

$$\overline{R}^2 = 1 - \frac{n-1}{n-(k+1)}(1-R^2) = 1 - \frac{10-1}{10-3}(1-0.99) = 0.99 \tag{2.31}$$

D'après le résultat de R^2 et $\overline{R}^2$, on conclu que la qualité de l'ajustement est très importante.

5. Le test globale
   $H_0 : B_1 = B_2 = ... = B_k = 0$, aucune variable n'explique la variable Y
   $H_1 : \exists i = \overline{1,k} / B_i \neq 0$

$$F_c = \frac{SCE/k}{SCR/(n-(k+1))} = \frac{R^2/k}{(1-R^2)/(n-(k+1))} \tag{2.32}$$

$$F_c = \frac{0.99/2}{(1-0.99)/(10-3)} = 346.5$$

On accepte H_1, car $F_c = 346.5 > F_{\frac{\alpha}{2}}(k, n-(k+1)) = F_{0.05}(2,7) = 4.74$, ce qui signifie que le modèle est globalement significatif au seuil de 10%

6. On utilise le test de Student pour la significativité de B_i, qui est définie par :

   $H_0 : \ B_i = 0$ contre $H_1 : \ B_i \neq 0$
   Pour cela, on doit calculer $t_{c\ B_i}$

$$t_{c_{B_i}} = \left| \frac{\widehat{B_i} - B_i}{\widehat{\sigma}_{\widehat{B_i}}} \right| \tag{2.33}$$

On obtient alors les résultats suivants :

$$t_{c_{B_0}} = \frac{30.98 - 0}{1.63} = 19.62$$

$$t_{c_{B_1}} = \frac{0.65 - 0}{0.25} = 2.62$$

$$t_{c_{B_2}} = \frac{1.11 - 0}{0.26} = 4.27$$

On constate d'après la valeur de la Statistique de Student calculer pour chaque B_i qu'elle est superieure à la valeur de la statistique de Student ($t_T(7, 0.025) = 2.365$) lu sur la table avec 7 est son degré de liberté et 0.025 est le seuil de confiance. On conclut donc que tous les paramètres sont significatifs et que les deux variables sont significatives. Ce qui nous permis de dire que ce modèle est validé.

7. Le calcul de la prévision à la date 11 sachant que $X_{1\ 11} = 36$ et $X_{2\ 11} = 27$

   Du moment le modèle est validé, on peut donc calculer la prévision à la date 11 .

   ♣ La valeur ponctuelle ajustée de la prévision est donné par :

$$\widehat{Y}_{n+h} = \widehat{B}_0 + \widehat{B}_1 X_{1\ n+h} + \widehat{B}_2 X_{2\ n+h} + ...\widehat{B}_k X_{k\ n+h} \tag{2.34}$$

   Donc

$$\widehat{Y}_{11} = \widehat{B}_0 + \widehat{B}_1 X_{1\ 11} + \widehat{B}_2 X_{2\ 11} \tag{2.35}$$

   Donc

$$\widehat{Y}_{11} = 30.98 + 0.65 X_{1\ 11} + 1.11 X_{2\ 11} = 30.98 + 0.65(36) + 1.11(27) = 85.35$$

   ♣ L'intervalle de prédiction est défini par :

$$Y_{11} = \widehat{Y}_{11} \pm t_{0.025}(7)\ \sigma_{e_{n+h}} \tag{2.36}$$

Où $\sigma^2_{e_{n+h}}$ est la variance de l'erreur de prévision qui est donnée par :

$$\sigma^2_{e_{n+h}} = Var(e_{n+h}) = \widehat{\sigma}^2_{\varepsilon}\left[X'_{n+h}(X'X)^{-1}X_{n+h} + 1\right] \tag{2.37}$$

$$\sigma^2_{e_{n+h}} = \widehat{\sigma}^2_{\varepsilon}\left[X'_{11}(X'X)^{-1}X_{11} + 1\right]$$

Avec $X'_{11} = (1, 36, 27)$. Alors

$$\sigma^2_{e_{n+h}} = 1.94\left[\frac{1}{157280}(1, 36, 27)\begin{pmatrix} 214448 & -27840 & 25200 \\ -27840 & 5040 & -5240 \\ 25200 & -5240 & 5760 \end{pmatrix}\begin{pmatrix} 1 \\ 36 \\ 27 \end{pmatrix} + 1\right]$$

$$\sigma^2_{e_{n+h}} = 1.94\left[\frac{1}{157280}(1, 36, 27)\begin{pmatrix} -107392 \\ 12120 \\ -7920 \end{pmatrix} + 1\right] = 1.94\,[0.823 + 1] = 3.537$$

♣ L'intervalle de prédiction est défini par :

$$Y_{11} = 85.35 \pm 2.62(1.881) = [81.02\ ;\ 89.68]. \tag{2.38}$$

8. Le tableau de l'analyse de la variance

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression X_1, X_2	$SCE = \underset{=1634-13.6=1620.4}{SCT - SCR}$	2	$\frac{SCE}{2} = \frac{1620.4}{2} = 810.2$
Résiduelle	$SCR = 13.6$	$n - 3 = 10 - 3 = 7$	$\frac{SCR}{7} = \frac{13.6}{7} = 1.94$
Totale	$SCT = 1634$	$n - 1 = 10 - 1 = 9$	

ùe bien le modèle et que ce modèle contient une constante.

2.9 *Exercices*

exercice 4.

1. Soit le modèle : $y_t = B_0 + B_1 x_{1\ t} + B_2 x_{2\ t} + B_3 x_{3\ t} + \varepsilon_t, \quad i = \overline{1,8}$
   Estimer par la méthode des moindres carrés ordinaires les paramètres du modèle et interpréter les résultats, sachant que :
   $X'X = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}$, $X'Y = \begin{pmatrix} 974 \\ 46 \\ 144 \\ 148 \end{pmatrix}$ et $Y'Y = 124252$

2. Calculer le SCT, SCR, SCE et le coefficient de détermination. Conclusion ?

3. Les variables x_1, x_2 et x_3 sont-elles significatives au seuil 5% ?

exercice 5.

1. Soit le modèle :
   $y_t = \underset{(0.926)}{23} + \underset{(1)}{4}\, x_{1t} + \underset{(1)}{6}\, x_{2t} + \underset{(0.926)}{5}\, x_{3t}$

   $R^2 = 0.857$
   $SCR = 120$
   (.) Ecart-type
   $n = 14$

2. Calculer $\widehat{\sigma}_\varepsilon^2$ et le coefficient de détermination corrigé. Conclusion ?

3. Ce modèle est-il globalement significatif au seuil de 5% ?

4. Les variables x_1, x_2 et x_3 sont-elles significatives au seuil 5% ?

5. Prédire la valeur de y pour un individu tel que $x_{1\ 15} = 1$, $x_{2\ 15} = 1.5$ et $x_{3\ 15} = 2$.

6. Calculer l'intervalle de prédiction associé au seuil de confiance 5% .

exercice 6.
Le résultat de l'estimation du modèle de régression linéaire multiple est donné par le tableau suivant :

Dependent Variable: Y
Method: Least Squares
Date: 06/03/14 Time: 19:24
Sample: 1 5
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.000000	1?	-2.886751	0.1020
X1	3.000000	0.316228	2?	0.0109
X2	-0.400000	0.141421	-2.828427	0.1056
R-squared	0.980000	Mean dependent var		4.000000
Adjusted R-squared	3?	S.D. dependent var		3.162278
S.E. of regression	4?	Akaike info criterion		2.205296
Sum squared resid	0.800000	Schwarz criterion		1.970958
Log likelihood	-2.513239	F-statistic		5?
Durbin-Watson stat	2.250000	Prob(F-statistic)		0.020000

1. Ecrire l'équation du modèle de régression.

2. Compléter les valeurs manquantes dans le tableau.

3. Ce modèle est-il globalement significatif au seuil de 5% ?

4. Testez au seuil de 5% la signification de B_0 , B_1 et B_2. Conclusion.

5. Donner le tableau de l'analyse de la variance.

exercice 7.
Nous utilisons le modèle de régression linéaire multiple :

$$Y_t = B_0 + B_1 X_{1\ t} + B_2 X_{2\ t} + \varepsilon_t$$

1. Compléter le tableau d'analyse de variance suivant :

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen
Régression X_1, X_2	1504.4		
Résiduelle			19.6
Totale	1680.8		

2. Tester l'hypothèse nulle $H_0 : B_1 = B_2 = 0$ contre l'hypothèse alternative H_1 : *au moins un des* $B_i \neq 0$.
3. Quel est le coefficient de détermination R^2 du modèle ?
4. Donner une estimation de la variance de ε.

exercice 8.

1. Soit le modèle : $y_t = B_0 + B_1 x_{1\ t} + B_2 x_{2\ t} + B_3 x_{3\ t} + \varepsilon_t, \quad i = \overline{1, 50}$
   Estimer par la méthode des moindres carrés ordinaires les paramètres du modèle et interpréter les résultats, sachant que :
   $$X'X = \begin{pmatrix} ? & ? & 0 & ? \\ 10 & 10 & ? & 0 \\ ? & 0 & 5 & ? \\ 0 & ? & 0 & 2 \end{pmatrix}, \; X'Y = \begin{pmatrix} -30 \\ 10 \\ 5 \\ 2 \end{pmatrix} \quad \text{et} \quad Y'Y = 59.984$$
2. Donner les valeurs manquantes.
3. Estimer par la méthode des moindres carrés ordinaires les paramètres du modèle.
4. Calculer le SCT, SCR, SCE et le coefficient de détermination. Conclusion ?
5. Donner la valeur de $\widehat{\sigma}_\varepsilon^2$.
6. Les variables x_1, x_2 et x_3 sont-elles significatives au seuil 5% ?

Bibliographie

[1] Y. Dodge, V.Rousson, *Analyse de régression appliquée* . Dunod, (2004).

[2] R. Bourbonnais, *Économétrie* . Dunod, (1998).

[3] M. Tenenhaus, *Statistique : Méthodes pour décrire, expliquer et prévoir* , Dunod, (2007).

[4] D. N. GUJARAT, *économétrie* de boeck université. Paris, (2004).

[5] E. DOR, *économétrie, Synthèse de cours et exercices corrigés, collection synthex*. Pearson. Education France, (2004) ;

[6] S. Khedhiri, *Cours d'économétrie, méthodes et applications*. Lavoisier. Paris, (2007) ;

[7] J. LABARERES, *Corrélation - Régression Exercices commentés*. Université Joseph Fourier de Grenoble. Paris, (2012) ;

[8] A. Guyader, *Régression linéaire*. Université Rennes 2. Paris, (2013) ;

[9] P. A. Cornillon, E. M. Lober, *Régression théorie et applications*. Springer. France, (2006) ;

[10] J. Lenoir, *Régression linéaire*. Université de Picardie. http ://www.u-picardie.fr/edysan/.

Table des matières

1 *Modèle de régression linéaire simple* 3

1.1 *Présentation du modèle* . . . 3
1.2 *Estimation des paramètres par la méthode Moindres carrés ordinaires (MCO)* 4
1.3 *Propriétés des estimateurs* $\widehat{B_0}$ *et* $\widehat{B_1}$ . . . 5
1.4 *Analyse de la variance* . . . 6
1.4.1 *Décomposition de la variance* . . . 7
1.5 *Tests de significativités du modèle* . . . 8
1.5.1 *Test de significativité globale du modèle (Fisher)* . . . 8
1.5.2 *Test de Student* . . . 8
1.6 *Prévision* . . . 10
1.7 *Quelques exemples* . . . 11
1.8 Exercices . . . 11
1.9 *Énoncés des exercices* . . . 23

2 *Modèle de régression linéaire multiple* 25

2.1 *Présentation du modèle* . . . 25
2.2 *la forme matricielle* . . . 26
2.3 *Estimation des paramètres par la méthode Moindres carrés ordinaires (MCO)* 27
2.4 *Propriétés des estimateurs* . . . 29
2.5 *Equation d'analyse de la variance et la qualité de l'ajustement* . . . 31
2.5.1 *Equation d'analyse de la variance* . . . 31
2.5.2 *La qualité de l'ajustement* . . . 31
2.6 *Tests Statistiques* . . . 33
2.6.1 *Test de significativité globale du modèle (Fisher)* . . . 33
2.6.2 *Test de Student* . . . 33
2.7 *Prévision* . . . 35
2.8 *Exemple* . . . 37
2.9 *Exercices* . . . 46