

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
UNIVERSITÉ MOHAMED KHIDER, BISKRA  
Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie  
DÉPARTEMENT DE MATHÉMATIQUES



Mémoire présenté en vue de l'obtention du Diplôme de :  
MASTER en Mathématiques

Option : Statistiques

Par

SEID Zineb

Titre :

## Estimation des paramètres dans une regression multiple

Membres du Jury :

Pr. SAYAH Abd Allah	UMKB	Président
Pr. BENATIA Fatah	UMKB	Encadreur
Dr. SOLTANE Luiza	UMKB	Examinatrice

Juin 2024

## Dédicace

Je dédie ce humble travail

À mes chers parents

qui ont travaillé dur pour moi et m'ont donné tout ce que je voulais

Qu'ALLAH les protège .

À mes très chers frères, mes très chères Sœurs qui m'ont encouragé sur le long de  
mon

parcours universitaire, et à leurs enfants surtout à l'adorable ma petite nièce **INES**

À mes amis durant mes années d'études qui ma beacoup encouragé.

À toutes mes profeseurs que j'ai connus durant mes études.

À mes connaissances de proche ou de loin .

À tous ceux et celle que me souhaitent la réussite pour toute ma vie

À tout la promotion de deusième Master mathématique de toutes spécialité

2019-2024

À tous ceux que j'ai oublié de mentionner leurs noms.

**SEID Zineb**

## REMERCIEMENTS

Je voudrais dans un premier temps remercier mon encadrant le professeur **Mr. BENATIA Fatah** pour sa disponibilité, son aide sa patience avec moi et surtout ses judicieuses conseils qui ont contribué à alimenter ma réflexion.

Mes remerciements s'adressent aussi à tous nos professeurs pour leurs générosités et la grande patience dont ils ont su faire preuve malgré leur charges académiques et professionnelles.

Je tiens à remercier **Dr SAYAH Abd Allah**, maître de conférence à l'université de Biskra, qui m'a fait l'honneur de présider mon jury et d'avoir accepté d'évaluer mon mémoire.

Je remercie très vivement **Dr SOLTANE Luiza**, maître de conférence à l'université de Biskra, d'avoir accepté d'examiner ce travail.

Je tiens également à remercier spécialement mes chers amies **Asma Kabot, Mansoul Rokia, Djihan Gouicem**

qui ont toujours été là pour moi ,leur soutien inconditionnel et leurs encouragements ont été d'une grande aide.

Un grand merci aux professeurs **Mr. YAHIA. Djabrane** et **Mr. CHARFAOUI. Mouloud** pour leur aide, leurs conseils et leur encouragement à tous les étudiants qui sont en train d'obtenir leur diplôme de fin d'étude

# Table des matières

<b>Remerciements</b>	<b>ii</b>
<b>Table des matières</b>	<b>iii</b>
<b>Table des figures</b>	<b>iv</b>
<b>Liste des tables</b>	<b>v</b>
<b>Introduction</b>	<b>1</b>
<b>1 Régression linéaire simple</b>	<b>3</b>
<b>1.1 Modélisation</b> . . . . .	3
<b>1.1.1 Modélisation statistique</b> . . . . .	3
<b>1.1.2 Modèle linéaire simple sous forme matricielle</b> . . . . .	4
<b>1.1.3 Hypotèses sur le modèle</b> . . . . .	4
<b>1.2 Estimation des paramètres</b> . . . . .	5
<b>1.2.1 Estimation par la méthode des moindres carrés ordinaires (MCO)</b> . . . . .	6
<b>1.3 Estimation par la méthode du maximum de vraisemblance MV</b> . . . . .	8
<b>1.4 Interprétations géométriques</b> . . . . .	10
<b>1.4.1 Représentation des individus</b> . . . . .	10

1.4.2	Représentation des variables	11
1.5	Lois des estimateurs	12
1.5.1	Propriétés	13
1.6	Intervalles et régions de confiance	14
1.7	Qualité d'ajustement	16
1.7.1	Décomposition de la variance des $Y_i$	16
1.7.2	Le coefficient de détermination	17
1.8	Tests des coefficients	18
1.9	Construction du test des paramètres et règles de décision	20
1.9.1	test des paramètres	20
1.9.2	Règles de décision	21
1.10	Analyse de la variance et test de Fisher	22
1.10.1	Règle de décision	23
1.11	Prévision	24
<b>2</b>	<b>Régression linéaire multiple</b>	<b>26</b>
2.1	Modalisation	26
2.2	Les hypothèses du modèle	28
2.3	Estimation des paramètres	29
2.3.1	Estimation par la méthode des moindres carrés ordinaires (MCO)	29
2.3.2	Propriétés des Estimateurs :	31
2.4	Estimation par la méthode MV	34
2.5	Interprétation géométrique	37
2.6	Qualité de l'ajustement	38
2.7	Lois des estimateurs	39

2.8 Tests et Intervalles de confiances . . . . .	40
2.8.1 Intervalles et région de confiances . . . . .	40
2.8.2 La région de confiance . . . . .	42
2.9 Tests d'hypothèses . . . . .	42
2.9.1 Tests d'hypothèses simple (Test de signification des paramètres)	42
2.9.2 Tests d'hypothèses globale (Test de signification globale) . . .	43
2.9.3 Test sur le modèle réduit . . . . .	44
2.10 Prévission . . . . .	45
<b>3 Application sous R</b>	<b>46</b>
3.1 Initialisation de la régression linéaire multiple avec R . . . . .	46
3.1.1 Modèle de régression linéaire multiple . . . . .	47
3.1.2 Construction du modèle . . . . .	47
3.1.3 Représentation graphique . . . . .	49
3.1.4 Intervalles de confiance . . . . .	50
3.1.5 Interprétation . . . . .	53
3.1.6 Interprétation des coefficients . . . . .	54
3.1.7 Qualité du modèle . . . . .	56
3.1.8 Prediction . . . . .	58
<b>Conclusion</b>	<b>60</b>
<b>Bibliographie</b>	<b>61</b>
<b>Annexe A : Logiciel R</b>	<b>63</b>
<b>Annexe B : Abréviations et Notations</b>	<b>65</b>



# Table des figures

1.1 Ajustement du nuage de point par une droite de régression	11
1.2 Comparaison entre ellipse de confiance et rectangle de confiance.	14
3.1 Présentations des données	48
3.2 Constrction du modèle de régression multiple	48
3.3 Coefficient du modèle	48
3.4 La droite de régression du prix de vente en tèreme de surface	49
3.5 la droite de régression de prix de vente en tèreme de nombre de chambres	50
3.6 Intervalle de confiance à 95%	51
3.7 Valeurs prédites	51
3.8 Les résidus(Valeur prédite -Valeur réelle)	51
3.9 Présentation des résidus	53
3.10 Résidus vs surface	54
3.11 Résidus vs prix de vente	55
3.12 Residus vs. valeurs prédites	56
3.13 Test de significativité globale du modèle	56
3.14 Test de significativité	57

<b>3.15</b> <b>Signe du coefficient</b> . . . . .	57
<b>3.16</b> <b>Qualité du modèle</b> . . . . .	58
<b>3.17</b> <b>Prédiction des valeurs</b> . . . . .	58

# Liste des tableaux

1.1 (Table d'ANOVA)	23
2.1 (Table d'ANOVA à deux facteurs)	44

# Introduction

L'origine du mot régression vient de Sir Francis Galton. En 1885, travaillant sur l'hérédité, il chercha à expliquer la taille des fils en fonction de celle des pères, les résultats obtenus l'ont conduit à considérer sa théorie dite théorie de régression.

L'analyse de régression peut être définie comme la recherche de la relation stochastique qui lie deux ou plusieurs variables. Son champ d'application recouvre de multiples domaines, parmi les quels on peut citer la physique, l'astronomie, la biologie, la chimie, la médecine, la géographie, la sociologie, l'histoire, l'économie, la linguistique et le droit.

La régression est l'une des méthodes les plus connues et les plus appliquées en statistique pour l'analyse de données quantitatives, grâce à son objectif double Elle permet tout d'abord d'établir la relation entre une variable quantitative expliquée (ou dépendante) à une ou plusieurs autres variables quantitatives explicatives (ou indépendantes) sous la forme d'un modèle et aussi d'effectuer des prévisions de l'une des variables en fonction de l'autre, si on s'intéresse à la relation entre deux variables, on parlera de régression simple en exprimant une variable en fonction de l'autre. Si la relation porte entre une variable et plusieurs autres variables, on parlera de régression multiple.

La régression linéaire simple et multiple est une classe particulière de modèle de régression qui sert à expliquer une variable  $Y$ , appelée variable endogène par une

ou plusieurs variables explicatives dites exogènes a travers une fonction affine. Elle est dite linéaire si elle impose une forme fonctionnelle linéaire dans les paramètres du modèle.

Les paramètres de ce modèle sont estimés par deux méthodes différentes : MCO et MV.

L'objectif de ce mémoire est de présenter les différentes méthodes de régression linéaire multiple.

Ce travail se situe dans le cadre de l'analyse des données quantitatives, il est composé de trois chapitres :

**Chapitre 1** (Régression Linéaire Simple) : Ce chapitre est consacré à étudier la régression linéaire simple où on va estimer les paramètres  $\beta_0$  et  $\beta_1$  du modèle

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

par la méthode des moindres carrés ordinaire (MCO) et la méthodes des maximum de vraisemblance(MV) et tester sa significativité. En présentant les principales notions et propriétés de cette régression linéaire simple.

**Chapitre 2** (Régression Linéaire Multiple) : dans ce chapitre on généralise les résultats du chapitre 1 en traitant la régression multiple, dans ce cas le nombre des variables explicatives est supérieure ou égal à deux. représentée par le modèle suivant :

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

pour  $i = 1, \dots, n$ , où :  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  sont les paramètres du modèle.

**Chapitre 3** (application et simulation) : On termine ce chapitre avec un exemple illustratif sur un modèle de régression linéaire multiple, à l'aide du logiciel R, et on arrive finalement à une conclusion suivi de la liste des références bibliographiques.

# Chapitre 1

## Régression linéaire simple

Au cours de ce chapitre nous allons examiner la modélisation linéaire par le concept le plus élémentaire, comme une méthode statistique permettant de représenter la relation linéaire entre une seule variable explicative ( $X$ ) et une variable à expliquer ( $Y$ ) par une fonction affine. Après avoir explicité les hypothèses nécessaires et les termes du modèle, on discute les méthodes d'estimation des paramètres, les lois des estimateurs, les intervalles de confiance puis la signification des tests d'hypothèse et la qualité d'ajustement du modèle dont le but est la prévision. Pour plus de détails on peut consulter les livres de : Cornillon et Matzner [8]

### 1.1 Modélisation

#### 1.1.1 Modélisation statistique

**Définition 1.1.1** (*régression linéaire simple*) : Un Modèle de régression linéaire simple est défini par une équation de la forme  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \forall i \in \{1, \dots, n\}$

$n$  : est le nombre d'observation (la taille de l'échantillon).

$x_i$  : la  $i^{\text{ème}}$  observation de la variable explicative  $X$  (valeur fixée).

$y_i$  : la  $i^{\text{ème}}$  observation de la variable à expliquer  $Y$  (valeur observé et aléatoire).

$\beta_0$  et  $\beta_1$  : sont les paramètres inconnue du modèle ( les coefficients).

$\varepsilon_i$  : est une erreur ( qui viennent du fait que les points ne sont jamais parfaitement alignés sur une droite).

### 1.1.2 Modèle linéaire simple sous forme matricielle

Notons que le modèle de régression linéaire simple de la définition peut encore s'écrire de façon vectorielle :

$$Y = X\beta + \varepsilon$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & \vdots \\ 1 & x_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_1 \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

- $Y = [y_1, \dots, y_n]^t$  : vecteur aléatoire de dimension  $n$ .

- $\mathbb{1} = [1, \dots, 1]^t$  est le vecteur de  $\mathbb{R}^n$  dont les  $n$  composantes valent toutes 1.

- $X = [x_1, \dots, x_n]^t$  est un vecteur de dimension  $n$  donnée (non aléatoire ).

- $\beta_0$  et  $\beta_1$  sont les paramètres inconnus (mais non aléatoire ) du modèle.

- $\varepsilon = [\varepsilon_1, \dots, \varepsilon_n]^t$  est aléatoire de dimension  $n$ .

Cette notation vectorielle sera commode notamment pour l'interprétation géométrique du problème.

### 1.1.3 Hypotèses sur le modèle

Les tests d'hypothèses statistiques sur les paramètres (les coefficients) du modèle de la régression linéaire simple et l'établissement des intervalles de confiance qui permettent de vérifier la validité du modèle. Pour tester des hypothèses sur les pa-

paramètres  $\beta_0$  et  $\beta_1$ , on suppose que les erreurs sont indépendantes et normalement distribuées. C'est-à-dire les erreurs suivent la loi normale  $\mathcal{N}(0, \sigma^2)$ . Soit le modèle de régression simple

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1.1)$$

Les hypothèses du modèle : Le modèle [1.1](#) possède les propriétés suivantes :

( $H_1$ ) :  $\mathbb{E}(\varepsilon_i) = 0 \forall_i$ , L'erreur est centrée.

( $H_2$ ) :  $\mathbb{E}(\varepsilon_i^2) = \sigma^2$ ,  $\forall_i$  : La variance de l'erreur est constante.

( $H_3$ ) :  $Cov(\varepsilon_i, \varepsilon_j) = \delta_{ij}\sigma_\varepsilon^2$ ,  $\forall_i, \forall \varepsilon_i \neq \varepsilon_j$  : Les erreurs ne sont pas auto-correlées.

( $H_4$ ) :  $Cov(X, \varepsilon) = 0$  : L'erreur n'est pas corrélée avec la variable  $X$ .

( $H_5$ ) : La variable  $X$  n'est pas aléatoire.

( $H_6$ ) : Le modèle est linéaire en  $X$  par rapport aux paramètres.

Les erreurs sont donc supposées centrées, de même variance (homoscédasticité) et non corrélées entre elles ( $\delta_{ij}$  est le symbole de Kronecker, i.e  $\delta_{ij} = 1$  si  $i = j$ ,  $\delta_{ij} = 0$  si  $i \neq j$ ), voir [\[1\]](#), page 2]

## 1.2 Estimation des paramètres

On a deux méthodes pour estimer les paramètres du modèle, la méthode des moindres carrés ordinaires (MCO) qui ne nécessite pas d'hypothèse supplémentaire sur la distribution de  $\varepsilon_i$ , et par contre la méthode du Maximum de Vraisemblance (MV) qui est fondée sur la normalité de  $\varepsilon_i$

### 1.2.1 Estimation par la méthode des moindres carrés ordinaires (MCO)

**Définition 1.2.1** (MCO) : On appelle estimateurs des MCO de  $\beta_0$  et  $\beta_1$ , les estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$  obtenus par minimisation de la quantité voir [14]

$$\Phi(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^n \varepsilon_i^2$$

Les estimateurs des MCO peuvent également s'écrire sous la forme suivante :

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1) \in \mathbb{R} \times \mathbb{R}}{\operatorname{arg\,min}} \Phi(\beta_0, \beta_1)$$

#### calcul des estimateurs de $\beta_0, \beta_1$

La fonction  $\Phi(\beta_0, \beta_1)$  est strictement convexe. Elle possède un unique minimum si elle admet un point singulier. En annulant les dérivées partielles, nous obtenons un système d'équation normales :

$$\left\{ \begin{array}{l} \frac{\partial \Phi(\beta_0, \beta_1)}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0, \\ \frac{\partial \Phi(\beta_0, \beta_1)}{\partial \beta_1} = -2 \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) = 0, \end{array} \right. \left| \begin{array}{l} \beta_0 = \hat{\beta}_0 \\ \beta_1 = \hat{\beta}_1 \end{array} \right.$$

après quelques calculs on déduit :

$$\left\{ \begin{array}{l} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \frac{S_{xy}}{S_x} \end{array} \right.$$

où, les moyennes empiriques des  $x_i$  et des  $y_i$  (respectivement),

$$\begin{cases} \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \end{cases}$$

ou :

$$S_x = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}$$

### Quelques propriétés des estimateurs

#### **Théorème 1.2.1** : *(Biais des estimateurs)*

sous les hypothèses  $(H_1)$  et  $(H_2)$ . Les estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$  sont sans biais de  $\beta_0$  et  $\beta_1$ , c-à-d :

$$\mathbb{E}(\hat{\beta}_0) = \beta_0 \text{ et } \mathbb{E}(\hat{\beta}_1) = \beta_1$$

#### **Théorème 1.2.2** *(Variance et covariance)*

Sous les hypothèses  $(H_1)$  et  $(H_2)$ . Les variances et la covariance des estimateurs sont

$$Var(\hat{\beta}_0) = \frac{\sigma_\varepsilon^2 \sum_{i=1}^n x_i^2}{nS_x}, Var(\hat{\beta}_1) = \frac{\sigma_\varepsilon^2}{S_x} \text{ et } Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma_\varepsilon^2 \bar{x}}{S_x}$$

La matrice de variance covariance de  $\hat{\beta}_0$  et  $\hat{\beta}_1$  s'écrit comme suit :

$$\Psi(\hat{\beta}_0; \hat{\beta}_1) = \frac{\sigma_\varepsilon^2}{S_x} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

qui est estimée en remplaçant  $\sigma_\varepsilon^2$  par son estimateur  $\hat{\sigma}_\varepsilon^2$

**Théorème 1.2.3** (*l'estimateur de  $\sigma_\varepsilon^2$* ) *L'estimateur sans biais de la variance résiduelle (variance des erreurs)  $\sigma_\varepsilon^2$  est :*

$$S^2 = \hat{\sigma}_\varepsilon^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2$$

avec les résidus définis par  $\varepsilon_i = y_i - \hat{y}_i$  où  $\hat{y}_i$  est la valeur ajustée de  $y_i$  par le modèle, c'est-à-dire  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

## 1.3 Estimation par la méthode du maximum de vraisemblance MV

### Cas d'erreurs gaussiennes

Mieux que les expressions des estimateurs et celles de leurs variances, on aimerait connaître leurs lois : ceci permettrait par exemple d'obtenir des régions de confiance et d'effectuer des tests d'hypothèses. Dans cette optique, il faut bien entendu faire une hypothèse plus forte sur notre modèle, à savoir préciser la loi des erreurs. Nous supposons ici que les erreurs sont gaussiennes. Les hypothèses  $(H_1)$  et  $(H_2)$  deviennent dès lors voir [1] :

$$(H) = \begin{cases} (\mathcal{H}_1) : \varepsilon_i \rightsquigarrow \mathcal{N}(0, \sigma_\varepsilon^2) \\ (\mathcal{H}_2) : \varepsilon_i \text{ mutuellement indépendantes} \end{cases}$$

Le modèle de régression simple devient un modèle paramétrique, où les paramètres  $(\beta_0, \beta_1, \sigma_\varepsilon^2)$  sont à valeurs dans

$\mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^*$ . La loi des  $\varepsilon_i$  étant connue, les lois des  $Y_i$  s'en déduisent :

$$\forall_i \in \{1, \dots, n\} \quad y_i \sim N(\beta_1 + \beta_2 x_i, \sigma_\varepsilon^2),$$

et les  $y_i$  sont mutuellement indépendantes puisque les  $\varepsilon_i$  le sont. Nous pouvons donc calculer l'estimation des paramètres  $\beta_0$ ,  $\beta_1$  et  $\sigma_\varepsilon^2$  en maximisant la vraisemblance par rapport à ces paramètres, sous les hypothèses supplémentaires précédentes de normalité des erreurs :

la fonction de vraisemblance de l'échantillon  $(y_1, y_2, \dots, y_n)$  s'écrit

$$\begin{aligned} L(\beta_0, \beta_1, \sigma_\varepsilon^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\Pi\sigma_\varepsilon^2}} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} [y_i - (\beta_0 + \beta_1 x_i)]^2 \right\} \\ &= \left( \frac{1}{\sqrt{2\Pi\sigma_\varepsilon^2}} \right)^n \exp \left[ -\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \right] \\ L(\beta_0, \beta_1, \sigma_\varepsilon^2) &= \left( \frac{1}{\sqrt{2\Pi\sigma_\varepsilon^2}} \right)^n \exp \left[ -\frac{1}{2\sigma_\varepsilon^2} \Phi(\beta_0, \beta_1) \right] \end{aligned}$$

ce qui donne par la fonction log-vraisemblance :

$$\begin{aligned} \text{Log } L(\beta_0, \beta_1, \sigma_\varepsilon^2) &= -\frac{n}{2} \log(2\Pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \\ \text{Log } L(\beta_0, \beta_1, \sigma_\varepsilon^2) &= -\frac{n}{2} \log(2\Pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} \Phi(\beta_0, \beta_1) \end{aligned}$$

Nous voulons maximiser cette quantité par rapport aux trois variables  $(\beta_0, \beta_1, \sigma_\varepsilon^2)$ . Les deux premières variables n'apparaissent que dans le terme en  $-\Phi(\beta_0, \beta_1)$ , qu'il faut donc minimiser. Or on a déjà vu que cette quantité est minimale lorsqu'on considère les estimateurs des moindres carrés, c'est-à-dire pour  $\beta_0 = \hat{\beta}_0$  et  $\beta_1 = \hat{\beta}_1$ , où les estimateurs du maximum de vraisemblance de  $\beta_0$  et  $\beta_1$  sont égaux aux estimateurs

des MCO et ils ont les mêmes propriété (voir) [1.2.1](#) (voir) [1.2.2](#)

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \frac{S_{xy}}{S_x} \end{cases}$$

Ceci étant vu, il reste simplement à maximiser  $\text{Log } L(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_\varepsilon^2)$  par rapport à  $\sigma_\varepsilon^2$ .

Calculons donc la dérivée par rapport à  $\sigma_\varepsilon^2$  :

$$\begin{aligned} \frac{\partial \text{Log} L(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_\varepsilon^2)}{\partial \sigma_\varepsilon^2} &= -\frac{n}{2\sigma_\varepsilon^2} - \frac{1}{2\sigma_\varepsilon^4} \Phi(\hat{\beta}_0, \hat{\beta}_1) \\ &= -\frac{n}{2\sigma_\varepsilon^2} - \frac{1}{2\sigma_\varepsilon^4} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{aligned}$$

D'où l'on déduit que l'estimateur du maximum de vraisemblance de  $\sigma_\varepsilon^2$  est différent de l'estimateur  $\hat{\sigma}_\varepsilon^2$  vu précédemment et vaut :

$$\hat{\sigma}_{\varepsilon_{MV}}^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$$

L'estimateur du maximum de vraisemblance de  $\sigma_\varepsilon^2$  est donc biaisé. On a en effet  $\mathbb{E} \left[ \hat{\sigma}_{\varepsilon_{MV}}^2 \right] = \frac{n-2}{n} \sigma_\varepsilon^2$  mais ce biais est d'autant plus négligeable que le nombre d'observations est grand.

## 1.4 Interprétations géométriques

### 1.4.1 Représentation des individus

La représentation graphique consiste à présenter les observations  $(x_i, y_i)$  dans le plan  $\mathbb{R}^2$  ( nuage de points) par une droite qui s'appelle la droite de régression linéaire simple ou la droite ajustée . Cette droite minimise la somme des carrés des distances

verticales des points du nuage à la droite ajustée, où  $\hat{\beta}_0$  correspond à l'ordonnée à l'origine et  $\hat{\beta}_1$  représente la pente de la droite ajustée d'équation  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

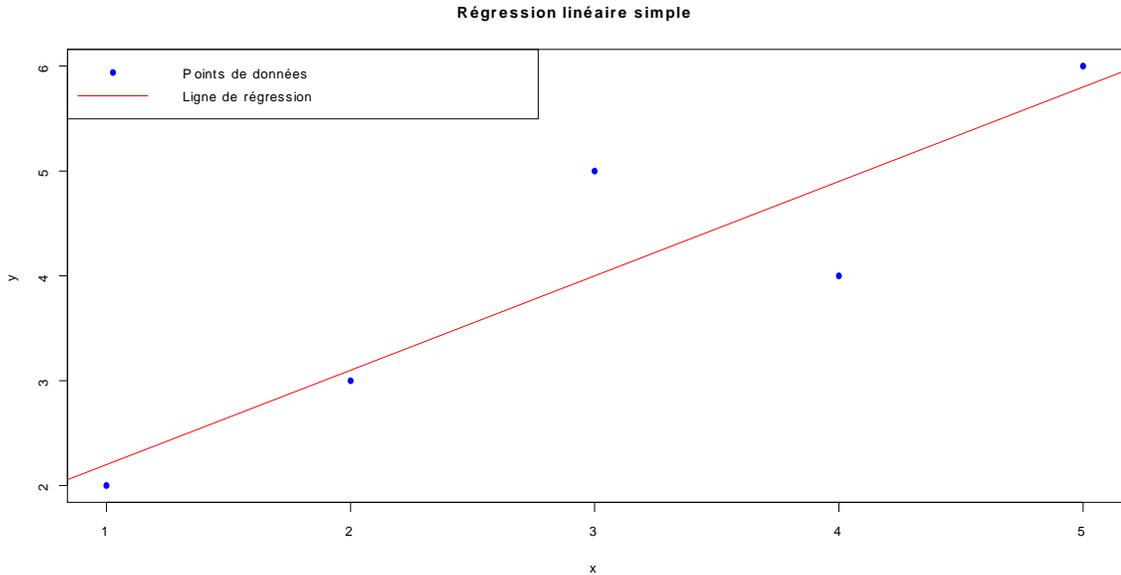


FIG. 1.1 – Ajustement du nuage de point par une droite de régression

### 1.4.2 Représentation des variables

On peut étudier le problème d'un point de vue vectoriel, de tel sort on a deux vecteurs à la disposition : le vecteur  $X = [x_1, \dots, x_n]^t$  des  $n$  observations pour la variable explicative et le vecteur  $Y = [y_1, \dots, y_n]^t$  des  $n$  observations pour la variable à expliquer. Ces deux vecteurs appartiennent au même espace  $R^n$  : l'espace des variables.

Si on ajoute à cela le vecteur  $\mathbb{1} = [1, \dots, 1]^t$ , on voit tout d'abord que par l'hypothèse selon laquelle tous les  $X_i$  ne sont pas égaux, les vecteurs  $\mathbb{1}$  et  $X$  ne sont pas colinéaires (il existe au moins deux points d'abscisses différentes), les deux vecteurs engendrent donc un sous-espace de  $R^n$  de dimension 2, noté  $M(X)$ . ,mais ces vecteurs ne sont pas nécessairement orthogonaux. Ces vecteurs sont orthogonaux lorsque la moyenne des observations  $x_1, \dots, x_n$  vaut zéro. On peut projeter orthogonalement le vecteur  $Y$

sur le sous-espace  $M(X)$  notons provisoirement  $\tilde{Y}$  ce projeté. Puisque  $(\mathbb{I}, X)$  forme une base de  $M(X)$ , il existe une unique décomposition de la forme  $\tilde{Y} = \tilde{\beta}_0 \mathbb{I} + \tilde{\beta}_1 X$ . Par définition du projeté orthogonal,  $\tilde{Y}$  est l'unique vecteur de  $M(X)$  minimisant la distance euclidienne  $\|Y - \tilde{Y}\|$ , ce qui revient au même que de minimiser son carré. Or, par définition de la norme euclidienne, cette quantité vaut :

$$\|Y - \tilde{Y}\|^2 = \sum_{i=1}^n (y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i))^2,$$

ce qui nous ramène à la méthode des moindres carrés ordinaires. On en déduit que  $\tilde{\beta}_0 = \hat{\beta}_0, \tilde{\beta}_1 = \hat{\beta}_1$

et  $\tilde{Y} = \hat{Y} = [\hat{y}_1, \dots, \hat{y}_n]^t$ , avec les expressions de  $\hat{\beta}_0, \hat{\beta}_1$  et  $\hat{Y}$  vues précédemment.

## 1.5 Lois des estimateurs

Nous allons maintenant voir comment les lois précédentes interviennent dans nos estimateurs. Afin de faciliter la lecture de cette partie, fixons les notations suivantes :

$$C = \frac{-\sigma_\varepsilon^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \sigma_\varepsilon^2 = \frac{1}{n-2} \sum_{i=1}^n \varepsilon_i^2$$

$$\sigma_0^2 = \text{Var}(\hat{\beta}_0) = \sigma_\varepsilon^2 \left( \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \right), \quad \sigma_0^2 = \sigma_\varepsilon^2 \left( \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

$$\sigma_1^2 = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \sigma_1^2 = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Comme nous l'avons vu,  $\sigma_0^2, \sigma_1^2$  et  $C$  sont les variances et covariance des estimateurs des moindres carrés ordinaires. les quantités  $\hat{\sigma}_0^2$  et  $\hat{\sigma}_1^2$  correspondent quant à elles aux estimateurs des variances de  $\hat{\beta}_0$  et  $\hat{\beta}_1$ .

### 1.5.1 Propriétés

#### Lois des estimateurs avec variance connue ( $\sigma_\varepsilon^2$ )

Les lois des estimateurs des MCO avec variance  $\sigma_\varepsilon^2$  connue sont :

(i)

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \sim N(\beta, \sigma_\varepsilon^2 V) \text{ où } \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

et

$$V = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{bmatrix} \sum_{i=1}^n x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} = \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} \sigma_0^2 & c \\ c & \sigma_1^2 \end{bmatrix}$$

(ii)  $\frac{(n-2)}{\sigma_\varepsilon^2} \hat{\sigma}_\varepsilon^2 \sim \chi_{n-2}^2$ , loi du  $\chi^2$  à  $(n-2)$  degrés de liberté.

(iii)  $\hat{\beta}$  et  $\hat{\sigma}_\varepsilon^2$  sont indépendants.

#### Remarque.

· Ces propriétés, comme celles à venir, ne sont pas plus faciles à montrer dans le cadre de la régression linéaire simple que dans celui de la régression linéaire multiple.

· Le problème des propriétés ci-dessus vient de ce qu'elles font intervenir la variance théorique  $\sigma_\varepsilon^2$ , généralement inconnue. La façon naturelle de procéder est de la remplacer par son estimateur  $\hat{\sigma}_\varepsilon^2$ .

· Les lois intervenant dans les estimateurs s'en trouvent de fait légèrement modifiées.

#### Lois des estimateurs avec variance estimée ( $\sigma_\varepsilon^2$ )

Les lois des estimateurs des MCO avec variance  $\hat{\sigma}_\varepsilon^2$  estimée sont :

(i)  $\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_0} \sim T_{n-2}$ , où  $T_{n-2}$  est une loi de Student à  $(n-2)$  degrés de liberté.

(ii)  $\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_1} \sim T_{n-2}$ .

(iii)  $\frac{1}{2\sigma_\varepsilon^2}(\hat{\beta} - \beta)^t V^{-1}(\hat{\beta} - \beta) \sim F_{n-2}^2$ , loi de Fisher de paramètres  $(2, n - 2)$ .

Ces dernières propriétés nous permettent de donner des intervalles de confiance (IC) ou des régions de confiance (RC) des estimateurs (paramètres inconnues). En effet, de tester la signification des paramètres et la validation du modèle. la valeur ponctuelle d'un estimateur est de peu d'intérêt en général et il est intéressant de lui associer un intervalle de confiance. Les résultats sont donnés pour un  $\alpha$  général, en pratique on prend typiquement  $\alpha = 0, 05$ .

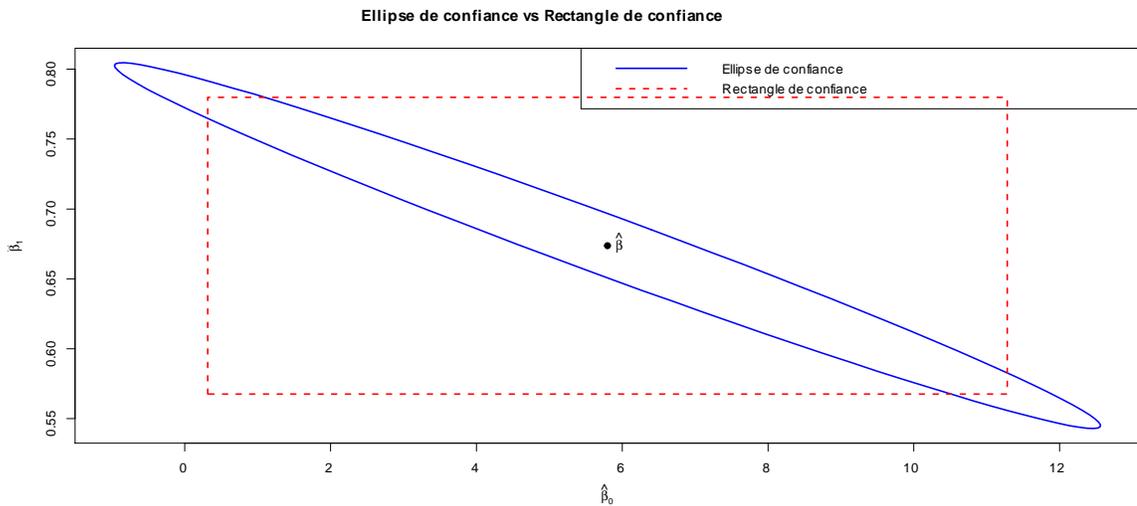


FIG. 1.2 – Comparaison entre ellipse de confiance et rectangle de confiance.

## 1.6 Intervalles et régions de confiance

On cherche les intervalles et région de confiance des paramètres  $\beta_0$  et  $\beta_1$  au niveau de confiance  $(1 - \alpha)$ , tel que  $\alpha \in [0, 1]$ .

(i) d'après la propriété [1.5.1](#), on a :

$$P\left(\frac{|\hat{\beta}_j - \beta_j|}{\hat{\sigma}_{\hat{\beta}_j}} \prec t\right) = 1 - \alpha, \quad j = 0, 1,$$

on obtient les (I.C) des paramètres  $\beta_0$  et  $\beta_1$

$$\beta_j \in \left[ \hat{\beta}_j - t \hat{\sigma}_{\hat{\beta}_j}, \hat{\beta}_j + t \hat{\sigma}_{\hat{\beta}_j} \right]$$

où  $t = T_{n-2}(1 - \alpha/2)$  est le fractile d'ordre  $(1 - \alpha/2)$  d'une loi de Student  $T_{n-2}$ .

(ii) De même, pour l'IC de  $\sigma_\varepsilon^2$

$$P(t_1 \prec \frac{(n-2)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \prec t_2) = 1 - \alpha,$$

donc,

$$\sigma_\varepsilon^2 \in \left[ \frac{(n-2)\hat{\sigma}_\varepsilon^2}{t_2}, \frac{(n-2)\hat{\sigma}_\varepsilon^2}{t_1} \right]$$

où  $t_1 = T_{n-2}(\alpha/2)$  (respt  $t_2 = T_{n-2}(1 - \alpha/2)$ ) est le fractile d'ordre  $(\alpha/2)$  (respt  $(1 - \alpha/2)$ ) de la loi de  $\chi_{n-2}^2$

(iii) la région de confiance de  $\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$  au niveau  $(1 - \alpha)$  est

$$\frac{1}{2\hat{\sigma}_\varepsilon^2} \left\{ n(\hat{\beta}_0 - \beta_0)^2 + 2n\bar{x}(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) + (\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^n x_i^2 \right\} \leq f$$

où  $f$  est le fractile d'ordre  $(1 - \alpha)$  d'une loi de Fisher  $F_{n-2}^2$ .

Le point (iii) donne la région de confiance simultanée des paramètres  $(\beta_0, \beta_1)$  de la régression, appelée ellipse de confiance, tandis que (i) et (ii) donnent des intervalles de confiance pour  $\beta_0$  et  $\beta_1$  pris séparément. La figure [1.2](#) montre la différence entre

ces deux notions.

## 1.7 Qualité d'ajustement

### 1.7.1 Décomposition de la variance des $Y_i$

lorsque le modèle est établi, il faut juger sa qualité d'ajustement et sa fiabilité, on utilise l'équation de l'analyse de la variance, c-à-d cherchons tout d'abord à décomposer la variance des  $Y_i$  autour de leur moyenne en une somme de deux autres variances. Pour cela nous conservons les notations du paragraphe précédent, avec  $\hat{Y} = [\hat{y}_1, \dots, \hat{y}_n]^t$  la projection orthogonale du vecteur  $Y$  sur  $M(X)$  et

$$\hat{\varepsilon} = Y - \hat{Y} = [\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n]^t$$

Le théorème de Pythagore donne alors directement :

$$\begin{aligned} \left\| Y - \bar{Y}\mathbb{I} \right\|^2 &= \left\| \hat{Y} - \bar{Y}\mathbb{I} \right\|^2 + \left\| \hat{\varepsilon} \right\|^2 \\ \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2 \end{aligned}$$

$$\text{où } \sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SCT = SCE + SCR.$$

où :

$SCT$  : somme des carrés totale (variation totale des  $Y_i$ ),

$SCR$  : somme des carrés des résidus (variation des résidus  $\hat{\varepsilon}_i$ ),

$SCE$  : somme des carrés expliqués (variation expliquée).

Ceci peut se voir comme une formule typique de décomposition de la variance. Elle

permet en outre d'introduire le coefficient de détermination de façon naturelle.

### 1.7.2 Le coefficient de détermination

le coefficient de détermination  $R^2$  est un indicateur de la qualité d'ajustement qui exprime le rapport entre la variation expliquée et la variation totale, dont :

$$R^2 = \frac{SCE}{SCT} = \frac{\left\| \hat{Y} - \bar{Y} \mathbf{1} \right\|^2}{\left\| Y - \bar{Y} \mathbf{1} \right\|^2} = 1 - \frac{\left\| \hat{\varepsilon} \right\|^2}{\left\| Y - \bar{Y} \mathbf{1} \right\|^2} = 1 - \frac{SCR}{SCT} = \frac{S_{xy}^2}{S_x S_y} \leq 1.$$

Dans le cas où  $R^2$  est proche de 1 alors l'ajustement est meilleure, c-à-d la variation expliquée est proche de la variation totale alors la variation résiduelle est proche de 0.

On voit sur la figure ?? que  $R^2$  correspond au cosinus carré de l'angle  $\theta$ . De façon schématique, on peut différencier les cas suivants :

- Si  $R^2 = 1$ , le modèle explique tout, l'angle  $\theta$  vaut zéro et  $Y$  est dans  $M(X)$ , c'est-à-dire que  $Y_i = \beta_0 + \beta_1 X_i$  pour tout  $i$  : les points de l'échantillon sont parfaitement alignés sur la droite des moindres carrés ;
- Si  $R^2 = 0$ , cela veut dire que  $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 0$ , donc  $\hat{y}_i = \bar{y}$  pour tout  $i$ . Le modèle de régression linéaire est inadapté puisqu'on ne modélise rien de mieux que la moyenne,
- Si  $R^2$  est proche de zéro, cela veut dire que  $Y$  est quasiment dans l'orthogonal de  $M(X)$ , le modèle de régression linéaire est inadapté, la variable  $X$  n'explique pas bien la variable réponse  $Y$  (du moins pas de façon affine).

## 1.8 Tests des coefficients

Soit le modèle de régression linéaire simple voir [3] :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ où } i = 1, \dots, n$$

on sait que  $\varepsilon_i$  suit la loi normale de moyenne nulle et de variance  $\sigma_\varepsilon^2$ ,

c-à-d  $\varepsilon_i \sim N(0, \sigma^2)$

$$S^2 = \hat{\sigma}_\varepsilon^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2 \implies (n-2)\hat{\sigma}_\varepsilon^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2$$

$$(n-2) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} = \sum_{i=1}^n \left( \frac{\hat{\varepsilon}_i}{\sigma_\varepsilon} \right)^2$$

on a :  $\frac{\hat{\varepsilon}_i}{\sigma_\varepsilon} \sim N(0, 1)$

donc,

$$(n-2) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{n-2}^2,$$

où  $\chi_{n-2}^2$  est la loi de Khi-deux de  $(n-2)$  degrés de liberté.

D'autre part on a vu aussi

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma_\varepsilon^2}{S_{x^2}}\right) \text{ et } \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_\varepsilon^2}{S_{x^2}}\right) \quad (1.2)$$

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}} \sim T_{n-2}$$

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}} \sim T_{n-2}, \text{ où } T_{n-2} \text{ est une loi de Student à } (n-2) \text{ degrés de liberté.}$$

soit  $z_1 = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$

$$z_1 = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_\varepsilon / S_x} = \frac{\hat{\beta}_1 - \beta_1}{\sigma_\varepsilon / S_x} * \frac{\sigma_\varepsilon}{\hat{\sigma}_\varepsilon} = \frac{\hat{\beta}_1 - \beta_1}{\sigma_\varepsilon / S_x} * \frac{1}{\frac{\hat{\sigma}_\varepsilon}{\sigma_\varepsilon}}$$

on a si  $z \sim T_n \iff z = \frac{x}{\sqrt{\frac{y}{n-2}}}$  telque :

$$\begin{cases} x \sim N(0, 1) \\ y \sim \chi_n^2 \end{cases}, \text{ où } n = d.d.l$$

ce qui implique

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_\varepsilon / S_x} * \frac{1}{\frac{\hat{\sigma}_\varepsilon}{\sigma_\varepsilon}} = \frac{x}{\sqrt{\frac{y}{n-2}}}$$

donc,

D'après [1.2](#)

on trouve

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_\varepsilon / S_x} \sim N(0, 1)$$

en plus

$$\frac{\hat{\sigma}_\varepsilon}{\sigma_\varepsilon} = \sqrt{\frac{(n-2)}{n-2} * \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2}} = \sqrt{\frac{\chi_{n-2}^2}{n-2}}$$

D'après la définition de la loi de Student  $T$  qui est : le rapport d'une loi centrée réduite et la racine carrée d'une loi de khi-deux divisée par le nombre de ses degrés de liberté. On a :

$$z_1 = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}} \sim T_{n-2} \tag{1.3}$$

où  $T_{n-2}$  est une loi de Student à  $(n - 2)$  degrés de liberté.

même démarche pour  $\beta_0$ , on obtient

$$z_2 = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}} \sim T_{n-2} \tag{1.4}$$

à partir des résultats précédents 1.3 et 1.4 on peut effectuer d'hypothèse suivant

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}, \text{ et } \begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

## 1.9 Construction du test des paramètres et règles de décision

### 1.9.1 test des paramètres

► On considère un test bilatéral et on commence par le paramètre  $\beta_1$ , Tester  $\beta_1$  consiste à tester l'hypothèse suivante voir 3 :

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases},$$

sous l'hypothèse  $H_0 : \beta_1 = 0$ , on obtient la valeur critique ( $T_C$ ) calculée à partir de l'expression

$$T_C = \frac{|\hat{\beta}_1 - \beta_1|}{\hat{\sigma}_{\beta_1}} \sim T_{n-2}, \text{ où}$$

·  $T_C$  = la valeur critique liée la statistique  $\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$  (dite calculée).

·  $\hat{\beta}_1$  : Désigne la valeur estimée du paramètre  $\beta_1$ .

·  $\hat{\sigma}_{\beta_1}$  : Désigne la valeur de l'écart-type du paramètre  $\beta_1$ .

·  $\alpha$  : Le seuil donné. (En général, on prend  $\alpha = 5\%$ ).

·  $n - 2$  : Degré de liberté.

·  $T_{n-2}$  : Désigne la valeur de la statistique Student  $T$  lue à partir de la table statistique

Donc,

Le test est effectué au seuil de confiance  $\alpha$ ,  $\alpha \in ]0, 1[$ , au encore au niveau de confiance  $(1 - \alpha)$ , on note  $T_{n-2}(\alpha/2)$  la valeur sur la table de student à  $(n - 2)$  degré de liberté

au seuil  $\alpha/2$

- De même, le test du paramètre  $\beta_0$  consiste à tester l'hypothèse suivante :

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}, \text{ on obtient la valeur critique } (T_C) \text{ calculée à partir de l'expression :}$$

$$T_C = \frac{|\hat{\beta}_0 - \beta_0|}{\hat{\sigma}_{\beta_0}} \sim T_{n-2}, \text{ où}$$

•  $T_C$  = la valeur critique liée la statistique  $\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$  (dite calculée).

•  $\hat{\beta}_0$  : Désigne la valeur estimée du paramètre  $\beta_0$ .

•  $\hat{\sigma}_{\beta_0}$  : Désigne la valeur de l'écart-type du paramètre  $\beta_0$ .

•  $\alpha$  : Le seuil donné. (En général, on prend  $\alpha = 5\%$ ).

•  $n - 2$  : Degré de liberté.

•  $T_{n-2}$  : Désigne la valeur de la statistique Student  $T$  lue à partir de la table statistique

Donc,

Le test est effectué au seuil de confiance  $\alpha$ ,  $\alpha \in ]0, 1[$ , au encore au niveau de confiance  $(1 - \alpha)$ , on note  $t_{n-2}(\alpha/2)$  la valeur sur la table de student à  $(n - 2)$  degré de liberté au seuil  $\alpha/2$

## 1.9.2 Règles de décision

• La règles de décision du paramètre  $\beta_1$  est :

- si  $|T_C| < T_{n-2}(\alpha/2)$ , on accept  $H_0$  (on rejette  $H_1$ ) : la variable  $X_i$  n'est pas contributive à l'explication  $Y$ .

- si  $|T_C| > T_{n-2}(\alpha/2)$ , on accept  $H_1$  (on rejette  $H_0$ ) : la variable  $X_i$  est contributive à l'explication  $Y$ .

• De même, La règles de décision du paramètre  $\beta_0$  est :

- si  $|T_C| < T_{n-2}(\alpha/2)$ , on accepte  $H_0$  (on rejette  $H_1$ ) : le modèle ne contient pas de constante.

- si  $|T_C| > T_{n-2}(\alpha/2)$ , on accepte  $H_1$  (on rejette  $H_0$ ) : le modèle contient la constante.

## 1.10 Analyse de la variance et test de Fisher

• Nous avons précédemment défini les sommes de l'équation d'analyse de variance  $SCT$ ,  $SCR$  et  $SCE$ ; ces sommes peuvent être utilisées pour tester l'hypothèse sui-

$$\text{vante : } \begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}, j = 0, 1$$

Sous l'hypothèse  $H_0 : \beta_j = 0$ , On a

$$\mathbb{E}(SCT) = (n-1)\sigma_\varepsilon^2, \quad \mathbb{E}(SCE) = (1)\sigma_\varepsilon^2 \quad \text{et} \quad \mathbb{E}(SCR) = (n-2)\sigma_\varepsilon^2$$

avec  $(n-1)$ ,  $(1)$  et  $(n-2)$  sont des degrés de libertés de  $SCT$ ,  $SCE$  et  $SCR$  respectivement.

D'autre part, lorsque  $H_0$  est vérifiée on peut finalement vérifier que

$$\frac{SCT}{n-1} \sim \chi_{n-1}^2, \quad \frac{SCE}{1} \sim \chi_1^2, \quad \frac{SCR}{n-2} \sim \chi_{n-2}^2$$

Tant que  $\frac{SCE}{\sigma_\varepsilon^2}$  et  $\frac{SCR}{\sigma_\varepsilon^2}$  sont indépendants, on peut déduire donc la valeur critique de Fisher  $F$  qui se définit comme suit : le rapport entre deux Khi-deux ( $\chi$ ) indépendants et leurs degrés de libertés.

Alors, on obtient :

$$F_c = \frac{\frac{SCE}{\sigma_\varepsilon^2 \times 1}}{\frac{SCR}{(n-2) \times \sigma_\varepsilon^2}} = \frac{\frac{SCE}{1}}{\frac{SCR}{n-2}} = \frac{SCE \times (n-2)}{SCR \times 1} \sim F_T(1, n-2)$$

avec

Source de la variance	<i>ddl</i>	Somme de scarrés	Moyenne de scarrés	$F_c$
<i>Expliquée</i>	1	<i>SCE</i>	$MCE = SCE / 1$	$F_c = \frac{MCE}{MCR}$
<i>Résiduelle</i>	$n - 2$	<i>SCR</i>	$MCR = SCR / (n - 2)$	
<i>Totale</i>	$n - 1$	<i>SCT</i>		

TAB. 1.1 – (Table d’ANOVA)

·  $F_C$  : désigne la valeur critique de Fisher calculée qui sera comaprer à la valeur tabulée

·  $(1, n - 2, \alpha\%)$  : se sont des degrés de libertés.

·  $F_T$  : désigne la valeur de Fisher lue à partir de la table statistique de Fisher aux degrés de liberté  $(1, n - 2)$ .

·  $\alpha\%$  : Le seuil donné

cette modalisation nous permet de faire un test de signification globale

### 1.10.1 Règle de décision

1) Si  $|F_C| > F_T(1, n - 2; \%)$  on rejette l’hypothèse  $H_0$ , cela signifie que,  $H_1$  est acceptée : c’est-à-dire le modèle est globalement significatif.

2) Si  $|F_C| < F_T(1, n - 2; \%)$  on rejette l’hypothèse  $H_1$  ; cela signifie que,  $H_0$  est acceptée : c’est-à-dire le modèle n’est pas globalement significatif.

### Tableau d’analyse de la variance

On construisant le tableau d’analyse de la variance comme suit (la table d’ANOVA)

On accepte  $H_0$  si :

$$F_c = \frac{MCE}{MCR} \leq F_T$$

où  $F_T$  est le fractile d’ordre  $(1 - \alpha)$  de la loi de Fisher  $F(1, n - 2)$

## 1.11 Prévision

Un des buts de la régression est de faire de la prévision, c'est-à-dire de prévoir la variable à expliquer  $Y$  en présence d'une nouvelle valeur de la variable explicative  $X$ . Soit donc  $X_{n+1}$  une nouvelle valeur, pour laquelle nous voulons prédire  $Y_{n+1}$ . Le modèle est toujours le même :

$$Y_{n+1} = \beta_0 + \beta_1 X_{n+1} + \varepsilon_{n+1}$$

avec  $\mathbb{E}[\varepsilon_{n+1}] = 0, \text{Var}(\varepsilon_{n+1}) = \sigma^2$  et  $\text{Cov}(\varepsilon_{n+1}, \varepsilon_i) = 0$  pour  $i = 1, \dots, n$ . Il est naturel de prédire la valeur correspondante via le modèle ajusté :

$$\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{n+1}$$

Deux types d'erreurs vont entacher notre prévision : la première est due à la non-connaissance de  $\varepsilon_{n+1}$ , la seconde à l'incertitude sur les estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$ .

**Proposition 1.11.1** (*Erreur de prévision*) *L'erreur de prévision  $\hat{\varepsilon}_{n+1} = (y_{n+1} - \hat{y}_{n+1})$  satisfait les propriétés suivantes :*

$$\left\{ \begin{array}{l} \mathbb{E}[\hat{\varepsilon}_{n+1}] = 0 \\ \text{Var}(\hat{\varepsilon}_{n+1}) = \sigma^2 \left( 1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \end{array} \right.$$

*Pour l'espérance, il suffit d'utiliser le fait que  $\varepsilon_{n+1}$  est centrée et que les estimateurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$  sont sans biais :*

$$\mathbb{E}[\varepsilon_{n+1}] = \mathbb{E}[\beta_0 - \hat{\beta}_0] + \mathbb{E}[\beta_1 - \hat{\beta}_1]x_{n+1} + \mathbb{E}[\varepsilon_{n+1}] = 0.$$

Ainsi la variance augmente lorsque  $x_{n+1}$  s'éloigne du centre de gravité du nuage. Autrement dit, faire de la prévision lorsque  $x_{n+1}$  est 'loin' de  $\bar{x}$  est périlleux, puisque la variance de l'erreur de prévision peut être très grande, Ceci s'explique intuitivement par le fait que plus une observation  $x_{n+1}$  est éloignée de la moyenne  $\bar{x}$  et moins on a d'information sur elle.

# Chapitre 2

## Régression linéaire multiple

Le modèle de régression linéaire multiple est une généralisation du modèle de régression linéaire simple C'est- à-dire qu'au lieu d'une seule variable explicative  $x_i$ , on va voir  $(p)$  variables explicatives,  $x_{ki}$ ,  $k = (1, \dots, p)$ . Alors dans ce chapitre nous considérons que la variable aléatoire à expliquer  $Y$  est modélisée par plusieurs variables explicatives  $X_k$ , alors on considère l'un des outils statistique le plus habituellement mis en œuvre pour l'étude des données multidimensionnelles, le traitement se fait uniquement à l'aide d'écritures matricielles. Pour plus de détails consulter les livres de : Cornillon et Matzner voir [\[8\]](#)

### 2.1 Modalisation

Le modèle de régression linéaire multiple est une généralisation du modèle de régression simple dans lequel les variables explicatives sont en nombre quelconque. il est défini par la formule suivante :

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{i,k} + \varepsilon_i, i = 1, \dots, n, n > p$$
$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n \quad (2.1)$$

où :

- $y_i$  :est la  $i$  <sup>ème</sup> variable à expliquée
- $x_{i,k}$  :est la  $i$  <sup>ème</sup> variable explicative(non aléatoire) pour la  $k$  <sup>ème</sup> individu (non aléatoire),
- Les erreurs  $\varepsilon_i$  sont des variables aléatoires inconnues,
- Les paramètres du modèle  $\beta_k$  ( $k = 0, \dots, p$ ) sont des constantes inconnus(paramètres du modèle).

L'écriture du modèle [2.1](#) est n'est pas très pratique. Afin de faciliter l'expression de certains résultats, on a habituellement recours aux notations matricielles. En écrivant le modèle [2.1](#), observation par observation, on obtient

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \dots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \dots + \beta_p x_{2p} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \dots + \beta_p x_{np} + \varepsilon_n \end{cases}$$

notons que le modèle de régression linéaire multiple peut encore s'écrire sous forme matricielle, comme suit :

$$Y = X\beta + \varepsilon,$$

telque :

$$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}$$

$Y$  est un vecteur aléatoire de dimension  $n$ ,

Les données sont rangées dans une matrice  $X$  de dimension  $(n, p + 1) \in R^n \times R^{p+1}$  dont la première colonne contient le vecteur unitaire indique la constante  $\beta_0$ ,

$\beta$  dans l'équation, est le vecteur des paramètres du modèle de dimension  $p + 1$ ,

$\varepsilon$  est le vecteur des erreurs de dimension  $n$ .

D'où :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} \dots x_{1,p} \\ 1 & x_{2,1} \dots x_{2,p} \\ \vdots & \ddots \\ 1 & x_{n,1} \dots x_{n,p} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

avec

- $\text{Dim } X = (n, p + 1), \text{Dim } Y = (n + 1), \text{Dim } \beta = (n, p + 1), \text{Dim } \varepsilon = (n, p + 1)$
- Les paramètres du modèle à estimer sont les coefficients  $\beta_1, \beta_2, \beta_3, \dots, \beta_{p+1}$  et la variance  $\sigma^2$ .
- On remarque que la première colonne de la matrice  $X$ , est le vecteur  $\mathbb{1}_n$  (tous les termes sont  $\mathbf{1}$ ), qui correspond au coefficient  $\beta_0$  (coefficient du terme constant).
- La matrice  $X$  est de  $n$  lignes et  $p + 1$  colonnes ( $p$  est le nombre de variables explicatives réelles, c'est-à-dire constante exclue) et  $n$  est le nombre d'observations.

## 2.2 Les hypothèses du modèle

$(H_1)$  : La matrice  $X$  est de plein rang, c-à-d  $\text{rang}(X) = p + 1$

$(H_2)$  :  $\mathbb{E}(\varepsilon) = 0_n \in R^n, \text{Var}(\varepsilon) = \sigma_\varepsilon^2 \mathbb{I}_n$  : Les erreurs sont centrées, ont la même variance, et non corrélées entre elles.

avec  $\mathbb{I}_n$  la matrice identité d'ordre  $n$  et  $0_n = (0, 0, \dots, 0)^t \in R^n$  :

$(H_3)$  : Les erreurs  $(\varepsilon_i)$  sont indépendantes des  $X_k$  ( $k = \overline{1, p}$ ).

$(H_4)$  : Les valeurs  $x_{ki}$  sont observées sans erreurs.

$(H_5)$  :  $\mathbb{E}(\varepsilon_i, \varepsilon_j^t) = 0$  si  $i \neq j$  (indépendance des erreurs).

$(H_6)$  :  $\text{Cov}(x_{ki}, \varepsilon_i) = 0$ , l'erreur est indépendante de  $x_{ki}$ .

( $H_7$ ) : Absence de colinéarité entre les variables explicatives  $\implies (X^t X)$  régulière et  $(X^t X)^{-1}$  existe.

( $H_8$ ) :  $(\frac{X^t X}{n})$  : tend vers une matrice finie non singulière.

( $H_9$ ) :  $n > p + 1$  : nombre d'observations est supérieur aux nombre des explicatives.

## 2.3 Estimation des paramètres

pour estimer le vecteur  $(\beta)$  composé des coefficients  $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ , et  $\sigma_\varepsilon^2$ , on a deux méthodes à suivre, la méthode des moindres carrés ordinaire (MCO) qui consiste à minimiser l'erreur quadratique, et la méthode du MV qui consiste à maximiser la quantité du vraisemblance par rapport aux paramètres inconnus du vecteur  $(\beta)$  et  $\sigma_\varepsilon^2$ , sous l'hypothèse d'indépendance et de normalité des erreurs voir [14].

### 2.3.1 Estimation par la méthode des moindres carrés ordinaires (MCO)

Soit le modèle sous forme matricielle à  $(p)$  variables explicatives et  $(n)$  observations

$$Y = X\beta + \varepsilon,$$

notons

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \left( \sum_{i=1}^n \varepsilon_i^2 \right) = \arg \min_{\beta \in \mathbb{R}^{p+1}} \Phi(\beta)$$

Afin de chercher la statistique  $\hat{\beta}$  qui minimise

$$\Phi(\beta) = \Phi(\beta_0, \dots, \beta_p) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \dots - \beta_p x_{pi}]^2$$

Pour calculer  $\hat{\beta}$ , nous proposons l'écriture matricielle suivante :

$$\begin{aligned}\Phi(\beta) &= \varepsilon^t \varepsilon = \|Y - X\beta\|^2 \\ &= (Y^t - X^t \beta^t)(Y - X\beta) \\ &= (Y - X\beta)^t (Y - X\beta) \\ &= Y^t Y - Y^t X\beta - \beta^t X^t Y + \beta^t X^t X\beta\end{aligned}$$

et puisque  $Y^t X\beta = \beta^t X^t Y \in \mathbb{R}$ , on obtient :

$$\Phi(\beta) = Y^t Y - 2\beta^t X^t Y + \beta^t X^t X\beta$$

Pour obtenir le minimum de la somme nous annulons les dérivées partielles de  $\Phi(\beta)$  par rapport aux paramètres  $\beta_k$ , et on obtient un système d'équation suivant :

$$\frac{\partial \Phi(\hat{\beta})}{\partial \beta} = 0 \implies -2X^t Y + 2X^t X \hat{\beta} = 0$$

où,

$$\frac{\partial(\beta^t X^t X\beta)}{\partial \beta} = \frac{\partial((\beta X)^t (X\beta))}{\partial \beta} = 2X^t X\beta$$

$$X^t X \hat{\beta} = X^t Y$$

Alors l'expression de l'estimateur est

$$\hat{\beta} = (X^t X)^{-1} X^t Y$$

où  $X^{-1}$  la matrice inverse de  $X$

1) La matrice  $X^t X$  est inversible, symétrique et rang  $(p + 1)$ .

2)  $X^t X$  est définie positive.

3) La valeur ajustée de  $Y$  est  $\hat{Y} = X\hat{\beta}$

4) Le vecteurs des résidus :

Le vecteur des résidus s'obtient simplement en faisant la différence entre le vecteur des valeurs observées de  $Y$  et le vecteur de valeurs prédites  $\hat{Y}$

$$\hat{\varepsilon} = Y - \hat{Y} = Y - X\hat{\beta} = Y - X(X^t X)^{-1} X^t Y = [\mathbb{I}_n - X(X^t X)^{-1} X^t] Y = (I_n - H)Y$$

$$\hat{\varepsilon} = PY$$

Le vecteur des résidus  $\varepsilon$  est orthogonal à la fois au vecteur des valeurs prédites  $\hat{Y}$  et à la matrice  $X$ . Ce qui signifie mathématiquement que  $\varepsilon^t \hat{Y} = 0$  et  $\varepsilon^t X = 0$

$$P = I_n - H$$

$$H = X(X^t X)^{-1} X^t$$

### 2.3.2 Propriétés des Estimateurs :

**Proposition 2.3.1** (*Biais et matrice de variance covariance*)

1/  $\hat{\beta}$  est un estimateur sans biais :

A) l'estimateur  $\hat{\beta}$  de  $\beta$  est sans biais c-à-d :

$$E(\hat{\beta}) = \beta$$

B) et de variance

$$Var(\hat{\beta}) = \sigma_\varepsilon^2 (X^t X)^{-1}$$

on a :

$$\begin{aligned}\hat{\beta} &= (X^t X)^{-1} X^t Y = (X^t X)^{-1} X^t \beta + (X^t X)^{-1} X^t \varepsilon = \beta + (X^t X)^{-1} X^t \varepsilon \\ \hat{\beta} - \beta &= (X^t X)^{-1} X^t \varepsilon\end{aligned}\tag{2.2}$$

d'où

$$E(\hat{\beta}) = E(\beta + (X^t X)^{-1} X^t \varepsilon) = \beta + (X^t X)^{-1} X^t E(\varepsilon)$$

Donc,

$$E(\hat{\beta}) = \beta$$

2/ Estimateur de la matrice de variance-covariance de  $\hat{\beta}$  :

$$Var(\hat{\beta}) = \mathbb{E} \left[ (\hat{\beta} - \mathbb{E}(\hat{\beta})) (\hat{\beta} - \mathbb{E}(\hat{\beta}))^t \right] = \mathbb{E} \left[ (\hat{\beta} - \beta) (\hat{\beta} - \beta)^t \right]\tag{2.3}$$

$$Var(\varepsilon) = \mathbb{E}(\varepsilon \varepsilon^t) = \sigma_\varepsilon^2 \mathbb{I}$$

**Proposition 2.3.2** (*Propriétés de  $\hat{\varepsilon}$  et  $\hat{Y}$* ) Sous les hypothèses du modèle ( $H_1$ ) et ( $H_2$ ), on a :

1.  $\mathbb{E}(\hat{\varepsilon}) = 0$ .
2.  $Var(\hat{\varepsilon}) = \sigma_\varepsilon^2 P$ .

Où :

$$\left\{ \begin{array}{l} \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \\ \varepsilon \sim N(0, \mathbb{I}_n \sigma_\varepsilon^2) \end{array} \right\}, \left\{ \begin{array}{l} \sigma_\varepsilon = \frac{1}{n-p-1} \varepsilon^t \varepsilon \\ \sigma_\varepsilon = \frac{1}{n-p-1} \sum_{i=1}^n \varepsilon_i^2 \end{array} \right.$$

3.  $\mathbb{E}(\hat{Y}) = X\beta$ .
4.  $Var(\hat{Y}) = \sigma_\varepsilon^2 H$ .

5.  $Cov(\hat{\beta}, \hat{\varepsilon}) = 0$ .

**Proposition 2.3.3** (Biais de  $\sigma_\varepsilon^2$ ) :

tanque la variance résiduelle  $\sigma_\varepsilon^2$  est inconnu, alors son estimateur sans biais  $\hat{\sigma}_\varepsilon^2$  est donné par :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-p-1} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n-p-1} \hat{\varepsilon}^t \hat{\varepsilon}$$

L'estimateur sans biais de la variance  $\hat{\sigma}_\varepsilon^2$  est donné par :

$$S^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{n-p-1} = \frac{SCE}{n-p-1}$$

où  $SCE$  est la somme des carrés des résidus

$$SCE = \left\| \hat{Y}_i - \bar{Y} \mathbb{I}_n \right\|^2$$

On définit également la somme totale des carrés

$$SCT = \left\| Y_i - \bar{Y} \mathbb{I}_n \right\|^2 = Y^t Y - n(\bar{Y})^2$$

et la somme des carrés de la régression

$$SCR = \left\| \hat{\varepsilon} \right\|^2$$

On peut aisément vérifier que :

$$SCT = SCR + SCE$$

Le vecteur des valeurs ajustées?? peut être interprété comme la projection de  $Y$  sur

le sous-espace engendré par les colonnes de la matrice  $X$

$$\hat{Y} = P_x Y$$

Où  $P_x$  est l'opérateur de projection.

$$P_x = X(XX)^{-1}X^t.$$

## 2.4 Estimation par la méthode MV

La méthode de l'estimation par maximum de vraisemblance consiste à estimer les paramètres inconnus  $\beta$  et  $\sigma_\varepsilon^2$  à faire une hypothèse sur la distribution de probabilité de  $\varepsilon_i$ . En effet, on suppose, que les  $\varepsilon_i$  sont indépendants et suivent une loi normale de moyennes nulle et de variance  $\sigma_\varepsilon^2$

En reprenant, l'équation du modèle linéaire initiale, on a :

$$Y = X\beta + \varepsilon,$$

L'hypothèse de normalité du vecteur des résidus se présente comme suit :

$$\varepsilon \sim N_n(0_n, \sigma_\varepsilon^2 \mathbb{I}_n)$$

Il faut aussi noter que comme  $Y = X\beta + \varepsilon$  et que  $\varepsilon \sim N_n(0_n, \sigma_\varepsilon^2 I_n)$  alors on aura :

$$Y \sim N_n(X\beta, \sigma_\varepsilon^2 I_n)$$

$N_n$  : désigne la loi normale multidimensionnelle (de dimension  $n$ ). Cette méthode est

basée sur la maximisation de la fonction de vraisemblance, définie comme suit :

$$L(\beta, \sigma_\varepsilon^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\Pi\sigma_\varepsilon^2}} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} \left[ y_i - \beta_0 - \sum_{k=1}^p \beta_k x_{ki} \right]^2 \right\}$$

ce qui revient à maximiser la fonction vraisemblance, donnée par :

$$L(\beta, \sigma_\varepsilon^2) = \left( \frac{1}{\sqrt{2\Pi\sigma_\varepsilon^2}} \right)^n \exp \left[ -\frac{1}{2\sigma_\varepsilon^2} \sum_{k=1}^n (y_i - \beta_0 - \sum_{k=1}^p \beta_k x_{ki})^2 \right]$$

Cependant, étant donné le nombre de paramètre à estimer, pour alléger l'étape des dérivations, il est judicieux de présenter la fonction de vraisemblance sous la forme matricielle afin de faciliter la dérivation. En effet, la forme matricielle de la fonction de vraisemblance est :

$$L(\beta, \sigma_\varepsilon^2) = \left( \frac{1}{\sqrt{2\Pi\sigma_\varepsilon^2}} \right)^n \exp \left[ -\frac{1}{2\sigma_\varepsilon^2} (Y - X\beta)^t (Y - X\beta) \right]$$

Car il facile de chercher à maximiser le logarithme de la fonction de vraisemblance plutôt que la fonction elle-même, on prend le logarithme de la fonction de vraisemblance :

$$\text{Log } L(\beta, \sigma_\varepsilon^2) = -\frac{n}{2} \log(2\Pi) - \frac{n}{2} \log(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} (Y - X\beta)^t (Y - X\beta)$$

La méthode du maximum de vraisemblance consiste à choisir les paramètres  $\beta$  et  $\sigma_\varepsilon^2$  de sorte à maximiser cette fonction de vraisemblance. Pour cela, il faut dériver la dérivée et retrouver les conditions de premier ordre afin d'en déduire chacun des paramètres. En dérivant cette fonction par rapport à  $\beta$  et à  $\sigma_\varepsilon^2$ , on trouve :

$$\begin{cases} \frac{\partial \text{Log}L(\hat{\beta}, \hat{\sigma}_\varepsilon^2)}{\partial \hat{\beta}} = \frac{1}{\hat{\sigma}_\varepsilon^2} (X^t Y - X^t X \hat{\beta}) = 0 \\ \frac{\partial \text{Log}L(\hat{\beta}, \hat{\sigma}_\varepsilon^2)}{\partial \hat{\sigma}_\varepsilon^2} = -\frac{n}{2\hat{\sigma}_\varepsilon^2} - \frac{1}{2(\hat{\sigma}_\varepsilon^2)^2} (Y - X\hat{\beta})^t (Y - X\hat{\beta}) = 0 \end{cases}$$

En résolvant ce système on trouve :

$$\hat{\beta}_{mv} = (X^t X)^{-1} X^t Y$$

On constate alors que sous l'hypothèse de normalité, l'estimateur de maximum de vraisemblance (MV) est égal à l'estimateur des moindres carrés ordinaires.

Si

$$\varepsilon \sim N_n(0_n, \sigma_\varepsilon^2 \mathbb{I}_n) \Leftrightarrow Y \sim N_n(X\beta, \sigma_\varepsilon^2 \mathbb{I}_n) \implies \hat{\beta}_{mv} = \hat{\beta}_{mco} = (X^t X)^{-1} X^t Y$$

On montre par ailleurs que la variance de l'estimateur MV est :

$$Var(\hat{\beta}_{mv}) = \sigma_\varepsilon^2 (X^t X)^{-1}$$

Cependant  $\sigma_\varepsilon^2$  étant, en général inconnu, il faut considérer sa valeur estimée  $\hat{\sigma}_\varepsilon^2$  :

$$\hat{\sigma}_\varepsilon^2 = \frac{\varepsilon^t \varepsilon}{n - p}$$

Où  $(n)$  est le nombre d'observations et  $(p)$  est le nombre de paramètres à estimer ( $y$  compris la constante).  $(p)$  correspond au rang de la matrice. Ainsi  $(n - p)$  est le nombre de degrés de liberté. Avec cette variance estimée, la variance de l'estimateur MV est :

$$Var(\hat{\beta}_{mv}) = \hat{\sigma}_\varepsilon^2 (X^t X)^{-1} = \frac{1}{n} \varepsilon^t \varepsilon$$

Cette expression représente la variance estimée de l'estimateur de MV. Il faut aussi noter son espérance qui est égale à  $\beta$  car :

$$\mathbb{E}(\hat{\beta}_{mv}) = \beta$$

· Si l'estimateur du maximum de vraisemblance admet une solution unique, alors cet estimateur est convergent et asymptotiquement efficace du paramètre. De plus, cet estimateur converge en loi vers une normale

$$(\hat{\beta}_{mv}) \sim \mathcal{N}(\beta, \hat{\sigma}_\varepsilon^2 (X^t X)^{-1})$$

· Cependant, l'estimateur du maximum de vraisemblance n'est pas nécessairement sans biais. L'estimateur du maximum de vraisemblance de est en effet biaisé.

· On remarque que les deux méthodes MCO et MV donnent les mêmes estimateurs pour  $\beta$ , alors que les estimateurs de  $\sigma_\varepsilon^2$  sont différents

## 2.5 Interprétation géométrique

Dans la régression linéaire multiple, au lieu de travailler dans un plan (comme dans le cas de la régression linéaire simple quand on a une seule variable explicative  $X^t$ ), nous travaillons dans un espace de dimension supérieure. Le plan de projection formé par  $(X^t, \mathbb{I})$  dans la régression linéaire simple est remplacé par un hyperplan formé par les vecteurs  $X_1^t, \dots, X_p^t, \mathbb{I}$ . Cet hyperplan est un sous-espace de l'espace des variables explicatives. Régresser  $Y$  sur les  $(p)$  variables explicatives consisterait à projeter orthogonalement  $Y$  sur l'hyperplan déterminé par  $X_1^t, \dots, X_p^t, \mathbb{I}$ .

Autrement dit, nous ajustons les coefficients du modèle de sorte que la somme des carrés des distances entre les valeurs observées de  $Y$  et les valeurs prédites sur cet hyperplan soit minimisée.

·  $P$  est la matrice de projection orthogonale sur  $M^\perp(X)$  (l'espace orthogonal à  $M(X)$ )

et  $H$  est la matrice de projection orthogonale sur  $M(X)$  (le sous espace de  $R^n$  engendré par les  $(p + 1)$  vecteurs de la matrice  $X$ ) voir [14].

## 2.6 Qualité de l'ajustement

De même que la régression simple, on a :

$$\begin{aligned}\sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (\hat{y}_i - y_i)^2 \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (\varepsilon_i)^2 \\ SCT &= SCE + SCR\end{aligned}$$

· La qualité d'ajustement est mesurée par le coefficient de détermination  $R^2$  (La mesure de l'ajustement (empirique)), défini par :

$$D = R^2 = \frac{\left\| \hat{Y} - \bar{y}\mathbf{1} \right\|^2}{\left\| Y - \bar{y}\mathbf{1} \right\|^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \leq 1$$

où  $0 \leq R^2 \leq 1$

· Le coefficient de détermination  $R^2$  est un indicateur de la qualité d'ajustement qui exprime le rapport entre la variation expliquée et la variation totale, dont

Dans le cas où  $R^2$  est proche de 1 alors l'ajustement est meilleur, c-à-d la variation expliquée est proche de la variation totale alors la variation résiduelle est proche de 0.

· Cette propriété tient uniquement lorsqu'il existe une constante dans la régression.

Mais lorsqu'il n'y a pas de constante dans l'équation estimée, la propriété sur la somme des carrés totale n'est plus vérifiée car  $\sum_{i=1}^n \varepsilon_i \neq 0$ .

· Il faut aussi savoir que lorsque cette propriété n'est pas vérifiée, le test de significativité globale ou le calcul du  $R^2$  ne sont plus valables.

· Puisque  $R^2$  dépend fortement de  $p$ , on ne peut pas être utilisé pour comparer la qualité de deux modèles de la régression linéaire multiple qui diffèrent quant

aux nombre de variables explicatives, alors on lui préfère d'utiliser le coefficient de  $\overline{R^2}$ détermination ajusté , qui définit par :

avec  $n > p$ , où :

$n$  : la taille de l'échantillon

$p$  : nombre des paramètres

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \overline{R^2} = 1 - \frac{\frac{1}{n-p-1} \left( \sum_{i=1}^n \hat{\varepsilon}_i^2 \right)}{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\frac{1}{n-p-1} SCR}{\frac{1}{n-1} SCT}$$

$$\overline{D} = \overline{R^2} = 1 - \frac{n-1}{n-p-1} (1 - R^2)$$

où

- $n$  : est le nombre d'observations
- $(p + 1)$  est nombre de séries explicatives.
- Le coefficient de détermination corrigé  $\overline{R^2}$  vérifie les propriétés suivantes :
  - 1)  $\overline{R^2} < R^2$  si  $p > 1$ .
  - 2)  $\overline{R^2} = R^2$  si  $p = 1$ .
  - 3)  $\overline{R^2}$  peut prendre des valeurs négatives.

## 2.7 Lois des estimateurs

Nous avons ce qui précède, la gaussianité des erreurs voir [3] :

$$\varepsilon \sim N_n(0, \sigma_\varepsilon^2 I_n)$$

donc  $Y$  suit la loi normale dans  $R^n$  :

$$Y \sim N_n(X\beta, \sigma_\varepsilon^2 \mathbb{I}_n)$$

$N_n$  : la Gaussien de menssion (n)

la densité du vecteur  $Y$  est :

$$f_{(y_1, \dots, y_n)}(Y_1, \dots, Y_n) = \frac{1}{\sigma^n (\sqrt{2\pi})^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} (Y - X\beta)^t (Y - X\beta) \right\}$$

### **Théorème 2.7.1**

Sous les hypothèses précédentes, nous avons

- i)  $\hat{\beta} \sim N_{p+1}(\beta, \sigma_\varepsilon^2 (X^t X)^{-1})$
- ii)  $(n - p - 1) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{n-p-1}^2$ .
- iii)  $\hat{\beta}$  et  $\hat{\sigma}_\varepsilon^2$  sont deux variablse aléatoire indépendantes.

i) est déjà vu

$$\text{ii) } \hat{\sigma}_\varepsilon^2 = \frac{1}{n-p-1} \sum_{i=1}^n \hat{\varepsilon}_i^2.$$

## **2.8 Tests et Intervalles de confiances**

### **2.8.1 Intervalles et région de confiances**

on a d'après le théorème 2.7.1 :

Pour chaque  $j = \overline{0, p}$ ,  $\mathbb{E}(\hat{\beta}_j) = \beta_j$

Et soit  $V_j, j = 0, \dots, p$ , les éléments de la diagonale de  $(X^t X)^{-1}$

- i) Si  $\sigma_\varepsilon^2$  est connue. Notons  $V_j$  est le  $j^{eme}$  terme diagonal de la matrice  $(X^t X)^{-1}$ ,

alors

$$\frac{|\hat{\beta}_j - \beta_j|}{\sigma\sqrt{V_j}} \sim N(0, 1), j = \overline{0, p}$$

Donc,

$$Var(\hat{\beta}_j) = V_j\sigma_\varepsilon^2$$

ii) Si  $\sigma_\varepsilon^2$  est inconnue. Pour  $j = \overline{0, p}$ .

Alors

$$\frac{|\hat{\beta}_j - \beta_j|}{\sigma\sqrt{V_j}} \sim T_{n-p-1}$$

où  $T_{n-p-1}$  est une loi de Student à  $(n - p - 1)$  ddl.

iii) Si  $\sigma_\varepsilon^2$  est inconnue, alors :

$$\frac{1}{(p+1)\sigma_\varepsilon^2} (R(\hat{\beta} - \beta))^t (R(X^t X)^{-1} R^t)^{-1} (R(\hat{\beta} - \beta)) \sim F(p+1; n-p-1)$$

où  $R$  est la matrice de taille  $(p+1 \times p)$  dont tous les éléments sont nuls sauf les  $R_{ij}$  qui valent 1, et  $F(p+1, n-p-1)$  est une loi de Fisher à  $(p+1)$  et  $(n-p-1)$  ddl.

Ceci nous permet de donner des intervalles de confiance pour  $\beta_j, j = \overline{0, p}$ .

En effet :

$$P \left\{ \frac{|\hat{\beta}_j - \beta_j|}{\hat{\sigma}\sqrt{V_j}} \leq t \right\} = 1 - \alpha$$

où  $(t)$  est la valeur tabulés de la loi de student  $T_{1-\frac{\alpha}{2}}$

$$P \left\{ -t\hat{\sigma}\sqrt{V_j} \leq \hat{\beta}_j - \beta_j \leq t\hat{\sigma}\sqrt{V_j} \right\} = 1 - \alpha$$

$$P \left\{ \hat{\beta}_j - t\hat{\sigma}\sqrt{V_j} \leq \beta_j \leq \hat{\beta}_j + t\hat{\sigma}\sqrt{V_j} \right\} = 1 - \alpha$$

L'intervalle de confiance au niveau  $(1 - \alpha)$  pour chaque coefficient  $\beta_j$  du modèle est :

$$\beta_j \in \left[ \hat{\beta}_j - t \hat{\sigma} \sqrt{V_j}; \hat{\beta}_j + t \hat{\sigma} \sqrt{V_j} \right]$$

où  $(t)$  étant le fractile d'ordre  $(1 - \frac{\alpha}{2})$  de la loi de Student à  $(n - p - 1)$  ddl.

De même, pour l'intervalle de confiance de  $\sigma_\varepsilon^2$ .

$$P \left\{ t_2 \leq \frac{(n-p) \hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \leq t_1 \right\} = 1 - \alpha$$

$$\sigma_\varepsilon^2 \in \left[ \frac{(n-p) \hat{\sigma}_\varepsilon^2}{t_1}; \frac{(n-p) \hat{\sigma}_\varepsilon^2}{t_2} \right]$$

où  $t_1$  (respt  $t_2$ ) est le fractile d'ordre  $(1 - \alpha/2)$  (respt  $\alpha/2$ ) de la loi de  $\chi_{n-2}^2$ .

## 2.8.2 La région de confiance

Finalement, la région de confiance au niveau  $(1 - \alpha)$  pour  $p+1$  ( $p+1 \leq p$ ) paramètres  $\beta_j$  notés  $(\beta_{j_1}, \dots, \beta_{j_q})$  est :

$$RC = \left\{ R\beta \in \mathbb{R}^{p+1}, \frac{1}{(p+1) \hat{\sigma}_\varepsilon^2} (R(\hat{\beta} - \beta))^t (R(X^t X)^{-1} R^t)^{-1} (R(\hat{\beta} - \beta)) \leq f \right\}$$

où  $(f)$  étant le fractile d'ordre  $(1 - \frac{\alpha}{2})$  de la loi de Fisher  $F(p+1, n-p)$  ddl.

## 2.9 Tests d'hypothèses

### 2.9.1 Tests d'hypothèses simple (Test de signification des paramètres)

pour tester

$$\begin{cases} H_0 : \beta_j = \beta \\ H_1 : \beta_j \neq \beta \end{cases}, \beta \in \mathbb{R}$$

si  $\alpha$  est le niveau de confiance  $\alpha \in ]0, 1[$ , alors on accepte l'hypothèse  $H_0$

si:

$$\frac{|\hat{\beta}_j - \beta_j|}{\hat{\sigma} \sqrt{V_j}} \leq T_{1-\alpha/2}$$

si  $\beta = 0$ , on dit qu'on fait un test de signification de student

## 2.9.2 Tests d'hypothèses globale (Test de signification globale)

test d'hypothèse globale

$$\begin{cases} H_0 : \hat{\beta}_j = \beta_0 \\ H_1 : \hat{\beta}_j \neq \beta_0 \end{cases}, \beta_0 \in \mathbb{R}^{p+1}$$

Le modèle dans ce cas sera tester globalement :

on sait :

$$\frac{(\hat{\beta} - \beta)^t (X^t X)^{-1} (\hat{\beta} - \beta)}{(p+1) \sigma_\varepsilon^2} \sim \chi_{p+1}^2 / p + 1$$

on a :

$$(n-p-1) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{(n-p-1)}^2 \implies \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{(n-p-1)}^2 / n - p - 1$$

$$\begin{aligned} Q(\hat{\beta}, \hat{\sigma}_\varepsilon^2) &= \frac{(\hat{\beta} - \beta)^t (X^t X)^{-1} (\hat{\beta} - \beta)}{(p+1) \sigma_\varepsilon^2} \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \\ &= \frac{(\hat{\beta} - \beta)^t (X^t X)^{-1} (\hat{\beta} - \beta) (n-p-1)}{(p+1) \hat{\sigma}_\varepsilon^2} \end{aligned}$$

Source de la variance	<i>ddl</i>	Somme de scarrés	Moyenne des carrés	$F_c$
<i>Expliquée</i>	$P$	$SCE$	$MCE = \frac{MCE}{p}$	$F_c = \frac{MCE}{MCR}$
<i>Résiduelle</i>	$n - p - 1$	$SCR$	$MCR = \frac{SCR}{n-p-1}$	
<i>Totale</i>	$n - 1$	$SCT$		

TAB. 2.1 – (Table d’ANOVA à deux facteurs )

avec  $\hat{\beta}$  et  $\hat{\sigma}_\varepsilon^2$  sont indépendante, on accepte  $H_0$  au niveau  $\alpha$  , si

$$Q(\hat{\beta}, \hat{\sigma}_\varepsilon^2) \leq f.$$

Où  $f$  est le fractile d’ordre  $(1 - \alpha)$  de la loi de Fisher  $F(p + 1, n - p - 1)$ .

Les résultats sont présentés dans un tableau appelé le tableau d’analyse de la variance à deux facteur (ANOVA) sous forme suivante :

$$F_c = \frac{MCE}{MCR} \leq f$$

Où  $f$  est le fractile d’ordre  $(1 - \alpha)$  de la loi de Fisher  $F(p + 1, n - p - 1)$ .

### 2.9.3 Test sur le modèle réduit

Pour tester la nullité de certains termes on écrit :

$$\begin{cases} H_0 : \beta_j = 0, j = \overline{0, p} \\ H_1 : \beta_j \neq 0, j = \overline{0, q} \end{cases}, q \prec p$$

Où  $(p)$  est le nombre exact des paramètres du modèle. Soit  $R_q^2$  le coefficient de détermination du modèle réduit à  $(p - q)$  variables.

Sous  $H_0$ , la statistique de test est :

$$Q_q = \frac{(R^2 - R_q^2)(n - p - 1)}{(1 - R^2)q} = \frac{(SCE - SCE_q)/q}{(SCR)/(n - p - 1)}$$

Où  $SCE_q$  est la somme des carrés expliquées du modèle réduit.

On compare la valeur de  $Q_q$  avec la valeur tabulée de la loi de Fisher à  $(p, n - p - 1)$  degré de liberté :  $f_{(n-p-1)}^{1-\alpha}$

On accepte  $H_0$  si :

$$Q_q \leq f$$

Où  $f$  est le fractile d'ordre  $(1 - \alpha)$  de la loi de Fisher  $F(q, n - p - 1)$ .

## 2.10 Prédiction

Comme mentionné précédemment, que la prédiction est l'objectif le plus fondamentale dans la régression. soit les valeurs des variables  $X^j$  pour une nouvelle observation :  $X_0^j = [1, X_0^1, X_0^2, \dots, X_0^p]$  appartenant au domaine dans lequel l'hypothèse de linéarité reste valide, une prédiction notée  $\hat{Y}_0$  de  $Y$  ou  $\mathbb{E}(Y)$  est donnée par :

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0^1 + X_0^2 + \dots + \hat{\beta}_q X_0^p.$$

Les intervalles de confiance des prévisions de  $Y$  et  $\mathbb{E}(Y)$ , au niveau de confiance  $(1 - \alpha)\%$ , sont données respectivement par :

$$Y_p = \hat{Y}_0 \pm t \hat{\sigma}_\varepsilon (1 + X_0 (X^t X)^{-1} X_0^t)^{1/2}$$

$$\mathbb{E}(Y_p) = \hat{Y}_0 \pm t \hat{\sigma}_\varepsilon (X_0 (X^t X)^{-1} X_0^t)^{1/2}$$

Où  $t$  est le fractile d'ordre  $(1 - \alpha/2)$  de la loi de student  $T(n - p - 1)$ .

# Chapitre 3

## Application sous R

Dans ce chapitre on va mettre en Œuvre une regression linéaire multiple sous le logiciel R, en utilisant un exemple numérique réalisé sur des données réelles pour plus de détails voir ([12]).

### 3.1 Initialisation de la régression linéaire multiple avec R

#### Objectif

On peut vouloir expliquer le prix de vente des maisons, en fonction de leur surface en présumant que plus la surface est élevée, et plus le prix de vente sera élevé. Il s'agit là d'une regression dite simple car elle ne comporte qu'une seule variables explicative. De plus, on peut aussi supposer que le nombre de chambres influence le prix de la maison à la hausse , il s'agira là d'une regression multiple avec deux variables explicatives.

L'objectif est d'évaluer si chacune des deux variables influence le prix, et, si tel est le cas, de tenter de quantifier cet effet.

### 3.1.1 Modèle de régression linéaire multiple

Soit le modèle suivant :

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

· où  $\beta_0, \beta_1, \beta_2$  sont des coefficients réels inconnus et est une variable quantitative de valeur moyenne nulle,

·  $Y$  est la variable à expliquer ( prix de vente)

·  $x_1$  et  $x_2$  sont les variables explicatives ( la surface et le nombre de chambres)

·  $\varepsilon$  est le terme d'erreur, assumé être distribué selon une loi Normale

Sous un certain nombre d'hypothèses, la méthode des moindres carrés ordinaires permet d'obtenir des estimateurs BLUE (Best Linear Unbiased Estimators), c'est-à-dire qui sont sans biais et à variance minimale. Les coefficients estimés sont issus de la minimisation de la somme des carrés des résidus entre les valeurs prédites et les valeurs observées

#### Présentation des données

Les données dont on dispose sont (15) observations se présentent sous forme d'un tableau, composé de quinze observations et trois variables :

sale.price = prix de vente de la maison (en milliers de dollars australiens).

area = surface au sol de la maison.

bedrooms = nombre de chambres dans la maison.

#### Importation les données

### 3.1.2 Construction du modèle

La construction du modèle de régression multiple on utilisant la fonction `lm()`

```

      area bedrooms sale.price
9      694         4      192.0
10     905         4      215.0
11     802         4      215.0
12    1366         4      274.0
13     716         4      112.7
14     963         4      185.0

```

FIG. 3.1 – Présentations des données

`pricereg2<-lm(sale.price ~area + bedrooms , data=houseprices)`

ou bien la fonction `summary` : `summary(pricereg2)`, qui nous donne les resultats suivants

```

Call:
lm(formula = sale.price ~ area + bedrooms, data = houseprices)

Residuals:
    Min       1Q   Median       3Q      Max
-80.897  -4.247   1.539  13.249  42.027

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -141.76132    67.87204  -2.089  0.05872 .
area          0.14255     0.04697   3.035  0.01038 *
bedrooms     58.32375    14.75962   3.952  0.00192 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.06 on 12 degrees of freedom
Multiple R-squared:  0.731,    Adjusted R-squared:  0.6861
F-statistic: 16.3 on 2 and 12 DF,  p-value: 0.0003792

```

FIG. 3.2 – Constrction du modèle de régression multiple

Et pour extraire que les coefficients estimés on utilise `coef()` : `coef(pricereg2)`, qui nous donne

```

      (Intercept)      area      bedrooms
-141.7613221    0.1425469    58.3237508

```

FIG. 3.3 – Coéfficient du modèle

### 3.1.3 Représentation graphique

On obtient donc l'équation de la droite de regression suivant :

$$Y = -141.76 + 0.14X_1 + 58.32X_2$$

que l'on peut tracer avec le nuage de points, ou nous pouvons visualiser l'ajustement des données pour chaque variable indépendamment en utilisant des graphiques 2D

#### Le graphe de la surface vs. prix de vente

```
plot(data$area, data$sale_price, main = "Surface vs. Prix de vente"
```

```
xlab = "Surface (area)", ylab = "Prix de vente (sale_price)", pch = 19)
```

et pour ajouter la ligne de régression pour surface en utilise la fonction abline :

```
abline(lm(sale_price ~ area, data = data), col = "blue")
```

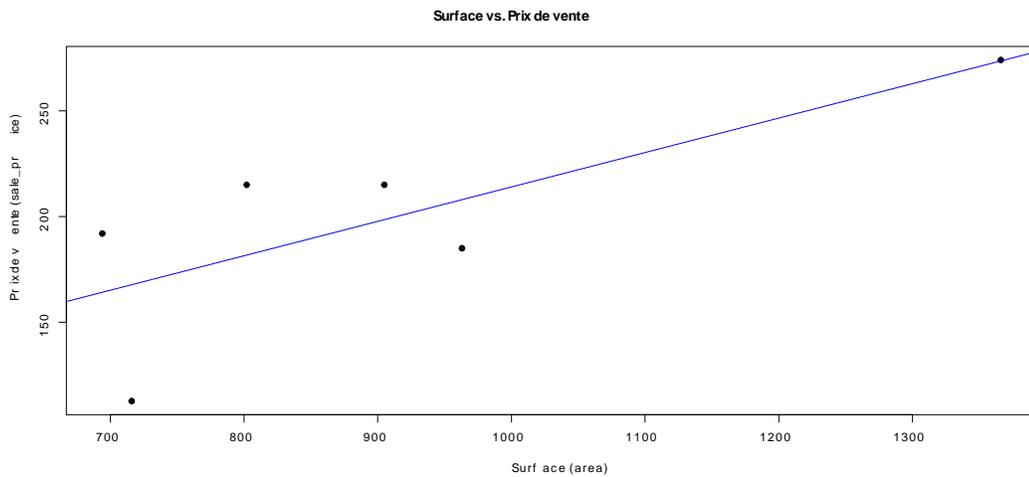


FIG. 3.4 – La droite de régression du prix de vente en terme de surface

Les coefficients associés aux variables 'surface' et 'chambres' sont significatifs (respectivement à 95% et à 99%) on peut donc les interpréter. Toutes choses égales par ailleurs, une chambre supplémentaire augmente par exemple le prix de la maison

environ de 58 milles dollars , et 100 unités de surface (inconnue) supplémentaires vont l'augmenter de 14 mille dollars, toutes choses égales par ailleurs.

### Le graphe du nombre de chambres vs. prix de vente

```
plot(data$bedrooms, data$sale_price, main = "Nombre de chambres vs. Prix
de vente"
xlab = "Nombre de chambres (bedrooms)", ylab = "Prix de vente (sale_price)",
pch = 19)
```

La ligne de régression pour chambres se fait par la fonction abline

```
abline(lm(sale_price ~bedrooms, data = data), col = "red")
```

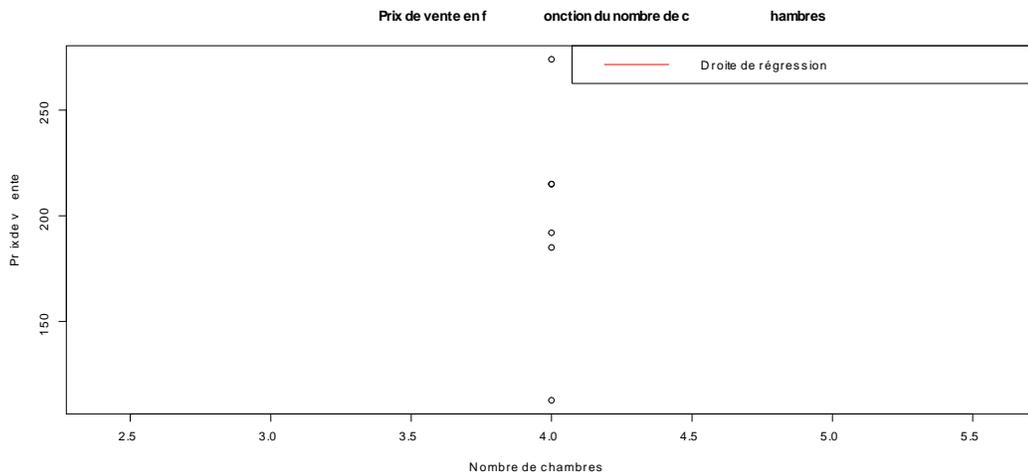


FIG. 3.5 – la droite de régression de prix de vente en terme de nombre de chambres

### 3.1.4 Intervalles de confiance

pour afficher l'intervalle de confiance à 95% pour les coefficients estimés en utilise la fonction confint() :

confint(pricereg2), qui nous donne

```

                2.5 %    97.5 %
(Intercept) -289.64179643  6.1191523
area         0.04019939  0.2448945
bedrooms     26.16530118 90.4822003
    
```

FIG. 3.6 – Intervalle de confiance à 95%

### Valeurs prédites

On utilise la fonction `fitted()` qui permet d'extraire les valeurs prédites qui se définie comme suit :

`fitted(pricereg2)`, dont les valeurs prédites sont

```

     9     10     11     12     13     14     15     16
190.4612 220.5387 205.8563 286.2528 193.5973 228.8064 208.5647 193.3122
     17     18     19     20     21     22     23
236.6465 217.9728 204.1458 249.0701 259.7611 293.2596 377.9546
    
```

FIG. 3.7 – Valeurs prédites

### Residuels

`resid()` permet d'extraire les résidus (Valeur prédite - Valeur réelle) :

`resid(pricereg2)`, qui donne les resultats suivantes

```

     9     10     11     12     13     14
1.5387504 -5.5386516  9.1436820 -12.2527859 -80.8972821 -43.8063735
     15     16     17     18     19     20
3.4352904 26.6878118 39.3535454 42.0271931 17.3542452  5.9299058
     21     22     23
0.2388861 -0.2596422 -2.9545748
    
```

FIG. 3.8 – Les résidus(Valeur prédite -Valeur réelle)

Plusieurs hypothèses sous-jacentes au modèle de regression linéaire portent sur les

résidus de la regression : indépendance, homoscedasticité (même variance) et normalité. On peut représenter graphiquement les résidus, afin d'observer si ces conditions sont (plus ou moins) respectées. Par exemple, on peut tracer :

- les résidus en ordonnées
- résidus en ordonnées, valeurs prédites en abscisse.
- résidus en ordonnées, variables explicatives en abscisse.

Si les résidus ne sont pas répartis de manière relativement indépendantes, de part et d'autre de la droite d'équation  $Y = 0$  (la moyenne des résidus est nulle par hypothèse), alors on suspecte un problème (corrélations, hétéroscedasticité ...) que l'on doit corriger avant de pouvoir interpréter un quelconque résultat

### **Le graphe de résidus**

en appliquant

```
res<-resid(pricereg2)
plot(res,main="Résidus")
abline(h=0,col="red")
```

pour avoir la figure suivant

On observe que les résidus semblent suivre une tendance. Sur des données en série temporelle (traditionnellement ordonnées chronologiquement), cela pourrait indiquer une auto-corrélation des erreurs (contraire à l'hypothèse d'indépendance), et donc de phénomène non pris-en compte (ex : phénomène de mode, effet saisonnier, etc).

et avec ces commandes on avoir la figure suivante :

```
plot(houseprices$area,res,main="Résidus")
abline(h=0,col="red")
```

On peut être tenté de voir une diminution des résidus (négatifs) lorsque la surface

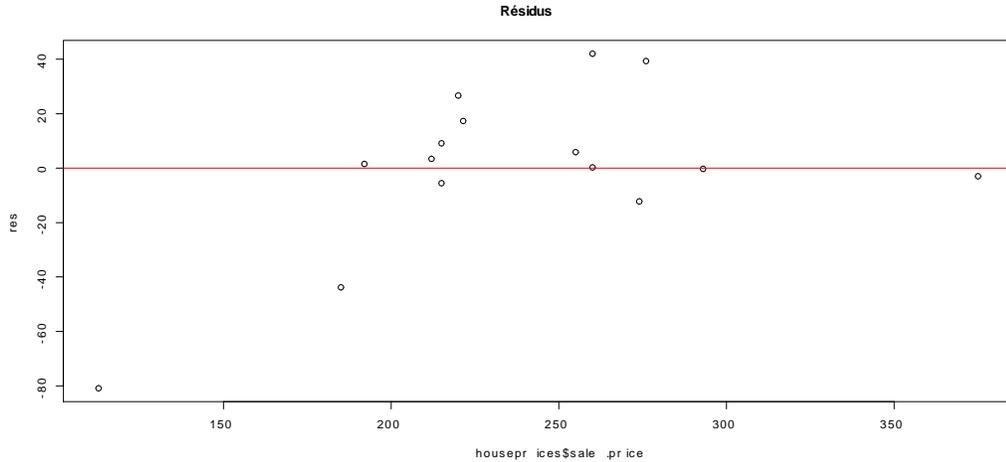


FIG. 3.9 – Présentation des résidus

de la maison augmente (mais ce n'est pas très marqué, d'autant plus que l'on n'a pas beaucoup d'observations). Comme il est montré dans la figure suivante

```
plot(houseprices$sale_price,res,main="Résidus")
```

```
abline(h=0,col="red")
```

Comme sur le graphique précédent, une valeur extrême (petite surface, petit prix) semble perturber un peu l'analyse...

pour l'extraction des valeurs prédites et en traçant les residus vs. valeurs prédites

```
fit<-fitted(pricereg2)
```

```
plot(fit,res,main="Residuals vs. fitted")
```

```
abline(h=0,col="red")
```

Là aussi, on peut être tenté de remarquer une diminution des résidus avec l'augmentation de la valeur prédite, mais ce n'est toutefois toujours pas extrêmement marqué (d'autant plus qu'il n'y a que 15 observations).

### 3.1.5 Interprétation

Avant d'interpréter les résultats, il convient d'évaluer la significativité du modèle.

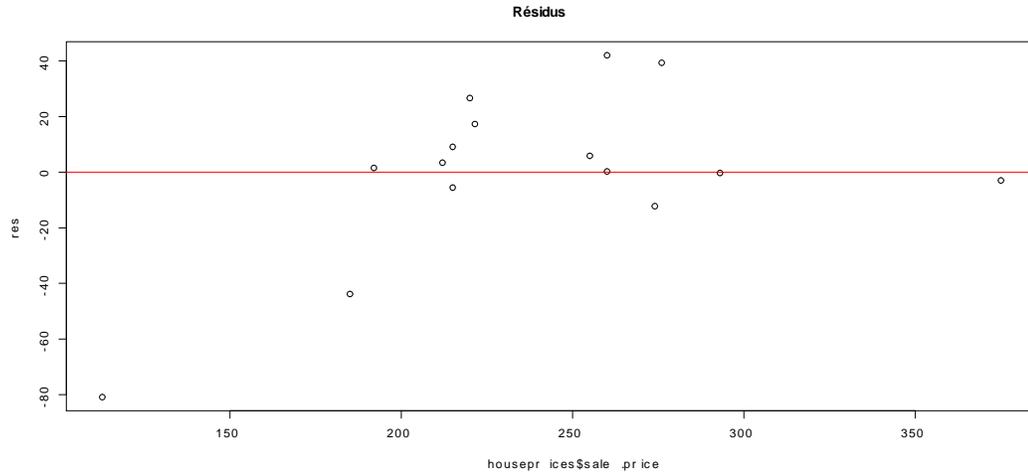


FIG. 3.10 – **Résidus vs surface**

### Test de significativité globale du modèle

$H_0$  : absence de significativité globale des variables, i.e au moins une variable n'est pas significativement différente de zéro.

Ce test ( $F$ -test) est basé sur la statistique de Fisher présentée en bas de la sortie R. Ici, comme la  $p$ -value associée est inférieure à 1%, on peut dire que l'on rejette fortement  $H_0$ , à savoir le modèle est bien globalement significatif. Si cela n'avait pas été le cas, il convient d'exclure une à une chaque variable, afin de déceler le problème.

### 3.1.6 Interprétation des coefficients

#### Significativité

Avant d'interpréter un coefficient (sens, magnitude de l'effet), il convient de s'assurer que celui ci est significatif, autrement dit, qu'il est significativement différent de zéro ( $H_0$ , soit une absence d'effet). Pour cela on utilise un test de Student. On calcule la statistique  $t$  pour chaque variable, assumée suivre une loi de Student ; que l'on compare ensuite avec la valeur théorique issue d'une table de Student (déterminée

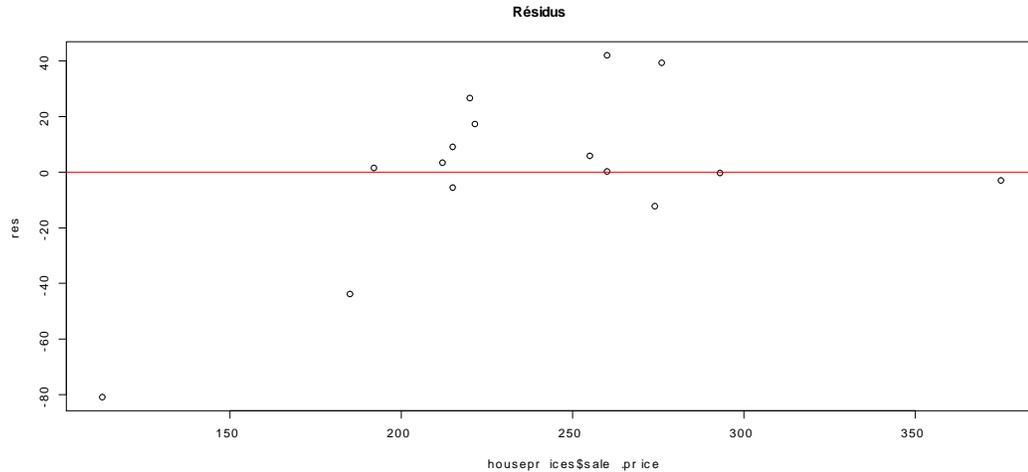


FIG. 3.11 – **Résidus vs prix de vente**

par le niveau du test, et la nombre d’observations).

On utilise souvent un niveau de 5% (soit un intervalle de confiance de 95%). Si la  $t$ -value calculée est supérieure (en valeur absolue) à la valeur théorique déterminée, alors on rejette  $H_0$  : Le coefficient est bien significativement différent de zéro, et on peut l’interpréter (signe, magnitude,..). Un autre moyen de réaliser le test est de regarder la  $p$ -value associée au coefficient, soit la probabilité pour que la valeur  $t$ -calculée soit supérieure en valeur absolue à la valeur théorique. Si cette probabilité est inférieure au seuil utilisé (ici 5%), alors le coefficient est significatif.

Ci-dessous, on constate que les deux coefficients sont bien significatifs, on va donc pouvoir les interpréter.

### Signe du coefficient

Avec un modèle linéaire, le signe du coefficient associé à une variable indique le sens de l’effet de cette variable sur la variable à expliquer. Par exemple, ici les coefficients de surface et chambres sont positifs, cela signifie que toutes choses égales par ailleurs (i.e nombre de chambres constant), augmenter la surface de la maison va avoir ten-

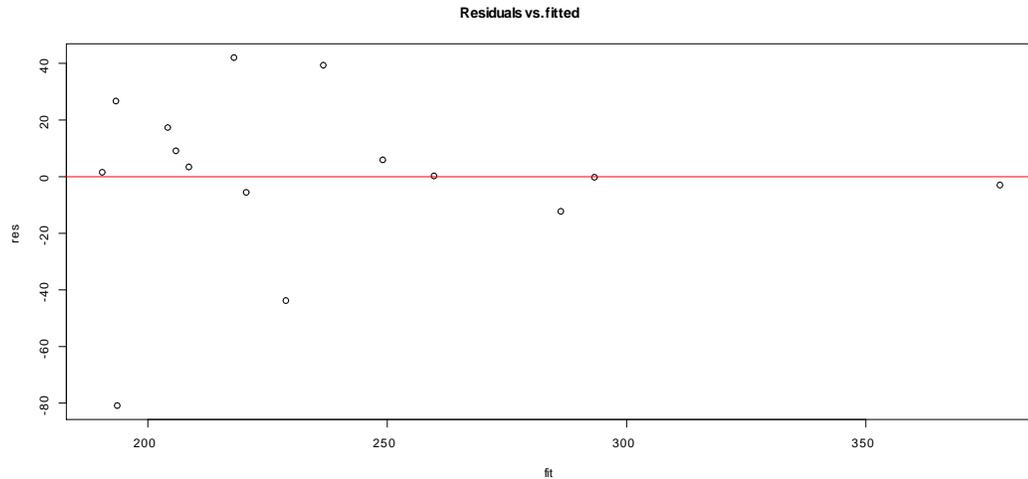


FIG. 3.12 – Residus vs. valeurs prédites

```
Call:
lm(formula = sale.price ~ area + bedrooms, data = houseprices)

Residuals:
    Min       1Q   Median       3Q      Max
-80.897  -4.247   1.539  13.249  42.027

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -141.76132   67.87204  -2.089  0.05872 .
area          0.14255    0.04697   3.035  0.01038 *
bedrooms     58.32375   14.75962   3.952  0.00192 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.06 on 12 degrees of freedom
Multiple R-squared:  0.731,    Adjusted R-squared:  0.6861
F-statistic: 16.3 on 2 and 12 DF,  p-value: 0.0003792
```

FIG. 3.13 – Test de significativité globale du modèle

dance à faire augmenter son prix. De même, augmenter le nombre de chambres de la maison, à surface constante va avoir tendance à augmenter son prix.

### 3.1.7 Qualité du modèle

Enfin, on peut regarder la qualité de la régression (au regard des données), mesurée par le coefficient de détermination  $R^2$ , qui se définit comme la part de variation dans la variable  $y$  qui est expliquée par des variations dans les variables explicatives (souvent exprimé en %).

```

Call:
lm(formula = sale.price ~ area + bedrooms, data = houseprices)

Residuals:
    Min       1Q   Median       3Q      Max
-80.897  -4.247   1.539  13.249  42.027

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -141.76132   67.87204   -2.089  0.05872 .
area          0.14255    0.04697    3.035  0.01038 *
bedrooms     58.32375    14.75962    3.952  0.00192 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.06 on 12 degrees of freedom
Multiple R-squared:  0.731,    Adjusted R-squared:  0.6861
F-statistic: 16.3 on 2 and 12 DF,  p-value: 0.0003792

```

FIG. 3.14 – Test de significativité

```

Call:
lm(formula = sale.price ~ area + bedrooms, data = houseprices)

Residuals:
    Min       1Q   Median       3Q      Max
-80.897  -4.247   1.539  13.249  42.027

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -141.76132   67.87204   -2.089  0.05872 .
area          0.14255    0.04697    3.035  0.01038 *
bedrooms     58.32375    14.75962    3.952  0.00192 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.06 on 12 degrees of freedom
Multiple R-squared:  0.731,    Adjusted R-squared:  0.6861
F-statistic: 16.3 on 2 and 12 DF,  p-value: 0.0003792

```

FIG. 3.15 – Signe du coefficient

Plus sa valeur est proche de 1, et plus l'adéquation entre le modèle et les données observées va être forte. Cependant, cette valeur est fortement influencée, entre autres, par le nombre de variables explicatives incluses dans la régression. Le  $R^2$  ajusté va alors tenir compte de ce nombre et sera donc plus correct.

Ici, il est de 69%, ce qui est plus que correct, même si il n'existe aucune règle, ni aucune échelle précise qui indiquerait pour quelles valeurs du  $R^2$ , la qualité doit être considérée "mauvaise" ou au contraire comme "excellente", une valeur trop élevée ( $R^2$  ou  $R^2$  ajusté supérieurs à 85%) peut cacher un grave problème (notamment d'endogénéité), et donc des résultats totalement erronés. De plus, il n'est bien souvent pas possible d'atteindre des valeurs jugées "satisfaisantes", en raison des données à disposition pour l'analyse, et il n'est donc pas rare que l'économètre doive se contenter

```

Call:
lm(formula = sale.price ~ area + bedrooms, data = houseprices)

Residuals:
    Min       1Q   Median       3Q      Max
-80.897  -4.247   1.539  13.249  42.027

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -141.76132   67.87204  -2.089  0.05872 .
area          0.14255    0.04697   3.035  0.01038 *
bedrooms     58.32375    14.75962   3.952  0.00192 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.06 on 12 degrees of freedom
Multiple R-squared:  0.731,    Adjusted R-squared:  0.6861
F-statistic: 10.0 on 2 and 12 DF, p-value: 0.000702

```

FIG. 3.16 – Qualité du modèle

d'un  $R^2$  de "seulement" 40% par exemple (voire 30%), Ici notre 70% est donc plus que convenable, et la recherche d'un  $R^2$  élevé ne doit pas être un but en soi.

### 3.1.8 Prediction

La fonction `predict()` permet de prédire la valeur de  $Y$  pour de nouvelles données (des variables explicatives). Seul les deux premiers arguments sont requis :

**se.fit** : permet d'afficher l'écart-type de la valeur prédite,

**interval** et **level** permettent afficher ici les valeurs de l'intervalle de confiance fixé à 99%.

```

predict(pricereg2, newdata=data.frame(area=800,bedrooms=2), se.fit=TRUE,
interval = "prediction", level = 0.99)

```

```

$fit
      fit      lwr      upr
1 88.92372 -57.53354 235.381

$se.fit
[1] 34.72881

$df
[1] 12

$residual.scale
[1] 33.05849

```

FIG. 3.17 – Prédiction des valeurs

Ainsi, le prix estimé d'une maison dont la surface au sol est de 800 unités et qui a 2 chambres serait de 88.92 mille dollars.

# Conclusion

En conclusion, l'analyse de régression est une méthode statistique utile pour modéliser et traiter des données quantitatives où l'objectif principal est de chercher une liaison linéaire entre une variable  $Y$  quantitative et une ou plusieurs variables  $X$  également quantitatives, soit dans le cas simple où elle ne considère qu'une seule variable indépendante, ou soit dans le cas d'une régression linéaire multiple qui permet de prendre en compte des relations plus complexes entre plusieurs variables, cette dernière nous permet d'obtenir des informations utiles sur les facteurs influençant sur la variable dépendante qui peut aussi de conduire à de meilleures prédictions lorsque les données sont multidimensionnelles. donc, afin d'expliquer les hypothèses nécessaires et les termes du modèle, les notions d'estimation des paramètres du modèle, il existe de nombreuses méthodes pour estimer ce modèle, telle que dans ce travail en utilise la méthode de moindre carré et la méthode de maximum de vraisemblance, avec les tests de signification du paramètres du modèle et l'estimation des intervalles de confiances.

# Bibliographie

- [1] Arnaud, G., (2010), Régression linéaire, (cours, Université Rennes 2).
- [2] Bel, L., Daudin, J.J., Etienne, M., Lebarbier, E., Mary-Huard, T., Robin, S., Vuillet, C., (2016), Le Modèle Linéaire et ses Extensions.
- [3] Benatia, F., (2023), Analyse des données, (cours, Université de Biskra).
- [4] Benoit, G. et Stéphane, C., (2022), Régression Linéaire Simple, (INSA Rouen Normandie - ITI).
- [5] Bernard, D. (2014), Régression, (cours, Université Rennes I, Campus de Beaulieu, 35042 Rennes cédex).
- [6] Champion, M. (2021), Régression linéaire multiple, (cours, Université de Paris).
- [7] Christophe, C., (2017), Modèles de régression, (cours, Université de Caen).
- [8] Cornillon, P.A., Matzner, E. Régression avec R, Springer.
- [9] Dodge, Y., Rousson, V. Analyse de régression appliquée, Dunod.
- [10] Dominique L., (2006), Régression multiple : principes et exemples d'application (Université de Pau et des Pays de l'Adour).
- [11] Gilbert, S., (2006), Probabilités Analyse des données et statistique, Professeur au-conservatoire national des Arts et Métiers, Paris.
- [12] [https://odr.inrae.fr/intranet/carto/cartowiki/index.php/Regression\\_lin%C3%A9aire\\_avec\\_](https://odr.inrae.fr/intranet/carto/cartowiki/index.php/Regression_lin%C3%A9aire_avec_)
- [13] Jean, J.R., (2014), REGRESSION LINEAIRE, Préparation à l'Agrégation (cours, Université de Bordeaux 1).

- [14] Pierre,A.C.,Eric,M.L.,(2010),Régression avec R,(Université Rennes 2).

# Annexe A : Logiciel R

## Qu'est-ce-que le langage R ?

- Le langage R est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests d'hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.
- R a été créé par Ross Ihaka et Robert Gentleman en 1993 à l'Université d'Auckland, Nouvelle Zélande, et est maintenant développé par la R Development Core Team. L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (Ross Ihaka et Robert Gentleman) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparenté.
- Parmi ses caractéristiques particulièrement intéressantes, on note :
  - langage basé sur la notion de vecteur, ce qui simplifie les calculs mathématiques et réduit considérablement le recours aux structures itératives (boucles for, while, etc.)
  - .
  - pas de typage ni de déclaration obligatoire des variables .
  - programmes courts, en général quelques lignes de code seulement .

- temps de développement très court.

**lm()** : pour la construction du modèle de régression multiple.

**summary()** : produire les sorties pour chaque régression présentées et afficher les coefficients estimés, leur écart-type, et la valeur de la statistique  $T$  de Student ainsi que la p-value (probabilité que le coefficient soit significativement différent de zéro) associées à chaque coefficient. aussi présentés le  $R^2$  et  $R^2$  ajusté, ainsi que la statistique  $F$  de Fisher (testant la significativité globales des variables), son degré de liberté, et la p-value associée.

**coef()** : Extraire que les coefficients estimés.

**plot()** : Tracer le nuage de points des données.

**abline()** : Pour ajouter la ligne de régression.

**confint()** : Afficher l'intervalle de confiance à 95% pour les coefficients estimés.

**fitted()** : Extraire les valeurs prédites.

**resid()** : Extraire les résidus (Valeur prédite - Valeur réelle).

**predict()** : Prédire la valeur de  $y$  pour de nouvelles données : Seul les deux premiers arguments sont requis :

**se.fit** : Permet d'afficher l'écart-type de la valeur prédite,

**interval et level** : permettent afficher ici les valeurs de l'intervalle de confiance fixé à 99%.

# Annexe B : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

$E(.)$	: Espérance mathématique.
$\sigma^2, V(.)$	: Variance mathématique.
$Cov(X, Y)$	: Covariance mathématique du couple $(X, Y)$ .
$IC$	: Intervalles de Confiance
$RC$	: Région de Confiance
$ddl$	: Degré De Liberté
$iid$	: indépendant identiquement distribué
$MCO$	: moindres carrés ordinaire
$MV$	: Maximum de Vraisemblance
$MCE$	: Moyenne des carrés Expliqué

$MCR$	: Moyenne des carrés résiduel
$SCT$	: Sommes des carrés totale
$SCE$	: Sommes des carrés expliqué
$SCR$	: Sommes des carrés résiduel
$L(\beta_0, \beta_1, \sigma_\varepsilon^2)$	: fonction de vraisemblance
$\varepsilon^t$	: transposé de la matrice $\varepsilon$
$ANOVA$	: d'analyse de la variance
$R^2$	: Le coefficient de détermination
$\bar{R}^2$	: Le coefficient de détermination ajusté
$V$	: les éléments de la diagonale de $(X^t X)^{-1}$

## Résumé

La régression est l'une des méthodes les plus connues en statistique appliquées pour l'analyse de données quantitatives, grâce à son objectif double. Elle permet tout d'abord d'établir la relation entre une variable quantitative expliquée (ou dépendante) à une ou plusieurs autres variables quantitatives explicatives (ou indépendantes) à travers une fonction affine. Elle est dite linéaire si elle impose une forme fonctionnelle linéaire dans les paramètres du modèle, et aussi d'effectuer des prévisions de l'une des variables en fonction de l'autre, si on s'intéresse à la relation entre deux variables, on parlera de régression simple en exprimant une variable en fonction de l'autre. Si la relation porte entre une variable et plusieurs autres variables, on parlera de régression multiple.

**Mots-clés :** Régression linéaire simple, Régression linéaire multiple, Moindres carrées, Maximum de vraisemblance.

## ملخص

يعد الانحدار أحد أفضل الطرق المعروفة في مجال الإحصاء المطبقة لتحليل البيانات الكمية ، و بفضل هدفه المزدوج ، فهو يسمح أولاً بإقامة العلاقة بين متغير كمي موضح (أو تابع ) و متغير أو عدة متغيرات كمية تفسيرية (أو مستقلة) من خلال دالة تألفيه. ويسمى خطياً إذا كان يشكل دالة خطية في معطيات النموذج ، كما يقوم أيضاً بالتنبؤ لأحد المتغيرات بدلالة الآخر ، فإذا كنا مهتمين بالعلاقة بين متغيرين، سنتحدث عن الانحدار البسيط من خلال التعبير عن متغير واحد بدلالة المتغير الآخر. أما إذا كانت العلاقة بين متغير واحد و عدة متغيرات أخرى فسننتحدث عن الانحدار المتعدد

**الكلمات المفتاحية:** الإنحدار الخط البسيط ، الانحدار الخط المتعدد ، المربعات الصغرى ، الإحتمال الأقصى

## Abstract

Regression is one of the best-known methods in statistics applied for the analysis of quantitative data, thanks to its dual objective. It first makes it possible to establish the relationship between an explained (or dependent) quantitative variable to one or more other explanatory (or independent) quantitative variable through an affinity function. It is called linear if it imposed a linear functional form in the parameters of the model, and also to make predictions of one of the variables as a function of the other, if we are interested in the relationship between two variables, we will talk about simple regression by expressing one variable as a function of the other. If the relationship is between one variable and several other variables, we will speak of multiple regression.

**Keywords :** Simple linear regression, Multiple linear regression, Least squares, Maximum likelihood.