

République Algérienne Démocratique et populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane Mira - Béjaia



Faculté des Sciences Exactes
Département de Mathématiques

Soutenance de mémoire de Master en Mathématiques
Option : Statistique et Analyse Décisionnelle

Thème

Régression linéaire multiple et modèle linéaire général

Réalisé par :

- DRIS Leila
- HACHEMI Warda épouse HIDER

Devant le jury composé de :

M^{elle} SAADI N. M.A.A. Université A. Mira- Béjaia

Mr BOURAINE M. M.A.A. Université A. Mira- Béjaia

M^{elle} TAMITI K. Doctorante Université A. Mira- Béjaia

Promotion 2015-2016

Remerciements

Je remercie le Dieu tout puissant pour la santé, la persévérance, l'intelligence, la force, l'énergie débordante qu'il nous a accordé tout au long de ce travail et durant tout mon cursus scolaire.

Nous remercions notre encadreur **Mr. BOURAINE**

Pour nous avoir facilité la réalisation de notre mémoire, pour avoir accepté de nous encadrer, et pour toute l'aide précieuse qu'il nous a toujours apporté.

À **Brahimi Nassim** pour ses grands efforts et pour son aide.

Mes remerciements sont aussi adressés M^{elle} SAADI N et M^{elle} TAMITI K qui nous font l'honneur de juger notre travail.

Et à toutes les personnes ayant contribué de près ou de loin à la réalisation de ce modeste travail.

Dédicaces

Je dédie ce travail à mes chers, adorables parents M^{me} **HACHEMI Ghania** et Mr. **HACHEMI Locif**. Papa et maman, merci d'avoir toujours été là pour moi, merci pour l'amour sans faille que vous n'avez cessé de m'apporter, ainsi que vos conseils qui m'ont toujours permis d'être sur le droit chemin. Voyez en ce mémoire un des fruits de tous les efforts et sacrifices que vous ne cessez de faire pour moi.

A mon mari Mr **HIDER Hamza** pour son soutien et pour l'amour qui me toujours donner.

A mon cher frère et meilleur ami, **Ahmed** et à ma petite sœur, ma petite princesse **Sara** que j'adore trop fort

A toute la famille **HACHEMI** et la famille **HIDER** petits et grands.

A toutes la promotion *master 2 SAD*.

Hachemi warda 😊

Dédicaces

Je dédie ce travail à mes chères, adorables parents Mme **Dris fatiha** et Mr.**Dris hocine**. Papa et maman, merci d'avoir toujours été là pour moi, merci pour l'amour sans faille que vous n'avez cessé de m'apporter, ainsi que vos conseils qui m'ont toujours permis d'être sur le droit chemin. Voyez en ce mémoire un des fruits de tous les efforts et sacrifices que vous ne cessez de faire pour moi.

A toute la famille **DRIS** petits et grands.

A tout mes chères amies

A toutes la promotion *SAD master 2* **SAD**.

Table des matières

Table de matière	5
Liste des Tableaux	5
Introduction générale	5
1 Rappel sur la régression linéaire simple	7
1.1 La régression lineaire simple	7
1.1.1 Modèle	7
1.1.2 Distribution des estimateurs des moindres carrés	9
1.1.3 Test et intervalle de confiance	16
1.1.4 Intervalle de prédiction	22
1.1.5 Test de signification global de Fisher	23
1.1.6 Coefficient de corrélation	25
1.1.7 Inerprétation du coefficient de corrélation	26
1.1.8 Cas particulier	26
2 La régression linéaire multiple	32
2.1 Régression linéaire multiple	32
2.1.1 Modèle	32
2.1.2 Modèle à deux variables explicatives	33
2.1.3 Propriétés	35
2.1.4 Sommes des carrés	35
2.1.5 Coefficient de détermination	36
2.1.6 Cas général	36
2.1.7 Propriétés des moindres carrés	39
2.1.8 Estimation de la variance des erreurs	41
2.2 Inférence sur les paramètres du modèle	43
2.2.1 Tests d'hypothèses	43
2.2.2 Intervalle de confiance	44
2.2.3 Analyse de la variance	44
2.2.4 Régression sans constante	46
2.2.5 Test F partiel	48

3	Le modèle linéaire général	51
3.1	Le modèle	51
3.1.1	Définition de modèle linéaire général	51
3.2	Régression avec les résidus normaux	52
3.2.1	Hypothèses du modèle linéaire général	52
3.2.2	Données observées, et formulation matricielle	52
3.2.3	Autre présentation du modèle linéaire général	53
3.3	Estimation du modèle	53
3.3.1	Estimateurs du maximum de vraisemblance	56
3.3.2	Propriétés des estimateurs du maximum de vraisemblance	58
3.3.3	Distribution de probabilité des estimateurs	58
3.3.4	Synthèse des résultats	61
3.3.5	Intervalle de confiance sur un coefficient de régression	61
3.4	Test d'un seul coefficient de régression	62
3.4.1	Construction du test	62
3.4.2	Modèle linéaire avec uniquement une constante	63
3.5	Tests de Wald sur les coefficients de régression	64
3.5.1	Test général d'une contrainte linéaire	64
3.5.2	Test global des coefficients de régression	65
3.5.3	Test de Fisher sur un coefficient de régression	68
3.5.4	R-carré et R-carré ajusté	69
3.5.5	Prévision ponctuelle d'une valeur	69
4	Application	71
4.1	Statistique descriptive	71
4.2	Estimation des paramètres de modèle par la méthode des moindres carrés	72
4.3	Tests sur les paramètres du modèle	74
4.3.1	Matrice de corrélation	75
4.4	Test d'ajustement	76
4.5	Analyse de la variance	77
4.6	Prédiction	78
	Bibliographie	83

Table des matières

Table des figures

Liste des tableaux

1.1	Forme général du Tableau ANOVA	25
2.1	Tableau de ANOVA pour la régression multiple	46
3.1	Tableau de ANOVA de régression général	68
4.1	Résumé des statistique.	71
4.2	Tableau représente la consommation d'essence pour les différentes véhicules.	73
4.3	Tableau des tests de chaque paramètre estimé	74
4.4	Tableau représentant les tests d'ajustement.	76
4.5	Tableau de ANOVA.	77
4.6	Tableau représente la prédiction et les résidus.	78
4.7	Tableau représente la prédiction (la borne inf et sup).	79

Introduction générale

L'analyse de régression peut être définie comme la recherche de la relation stochastique que lie deux ou plusieurs variables. Son champ d'application recouvre de multiples domaines, parmi lesquels 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 [1]. La régression est un des méthodes les plus connues et les plus appliquées en statistique pour l'analyse de données quantitatives. Elle est utilisée pour établir une liaison entre une variable quantitative et une ou plusieurs autres variables quantitatives,

sous la forme d'un modèle. 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. Le but est d'expliquer une variable Y , appelée variable endogène par une ou plusieurs variables explicatives dites exogènes à travers une fonction affine.

Ces modèles de régression font partie d'une classe plus large de modèle appelé modèle linéaire général.

En guise d'application, on va essayer d'expliquer la consommation d'un certain nombre de véhicules par le prix, la cylindrée, la puissance, et le poids.

Ce mémoire est organisé comme suit :

Le premier chapitre introduit la régression simple, la régression linéaire multiple fait l'objet du deuxième chapitre. Le troisième chapitre est consacré au modèle général, le mémoire se termine par un chapitre d'application.

Chapitre 1

Rappel sur la régression linéaire simple

1.1 La régression lineaire simple

Introduction

Dans ce chapitre, nous avons introduit le modèle de régression linéaire. Comme il s'agit d'un modèle avec une seule variable explicative, on parle de régression simple. Nous avons ainsi défini la droite de régression qu'il s'agit d'estimer à partir des données d'un échantillon par la méthode des moindres carrés [1].

1.1.1 Modèle

Considérons un échantillon de n observations paireses $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Le modèle de la régression linéaire, suppose, pour tout $i = 1, \dots, n$ la relation :

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (1.1)$$

où

- Y est la variable dépendante (variable endogène).
- β_0 et β_1 sont les coefficients (représenté par ordonnée à l'origine et pente du modèle).
- X est la variable indépendante (variable exogène).
- ϵ est une erreur aléatoire.

où les ϵ_i sont des quantités aléatoires inobservables qui représentent les erreurs. Ainsi les y_i sont des variables aléatoires (car elles dépendent des ϵ_i), alors que les x_i sont considérés

comme des nombres fixés. En supposant l'espérance des ϵ_i nulle, on a ainsi :

$$\begin{aligned} E(y_i) &= \beta_0 + \beta_1 x_i + E(\epsilon_i) \\ &= \beta_0 + \beta_1 x_i \end{aligned}$$

et

$$Var(y_i) = Var(\beta_0 + \beta_1 x_i + \epsilon_i) = Var(\epsilon_i).$$

Afin de pouvoir faire de l'inférence sur les paramètres β_0 et β_1 du modèle :

$$\mu_y(x) = \beta_0 + \beta_1 x$$

à partir de la droite des moindres carrés :

$$\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

Pour faire l'inférence sur les paramètres du modèle, on doit se fixer des hypothèses du base : $H_1 : E(\epsilon_i) = 0 \forall i$ (les erreurs sont en moyenne nulle).

$H_2 : Var(\epsilon_i) = \sigma_\epsilon^2 \forall i$, (l'amplitude de l'erreur est la même quelque soit la période.)

$H_3 : Cov(\epsilon_i, \epsilon_j) = 0 \forall i \neq j$, il suffit des erreurs indépendante.

H_4 : Les variables x_i sont non aléatoire.

H_5 :

$$\epsilon_i \text{ i.i.d. } N(0, \sigma^2), \forall i. \quad (1.2)$$

est utilisée lorsque'on veut déterminer la distance des estimateurs $\hat{\beta}_0$ $\hat{\beta}_1$ et les intervalles de confiance, les tests d'hypothèses.

Lorsque on utilise un modèle de régression simple on suppose donc que l'on a tiré un échantillon de ϵ_i distribué selon (1.2) cependant, on observe à la place les variables aléatoires y_i définies par (1.1) à partir de ces ϵ_i , pour des valeurs x_i fixées par avance et de paramètres β_0 et β_1 inconnus. Pour chaque valeur de x_i (fixe), y_i a une distribution normale d'espérance $\beta_0 + \beta_1 x_i$ et de variance σ^2 . Les espérances des différents y_i se trouvent ainsi alignées sur la droite de régression qu'il s'agit d'estimer par la méthode des moindres carrés.

1.1.2 Distribution des estimateurs des moindres carrés

Rappelons tout d'abord les formules pour les estimateurs obtenus par la méthode des moindres carrés. On a :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

où

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \text{ et } \bar{y} = \frac{\sum_{i=1}^n y_i}{n}.$$

Les estimateurs sont de type aléatoires car ils dépendent des y_i qui sont des variables aléatoires (contrairement aux x_i). De plus, ce sont des fonctions linéaires des y_i . Comme ces derniers sont par hypothèse normalement distribués, il s'ensuit que les variables aléatoires $\hat{\beta}_0$ sont également normalement distribuées. Afin de connaître leurs distributions, il reste donc à calculer leur espérance et leur variance.

Propriétés statistique des estimateurs

1. $E(\hat{\beta}_1) = \beta_1$ d'où $\hat{\beta}_1$ est sans biais.
2. $E(\hat{\beta}_0) = \beta_0$ d'où $\hat{\beta}_0$ est sans biais.
3. $Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$.
4. $Var(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$.
5. $Cov(\hat{\beta}_0, \hat{\beta}_1) = 0$

En effet

Démonstration

Propriétés statistique des estimateurs amener les résultats ensuite passer ou détermination chaque résultat.

Espérance mathématique de $\hat{\beta}_1$

On a :

$$\begin{aligned}
 E(\hat{\beta}_1) &= E\left(\frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
 &= E\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
 &= E\left(\frac{\beta_0 \sum_{i=1}^n (x_i - \bar{x}) + \beta_1 \sum_{i=1}^n (x_i - \bar{x})x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
 &= E\left(\frac{0 + \beta_1 \sum_{i=1}^n (x_i - \bar{x})x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)
 \end{aligned}$$

En effet :

$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$

De plus comme on a :

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

on obtient finalement :

$$E(\hat{\beta}_1) = \beta_1.$$

Ainsi, $\hat{\beta}_1$ est un estimateur sans biais de β_1 .

Espérance mathématique de $\hat{\beta}_0$

On a :

$$\begin{aligned}
 E(\hat{\beta}_0) &= E(\bar{y} - \hat{\beta}_1 \bar{x}) \\
 &= E(\bar{y}) - \bar{x}E(\hat{\beta}_1) \\
 &= E(\bar{y}) - \bar{x}\beta_1.
 \end{aligned}$$

Sachant que :

$$\begin{aligned}
 E(\bar{y}) &= E\left(\frac{\sum_{i=1}^n y_i}{n}\right) = \frac{\sum_{i=1}^n E(y_i)}{n} \\
 &= \frac{\sum_{i=1}^n E(\beta_0 + \beta_1 x_i)}{n} \\
 &= \frac{n\beta_0 + \beta_1 \sum_{i=1}^n x_i}{n} \\
 &= \beta_0 + \beta_1 \bar{x}.
 \end{aligned}$$

On a finalement :

$$E(\hat{\beta}_0) = \beta_0 + \beta_1 \bar{x} - \bar{x}\beta_1 = \beta_0.$$

Variance de $\hat{\beta}_1$

On a :

$$\begin{aligned}
 Var(\hat{\beta}_1) &= Var\left(\frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 Var(y_i)}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \\
 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \\
 &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.
 \end{aligned}$$

Variance de $\hat{\beta}_0$

On a :

$$Var(\hat{\beta}_0) = Var(\bar{y} - \hat{\beta}_1 \bar{x}) = Var(\bar{y}) + \bar{x}^2 Var(\hat{\beta}_1) - 2\bar{x} Cov(\bar{y}, \hat{\beta}_1).$$

Comme on a :

$$\begin{aligned}
 \text{Cov}(\bar{y}, \hat{\beta}_1) &= \text{Cov}\left(\frac{\sum_{i=1}^n y_i}{n}, \frac{\sum_{j=1}^n (x_j - \bar{x})y_j}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
 &= \frac{\sum_{i=1}^n (x_i - \bar{x})\text{Var}(y_i) + \sum_{i=1}^n \sum_{j=1, i \neq j}^n (x_j - \bar{x})\text{Cov}(y_i, y_j)}{n \sum_{i=1}^n (x_i - \bar{x})^2} \\
 &= \frac{\sigma^2 \sum_{i=1}^n (x_i - \bar{x}) + 0}{n \sum_{i=1}^n (x_i - \bar{x})^2} \\
 &= 0,
 \end{aligned}$$

et sachant par ailleurs que :

$$\text{Var}(\bar{y}) = \text{Var}\left(\frac{\sum_{i=1}^n y_i}{n}\right) = \frac{n^2 \sigma^2}{n} = \frac{n \sigma^2}{n},$$

En remplaçant dans $\text{Var}(\hat{\beta}_0)$ on obtient :

$$\begin{aligned}
 \text{Var}(\hat{\beta}_0) &= \text{Var}(\bar{y}) + \bar{x}^2 \text{Var}(\hat{\beta}_1) \\
 &= \frac{\sigma^2}{n} + \frac{\bar{x}^2 \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
 &= \frac{\sigma^2 (\sum_{i=1}^n (x_i - \bar{x})^2 + n \bar{x}^2)}{n \sum_{i=1}^n (x_i - \bar{x})^2}.
 \end{aligned}$$

En rappelant que :

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

on obtient finalement :

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Covariance entre $\hat{\beta}_1$ et $\hat{\beta}_0$

On a :

$$\begin{aligned} Cov(\hat{\beta}_1, \hat{\beta}_0) &= Cov(\hat{\beta}_1, \bar{y} - \hat{\beta}_1 \bar{x}) \\ &= Cov(\hat{\beta}_1, \bar{y}) - \bar{x} Var(\hat{\beta}_1) \\ &= 0 - \frac{\bar{x} \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \end{aligned}$$

et donc :

$$Cov(\hat{\beta}_1, \hat{\beta}_0) = \frac{-\bar{x} \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Estimation de la variance des erreurs

Dans les trois résultats de la section précédente, on voit apparaître la variance σ^2 dans les caractéristiques des estimateurs des ϵ_i . Cette quantité est toutefois inconnue et doit être estimée. Si les erreurs ϵ_i pouvaient être observées, un estimateur non biaisé de σ^2 serait donné par la formule habituelle :

$$\frac{\sum_{i=1}^n (\epsilon_i - \bar{\epsilon})^2}{n - 1},$$

où $\bar{\epsilon}$ est la moyenne des ϵ_i . Or ces quantités ne sont pas observables. Les erreurs ϵ_i peuvent être estimées par les résidus du modèle :

$$e_i = y_i - \hat{y}_i,$$

où les :

$$\begin{aligned} \hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i \\ &= \bar{y} + \hat{\beta}_1 (x_i - \bar{x}), \end{aligned}$$

sont les valeurs estimées des y_i . Ainsi, nous allons estimer σ^2 en utilisant la somme des carrés des résidus comme estimateur de la somme des carrés des erreurs :

$$\sum_{i=1}^n (e_i - \bar{e})^2 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

où \bar{e} désigne la moyenne des e_i (on a vu que la somme, donc la moyenne, des résidus est nulle). Comme :

$$\begin{aligned} E\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) &= \sum_{i=1}^n E(y_i^2) - \sum_{i=1}^n E(\hat{y}_i^2) \\ &= \sum_{i=1}^n (Var(y_i) + E^2(y_i)) - \sum_{i=1}^n (Var(\hat{y}_i) + E^2(\hat{y}_i)), \end{aligned}$$

on a :

$$\begin{aligned} E(\hat{y}_i) &= E(\hat{\beta}_0 + \hat{\beta}_1 x_i) = E(\hat{\beta}_0) + x_i E(\hat{\beta}_1) \\ &= \beta_0 + x_i \beta_1 \\ &= E(y_i). \end{aligned}$$

On obtient ainsi :

$$\begin{aligned} E\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) &= \sum_{i=1}^n Var(y_i) - \sum_{i=1}^n Var(\hat{y}_i) \\ &= n\sigma^2 - \sum_{i=1}^n Var(\hat{y}_i). \end{aligned}$$

D'autre part on a :

$$\begin{aligned}
\text{Var}(\hat{y}_i) &= \text{Var}(\hat{\beta}_0 + \hat{\beta}_1 x_i) \\
&= \text{Var}(\hat{\beta}_0) + x_i^2 \text{Var}(\hat{\beta}_1) + 2x_i \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\
&= \frac{\sigma^2 \sum_{j=1}^n x_j^2}{n \sum_{j=1}^n (x_j - \bar{x})^2} + \frac{x_i^2 \sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2} - \frac{2x_i \bar{x} \sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \\
&= \frac{\sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \left(\frac{\sum_{j=1}^n (x_j)^2}{n} + x_i^2 - 2x_i \bar{x} \right) \\
&= \frac{\sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \left(\frac{\sum_{j=1}^n x_j^2}{n} - \frac{n\bar{x}^2}{n} + x_i^2 - 2x_i \bar{x} + \bar{x}^2 \right) \\
&= \frac{\sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \left(\frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n} + (x_i - \bar{x})^2 \right) \\
&= \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \right).
\end{aligned}$$

On obtient ainsi :

$$E \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) = n\sigma^2 - \sigma^2 \left(\frac{n}{n} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_j - \bar{x})^2} \right) = \sigma^2(n - 2).$$

On peut ainsi définir un estimateur sans biais de σ^2 en posant :

$$\begin{aligned}
 S^2 &= \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \\
 &= \frac{SC_{res}}{n-2} \\
 &= \frac{SC_{tot} - SC_{reg}}{n-2} \\
 &= \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{n-2}.
 \end{aligned}$$

On estime par ailleurs l'écart type des erreurs σ par :

$$s = \sqrt{\frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{n-2}}.$$

1.1.3 Test et intervalle de confiance

Nous allons voir dans cette section comment effectuer des tests d'hypothèses sur les paramètres β_1 et β_0 du modèle de régression simple [11], ainsi que sur la manière de construire des intervalles de confiance.

Test sur la pente β_1

L'estimateur $\hat{\beta}_1$ était normalement distribué, d'espérance β_1 et de variance que l'on note ici par :

$$S^2(\hat{\beta}_1) = \hat{\sigma}^2(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (1.3)$$

Il s'ensuit que la quantité :

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma(\hat{\beta}_1)},$$

suit une loi normale centrée réduite. Or cette quantité ne peut pas être utilisée pour un problème de test d'hypothèses puisqu'on ne connaît pas la valeur de σ . En pratique, on estime $\sigma^2(\hat{\beta}_1)$ par :

$$\sigma^2(\hat{\beta}_1) = \frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

où S^2 est l'estimateur sans biais de σ^2 défini ci-dessus. On peut alors montrer que la quantité :

$$\frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)},$$

suit une loi de student avec $(n-2)$ degrés de liberté, où $s(\hat{\beta}_1)$ désigne la racine carrée de $s^2(\hat{\beta}_1)$. Il s'ensuit que dans un problème de test d'hypothèse bilatéral où l'on désire tester l'hypothèse nulle :

$$H_0 : \beta_1 = b_1$$

contre l'hypothèse alternative :

$$H_1 : \beta_1 \neq b_1$$

On utilise la statistique :

$$t_c = \frac{\hat{\beta}_1 - b_1}{s(\hat{\beta}_1)}.$$

On rejette H_0 au seuil de signification α si :

$$|t_c| > t_{\alpha/2, n-2},$$

où la valeur $t_{\alpha/2, n-2}$ est le $(1 - \alpha/2)$ quantile la loi de Student avec $(n - 2)$ degrés de liberté.

Un test d'hypothèse particulièrement intéressant est le test de l'hypothèse nulle :

$$H_0 : \beta_0 = 0$$

contre l'hypothèse alternative :

$$H_1 : \beta_0 \neq 0.$$

En effet, le non-rejet de l'hypothèse nulle implique un modèle avec un seul paramètre :

$$y_i = \beta_0 + \epsilon_i,$$

Par contre, si cette hypothèse H_0 est rejetée, c'est-à-dire si :

$$|t_c| = \frac{|\hat{\beta}_1|}{s(\hat{\beta}_1)} > t_{\alpha/2, n-2},$$

on dit que la relation entre les x_i et y_i est significative au seuil de signification α .

Intervalle de confiance pour β_1

L'intervalle de confiance au niveau $(1 - \alpha)$ pour le paramètre β_1 est définie par :

$$\left[\hat{\beta}_1 - t_{\alpha/2, n-2} \cdot s(\hat{\beta}_1), \hat{\beta}_1 + t_{\alpha/2, n-2} \cdot s(\hat{\beta}_1) \right];$$

où

$$\hat{\beta}_1 \pm t_{\alpha/2, n-2} \cdot s(\hat{\beta}_1).$$

Cet intervalle est construit de telle sorte qu'il contienne le paramètre inconnu β_1 avec une probabilité de $(1 - \alpha)$.

Rappelons que les hypothèses H_0 et H_1 peuvent également être testées à l'aide de l'intervalle de confiance ; on rejette alors l'hypothèse :

$$H_0 : \beta_1 = b_1,$$

pour l'hypothèse :

$$H_1 : \beta_1 \neq b_1,$$

au seuil de signification α si et seulement si la valeur testée b_1 ne se trouve pas dans l'intervalle de confiance de β_1 au niveau $(1 - \alpha)$.

Test sur l'ordonnée à l'origine β_0

On a vu que l'estimateur $\hat{\beta}_0$ est normalement distribué, d'espérance β_0 et de variance.

$$\sigma^2(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2},$$

Il s'ent suit que la quantité :

$$\frac{\hat{\beta}_0 - \beta_0}{\sigma(\hat{\beta}_0)},$$

suit une loi normale centrée réduite, où $\sigma(\hat{\beta}_0)$ désigne la racine carrée de $\sigma^2(\hat{\beta}_0)$. Or cette quantité ne peut pas être utilisée pour un problème de test d'hypothèses puisqu'on ne connaît pas la valeur de σ . En pratique, on estime $\sigma^2(\hat{\beta}_0)$ par :

$$\hat{\sigma}^2(\hat{\beta}_0) = \sigma^2(\hat{\beta}_0) = \frac{s^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2},$$

On peut alors montrer que la quantité :

$$\frac{\hat{\beta}_0 - \beta_0}{\sigma(\hat{\beta}_0)},$$

suit une loi de Student avec $(n - 2)$ degrés de liberté, où $s(\hat{\beta}_0)$ désigne la racine carrée de $s^2(\hat{\beta}_0)$.

Il s'ensuit que dans un problème de test d'hypothèses bilatéral où l'on désire tester l'hypothèse nulle :

$$H_0 : \beta_1 = b_1,$$

contre l'hypothèse alternative :

$$H_1 : \beta_1 \neq b_1,$$

on peut utiliser la statistique :

$$t_c = \frac{\hat{\beta}_1 - b_1}{s(\hat{\beta}_1)}.$$

On rejette H_0 au seuil de signification α si :

$$|t_c| > t_{\alpha/2, n-2}.$$

Un test d'hypothèses souvent effectué est le test de l'hypothèse nulle :

$$H_0 : \beta_1 = 0$$

contre l'hypothèse alternative :

$$H_1 : \beta_1 \neq 0$$

Le non-rejet de cette hypothèse nulle implique un modèle avec un seul paramètre :

$$y_i = \beta_1 x_i + \epsilon_i$$

Par contre, si cette hypothèse H_0 est rejetée, c'est-à-dire si :

$$|t_c| = \frac{|\hat{\beta}_0|}{s(\hat{\beta}_0)} > t_{\alpha/2, n-2},$$

La constante β_0 du modèle de régression (1.1) est dite significative au seuil de signification α .

Intervalle de confiance pour β_0

Un intervalle de confiance au niveau $(1 - \alpha)$ pour le paramètre β_0 est défini par :

$$\left[\hat{\beta}_1 - t_{\alpha/2, n-2} \cdot s(\hat{\beta}_1), \hat{\beta}_1 + t_{\alpha/2, n-2} \cdot s(\hat{\beta}_1) \right]$$

c'est-à-dire par :

$$\hat{\beta}_1 \pm t_{\alpha/2, n-2} \cdot s(\hat{\beta}_1)$$

Cet intervalle est construit de telle sorte qu'il contienne le paramètre inconnu β_0 avec une probabilité de $(1 - \alpha)$. On peut aussi tester les hypothèses H_0 et H_1 à l'aide de l'intervalle de confiance ; on rejete alors l'hypothèse :

$$H_0 : \beta_1 = b_0$$

pour l'hypothèse :

$$H_1 : \beta_1 \neq b_0$$

au seuil de signification α si et seulement si la valeur testée b_0 ne se trouve pas dans l'intervalle de confiance pour β_0 au niveau $(1 - \alpha)$.

Intervalle de confiance pour $\mu_y(x)$

Nous allons voir ici comment trouver un intervalle de confiance pour :

$$\mu_y(x) = \beta_0 + \beta_1 x,$$

c'est-à-dire pour l'ordonnée du point d'abscisse x se trouvant sur la droite de régression. On a vu que l'estimateur de $\mu_y(x)$ est donné par la droite des moindres carrés :

$$\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x,$$

Cet estimateur $\hat{y}(x)$ est lui aussi normalement distribué, étant une combinaison linéaire de deux estimateurs dont la distribution conjointe est normale bivariée. En outre on peut montrer que $\hat{y}(x)$ est un estimateur non-biaisé de $\mu_y(x)$ et de variance :

$$\sigma^2(\hat{y}(x)) = \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) = \sigma^2(y_i) \text{ pour } x = x_i,$$

On remarque que cette variance est d'autant plus grande que x est éloigné de \bar{x} , ce qui signifie que les estimateurs de $\mu_y(x)$ seront moins précis pour des valeurs de x éloignées de \bar{x} . Ainsi la quantité :

$$\frac{\hat{y}(x) - \mu_y(x)}{\sigma(\hat{y}(x))}$$

suit une loi normale standardisée, où $\sigma(\hat{y}(x))$ désigne la racine carrée de $\sigma^2(\hat{y}(x))$. Or, comme σ est inconnu, on estime $\sigma^2(\hat{y}(x))$ par :

$$S^2(\hat{y}(x)) = S^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right).$$

On peut alors montrer que la quantité :

$$\frac{\hat{y}(x) - \mu_y(x)}{s(\hat{y}(x))}.$$

où $s(\hat{y}(x))$ désigne la racine carrée de $S^2(\hat{y}(x))$, suit une loi de Student avec $(n - 2)$ degrés de liberté.

On peut aussi effectuer des tests d'hypothèses pour ces moyennes $\mu_y(x)$ de la même manière qu'on l'a fait pour β_0 ou β_1 . Généralement, on se contente ici de calculer un intervalle de confiance. Un tel intervalle au niveau $(1 - \alpha)$ pour $\mu_y(x)$ est défini par :

$$\left[\hat{y}(x) - t_{\alpha/2, n-2} \cdot s(\hat{y}(x)), \hat{y}(x) + t_{\alpha/2, n-2} \cdot s(\hat{y}(x)) \right],$$

c'est-à-dire par :

$$\hat{y}(x) \pm t_{\alpha/2, n-2} \cdot s(\hat{y}(x)).$$

Cet intervalle est construit de telle sorte qu'il contienne la moyenne $\mu_y(x)$ avec une probabilité de $(1 - \alpha)$.

1.1.4 Intervalle de prédiction

Lorsque les hypothèses sont vérifiées, on peut également calculer des *intervalles de prédiction*. Un intervalle de prédiction au niveau $(1 - \alpha)$ est construit de telle sorte qu'il contienne avec une probabilité de $(1 - \alpha)$ la valeur (future) de la variable Y d'un individu pour lequel on aura $X = x$. On note ici une telle quantité par $y_f(x)$. Si les hypothèses du modèle sont correctes, on a donc.

$$y_f(x) = \beta_0 + \beta_1 x + \epsilon_f = \mu_y(x) + \epsilon_f,$$

où ϵ_f est une erreur (future) distribuée selon une loi normale d'espérance nulle de variance σ^2 . Un intervalle de prédiction pour $y_f(x)$ au niveau $(1 - \alpha)$ est alors donné par

$$[\mu_y(x) - z_{\alpha/2}\sigma; \mu_y(x) + z_{\alpha/2}\sigma].$$

Cette formule est pourtant inutilisable en pratique, puisque β_0, β_1 et σ^2 sont inconnus. Notons alors que la quantité

$$\hat{y}_f(x) = \beta_0 + \beta_1 x - \epsilon_f = \hat{y}_f(x) - \epsilon_f$$

est distribuée selon une loi normale d'espérance $\mu_y(x)$ et de variance

$$\sigma^2(\hat{y}_f(x)) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right).$$

qui peut être approximée par

$$s^2(\hat{y}_f(x)) = s^2 \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right).$$

On peut montrer alors que la quantité :

$$\frac{\hat{y}_f(x) - \mu_y(x)}{s(\hat{y}_f(x))} = \frac{\hat{y}_f(x) - y_f(x)}{s(\hat{y}_f(x))}$$

où $s(\hat{y}_f(x))$ désigne la racine carrée de $s^2(\hat{y}_f(x))$, suit une loi de Student avec $(n - 2)$ degrés de liberté. Il s'en suit qu'un intervalle de prédiction au niveau $(1 - \alpha)$ pour $y_f(x)$ est donné par

$$[\hat{y}_f(x) - t_{\alpha/2, n-2} \cdot s(\hat{y}_f(x)), \hat{y}_f(x) + t_{\alpha/2, n-2} \cdot s(\hat{y}_f(x))].$$

c'est-à-dire

$$\hat{y}(x) \pm t_{\alpha/2, n-2} \cdot s(\hat{y}_f(x)).$$

Notons que contrairement à un intervalle de confiance, la longueur d'un intervalle de prédiction ne tend pas vers zéro lorsque la taille de l'échantillon augmente. Dans ce cas, l'intervalle de prédiction s'approche alors de l'intervalle

$$\mu_y(x) \pm z_{\alpha/2} \sigma.$$

1.1.5 Test de signification global de Fisher

Nous avons définis trois sommes des carrées $SC_{tot}, SC_{reg}, SC_{res}$. Dans cette section, nous allons voir comment on peut utiliser ces sommes de carrés pour tester l'hypothèse :

$$H_0 : \beta_1 = 0,$$

Si cette hypothèse est vérifiée, on peut montrer que les espérances de ces sommes de carrés valent respectivement :

$$E(SC_{tot}) = (n - 1)\sigma^2,$$

$$E(SC_{reg}) = \sigma^2,$$

$$E(SC_{res}) = (n - 2)\sigma^2.$$

où

SC_{tot} est la somme des carrés totale.

SC_{reg} est la somme des carrés régression.

SC_{res} est la somme des carrés régression.

Les quantités suivantes sont alors des estimateurs sans biais de σ^2 :

$$MC_{tot} = \frac{SC_{tot}}{n - 1},$$

$$MC_{reg} = \frac{SC_{reg}}{1},$$

$$MC_{res} = \frac{SC_{res}}{n - 2}.$$

où MC symbolise *moyenne des carrés*. Les nombres $(n - 1)$, 1 et $(n - 2)$ formant le dénominateur de ces estimateurs représentent les degrés de liberté associés à ces sommes de

carrés. c'est-à-dire le nombre de termes linéairement indépendants impliqués dans chacune de ces sommes. Les degrés de liberté représentent aussi le nombre d'éléments dans la somme des carrés moins le nombre de paramètres estimés dans cette somme. Notons que lorsque H_0 n'est pas vérifiée, seul MC_{res} est encore un estimateur sans biais de σ^2 . Il s'agit en fait de l'estimateur s^2 défini auparavant. Lorsque l'hypothèse H_0 est vérifiée, les quantités $\frac{SC_{tot}}{\sigma^2}$, $\frac{SC_{reg}}{\sigma^2}$, $\frac{SC_{res}}{\sigma^2}$ suivent des lois de χ^2 avec respectivement $(n-1)$, 1 et $(n-2)$ degrés de liberté. De plus, les quantités $\frac{SC_{reg}}{\sigma^2}$, $\frac{SC_{res}}{\sigma^2}$ sont indépendantes. Il s'ensuit que lorsque l'hypothèse H_0 est vérifiée, la statistique :

$$F_c = \frac{\left(\frac{SC_{reg}}{1 \cdot \sigma^2}\right)}{\left(\frac{SC_{res}}{(n-2) \cdot \sigma^2}\right)} = \frac{SC_{reg}}{SC_{res}} \cdot \frac{n-2}{1} = \frac{MC_{reg}}{MC_{res}}$$

suit une loi de Fisher avec 1 et $(n-2)$ degrés de liberté. On rejette ainsi H_0 au seuil de signification α si :

$$F_c > F_{(\alpha, 1, n-2)},$$

où la valeur critique $F_{(\alpha, 1, n-2)}$ est le $(1-\alpha)$ quantile d'une loi de Fisher avec 1 et $(n-2)$ degrés de liberté que l'on trouve dans une table de Fisher. Cette procédure est appelée *analyse de la variance*. On présente généralement les quantités nécessaires au calcul de la statistique de test F_c dans un tableau d'analyse de variance, abrégé par *tableau de ANOVA* et qui prend la forme de tableau 1.1.

Source de la variance	Degré de liberté	Somme des carrés	Moyenne des carrés
Régression	$n-1$	$SC_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$MC_{reg} = \frac{SC_{reg}}{1}$
Résiduelle	$n-2$	$SC_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$MC_{res} = \frac{SC_{res}}{n-2}$
Total	$n-1$	$SC_{res} = \sum_{i=1}^n (y_i - \bar{y})^2$	

TAB. 1.1 – Forme général du Tableau ANOVA

Remarquons que cet test qui utilise la statistique F_c est équivalent au test présenté dans la section précédente pour la même hypothèse qui utilise la statistique t_c . En effet, comme :

$$\begin{aligned}
 SC_{reg} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\
 &= \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= \sum_{i=1}^n (\bar{y} + \hat{\beta}_1(x_i - \bar{x}) - \bar{y})^2
 \end{aligned}$$

alors

$$\begin{aligned}
 F_c &= \frac{SC_{reg}}{SC_{res}/(n-2)} \\
 &= \frac{\hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{s^2} \\
 &= \frac{\hat{\beta}_1^2}{s^2(\hat{\beta}_1)} \\
 &= t_c^2
 \end{aligned}$$

L'équivalence des deux tests est alors la conséquence du fait que :

$$F_{(\alpha,1,n-2)} = t_{(\alpha/2,n-2)}^2.$$

1.1.6 Coefficient de corrélation

La corrélation entre deux variables aléatoires X et Y est mesurée par le coefficient

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{V(X), V(Y)}}$$

Le coefficient de corrélation échantillonnal est

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

1.1.7 Interprétation du coefficient de corrélation

On peut montrer que $-1 < r < 1$:

- Si $r = -1$ alors il y a corrélation parfaite entre X et Y et les points $(X_i; Y_i)$ sont tous sur la droite de régression.
- Si $r = 0$ alors il n'y a pas de corrélation entre X et Y et les points $(X_i; Y_i)$ sont dispersés au hasard.
- Si $0 < r < 1$ alors il y a corrélation positive faible, moyenne ou forte entre X et Y . Dans ce cas, une augmentation de X entraîne une augmentation de Y .
- Si $-1 < r < 0$ alors il y a corrélation négative faible, moyenne ou forte entre X et Y . Dans ce cas, une augmentation de X entraîne une diminution de Y .

Remarque 1.1.1. Pour décider de la qualité de l'ajustement en plus de la valeur du coefficient de corrélation, il faut tenir compte de la nature du nuage de points.

1.1.8 Cas particulier

Nous avons vu dans ce chapitre que lorsque l'on ne rejette pas les hypothèses $\beta_0 = 0$ ou $\beta_1 = 0$, on peut être amené à considérer des modèles avec un seul paramètre. Nous allons montrer maintenant comment ces modèles se résolvent avec la méthode des moindres carrés.

Modèle sans variable explicative

Partons du modèle :

$$y_i = \beta_0 + \epsilon_i,$$

En considérant également les hypothèses émises, on a donc les y_i qui sont indépendants et normalement distribués avec :

$$E(y_i) = \beta_0,$$

$$Var(y_i) = \sigma^2.$$

À partir d'un estimateur $\hat{\beta}_0$ de β_0 , on définit les valeurs estimées :

$$\hat{y}_i = \hat{\beta}_0,$$

et les résidus :

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0.$$

L'estimateur $\hat{\beta}_0$ obtenu par la méthode des moindres carrés est le point qui minimise la fonction définie par la somme des carrés des résidus :

$$z = f(\beta_0) = \sum_{i=1}^n (y_i - \beta_0)^2$$

c'est-à-dire le point qui annule sa dérivée :

$$\frac{\delta z}{\delta \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0).$$

$\hat{\beta}_0$ est donc la solution de :

$$\sum_{i=1}^n y_i = n\beta_0.$$

On trouve ainsi :

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}.$$

On a donc ici :

$$SC_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = SC_{tot},$$

et

$$SC_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - y_i)^2.$$

L'estimateur $\hat{\beta}_0 = \bar{y}$ est normalement distribué, d'espérance :

$$E(\hat{\beta}_0) = E(\bar{y}) = E(y_i) = \beta_0$$

et de variance :

$$\sigma^2(\hat{\beta}_0) = Var(\bar{y}) = \frac{Var(y_i)}{n} = \frac{\sigma^2}{n}.$$

Par ailleurs, l'espérance de la somme des carrés des résidus vaut :

$$E\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) = E\left(\sum_{i=1}^n (y_i - \bar{y})^2\right) = \sigma^2(n-1).$$

On obtient ainsi un estimateur sans biais de σ^2 en posant :

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

est un estimateur sans biais de $\sigma^2(\hat{\beta}_0)$ qui vaut :

$$s^2(\hat{\beta}_0) = \frac{s^2}{n}.$$

On peut montrer que la quantité :

$$\frac{\hat{\beta}_0 - \beta_0}{s(\hat{\beta}_0)}$$

où $s(\hat{\beta}_0)$ est la racine carrée de $s^2(\hat{\beta}_0)$, suit une loi de student avec $(n-1)$ degrés de liberté.

On rejette ainsi l'hypothèse nulle :

$$H_0 : \beta_0 = b_0,$$

pour l'hypothèse alternative :

$$H_0 : \beta_0 \neq b_0$$

au seuil de signification α si :

$$|t_c| = \frac{|\hat{\beta}_0 - b_0|}{s(\hat{\beta}_0)},$$

Si

$$|t_c| > t_\alpha(2, n-1)$$

où la valeur critique $t_{(\alpha/2, n-1)}$ est le $(1 - \alpha/2)$ quantile d'une loi de Student avec $(n-1)$ degrés de liberté que l'on trouve dans une table de Student. On voit ainsi que cette procédure est identique à la procédure du t-test classique utilisée pour inférer sur la moyenne β_0 des y_i .

Modèle sans constante

Partons maintenant du modèle :

$$y_i = \beta_1 x_i + \epsilon_i.$$

En considérant également les hypothèses émises, on a donc les y_i qui sont indépendants et normalement distribués avec :

$$E(y_i) = \beta_1 x_i,$$

$$Var(y_i) = \sigma^2.$$

À partir d'un estimateur $\hat{\beta}_1$ de β_1 , on définit les valeurs estimées :

$$\hat{y}_i = \hat{\beta}_1 x_i.$$

et les résidus :

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 x_i.$$

L'estimateur $\hat{\beta}_1$ obtenu par la méthode des moindres carrés est le point qui minimise la fonction définie par la somme des carrés des résidus :

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

c'est-à-dire le point qui annule sa dérivée :

$$\frac{\delta z}{\delta \beta_1} = -2 \sum_{i=1}^n x_i (y_i - \beta_1 x_i).$$

$\hat{\beta}_1$ est donc la solution de :

$$\sum_{i=1}^n x_i y_i = \beta_1 \sum_{i=1}^n x_i^2.$$

On trouve ainsi :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

La particularité d'un modèle sans constante est que la droite des moindres carrés :

$$\hat{y}(x) = \hat{\beta}_1 x$$

ne passe pas par le point (\bar{x}, \bar{y}) et que l'on n'a pas :

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i.$$

Par conséquent, la somme des résidus n'est pas nulle. Il en résulte également que l'on n'a pas non plus :

$$SC_{tot} = SC_{reg} + SC_{res}.$$

L'estimateur $\hat{\beta}_1$ est normalement distribué d'espérance :

$$\begin{aligned} E(\hat{\beta}_1) &= \frac{\sum_{i=1}^n x_i E(y_i)}{\sum_{i=1}^n x_i^2} \\ &= \frac{\beta_1 \sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} \\ &= \beta_1. \end{aligned}$$

et de variance :

$$\begin{aligned} \sigma^2(\hat{\beta}_1) &= \frac{\sum_{i=1}^n x_i^2 \text{Var}(y_i)}{\sum_{i=1}^n x_i^2} \\ &= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^n x_i^2}. \end{aligned}$$

Les valeurs estimées \hat{y}_i sont également normalement distribué d'espérance :

$$\begin{aligned} E(\hat{y}_i) &= E(\hat{\beta}_1 x_i) \\ &= x_i E(\hat{\beta}_1) \\ &= x_i \beta_1 \\ &= E(y_i). \end{aligned}$$

et de variance :

$$\text{Var}(\hat{y}_i) = \text{Var}(\hat{\beta}_1 x_i) = x_i^2 \text{Var}(\hat{\beta}_1) = \frac{x_i^2 \sigma^2}{\sum_{i=1}^n x_i^2}$$

Par ailleurs on peut calculer l'espérance de la somme des carrés des résidus :

$$\begin{aligned}
 E\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) &= E\left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n \hat{y}_i^2\right) \\
 &= \sum_{i=1}^n \text{Var}(y_i) - \sum_{i=1}^n \text{Var}(\hat{y}_i) \\
 &= n\sigma^2 - \frac{\sigma^2 \sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} \\
 &= \sigma^2(n-1).
 \end{aligned}$$

On obtient ainsi un estimateur sans biais de σ^2 en posant :

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}.$$

On a ainsi un estimateur :

$$s^2(\hat{\beta}_1) = \frac{s^2}{\sum_{i=1}^n x_i^2}$$

de $\sigma^2(\hat{\beta}_1)$. On peut montrer que la quantité :

$$\frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)}.$$

où $s(\hat{\beta}_1)$ suit une loi de Student avec $(n-1)$ degrés de liberté. On rejette ainsi l'hypothèse :

$$H_0 : \beta_1 = b_1$$

pour l'hypothèse alternative :

$$H : \beta_1 \neq b_1$$

au seuil de signification α si :

$$|t_c| = \frac{|\hat{\beta}_1 - b_1|}{s(\hat{\beta}_1)} > t_{(\alpha/2, n-1)}.$$

Chapitre 2

La régression linéaire multiple

2.1 Régression linéaire multiple

Introduction

En pratique la variable expliquée ne dépend pas que d'une variable explicative. Nous introduisons dans ce chapitre des modèles de régression linéaire avec plusieurs variables explicatives. On parle alors de la régression linéaire multiple. La droite de régression que l'on avait en régression simple est ici remplacée par un hyperplan des régressions. Afin d'estimer les paramètres de cet hyperplan, on peut comme en régression simple, utiliser la méthode des moindres carrés.

2.1.1 Modèle

Une variable quantitative Y dite *expliqué* (ou encore, exogène, dépendante) est mise en relation avec p variables quantitatives X^1, \dots, X^p dites *explicatives* (ou encore de contrôle, endogène, indépendantes, régresseurs).

Les données sont supposées provenir de l'observation d'un échantillon statistique de taille n ($n > p+1$) de \mathbb{R}^{p+1}

$$(x_i^1, \dots, x_i^j, \dots, x_i^p, y_i), i = 1, 2, \dots, n.$$

L'écriture du *modèle linéaire* dans cette situation conduit à supposer que l'espérance de Y appartient au sous-espace de \mathbb{R}^n engendré par $\mathbf{1}, X^1, \dots, X^p$ où $\mathbf{1}$ désigne le vecteur de \mathbb{R}^n constitué de 1. C'est-à-dire que les $p+1$ variables aléatoires vérifient :

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \dots + \beta_{p-1} x_i^{p-1} + \epsilon. \quad (2.1)$$

où la variable ϵ représente le comportement individuel avec les hypothèses suivantes :

1. Les ϵ_i sont des termes d'erreurs d'une variable, non observés, indépendants distribués :
 $E(\epsilon_i) = 0, \text{Var}(\epsilon) = \sigma_\epsilon^2 I.$
2. Les termes X^j sont supposés déterministes (facteurs contrôlés) ou bien l'erreur ϵ est indépendante de la distribution conjointe de $X^1 \dots X^p$. On écrit dans ce dernier cas que : $E(Y|X^1, \dots, X^p) = \beta_0 + \beta_1 X^1 + \beta_2 X^2 + \dots + \beta_p X^p$
 et $\text{Var}(Y|X^1, \dots, X^p) = \sigma_\epsilon^2.$
3. Les paramètres inconnus sont supposés constants.
4. En option, l'étude spécifique des lois des estimateurs, une quatrième hypothèse considère la normalité de la variable d'erreur $U(N(0, \sigma_\epsilon^2 I))$. Les ϵ_i sont alors i.i.d de loi $N(0, \sigma_\epsilon^2)$. Les données sont rangées dans une matrice $X(n(p+1))$ de terme général x_i^j , dont la première colonne contient le vecteur 1, ($x_0^i = 1$), et dans un vecteur Y de terme général y_i . En notant les vecteurs $\epsilon = [\epsilon_1, \dots, \epsilon_p]$ et $\beta = [\beta_0, \beta_1, \dots, \beta_p]$, le modèle s'écrit matriciellement :

$$y = X\beta + \epsilon$$

2.1.2 Modèle à deux variables explicatives

Considérons d'abord le cas d'un modèle avec deux variables explicatives. On suppose donc que l'on a pour tout $i = 1, \dots, n$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

Il s'agit d'estimer les paramètres β_0, β_1 et β_2 par $\hat{\beta}_0, \hat{\beta}_1$ et $\hat{\beta}_2$. Les valeurs estimés seront alors :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}.$$

et les résidus :

$$\begin{aligned} e_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} \end{aligned}$$

En utilisant la méthode des moindres carrés, on cherche à minimiser la quantité :

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2,$$

c'est-à-dire minimiser la fonction objectif :

$$z = f_{(\beta_0, \beta_1 \text{ et } \beta_2)} = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2.$$

Les dérivées partielles de z sont données par :

$$\frac{\delta z}{\delta \beta_0} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}),$$

$$\frac{\delta z}{\delta \beta_1} = -2 \sum_{i=1}^n x_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}),$$

$$\frac{\delta z}{\delta \beta_2} = -2 \sum_{i=1}^n x_{i2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}).$$

En annulant ces dérivées partielles, on abouti au système d'équations suivant dont les estimateurs $\hat{\beta}_0$, $\hat{\beta}_1$ et $\hat{\beta}_2$ sont les solutions :

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}) &= 0, \\ \sum_{i=1}^n x_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}) &= 0, \\ \sum_{i=1}^n x_{i2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}) &= 0. \end{aligned}$$

c'est-à-dire

$$\sum_{i=1}^n y_i = n\beta_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2},$$

$$\sum_{i=1}^n x_{i1}y_i = \hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1}x_{i2},$$

$$\sum_{i=1}^n x_{i2}y_i = \hat{\beta}_0 \sum_{i=1}^n x_{i2} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}x_{i2} + \hat{\beta}_2 \sum_{i=1}^n x_{i2}^2.$$

2.1.3 Propriétés

Les estimateurs des M.C sont des estimateurs sans biais, qui coïncident avec ceux des moindres carrés, sont uniformément meilleurs. On montre que la matrice de covariance des estimateurs se met sous la forme

$$E[(b - \beta)(b - \beta)'] = \sigma_e^2(X'X)^{-1}$$

celle des prédicteurs est

$$E[(\hat{y} - X\beta)(\hat{y} - X\beta)'] = \sigma_e^2 H$$

et celles des estimateurs des résidus est

$$E[(e - \mu)(e - \mu)'] = \sigma_e^2 [I - H]$$

tandis qu'un estimateur sans biais est fournis par :

$$S^2 = \frac{\|e\|^2}{n - p - 1} = \frac{\|y - X\beta\|^2}{n - p - 1} = \frac{SSE}{n - p - 1}$$

Ainsi les termes S^2 sont des estimateurs des variances des prédicteurs \hat{y}_i

2.1.4 Sommes des carrés

SSE est la somme des carrés des résidus (sum of squared errors).

$$SSE = \|y - \hat{y}\|^2 = \|e\|^2$$

On définit également la somme totale des carrés (total sum of square) par :

$$SST = \|y - \bar{y}1\|^2 = y'y - n\bar{y}^2$$

et la somme des carrés de la régression (régression sum of square) par :

$$SST = \|y - \bar{y}1\|^2 = \hat{y}' - n\bar{y} = y'Hy - n\bar{y}^2 = b'X'y - n\bar{y}^2$$

2.1.5 Coefficient de détermination

On appelle coefficient de détermination le rapport

$$R^2 = \frac{SSR}{SST}$$

qui est part de part de variance de Y expliquée par le modèle de régression. Géométriquement, c'est un rapport de carrés de longueur de deux vecteurs. C'est donc le cosinus carré de l'angle entre ces vecteurs : y et sa projection \hat{y} sur Vecteur(X) dans le cas extrême où $n = (p + 1)$, c'est-à-dire si le nombre de variable explicatives est grand comparativement au nombre d'observation, $R^2 = 1$ ou encors, il est géométriquement facile de voir que l'ajout de variable explicative ne peut que faire croître le coefficient de détermination. La quantité R est appelée coefficient de corrélation multiple entre Y est la variable explicative, c'est le coefficient de corrélation usuel entre y et sa prédiction \hat{y} .

2.1.6 Cas général

Considérons à présent un modèle avec un nombre arbitraire $(p - 1)$ de variables explicatives. Nous supposons donc que nos données sont liées par 4.2. Il s'agit d'estimer les p paramètres inconnus $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$ par $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{p-1}$ En utilisons la méthode des moindres carrés, on cherche à minimiser la quantité :

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_{p-1} x_{ip-1} \right)^2 ;$$

c'est-à-dire à minimiser la fonction objectif :

$$z = f(\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}) = \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_{p-1} x_{ip-1} \right).$$

Les estimateurs $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{p-1}$ qui permettent de définir l'hyperplan des moindres carrés sont donnés par le minimum de cette fonction Afin de pouvoir résoudre plus facilement ce problème, nous allons introduire ici la notation matricielle. Il s'agit de remplacer les n équations de 4.2 par la seule équation matricielle :

$$\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{p-1} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2p-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & \cdot & \cdot & \cdot & x_{np-1} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}$$

c'est-à-dire par :

$$y = X\beta + \epsilon,$$

où y est le $(n \times 1)$ vecteur des observations relatives à la variable expliquée, X est la $(n \times p)$ matrice relative aux $(p-1)$ variables explicatives avec en plus une colonne de 1 qui correspond au paramètre β_0 , β_1 est le $(p \times 1)$ vecteur des paramètres inconnus à estimer et ϵ est le $(n \times 1)$ vecteurs des erreurs. le problème consiste donc à estimer le vecteur β par un vecteur d'estimateurs :

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdot \\ \cdot \\ \hat{\beta}_{p-1} \end{bmatrix}$$

le vecteur des valeurs estimées :

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \cdot \\ \cdot \\ \hat{y}_{p-1} \end{bmatrix}$$

est alors obtenu en posant :

$$\hat{y} = X\hat{\beta}$$

alors que le vecteur des résidus :

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix}.$$

est obtenu en posant :

$$e = y - \hat{y}.$$

La méthode des moindres carrés consiste donc à minimiser la quantité :

$$\sum_{i=1}^n e_i^2 = e'e = \left((y - \hat{y})'(y - \hat{y}) = (y - X\hat{\beta})'(y - X\hat{\beta}) \right),$$

En développant cette expression, on obtient :

$$\begin{aligned} e'e &= y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\ &= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}. \end{aligned}$$

Dans ce développement, on a utilisé le fait que $\hat{\beta}' X' y$ est un scalaire, ce qui implique que :

$$\hat{\beta}' y = (\hat{\beta}' X' y)' = y' X \hat{\beta}.$$

On peut montrer en outre que la dérivée partielle de la fonction objectif $e'e$ par rapport à $\hat{\beta}$ est alors donnée par :

$$-2X' y + 2X' X \hat{\beta}.$$

l'estimateur $\hat{\beta}$ que l'on cherche est donc la valeur qui annule cette expression. On a ainsi une seule équation normale sous la forme matricielle :

$$(X' X) \hat{\beta} = X' y,$$

et on obtient l'estimateur des moindres carrés défini par :

$$\hat{\beta} = (X' X)^{-1} X' y.$$

Notons que dans le cas d'un modèle de régression simple, c'est-à-dire le modèle (2.1) avec $p = 2$, on a :

$$X' X = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

et donc

$$\begin{aligned} (X' X)^{-1} &= \frac{1}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i & -\sum_{i=1}^n x_i^2 \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \\ &= \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 / n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \end{aligned}$$

et finalement :

$$\hat{\beta} = \begin{pmatrix} \frac{\bar{y} \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{pmatrix}$$

Ce qui correspond aux estimateurs des moindres carrés.

2.1.7 Propriétés des moindres carrés

Lorsque on utilise la méthode des moindres carrés, on a toujours la propriété suivante :

$$\sum_{i=1}^n \hat{y}^2 = \sum_{i=1}^n \hat{y}_i y_i,$$

ou autrement dit sous forme matricielle :

$$\hat{y}' \hat{y} = \hat{y}' y.$$

En effet on a :

$$\begin{aligned} \hat{y}' \hat{y} &= (X\hat{\beta})' X\hat{\beta} \\ &= (X\hat{\beta})' X\hat{\beta} \\ &= \hat{\beta}' X' X\hat{\beta} \\ &= \hat{\beta}' X' X(X'X)^{-1} X' y \\ &= (X\hat{\beta})' y \\ &= \hat{y}' y, \end{aligned}$$

Il s'ensuit que :

$$\begin{aligned} e' e &= (y - \hat{y})' (y - \hat{y}) \\ &= (y' - \hat{y}') (y - \hat{y}) \\ &= y' y - y' \hat{y} - \hat{y}' y + \hat{y}' \hat{y} \\ &= y' y - \hat{y}' \hat{y}. \end{aligned}$$

Autrement dit, la somme des carrés des résidus peut s'exprimer par :

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n y_i^2 - \sum_{i=1}^n \hat{y}_i^2,$$

Par ailleurs, lorsque l'on considère un modèle avec une constante β_0 comme (2.1), on peut montrer que l'une des équations normales égal à :

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 - \dots - \hat{\beta}_{p-1} \bar{x}_{p-1},$$

Il s'ensuit que l'hyperplan des moindres carrés contient le point $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{p-1}, \bar{y})$. On a également :

$$\sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i,$$

En effet :

$$\begin{aligned} \sum_{i=1}^n y_i &= \sum_{i=1}^n \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_{p-1} x_{i(p-1)} \right) \\ &= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} + \dots + \hat{\beta}_{p-1} \sum_{i=1}^n x_{i(p-1)} \\ &= n \left(\bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 - \dots - \hat{\beta}_{p-1} \bar{x}_{p-1} \right) \\ &\quad + \hat{\beta}_1 n\bar{x}_1 + \hat{\beta}_2 n\bar{x}_2 + \dots + \hat{\beta}_{p-1} n\bar{x}_{p-1} \\ &= \sum_{i=1}^n y_i. \end{aligned}$$

Il s'ensuit que la somme (et donc la moyenne) des résidus est toujours nulle. En effet :

$$\sum_{i=1}^n e_i = \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{y}_i = 0$$

Nous en déduisons aussi la décomposition suivante :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

où, comme en régression simple, ces trois sommes de carrés sont appelées respectivement somme des carrés totale, somme des carrés due à la régression et somme des carrés des résidus et sont abrégées par :

$$SC_{tot} = SC_{reg} + SC_{res}.$$

On a en effet :

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

Ainsi :

$$\begin{aligned}
 \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 &= \left[\sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2 \right] + \left[\sum_{i=1}^n y_i^2 - \sum_{i=1}^n \hat{y}_i^2 \right] \\
 &= \sum_{i=1}^n y_i^2 - n\bar{y}^2 \\
 &= \sum_{i=1}^n (y_i - \bar{y})^2.
 \end{aligned}$$

À partir de ces sommes de carrés, on définit le coefficient de détermination par :

$$R^2 = \frac{SC_{reg}}{SC_{tot}}.$$

2.1.8 Estimation de la variance des erreurs

Comme en régression simple, nous allons estimer la variance des erreurs σ^2 à partir de la somme des carrés des résidus :

$$\sum_{i=1}^n (e_i - \bar{e})^2 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Or nous avons :

$$\begin{aligned}
 E \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) &= E \left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n \hat{y}_i^2 \right) \\
 &= \sum_{i=1}^n E(y_i^2) - \sum_{i=1}^n E(\hat{y}_i^2) \\
 &= \sum_{i=1}^n (Var(y_i) + E^2(y_i)) - \sum_{i=1}^n (Var(\hat{y}_i) + E^2(\hat{y}_i)).
 \end{aligned}$$

et le vecteur \hat{y} suit une loi multinormale avec :

$$E(\hat{y}) = E(X\hat{\beta}) = XE(\hat{\beta}) = X\beta = E(y).$$

c'est-à-dire pour $i = 1, \dots, n$:

$$E(\hat{y}_i) = E(y_i).$$

et que d'autre part, la matrice variance-covariance des valeurs estimés

$$Var(\hat{y}) = \begin{bmatrix} Var(\hat{y}_1) & Cov(\hat{y}_1, \hat{y}_2) & \dots & Cov(\hat{y}_1, \hat{y}_n) \\ Cov(\hat{y}_2, \hat{y}_1) & Var(\hat{y}_2) & \dots & Cov(\hat{y}_2, \hat{y}_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{y}_n, \hat{y}_1) & Cov(\hat{y}_n, \hat{y}_2) & \dots & Var(\hat{y}_n) \end{bmatrix}$$

est donné par :

$$\begin{aligned} Var(\hat{y}) &= Var(X\hat{\beta}) \\ &= Var(X(X'X)^{-1}Xy) \\ &= X(X'X)^{-1}X'Var(y)(X(X'X)^{-1}X')' \\ &= \sigma^2 X(X'X)^{-1}X'X(X'X)^{-1}X' \\ &= \sigma^2 X(X'X)^{-1}X'. \end{aligned}$$

Les variances des \hat{y}_i sont donc multiples de σ^2 . Ces multiples se trouvent sur la diagonale de la matrice :

$$H = X(X'X)^{-1}X'.$$

On obtient ainsi :

$$E\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) = \sum_{i=1}^n Var(y_i) - \sum_{i=1}^n Var(\hat{y}_i) = n\sigma^2 - \sigma^2 Tr(H),$$

où $Tr(H)$ désigne la trace de la matrice H, c'est-à-dire la somme de ses éléments diagonaux.

Or la trace d'une matrice est un opérateur commutatif. On a donc :

$$\begin{aligned} Tr(H) &= Tr(X(X'X)^{-1}X') \\ &= Tr((X'X)^{-1}XX') \\ &= Tr(I_p). \end{aligned}$$

où I_p est la matrice identité de dimension p dont la trace vaut p . On a donc :

$$E\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) = (n - p)\sigma^2$$

On peut ainsi définir un estimateur sans biais de σ^2 en posant :

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p} = \frac{SC_{res}}{n - p} = \frac{SC_{tot} - SC_{reg}}{n - p},$$

On estime par ailleurs l'écart type des erreurs σ par :

$$s = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p}}.$$

2.2 Inférence sur les paramètres du modèle

Comme pour la régression simple, on peut effectuer des tests d'hypothèses sur les paramètres d'un modèle de régression multiple et construire des intervalles de confiance.

2.2.1 Tests d'hypothèses

On a vu que le vecteur aléatoire $\hat{\beta}$ suit une loi multinormale, d'espérance β et de variance :

$$\sigma^2(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

Cela signifie que les estimateurs $\hat{\beta}_j$ sont normalement distribués d'espérance β_j et que leurs variance $\sigma^2(\hat{\beta}_j)$ se trouve sur la diagonale de la matrice $\sigma^2(\hat{\beta})$. Ces variances dépendent donc de la valeur inconnue σ^2 que l'on estime par s^2 . On estime ainsi les $\sigma^2(\hat{\beta}_j)$ par les $s^2(\hat{\beta}_j)$ qui se trouvent sur la diagonale de la matrice :

$$s^2(X'X)^{-1}.$$

Afin de tester l'hypothèse nulle :

$$H_0 : \beta_j = b_j$$

contre l'hypothèse alternative :

$$H_0 : \beta_j \neq b_j$$

pour un certain j entre 0 et $(p - 1)$, on calcule la statistique :

$$t_c = \frac{\hat{\beta}_j - b_j}{s(\hat{\beta}_j)}.$$

où $s(\hat{\beta}_j)$ est la racine carrée de $s^2(\hat{\beta}_j)$. Sous l'hypothèse nulle, on peut montrer que cette statistique suit une loi de Student avec $(n - p)$ degrés de liberté. On rejette ainsi H_0 au seuil de signification α si :

$$|t_c| > t_{(\alpha/2, n-p)}.$$

où la valeur critique $t_{(\alpha/2, n-p)}$ quantile d'une loi de Student avec $(n - p - 1)$ degrés de liberté que l'on trouve dans une table de Student. En particulier, si l'on veut tester l'hypothèse nulle :

$$H_0 : \beta_j = 0$$

contre l'hypothèse alternative :

$$H_1 : \beta_j \neq 0$$

pour un certain j entre 0 et $(n - p)$, la statistique à calculer est donnée par :

$$t_c = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)}.$$

Dans le cas où l'on ne rejette pas cette hypothèse nulle, on dit que les x_{ij} relatifs à la variable X_j ne sont pas significatifs au sein du modèle. Ceci implique qu'un modèle plus simple, sans cette variable, peut être considéré.

2.2.2 Intervalle de confiance

Un intervalle de confiance au niveau $(1 - \alpha)$ pour un paramètre β_j est défini par :

c'est-à-dire :

$$\hat{\beta}_j \neq t_{(\alpha/2, n-p-1).s(\hat{\beta}_j)}.$$

Cet intervalle est construit de telle sorte qu'il contienne le paramètre inconnu β_j avec une probabilité de $(1 - \alpha)$.

2.2.3 Analyse de la variance

Nous retrouvons ici des liens entre la régression multiple et la régression simple. En effet, si l'on considère un modèle de type (2.1), on peut montrer que si l'hypothèse :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$$

est vérifiée, c'est-à-dire que l'on a un modèle sans variable explicative :

$$y_i = \beta_0 + \epsilon_i$$

alors les espérances des trois sommes des carrés introduites précédemment sont respectivement :

$$E(SC_{tot}) = (n - 1)\sigma^2$$

$$E(SC_{reg}) = (p - 1)\sigma^2$$

$$E(SC_{res}) = (n - p)\sigma^2$$

Les quantités suivantes sont alors des estimateurs sans biais de la variance des erreurs σ^2 :

$$MC_{tot} = \frac{SC_{tot}}{(n - 1)}$$

$$MC_{reg} = \frac{SC_{reg}}{(p - 1)}$$

$$MC_{res} = \frac{SC_{res}}{(n - p - 1)}$$

Lorsque H_0 n'est pas vérifiée, seule MC_{res} est encore un estimateur sans biais de σ^2 . Lorsque l'hypothèse H_0 est vérifiée, on peut montrer, de manière analogue à ce que l'on a vu en régression simple, que la statistique :

$$F_c = \frac{MC_{reg}}{MC_{res}}$$

suit une loi de Fisher avec $(p - 1)$ et $(n - p)$ degrés de liberté. On rejette ainsi H_0 à un seuil de signification α si :

$$F_c > F_{(\alpha, p-1, n-p)}$$

où la valeur critique $F_{(\alpha, p-1, n-p)}$ est le $(1 - \alpha)$ quantile d'une loi de Fisher avec $(p - 1)$ et $(n - p)$ degrés de liberté que l'on trouve dans une table de Fisher. Le tableau 2.1 nous montre comment on peut construire un tableau d'analyse de variance (ANOVA) lors d'une régression multiple. On notera qu'en posant $p = 2$, on retrouve le tableau ANOVA de la régression simple.

Source de variation	Degré de liberté	Somme des carrés	Moyenne des carrés	F_C
Régression	$p - 1$	$SC_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\frac{SC_{reg}}{p-1}$	$\frac{MC_{reg}}{MC_{res}}$
Résiduelle	$n - p$	$SC_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$\frac{SC_{res}}{n-p}$	
Total	$n - 1$	$SC_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$		

TAB. 2.1 – Tableau de ANOVA pour la régression multiple

2.2.4 Régression sans constante

Comme en régression simple, le cas d'un modèle de régression multiple sans constante est particulier. Ainsi, lorsque l'on considère un modèle avec p variables explicatives où l'on a pour tout $i = 1, \dots, n$:

$$y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i. \quad (2.2)$$

(notons que p désigne ici à la fois le nombre de paramètres à estimer et le nombre de variables explicatives), c'est-à-dire lorsque :

$$y = X\beta + \epsilon,$$

où X est ici la matrice des variables explicatives sans la colonne de 1, la solution des moindres carrés est toujours donnée par :

$$\hat{\beta} = (X'X)^{-1}X'y,$$

et on définit de même le vecteur des valeurs estimées par :

$$\hat{y} = X\hat{\beta},$$

et le vecteur des résidus par :

$$e = y - \hat{y},$$

La somme des résidus n'est donc pas nulle, il s'ensuit que l'on n'a pas non plus :

$$SC_{tot} = SC_{reg} + SC_{res} \quad (2.3)$$

du moins si ces sommes de carrés sont définies comme auparavant. Par contre, on a tout de même :

$$SC_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i)^2 - \sum_{i=1}^n (\hat{y}_i)^2.$$

qui, on l'a vu, se déduisant de la définition de $\hat{\beta}$. C'est pourquoi, certains auteurs proposent de définir ici :

$$SC_{tot} = \sum_{i=1}^n y_i^2,$$

et :

$$SC_{reg} = \sum_{i=1}^n \hat{y}_i^2,$$

de manière à avoir tout de même 2.3 . De même, ces auteurs définissent ici un coefficient de détermination par :

$$R^2 = \frac{\sum_{i=1}^n \hat{y}_i^2}{\sum_{i=1}^n y_i^2}.$$

bien que l'interprétation de cette quantité soit délicate. Si l'on fait les mêmes hypothèses sur les ϵ_i , les formules pour l'espérance et la variance de $\hat{\beta}$ et \hat{y} demurent valables (de même que les procédures de tests d'hypothèses sur les différents paramètres β_j du modèle). En ce qui concerne l'estimation de la variance des erreurs σ^2 , on a plus ici :

$$\sum_{i=1}^n (e_i - \bar{e})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Pourtant, la somme des carrés des résidus est tout de même utilisée comme numérateur de l'estimateur de σ^2 . On a toujours :

$$E\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) = E\left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n \hat{y}_i^2\right) = \sum_{i=1}^n Var(y_i) - \sum_{i=1}^n Var(\hat{y}_i) = (n - p)\sigma^2.$$

Un estimateur sans biais de σ^2 est donc toujours donné par :

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p} = \frac{SC_{res}}{n - p}.$$

Afin de tester l'hypothèse que tous les coefficients β_j sont nuls :

$$H_0 : \beta_1 = \dots = \beta_p = 0$$

on calcule ici la statistique de test :

$$F_c = \frac{SC_{reg}}{SC_{res}} \cdot \frac{n - p}{p}.$$

en utilisant la nouvelle définition de SC_{reg} , qui lorsque H_0 est vérifiée suit une loi de Fisher avec p et $(n - p)$ degrés de liberté. Cette statistique est donc définie de manière analogue à celle utilisée pour un modèle constant, sauf que le nombre de degrés de liberté associé à SC_{reg} est ici de p et non de $(p - 1)$. On rejette ainsi H_0 à un seuil de signification α si :

$$F_c > F_{(\alpha, p, n-p)}.$$

où la valeur critique $F_{(\alpha, p, n-p)}$ est le $(1 - \alpha)$ quantile d'une loi de Fisher avec p et $(n - p)$ degrés de liberté que l'on trouve dans une table de Fisher.

2.2.5 Test F partiel

Les tests d'hypothèses présentés dans les trois dernières sections sont tous des cas particuliers du test présenté, le test du F partiel. Cette procédure nous permet de tester la nullité d'un certain nombre r de paramètres dans un modèle avec p paramètres. L'hypothèse nulle est donc celle d'un modèle réduit avec $(p - r)$ paramètres et l'hypothèse alternative celle d'un modèle complet avec p paramètres. En voici quelques exemples :

1. On peut tester la nullité d'un paramètre, par exemple β_1 , dans un modèle complet avec constante. On a donc :

$$H_0 : y_i = \beta_0 + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{p-1} + \epsilon_i$$

$$H_1 : y_i = \beta_0 + \beta_1 + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{p-1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{p-1} + \epsilon_i$$

2. On peut tester la nullité de tous les paramètres excepté la constante dans un modèle complet avec constante. On a donc :

$$H_0 : \beta_0 + \epsilon_i$$

$$H_1 : y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{p-1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{p-1} + \epsilon_i$$

3. On peut tester la nullité de tous les paramètres dans un modèle complet sans constante. On a donc :

$$H_0 : y_i = \epsilon_i$$

$$H_1 : y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{p-1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{p-1} + \epsilon_i$$

Afin de pouvoir décider d'un tel test, il s'agit de procéder comme suit :

- Il faut tout d'abord calculer les valeurs estimées \hat{y}_i en utilisant la méthode des moindres carrées pour chacun des deux modèles définis par H_0 et H_1 . On les note respectivement par $\hat{y}_i(H_0)$ et $\hat{y}_i(H_1)$. De même, on note par $SC_{res}(H_0)$ et $SC_{res}(H_1)$ la somme des carrés des résidus obtenus pour ces modèles.
- On calcule la statistique :

$$\begin{aligned} F_c &= \frac{\sum_{i=1}^n \hat{y}_i^2(H_1) - \sum_{i=1}^n \hat{y}_i^2(H_0)}{\sum_{i=1}^n \hat{y}_i^2 - \sum_{i=1}^n \hat{y}_i^2(H_1)} \cdot \frac{n-p}{r} \\ &= \frac{SC_{res}(H_0) - SC_{res}(H_1)}{SC_{res}(H_1)} \cdot \frac{n-p}{r}. \end{aligned}$$

- On rejette l'hypothèse nulle au seuil de signification α si :

$$F_c > F_{\alpha, r, n-p}$$

où la valeur critique $F_{\alpha, r, n-p}$ est $(1 - \alpha)$ quantile d'une loi de Fisher avec r et $(n - p)$ degrés de liberté que l'on trouve dans une table de Fisher. Reprenons à ce sujet les trois exemples donné ci-dessus :

On a $r = 1$ et :

$$\frac{SC_{res}(H_0) - SC_{res}(H_1)}{SC_{res}(H_1)} \cdot \frac{n-p}{1}$$

Cette statistique est égale au carré de la statistique t_c introduite pour tester la nullité d'un paramètre dans un modèle complet avec constante. Comme on a également :

$$t_{(\alpha/2, n-p)}^2 = F_{\alpha, 1, n-p}$$

cet test du F partiel est équivalent au test de Student présenté. On a $r = p - 1$ et :

$$\frac{SC_{res}(H_0) - SC_{res}(H_1)}{SC_{res}(H_1)} \cdot \frac{n-p}{p-1}$$

Comme on a ici :

$$\hat{y}_i(H_0) = \bar{y}$$

et donc :

$$SC_{res}(H_0) = \sum (y_i - \bar{y})^2$$

On a $r = p$:

$$\frac{SC_{res}(H_0) - SC_{res}(H_1)}{SC_{res}(H_1)} \cdot \frac{n - p}{p}$$

Lorsque les deux modèles définis par H_0 et H_1 admettent une constante et que l'on définit :

$$SC_{reg}(H_0) = \sum_{i=1}^n (\hat{y}_i(H_0) - \bar{y})^2$$

$$SC_{reg}(H_1) = \sum_{i=1}^n (\hat{y}_i(H_1) - \bar{y})^2$$

ainsi que :

$$R^2(H_0) = \frac{SC_{reg}(H_0)}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$R^2(H_1) = \frac{SC_{reg}(H_1)}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

on a alors :

$$\begin{aligned} F_c &= \frac{SC_{res}(H_1) - SC_{res}(H_0)}{SC_{res}(H_1)} \cdot \frac{n - p}{r} \\ &= \frac{R^2(H_1) - R^2(H_0)}{1 - R^2(H_1)} \cdot \frac{n - p}{r} \end{aligned}$$

Remarquons que :

$$R^2(H_1) \geq R^2(H_0)$$

vu que le minimum de la fonction objectif des moindres carrées est nécessairement plus petit lorsqu'il ya plus de variables explicatives dans le modèle (et donc $SC_{res}(H_1) \leq SC_{res}(H_0)$). Ainsi, ne peut qu'augmenter lorsque l'on introduit des variables supplémentaires dans le modèle. Un test du F partiel nous dit qu'il augment (significativement)ou non.

Chapitre 3

Le modèle linéaire général

Introduction

Nous allons aborder dans ce chapitre le modèle linéaire général et nous allons voir que sous certaines hypothèses on se ramène au modèle linéaire simple ou multiple.

3.1 Le modèle

3.1.1 Définition de modèle linéaire général

Le modèle linéaire général est une généralisation de la régression au cas où p variables explicatives sont utilisées pour expliquer la variable dépendante y . La relation la plus simple est une relation linéaire, entre les variables explicatives et la variable dépendante. Le modèle linéaire général s'écrit :

$$y_i = \sum_{j=1}^p x_{ij}\beta_j + \epsilon_i, \quad (3.1)$$

où

▷ y_i est la valeur prise par la variable dépendante sur l'individu i , les y_i sont des variables aléatoires,

▷ x_{ij} représente la valeur prise par la $j^{\text{ième}}$ variable explicative sur l'individu i , les x_{ij} sont supposés non aléatoires,

▷ β_j est la $j^{\text{ième}}$ composante du coefficient de régression, les ϵ_i sont des variables aléatoires telles que :

• $E(\epsilon_i) = 0$, pour tout i ,

- $E(\epsilon_i, \epsilon_k) = 0$, pour tout $i \neq k$,
- $E(\epsilon_i^2) = \sigma_\epsilon^2$, pour tout i .

3.2 Régression avec les résidus normaux

3.2.1 Hypothèses du modèle linéaire général

Avec le modèle linéaire général, on énonce un ensemble d'hypothèses qu'il est utile d'expliciter :

- La relation entre les variables explicatives et la variable dépendante y est linéaire.
- Il n'y a ni d'erreurs de mesure, ni d'erreurs d'échantillonnage sur les variables explicatives, autrement dit les x_{ij} ne sont pas aléatoires.
- Les termes d'erreur ϵ_i sont d'espérances nulles.
- Les termes d'erreur ϵ_i sont non-corrélés.
- Tous les ϵ_i ont la même variance (homoscédasticité).

3.2.2 Données observées, et formulation matricielle

En pratique, on observe n réalisations du modèle. On peut donc écrire le modèle sous forme matricielle.

$$y = X\beta + \epsilon$$

où

- X est une matrice de constantes (non-aléatoire) de plein rang de dimension $n \times p$ des x_{ij} .
- β est un vecteur (inconnu) de \mathbb{R}^P .
- ϵ est un vecteur (inconnu) de dimension n de variables aléatoires ϵ_i .

Les hypothèses du modèle linéaire général peuvent être reformulées :

- La matrice X n'est pas aléatoire,
- La matrice X est supposée de plein rang (Dans le cas contraire, on dit qu'il y a multicolinéarité, c'est-à-dire qu'au moins une des colonnes de la matrice peut s'exprimer comme une combinaison linéaire des autres colonnes),
- $E(\epsilon) = 0$
- $var(\epsilon_i) = \sigma_\epsilon^2$.
- $cov(\epsilon_i, \epsilon_j) = 0$ (toutes les corrélations sont nulles).

3.2.3 Autre présentation du modèle linéaire général

Une présentation plus synthétique du modèle linéaire général est la suivante : soit y un vecteur aléatoire de \mathbb{R}^n tel que :

- $E(y) = X\beta$ où X est une matrice $n \times p$ et $\beta \in \mathbb{R}^p$
- $Var(y) = I\sigma_\epsilon^2$ est une matrice identité $n \times n$ et σ_ϵ^2 est un scalaire. Cette formulation est équivalente à la précédente.

3.3 Estimation du modèle

L'objectif est d'estimer β et σ_ϵ^2 . La méthode des moindres carrés consiste à minimiser en β , l'expression

$$\epsilon' \epsilon = (y - \beta X)'(y - \beta X)$$

La solution fournit l'estimateur des moindres carrés (ordinaires) $\hat{\beta}$ de β , qui se note

$$\hat{\beta} = (X'X)^{-1}X'y$$

L'estimateur $\hat{\beta}$ est une variable aléatoire, car il dépend de y qui est une variable aléatoire.

Théorème : L'estimateur $\hat{\beta} = (X'X)^{-1}X'y$ est sans biais.

Démonstration

Comme

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + \epsilon) \\ &= (X'X)^{-1}X'(X\beta) + (X'X)^{-1}X'\epsilon \\ &= \beta + (X'X)^{-1}X'\epsilon \end{aligned}$$

On a :

$$\begin{aligned} E(\hat{\beta}) &= E(\beta + (X'X)^{-1}X'\epsilon) \\ &= \beta + (X'X)^{-1}X'E(\epsilon) \\ &= \beta \end{aligned}$$

Théorème : $Var(\hat{\beta}) = \sigma_\epsilon^2(X'X)^{-1}$

Démonstration : Comme

$$\begin{aligned} Var(\hat{\beta}) &= Var(\sigma_\epsilon^2(X'X)^{-1}) \\ &= (X'X)^{-1}X'Var(\epsilon)X(X'X)^{-1} \\ &= (X'X)^{-1}X'I\sigma_\epsilon^2X(X'X)^{-1} \\ &= \sigma_\epsilon^2(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma_\epsilon^2(X'X)^{-1} \end{aligned}$$

On appelle erreur standard ou erreur type de $\hat{\beta}_j$, l'écart type de cet estimateur.

$$\sigma(\hat{\beta}_j) = \sqrt{Var(\hat{\beta})} = \sigma_\epsilon \sqrt{[X'X]_{jj}}$$

où $[X'X]_{jj}$ est le scalaire correspondant à la $j^{\text{ième}}$ ligne et la $j^{\text{ième}}$ colonne de la matrice $(X'X)$

Théorème : (de Gauss Markov) l'estimateur $\hat{\beta} = (X'X)^{-1}X'y$ est le meilleur (au sens de la plus petite variance) estimateur linéaire en y sans biais de β .

Démonstration

Soit $\hat{\beta} = Cy$ est un estimateur linéaire. En posant $B = C - (X'X)^{-1}X'$.

on a $\beta^* = (B + (X'X)^{-1}X')y$. Comme :

$$\begin{aligned} E(\beta^*) &= E((B + (X'X)^{-1}X')(X\beta + \epsilon)) \\ &= B + (X'X)^{-1}X'(X\beta) \\ &= BX\beta + X \end{aligned}$$

pour que β^* soit sans biais, il faut que :

$$BX\beta + \beta = \beta$$

c'est-à-dire que

$$BX\beta = 0$$

pour tout $\beta \in \mathbb{R}^p$. Donc,

$$BX = 0$$

Calculons maintenant la variance de β^* :

$$\begin{aligned} \text{Var}(\beta^*) &= (B + (X'X)^{-1}X')\text{Var}(y)(B + (X'X)^{-1}X')' \\ &= (B + (X'X)^{-1}X')I\sigma_\epsilon^2(B + (X'X)^{-1}X')' \\ &= (BB' + BX(X'X)^{-1} + (X'X^{-1})X'B' + (X'X)^{-1})\sigma_\epsilon^2, \end{aligned}$$

On a finalement

$$\text{Var}(\beta^*) = (BB' + (X'X)^{-1})\sigma_\epsilon^2$$

Comme X est connu, il suffira d'estimer σ_ϵ^2 pour estimer la variance de $\hat{\beta}$. Le vecteur des termes d'erreur ϵ peut être estimé par :

$$e = \hat{\epsilon} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = P \perp_X Y$$

Notre objectif est de calculer $E(e'e)$. Pour obtenir les résultats, on utilisera le lemme général suivant.

Lemme :

Soit un vecteur u composé de n variables aléatoires d'espérance nulle, et tel que $\text{Var}(u) = I\sigma_u^2$ et A est une matrice symétrique non-aléatoire, alors

$$E(u' Au) = \sigma_u^2 \text{trace}(A)$$

Démonstration

$$E(u' Au) = \sum_{i=1}^n a_{ii} E(u_i^2) + \sum_{i=1}^n \sum_{j=1}^n a_{ij} E(u_i u_j)$$

Or $E(u_i u_j) = 0$ quand $i \neq j$. Donc,

$$E(u' Au) = \sum_{i=1}^n a_{ii} E(u_i^2) = \sum_{i=1}^n a_{ii} \sigma_u^2 = \sigma_u^2 \text{trace}(A)$$

Grâce au lemme, on peut calculer l'espérance de $e'e$.

Théorème : soit $e = y - X\hat{\beta}$, alors

$$E(e'e) = (n - p)\sigma_u^2.$$

On peut également écrire :

$$e = (I - P_X)y,$$

Où P_X est un projecteur (c'est-à-dire est une matrice idempotente) sur le sous-espace engendré par les colonnes de X .

$$P_X = X(X'X)^{-1}X'$$

Donc

$$\begin{aligned} e &= (I - P_X)y \\ &= (I - P_X)(X\beta + \epsilon) \\ &= X\beta + P_X X\beta + \epsilon - P_X \epsilon \end{aligned}$$

Or $P_X = X(X'X)^{-1}X'$, ce qui donne

$$e = \epsilon - P_X \epsilon = (I - P_X)\epsilon$$

On obtient

$$e'e = \epsilon'(I - P_X)'(I - P_X)\epsilon$$

et comme $(I - P_X)\epsilon$ est symétrique et idempotente, on a

$$E(e'e) = \sigma_\epsilon^2 \text{Trace}(I) - \sigma_\epsilon^2 \text{Trace}(P_X)$$

Or $\text{Trace}(I) = n$ et $\text{Trace}(P_X) = p$, car la trace d'une matrice idempotente est égale à son rang. Donc

$$E(e'e) = \sigma_\epsilon^2 n - \sigma_\epsilon^2 p = (n - p)\sigma_\epsilon^2$$

le théorème précédent nous permet de construire un estimateur sans biais pour σ_ϵ^2 .

$$\hat{\sigma}_\epsilon^2 = \frac{e'e}{n - p}$$

La quantité $(n - p)$ est appelée nombre de degrés de liberté, est le rang de $(I - P_X)$.

3.3.1 Estimateurs du maximum de vraisemblance

Une autre approche consiste à faire une hypothèse sur la distribution de probabilité de ϵ . On suppose que les ϵ_i sont des variables aléatoires indépendantes ayant des distributions normales de moyennes nulles et de variance σ_ϵ^2 . On peut donc écrire que le vecteur ϵ a une distribution multinormale :

$$\epsilon \sim N(0, I\sigma_\epsilon^2)$$

et, comme $y = X\beta + \epsilon$,

$$y \sim N(X\beta, I\sigma_\epsilon^2)$$

et donc

$$y - X\beta \sim N(0, I\sigma_\epsilon^2)$$

on a

$$\begin{aligned} f_y(u) &= \frac{1}{(2\pi)^{n/2} |I\sigma_\epsilon^2|^{1/2}} \exp\left[-\frac{1}{2\sigma_\epsilon^2 (u - X\beta)' I^{-1} (u - X\beta)}\right] \\ &= \frac{1}{2\pi\sigma_\epsilon^2} \exp\left[-\frac{1}{2\sigma_\epsilon^2 (u - X\beta)' I^{-1} (u - X\beta)}\right] \end{aligned}$$

On se trouve dans un problème paramétrique classique. Comme y et X sont observés, on va estimer les paramètres β et σ_ϵ^2 .

La méthode du maximum de vraisemblance consiste à estimer le paramètre par l'estimateur qui maximise la densité pour les données observées. La fonction de vraisemblance s'écrit :

$$L(\beta, \sigma_\epsilon^2) = \frac{1}{(2\pi\sigma_\epsilon^2)^{n/2}} \exp\left[-\frac{(y - X\beta)' (y - X\beta)}{2\sigma_\epsilon^2}\right]$$

Il est souvent plus facile (et c'est le cas ici) de chercher à maximiser le logarithme de la fonction de vraisemblance (le résultat sera le même) plutôt que la fonction elle-même. Le logarithme de la vraisemblance vaut :

$$l(\beta, \sigma_\epsilon^2) = \log L(\beta, \sigma_\epsilon^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_\epsilon^2) - \frac{(y - X\beta)' (y - X\beta)}{2\sigma_\epsilon^2}$$

On obtient le maximum en annulant les dérivées partielles par rapport aux paramètres.

On obtient

$$\frac{\delta l(\beta, \sigma_\epsilon^2)}{\delta \beta} = \frac{X'y - X'X\beta}{\sigma_\epsilon^2}$$

et

$$\frac{\delta l(\beta, \sigma_\epsilon^2)}{\delta \sigma_\epsilon^2} = -\frac{n}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4} (y - X\beta)' (y - X\beta) = 0$$

L'estimateur du maximum de vraisemblance de σ_ϵ^2 est donné par :

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n} (y - X\hat{\beta})' (y - X\hat{\beta}) = \frac{e'e}{n}$$

L'estimateur $\hat{\sigma}_\epsilon^2$ est biaisé.

3.3.2 Propriétés des estimateurs du maximum de vraisemblance

Rappelons quelques propriétés des estimateurs :

- Un estimateur $\hat{\theta}$ d'un paramètre θ est sans biais, si $E(\hat{\theta}) = \theta$ pour toute valeur de θ .
- Un estimateur est efficace ou de variance minimum si sa variance est plus petite ou égale que celles de tous les estimateurs du paramètre.
- Un estimateur $\hat{\theta}$ est convergent, s'il converge en probabilité vers le paramètre à estimer, c'est-à-dire :

$$\lim Pr(|\hat{\theta} - \theta| > \epsilon) = 0,$$

où ϵ est une quantité arbitrairement petite.

- Une statistique est exhaustive si elle épuise toute l'information relative au paramètre. La méthode du maximum de vraisemblance fournit des estimateurs ayant les propriétés suivantes :
- S'il existe une statistique exhaustive, alors l'estimateur du maximum de vraisemblance en dépend.
- Si $\hat{\theta}$ est un estimateur du maximum de vraisemblance de alors $f(\hat{\theta})$ est l'estimateur du maximum de vraisemblance de $f(\theta)$.
- Si l'estimateur du maximum de vraisemblance $\hat{\theta}$ admet une solution unique, alors cet estimateur est convergent et asymptotiquement efficace du paramètre. De plus, cet estimateur converge en loi vers une normale.

Cependant, l'estimateur du maximum de vraisemblance n'est pas nécessairement sans biais. L'estimateur du maximum de vraisemblance de σ_ϵ^2 est en effet biaisé.

3.3.3 Distribution de probabilité des estimateurs

Dans le modèle linéaire général avec des termes d'erreur normaux, on a

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \epsilon) = \beta + (X'X)^{-1}X'\epsilon$$

Donc, $\hat{\beta}$ est une combinaison linéaire de variables aléatoires normales i.i.d. Or une combinaison linéaire de variables normales indépendantes est aussi une variable normale. Donc

$$\hat{\beta} \sim N(\beta, X'X)^{-1}\sigma_\epsilon^2).$$

Lemme : Soient u un vecteur aléatoire de distribution normale de \mathbb{R}^n , de moyennes nulles et de variance I , et Γ une matrice orthogonale de dimension $n \times n$, alors

$$\Gamma u \sim N(0, \Gamma)$$

et

$$\Gamma' u \sim N(0, \Gamma)$$

Théorème :

Soit un vecteur aléatoire u de distribution normale, de moyennes nulles et de variance I . Si P est symétrique, idempotente et de rang p , alors $u'Pu$ est une variable χ_p^2 à p degrés de liberté.

Démonstration :

La matrice P admet une décomposition en valeurs propres et vecteurs propres. si A représente la matrice diagonale ayant les valeurs propres λ_i de P sur sa diagonale, et Γ est une matrice orthogonale contenant les n vecteurs propres de P , alors on peut écrire :

$$P = \Gamma A \Gamma'$$

La forme quadratique peut s'écrire

$$u'Pu = u'\Gamma A \Gamma' u = v'Av$$

où $v = \Gamma' u$, comme P est idempotente et de rang p , P a p valeurs propres égales à 1 et $n - p$ valeurs propres égales à 0. La forme quadratique

$$v'Av = \sum_{i=1}^n v_i^2 \lambda_i = \sum_{i=1, \lambda_i=1}^n v_i^2$$

peut donc s'écrire comme une somme de p carrés de variables aléatoires normales centrées réduites indépendantes, ce qui définit une χ_p^2 .

Dans le modèle linéaire général avec des termes d'erreur normaux,

$$(\hat{\beta} - \beta)' \frac{X'X}{\sigma_\epsilon^2} (\hat{\beta} - \beta) \sim \chi_p^2$$

et

$$\begin{aligned} \hat{\beta} - \beta &= (X'X)^{-1} X' y - \beta \\ &= (X'X)^{-1} X' (X\beta + \epsilon) - \beta \\ &= \beta + (X'X)^{-1} X' \epsilon - \beta \\ &= (X'X)^{-1} X' \epsilon \end{aligned}$$

Donc

$$\begin{aligned} (\hat{\beta} - \beta)' \frac{X'X}{\sigma_\epsilon^2} (\hat{\beta} - \beta) &= \epsilon' X (X'X)^{-1} \frac{X'X}{\sigma_\epsilon^2} (X'X)^{-1} X' \frac{\epsilon}{\sigma_\epsilon} \\ &= \frac{\epsilon'}{\sigma_\epsilon} X (X'X)^{-1} X' \frac{\epsilon}{\sigma_\epsilon} \end{aligned}$$

Comme la matrice $X(X'X)^{-1}X'$ est symétrique idempotente et de rang p et que $\frac{\epsilon'}{\sigma_\epsilon}$ est un vecteur multinormal non-corrélé.

Dans le modèle linéaire général avec des termes d'erreur normaux,

$$\frac{e'e}{\sigma_\epsilon^2} \sim \chi_{n-p}^2$$

En effet,

$$e = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = P_{\perp X}\epsilon$$

où $P_{\perp X} = I - X(X'X)^{-1}X'$. Or $P_{\perp X}$ est une matrice idempotente de rang $n - p$. On obtient

$$\frac{e'e}{\sigma_\epsilon^2} = \frac{\epsilon'}{\sigma_\epsilon} P_{\perp X}' P_{\perp X} \frac{\epsilon}{\sigma_\epsilon} = \frac{\epsilon'}{\sigma_\epsilon} P_{\perp X} \frac{\epsilon}{\sigma_\epsilon} \sim \chi_{n-p}^2$$

L'indépendance de $\hat{\beta}$ et $\hat{\sigma}_\epsilon^2$ se montre grâce au résultat suivant :

Théorème :

Soient les matrices $B(p \times n)$ et $A(n \times n)$ et un vecteur aléatoire $u \sim N(\mu, \sigma_u^2)$, alors les p formes linéaires Bu sont indépendantes de la forme quadratique $u' Au$ si $BA = 0$.

corollaire

Dans le modèle linéaire avec des termes d'erreur normaux,

- $\hat{\beta}$ est indépendant de $e'e$.
- $\hat{\beta}$ est indépendant de $\hat{\sigma}_\epsilon^2 = \frac{e'e}{n-p}$

En effet,

$e'e = \epsilon' P_{\perp X} \epsilon$ où $P_{\perp X} = I - X(X'X)^{-1}X'$ et

$(\hat{\beta} - \beta) = (X'X)^{-1}X'\epsilon$ or $(X'X)^{-1}X'P_{\perp X} = 0$

qui implique directement le corollaire.

3.3.4 Synthèse des résultats

En résumé, si $y = X\beta + \epsilon$ est un modèle linéaire général avec des termes d'erreur normaux :

- $\hat{\beta}$ et $\hat{\sigma}_\epsilon^2$ sont convergents, exhaustifs, efficaces et sans biais,
- $\hat{\beta}$ et $\hat{\sigma}_\epsilon^2$ sont indépendants,
- $\hat{\beta} \sim N(\beta, (X'X)^{-1}\sigma_\epsilon^2)$.
- $(\hat{\beta} - \beta)' \frac{X'X}{\sigma_\epsilon^2} (\hat{\beta} - \beta) \sim \chi_p^2$.

3.3.5 Intervalle de confiance sur un coefficient de régression

On a :

$$\hat{\beta}_j \sim N(\beta_j, [(X'X)^{-1}]_{jj}\sigma_\epsilon^2),$$

où $[(X'X)^{-1}]_{jj}$ est la composante correspondant à la j ème ligne et à la j ème colonne de la matrice $(X'X)^{-1}$ on obtient donc que

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{[(X'X)^{-1}]_{jj}\sigma_\epsilon^2}} \sim N(0, 1).$$

On a également que

$$\frac{n - p\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} = \frac{e'e}{\sigma_\epsilon^2} \sim \chi_{n-p}^2.$$

De plus $\hat{\beta}_j$ est indépendante de $\hat{\sigma}_\epsilon^2$.

La quantité

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{[(X'X)^{-1}]_{jj}\sigma_\epsilon^2}} \sqrt{\frac{(n-p)\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} / (n-p)},$$

peut donc trouver comme un rapport d'une normale centrée réduite sur la racine carrée d'une khi-carrée divisée par son nombre de degrés de liberté. La variable normale est indépendante de la chi-carré, ce qui définit une variable de Student à $n - p$ degrés de liberté. En simplifiant, on obtient que

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{[(X'X)^{-1}]_{jj}\sigma_\epsilon^2}} \sim t_{n-p},$$

où t_{n-p} est une variable aléatoire de Student à $n - p$ degrés de liberté, ce qui implique que

$$Pr \left(-t_{1-\alpha/2, n-p} \leq \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_\epsilon \sqrt{[(X'X)^{-1}]_{jj}}} \leq t_{1-\alpha/2, n-p} \right) = 1 - \alpha,$$

où $t_{1-\alpha/2, n-p}$ est le quantile d'ordre $1 - \alpha$ d'une variable aléatoire de Student à $n-p$ degrés de liberté. Après quelques calculs, on a

$$Pr \left(\hat{\beta}_j - t_{1-\alpha/2, n-p} \hat{\sigma}_\epsilon \sqrt{[(X'X)^{-1}]_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{1-\alpha/2, n-p} \hat{\sigma}_\epsilon \sqrt{[(X'X)^{-1}]_{jj}} \right) = 1 - \alpha,$$

ce qui définit l'intervalle de confiance de niveau α , donnée par :

$$IC(1 - \alpha) = \left[\hat{\beta}_j - t_{1-\alpha/2, n-p} \hat{\sigma}_\epsilon \sqrt{[(X'X)^{-1}]_{jj}}; \hat{\beta}_j + t_{1-\alpha/2, n-p} \hat{\sigma}_\epsilon \sqrt{[(X'X)^{-1}]_{jj}} \right].$$

3.4 Test d'un seul coefficient de régression

3.4.1 Construction du test

Le problème consiste à tester la valeur d'un coefficient de régression particulier

$$H_0 : \beta_j = \beta_{j0}$$

$$H_1 : \beta_j \neq \beta_{j0}$$

Sous H_0 , $\hat{\beta}_j \sim \mathbb{N}(\beta_{j0}, \sigma^2(\beta_{j0}))$ est l'erreur-type de $\hat{\beta}_j$:

$$\sigma^2(\beta_{j0}) = [(X'X)^{-1} \sigma_\epsilon^2]_{jj}.$$

est simplement la composante correspondante à la $j^{\text{ième}}$ ligne et la $j^{\text{ième}}$ colonne de $Var(\hat{\beta}) = (X'X)^{-1} \sigma_\epsilon^2$. On peut donc estimer $\sigma^2(\beta_{j0})$ par

$$\hat{\sigma}^2(\beta_{j0}) = [(X'X)^{-1} \hat{\sigma}_\epsilon^2]_{jj}.$$

Comme $\hat{\sigma}^2$, $\hat{\beta}_j$ sont indépendants, et que

$$\frac{(n-p)\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} \sim \chi_{n-p}^2.$$

on obtient Sous H_0 la statistique

$$t = \frac{\frac{\hat{\beta}_j - \beta_{j0}}{\sigma(\hat{\beta}_j)}}{\sqrt{(n-p)\hat{\sigma}_\epsilon^2/(n-p)}} = \frac{\hat{\beta}_j - \beta_{j0}}{\sigma(\hat{\beta}_j)}$$

a donc une distribution de Student à $n-p$ degrés de liberté. On rejette H_0 si

$$|t| > t_{1-\alpha/2, n-p}$$

où $t_{1-\alpha/2, n-p}$ représente le quantile d'ordre $\alpha/2$ d'une variable aléatoire de Student à $n-p$ degrés de liberté.

3.4.2 Modèle linéaire avec uniquement une constante

Le test d'hypothèse sur la moyenne peut être vu comme un cas particulier d'un test sur le coefficient de régression. Soit $y_1, \dots, y_i, \dots, y_n$ une suite de n variables aléatoires indépendantes, telles que $y_i \sim N(\mu, \sigma^2)$, ce qui peut s'écrire sous la forme d'un modèle linéaire.

$$y_i = \mu + \epsilon_i, i = 1, \dots, n.$$

avec $\epsilon_i \sim N(0, \sigma^2)$, et les ϵ_i indépendants. Sous forme matricielle, on écrit

$$y = 1\mu + \epsilon.$$

où 1 est un vecteur colonne de \mathbb{R}^n composé de uns, et $\epsilon \sim N(0, I\sigma^2)$. On obtient alors

$$\hat{\mu} = (1'1)^{-1}1'y = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}$$

Les valeurs ajustées valent $y_i^* = \bar{y}$ et les résidus $e_i = y_i - \bar{y}$. L'estimateur de σ^2 vaut

$$\hat{\sigma}^2 = \frac{e'e}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

$$\text{Var}(\hat{\mu}) = (1'1)^{-1}\sigma^2 = \frac{\sigma^2}{n},$$

$$\text{Var}(\hat{\mu}) = (1'1)^{-1}\hat{\sigma}^2 = \frac{\hat{\sigma}^2}{n}.$$

$\hat{\mu}$ et $\hat{\sigma}^2$ sont indépendants, de plus on a,

$$\hat{\mu} \sim \mathbb{N}\left(\mu, (1'1)^{-1}\hat{\sigma}^2\right) = \mathbb{N}\left(\mu, \frac{\hat{\sigma}^2}{n}\right)$$

Donc,

$$d = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim \mathbb{N}(0, 1)$$

En outre, on peut écrire

$$K = \frac{(n-1)\hat{\sigma}}{\sigma^2} = \frac{\epsilon' P_c \epsilon}{\sigma},$$

où P_c la matrice idempotente de rang $n-1$ qui centre les valeurs :

$$P_c = I - \frac{11'}{n} \begin{pmatrix} 1 - 1/n & -1/n & -1/n & \dots & -1/n \\ -1/n & 1 - 1/n & -1/n & \dots & -1/n \\ -1/n & -1/n & 1 - 1/n & \dots & -1/n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & -1/n & \dots & 1 - 1/n \end{pmatrix}.$$

Les variables aléatoires d et K sont indépendantes. Donc

$$\frac{d}{\sqrt{K/(n-1)}} = \frac{\frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)\hat{\sigma}^2}{\sigma^2}/n-1}} = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\hat{\sigma}} \sim t_{n-1}$$

Ce résultat fondamental permet de mener une inférence sur la moyenne.

3.5 Tests de Wald sur les coefficients de régression

3.5.1 Test général d'une contrainte linéaire

L'objectif est de tester une hypothèse linéaire assez générale sur les coefficients de régression du type :

$$H_0 : R\beta = r$$

contre l'hypothèse alternative

$$H_1 : R\beta \neq r$$

où R est une matrice $q \times p$, $q \leq p$, et r un vecteur colonne de dimension q . En outre on suppose que R est de rang q .

- Le test $H_0 : \beta_j = c$ s'obtient en prenant $R = (0 \dots 0 1 0 \dots 0)$, $r = c$.
- Le test $H_0 : \beta_j = c$ s'obtient en prenant $R = I_p$ (matrice identité de dimension p) et r est un vecteur de 0 de dimension p .

Sous l'hypothèse H_0 ,

$$\begin{aligned} R\hat{\beta} &= R(X'X)^{-1}X'y - r \\ &= R(X'X)^{-1}X'(X\beta + \epsilon) - r \\ &= R\beta + R(X'X)^{-1}X'\epsilon - r \\ &= R(X'X)^{-1}X'\epsilon. \end{aligned}$$

De plus,

$$\text{Var}(R\hat{\beta} - r) = \text{Var}(R\hat{\beta}) = R\text{Var}(R\hat{\beta})R' = \sigma_\epsilon^2 R(X'X)^{-1}R'.$$

Examinons maintenant la forme quadratique :

$$(R\hat{\beta} - r)' \text{Var}(R\hat{\beta})^{-1} (R\hat{\beta} - r) = \frac{1}{\sigma_\epsilon^2} \epsilon' W \epsilon.$$

où

$$W = X(X'X)^{-1}R' \{R(X'X)^{-1}R'\}^{-1} R(X'X)^{-1}X,$$

On vérifie facilement que W est une matrice idempotente, symétrique de rang q , on obtient donc que

$$\frac{1}{\sigma_\epsilon^2} \epsilon' W \epsilon \sim \chi_q^2.$$

et donc

$$(R\hat{\beta} - r)' Var(R\hat{\beta})^{-1} (R\hat{\beta} - r) = \frac{1}{\sigma_\epsilon^2} (R\hat{\beta} - r)' \{R(X'X)^{-1}R'\}^{-1} (R\hat{\beta} - r) \sim \chi_q^2$$

Si la forme quadratique est grande, on soupçonne H_0 d'être faux. Cependant, on ne peut réaliser directement un test khi-deux dépend de σ^2 qui est inconnu. On sait par ailleurs que

$$\frac{1}{\sigma_\epsilon^2} e' e \sim \chi_{n-p}^2$$

De plus, comme

$$e' e = \epsilon' (I - P_x) \epsilon$$

et que $(I - P_x)W = 0$, on a l'indépendance de $e' e / \sigma_\epsilon^2$ et de $\epsilon' W \epsilon$. On peut construire une statistique de test

$$F_c = \frac{(R\hat{\beta} - r)' \{R(X'X)^{-1}R'\}^{-1} (R\hat{\beta} - r) \frac{1}{q}}{e' e \frac{1}{n-p}}$$

Sous H_0 , le numérateur et le dénominateur de F_c sont indépendants, et ont, à une constante près, une distribution chi^2 . La statistique de test F_c a donc une distribution de Fisher à q et $n - p$ degrés de liberté. Donc, en notant α l'erreur de première espèce, on rejette l'hypothèse où $F_{(1-\alpha, q, 1-\alpha)}$ est le quantile d'ordre $1 - \alpha$ d'une variable aléatoire de Fisher à q et $n - p$ degrés de liberté.

3.5.2 Test global des coefficients de régression

Un cas particulier du problème précédent consiste à tester la nullité de tous les coefficients de régression (excepté la constante). On suppose que la première colonne de la matrice X est composée de uns, c'est-à-dire que $x_{i1} = 1$ pour tout $i = 1, \dots, n$. La matrice R est de dimension $(p - 1) \times p$ et vaut :

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Alors

$$R\beta = \tilde{\beta} = (\beta_2, \dots, \beta_p)'$$

le test devient alors

$$\begin{aligned} H_0 &: \beta_j \text{ pour tout } j = 1 \dots n \\ H_1 &= \text{au moins un des } \beta_j \neq j \end{aligned}$$

ce qui peut aussi s'écrire

$$\begin{aligned} H_0 &: R\beta = 0 \\ H_1 &= R\beta \neq 0 \end{aligned}$$

$$\begin{aligned} H_0 &: \tilde{\beta} = 0 \\ H_1 &= \tilde{\beta} \neq 0 \end{aligned}$$

Théorème 3.1. $\{R(X'X)^{-1}R\} = \tilde{X}'P_c\tilde{X}$

où

$$P_c = \mathbb{I} - \frac{11'}{n},$$

Σ est la matrice variance-covariance et \tilde{X} est la matrice de dimension $n \times (p-1)$ composée des $(p-1)$ dernières colonnes de X .

Démonstration

On peut écrire

$$X'X = \begin{pmatrix} n & u' \\ u & Z \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \dots & \sum_{i=1}^n x_{i2} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2}^2 & \sum_{i=1}^n x_{i2}x_{i3} & \dots & \sum_{i=1}^n x_{i2}x_{ip} \\ \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i2}x_{i3} & \sum_{i=1}^n x_{i3}^2 & \dots & \sum_{i=1}^n x_{i3}x_{ip} \\ \cdot & \cdot & \cdot & \cdot & \sum_{i=1}^n x_{i2} \\ \cdot & \cdot & \cdot & \cdot & \sum_{i=1}^n x_{i2} \\ \cdot & \cdot & \cdot & \cdot & \sum_{i=1}^n x_{i2} \\ \sum_{i=1}^n x_{ip} & \sum_{i=1}^n x_{i2}x_{ip} & \sum_{i=1}^n x_{i3}x_{ip} & \dots & \sum_{i=1}^n x_{ip}^2 \end{pmatrix}$$

où

$$u = \left(\sum_{i=1}^n x_{i2} \quad \sum_{i=1}^n x_{i3} \quad \dots \quad \sum_{i=1}^n x_{ip} \right)'$$

et

$$Z = \begin{pmatrix} \sum_{i=1}^n x_{i2}^2 & \sum_{i=1}^n x_{i2}x_{i3} & \dots & \sum_{i=1}^n x_{i2}x_{ip} \\ \sum_{i=1}^n x_{i2}x_{i3} & \sum_{i=1}^n x_{i3}^2 & \dots & \sum_{i=1}^n x_{i3}x_{ip} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{i=1}^n x_{i2}x_{ip} & \sum_{i=1}^n x_{i3}x_{ip} & \dots & \sum_{i=1}^n x_{ip}^2 \end{pmatrix}$$

Par la méthode d'inversion par partie, on a

$$(X'X)^{-1} = \begin{pmatrix} n & u' \\ u & Z \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{n} + \frac{1}{n}u'Qu & -\frac{1}{n}u'Q \\ -\frac{1}{n}uQ & Q \end{pmatrix}$$

où

$$Q = \left(Z - \frac{1}{n}u'u \right)^{-1}$$

De plus,

$$\left(R(X'X)^{-1}R' \right)^{-1} = Q^{-1} = Z - \frac{1}{n}u'u = n \sum$$

où \sum est la matrice variance-covariance. la statistique de test devient :

$$F_c = \frac{SC_{reg/(1-p)}}{SC_{res}/(n-p)}$$

Source de variation	Degré de liberté	Somme des carrés	Moyenne des carrés
Régression	$p - 1$	SC_{reg}	$CM_{reg} = \frac{SC_{reg}}{p-1}$
Résiduelle	$n - p$	SC_{res}	$CM_{res} = \frac{SC_{res}}{n-p}$
Total	$n - 1$	SC_{tot}	$CM_{tot} = \frac{SC_{tot}}{n-1}$

TAB. 3.1 – Tableau de ANOVA de régression général

ce qui peut également s'écrire

$$F_c = \frac{SC_{tot} - SC_{res/(1-p)}}{SC_{res}/(n-p)}$$

Ce test est généralement résumé au moyen du tableau d'analyse de la variance : La règle de décision consiste à rejeter H_0 si $F_c > F_{1-\alpha, p-1, n-p}$ où $F_{1-\alpha, p-1, n-p}$ est le quantile d'ordre $(1 - \alpha)$ d'une variable aléatoire de Fischer à $(p - 1)$ et $(n - p)$ degrés de liberté.

3.5.3 Test de Fisher sur un coefficient de régression

Il est également possible de réaliser un test de Fisher pour un coefficient de régression au moyen du test de Fisher :

$$H_0 : \beta_j = \beta_{j0}$$

$$H_1 = \beta_j \neq \beta_{j0}$$

Pour ce faire, on prend

- $q = 1$,
- $R = (0 \dots 1 \dots 0)$,
- $r = \beta_{j0}$,

on obtient

- $\mathbf{R}\hat{\beta} - r = \hat{\beta}_j - \beta_{j0}$
- $R(X'X)^{-1}R' = [(X'X)^{-1}]_{jj}$

d'où

$$F_c = \frac{(\hat{\beta}_j - \beta_{j0})^2}{[(X'X)^{-1}]_{jj} \hat{\sigma}_\epsilon^2}$$

Sous H_0 , F_c suit une distribution de Fischer à 1 et $(n - p)$ degrés de liberté. On rejette donc H_0 si

$$F_c > F_{(1-\alpha, 1, n-p)}$$

où $F_{(1-\alpha, 1, n-p)}$ est le quantile d'ordre $(1 - \alpha)$ d'une variable aléatoire de Fisher à 1 et $(n - p)$ degrés de liberté. En effet le carré d'une variable de Student à $(n - p)$ degrés de liberté est une variable de Fisher à 1 et $(n - p)$ degrés de liberté.

3.5.4 R-carré et R-carré ajusté

Dans une régression multivariée, le R-carré est défini par

$$R^2 = \frac{\sum_{i=1}^n (y_i^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{SC_{regr}}{SC_{tot}} = 1 - \frac{SC_{res}}{SC_{tot}}.$$

Le R^2 est la proportion de variance de la variable y expliquée par la régression. Le R^2 est une statistique biaisée. On applique en général une correction afin de diminuer le biais du R^2 , ce qui permet d'obtenir le R^2 ajusté :

$$R_{adj}^2 = 1 - (1 - R^2) \frac{1 - n}{n - p - 1}.$$

3.5.5 Prédiction ponctuelle d'une valeur

Cas général

Une fois le coefficient de régression estimé, il est possible de prédire une valeur pour y en fonction d'un ensemble de nouvelles variables explicatives

$$x_j = (x_{j1} \dots x_{jp})$$

La prédiction vient simplement et vaut :

$$\hat{y}_j = (x_{j1} \dots x_{jp}) \hat{\beta}$$

Le prédicteur peut également s'écrire

$$\begin{aligned} \hat{y}_j &= x_j \hat{\beta} \\ &= x_j (X'X)^{-1} X' y \\ &= x_j (X'X)^{-1} X' (X\beta + \epsilon) \\ &= x_j \beta + x_j (X'X)^{-1} X' \epsilon \end{aligned}$$

Comme la vraie valeur vaut

$$y_j = x_j\beta + \epsilon_j$$

l'erreur de prévision est

$$\hat{y}_j - y_j = x_j(X'X)^{-1}X'\epsilon - \epsilon_j$$

L'espérance de l'erreur de prédiction est nulle, en effet

$$E(\hat{y}_j - y_j) = E\{x_j(X'X)^{-1}X'\epsilon - \epsilon_j\} = x_j(X'X)^{-1}X'E(\epsilon) - E(\epsilon_j) = 0$$

Comme la valeur prédite se réfère à une nouvelle observation,

$$E(\epsilon_i, \epsilon_j) = 0$$

et donc

$$\begin{aligned} \text{var}(\hat{y}_j - y_j) &= \text{var}(\{x_j(X'X)^{-1}X'\epsilon\}) + \text{var}\{\epsilon_j\} \\ &= x_j(X'X)^{-1}X'\sigma_\epsilon^2 + x_j(X'X)^{-1}X'x'_j + \sigma_\epsilon^2 \\ &= \sigma_\epsilon^2\{x_j(X'X)^{-1}X' + 1\} \end{aligned}$$

On constate que la variance se décompose en deux parties. La première partie est due à l'instabilité des coefficients de régression, c'est-à-dire la dispersion de $\hat{\beta}$, et la seconde partie est due à l'erreur inconnue ϵ_j . On estime la variance simplement par :

$$\text{Var}(\hat{y}_j - y_j) = \sigma_\epsilon^2\{x_j(X'X)^{-1}X' + 1\}$$

où σ_ϵ^2 . Enfin, il est possible de construire un intervalle de confiance pour la prévision :

$$IC = (1 - \alpha) = \left[\hat{y}_j - t_{1-\alpha/2, n-p} \sqrt{(\hat{y}_i - y_i)}; \hat{y}_j + t_{1+\alpha/2, n-p} \sqrt{(\hat{y}_i - y_i)} \right]$$

Chapitre 4

Application

Le tableau 4.2 nous donne des caractéristiques calculées pour $n=31$ types des véhicules. Notre but est d'expliquer la consommation y_i en fonction de quatre variables explicatives, à savoir le prix x^1 , le nombre de cylindrées x^2 , la puissance x^3 et le poids x^4 données dans le tableau 4.2. Le modèle complet avec $p=5$ paramètres est ainsi donné par :

$$y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \epsilon_i = X\beta + \epsilon. \quad i = 1, \dots, n.$$

La matrice X est donc constitué d'une colonne de 1, ainsi que les données relatives à ces quatre variables explicatives.

4.1 Statistique descriptive

Le tableau 4.1 représente la moyenne et l'écart-type :

Variable	Observation	Minimum	Maximum	Moyenne	Ecart-type
y	31	5,8	21,3	10,097	39,54
x^1	31	2260,000	285000,000	44149,500	59079,121
x^2	31	658,000	5987,000	2135,267	1148,150
x^3	31	29,000	325,000	99,267	69,448
x^4	31	730,000	1276,000	1276,333	348,625

TAB. 4.1 – Résumé des statistique.

4.2 Estimation des paramètres de modèle par la méthode des moindres carrés

Le modèle estimé en utilisant le logiciel XLSTATISTICA est :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x^1 + \hat{\beta}_2 x^2 + \hat{\beta}_3 x^3 + \hat{\beta}_4 x^4,$$

avec :

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 3,32804 \\ 0.00002 \\ -0.000198958 \\ 0.00237304 \\ 0.00237304 \end{pmatrix}$$

on obtient alors :

$$\hat{y} = 3,32804 + 0.00002x_1 - 0.000198958x_2 + 0,0426412x_3 + 0.00237304x_4.$$

Obs	Type de voiture	Prix(x^1)	Cylindrées (x^2)	Puissance (x^3)	Poids (x^4)	Cons(y)
1	Daihatsu Cuore	11600	846	32	650	5,7
2	Suzuki Swift 1.0 GLS	12490	993	39	790	5,8
3	Flat Panda Mambo	10450	899	29	730	6,1
4	VW Polo 1.4 60	17140	1390	44	955	6,5
5	Opel Corsa 1.2i Eco	14825	1195	33	895	6,8
6	Subaru Vivio 4WD	13730	658	32	740	6,8
7	Toyota Corolla	19490	1331	55	1010	7,1
8	Ferrari	285000	5474	325	1690	21,3
9	Mercedes S600	183900	5987	300	2250	18,7
10	Maserati Ghibli	92500	2789	209	1485	14,5
11	Opel Astra	25000	1597	74	1080	7,4
12	Peugeot 30XS	22350	1761	74	1100	9
13	Renault Safrane	36600	2165	101	1500	11,7
14	Seat Ibiza	22500	1983	85	1075	9,5
15	VW GOLF	31580	1984	85	1155	9,5
16	Citron XZ Volcane	28750	1998	89	1140	8,8
17	Fiat tompra	22600	1580	65	1080	9,3
18	Ford	20300	1390	54	1110	8,
19	Honda civic Jocker	19900	1396	66	1140	7,7
20	Volvo	39800	2435	106	1370	10,8
21	Ford Fiesta	19740	1242	55	940	6,6
22	Hyundai	38990	2972	107	1400	11,7
23	Lancia	50800	2958	150	1550	11,9
24	Mazda Hatchback	36200	2497	122	1330	10,8
25	Mitsubishi	31990	1998	66	1300	7,6
26	Opel Omega	47700	2496	125	1670	11,3
27	Peugeot 806	3950	1998	89	1560	10,8
28	Nissan Primera	26950	1997	92	1240	9,2
29	Seat Alhambra	36400	1984	85	1635	11,
30	Toyota	50900	2438	97	1800	12,8
31	Volvo 960 Kombi aut	493006	2473	125	1570	12,7

TAB. 4.2 – Tableau représente la consommation d'essence pour les différentes véhicules.

Prix :(Frs).

Cylindrées :(Cm^3).

Puissance :(kW).

Poids :(kg).

Consommation :(100km/l).

4.3 Tests sur les paramètres du modèle

Variabes	Coef	t(26)	Ttab
$\hat{\beta}_0$	3,32804	3,916671	0,00005781
$\hat{\beta}_1$	0.00002	2,338940	0,027297
$\hat{\beta}_2$	-0.000198958	-0,8108	0,391797
$\hat{\beta}_3$	0.00237304	2,50143	0,018993
$\hat{\beta}_4$	0.00237304	4,234432	0,00008

TAB. 4.3 – Tableau des tests de chaque paramètre estimé

au seuil de signification $\alpha = 0.05$ on a :

pour $\hat{\beta}_0$

$$H_0 : \hat{\beta}_0 = 0 \text{ vs } H_1 : \hat{\beta}_0 \neq 0$$

$$t(0, 05; 26) = 3, 916671 > Texp = 0, 00005781$$

conclusion

on accepte l'hypothèse.

pour $\hat{\beta}_1$

$$H_0 : \hat{\beta}_1 = 0 \text{ vs } H_1 : \hat{\beta}_1 \neq 0$$

$$t(0, 05; 26) = 2, 338940 > Texp = 0, 027297$$

conclusion

alors on accepte l'hypothèse.

pour $\hat{\beta}_2$

$$H_0 : \hat{\beta}_2 = 0 \text{ vs } H_1 : \hat{\beta}_2 \neq 0$$

$$t(0, 05; 26) = -0, 8108 < Texp = 0, 391797$$

conclusion

alors on rejette l'hypothèse.

pour $\hat{\beta}_3$

$$H_0 : \hat{\beta}_3 = 0 \text{ vs } H_1 : \hat{\beta}_3 \neq 0$$

$$t(0, 05; 26) = 4, 234432 > Texp = 0, 00008$$

conclusion

alors on accepte l'hypothèse.

pour $\hat{\beta}_4$

$H_0 : \hat{\beta}_4 = 0$ vs $H_1 : \hat{\beta}_4 \neq 0$

$t(0, 05; 26) = 3,916671 > Texp = 0,00005781$

conclusion

alors on accepte l'hypothèse.

Les intervalles de confiance

Au seuil de significatfion $\alpha = 0.05$ on a les intervalles de confiance pour chacun des estimateurs : pour $\hat{\beta}_0$:

IC=[1,62806 ;1,06442].

pour $\hat{\beta}_1$:

IC=[0,00197455 ;0,002157667].

pour $\hat{\beta}_2$:

IC=[0,013909 ;0,0714914].

pour $\hat{\beta}_3$:

IC=[0,0008922 ;0,00385381].

pour $\hat{\beta}_4$:

IC=[2,09147 ;4,5646].

4.3.1 Matrice de corrélation

$$\begin{pmatrix} & 11600 & 846 & 32 & 650 & 5,7 \\ 11600 & 1 & 0,899 & 0,935 & 0,639 & 0,891 \\ 846 & 0,899 & 1 & 0,962 & 0,833 & 0,938 \\ 32 & 0,935 & 0,962 & 1 & 0,775 & 0,952 \\ 650 & 0,639 & 0,833 & 0,775 & 1 & 0,858 \\ 5,7 & 0,891 & 0,938 & 0,952 & 0,858 & 1 \end{pmatrix}$$

Interprétation

D'après la matrice de corrélation on remarque qu'il existe, en général une forte corrélation entre les variables du modèle. Néanmoins, la corrélation est moins forte entre le critère prix et poids.

4.4 Test d'ajustement

Observation	31
Somme des poids	31
DDL	25
R^2	0,950
R^2 ajusté	0,942
MCE	0,732
RMCE	0,85
MAPE	7,151
DW	2,17
Cp	5,000
AIC	-4,822
SBC	2,184
PC	0,071

TAB. 4.4 – Tableau représentant les tests d'ajustement.

Interprétation :

Coefficients d'ajustement : dans tableau 4.4 sont affichées les statistiques relatives à l'ajustement du modèle de régression :

- Observations : le nombre d'observations prises en compte dans les calculs.
- Somme des poids : la somme des poids des observations prises en compte dans les calculs.
- DDL : le nombre de degrés de liberté pour le modèle retenu (correspondant à la partie erreurs).
- R^2 : le coefficient de détermination du modèle. Le R^2 s'interprète comme la proportion de la variabilité de la variable dépendante expliquée par le modèle. Plus le R^2 est proche de 1, meilleur est le modèle. L'inconvénient du R^2 est qu'il ne prend pas en compte le nombre de variables utilisées pour ajuster le modèle.
- R^2 ajusté : le coefficient de détermination ajusté du modèle. Le R^2 ajusté peut être négatif si le R^2 est voisin de zéro. Le R^2 ajusté est une correction du R^2 qui permet de prendre en compte le nombre de variables utilisées dans le modèle.
- MCE : la moyenne des carrés des erreurs (MCE).
- RMCE : la racine de la moyenne des carrés des erreurs (RMCE) est la racine carrée de la MCE.
- MAPE : la Mean Absolute Percentage Error.

- DW : le coefficient de Durbin-Watson. C
- Cp : le coefficient Cp de Mallows.
- AIC : le critère d'information d'Akaike (Akaike's Information Criterion).
- SBC : le critère bayésien de Schwarz (Schwarz's Bayesian Criterion).
- PC : le critère de prédiction

4.5 Analyse de la variance

Sous test d'hypothèses :

$$H_0 : i = 1, \dots, 4, \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

vs

$$H_1 : \exists i = 1, \dots, 4 \text{ tq } \beta_i \neq 0$$

Source	DDL	Somme des carrées	Moyenne des carrées	Fc	Ftab
Régression	4	359,367	89,842	136,5413	0,00001
Résiduelle	26	22,769	0,8757		
Total	30	382,137			

TAB. 4.5 – Tableau de ANOVA.

Interprétation :

Comme la valeur critique $F_{(0,05;26)} = 136,5413$ est plus petite que la statistique de test F_c , on accepte H_0 à un seuil de signification $\alpha = 0,05$.

- La régression semble une très bonne qualité puisque que nous expliquons $R^2 = 0,9594$ et de la variance de l'endogène.
- Impression confirmé par le test de Fisher, $F = 136,5413 > 0,00001$: le modèle est globalement très significatif.
- Mise à part la variable cylindrée, toutes les variables sont significatives.

4.6 Prédiction

Observation	Poids	y_i	Préd (y_i)	Résidu	Résidu std
1	1	5,800	6,494	-0,94	0,811
2	1	6,100	5,991	0,109	0,127
3	1	6,500	7,189	-0,689	-0,805
4	1	6,800	6,699	0,101	0,118
5	1	7,100	6,275	0,525	0,613
6	1	21,300	7,771	-0,671	-0,785
7	1	18,700	20,527	0,773	0,903
8	1	14,500	20,103	-1,403	-1,640
9	1	7,400	14,439	0,061	0,071
10	1	9,00	8,541	-1,141	-1,333
11	1	11,700	8,502	0,498	0,582
12	1	9,500	10,888	0,812	1,591
13	1	8,800	8,605	0,895	1,046
14	1	9,300	9,079	0,421	0,492
15	1	8,800	9,073	-0,273	-0,319
16	1	9,300	7,913	1,387	1,621
17	1	8,600	8,129	0,471	0,550
18	1	7,700	8,563	-0,683	-1,009
19	1	10,800	10,443	0,357	0,417
20	1	7,600	7,537	-0,937	-1,095
21	1	11,700	10,388	1,162	1,591
22	1	11,900	12,307	-0,401	-0,476
23	1	10,800	10,628	0,172	0,201
24	1	7,600	9,141	-1, 541	-1,801
25	1	11,300	12,258	-0,958	-1,119
26	1	10,800	10,881	-0,081	-0,095
27	1	9,200	9,52	-0,320	-0,374
28	1	11,600	11,068	0,532	0,622
29	1	12,800	12,097	0,703	0,822
30	1	12,700	11,898	0,802	0,937
31	1	11,985	110,589	0,396	0,985

TAB. 4.6 – Tableau représente la prédiction et les résidus.

Observation	Poids	Ecart-type sur préd	Borne inf(0,95)	Borne sup(0,95)
1	1	0,280	5,916	7,071
2	1	0,311	5,350	6,633
3	1	0,257	6,660	7,718
4	1	0,288	6,105	7,293
5	1	0,344	5,566	6,984
6	1	0,206	7,347	8,196
7	1	0,793	18,893	22,161
8	1	0,602	18,864	21,342
9	1	0,684	13,031	15,847
10	1	0,186	8,157	8,925
11	1	0,192	8,106	8,898
12	1	0,239	10,396	11,380
13	1	0,276	8,073	9,173
14	1	0,189	8,689	9,469
15	1	0,213	8,633	9,512
16	1	0,248	7,403	8,423
17	1	0,214	7,689	8,569
18	1	0,214	10,190	9,004
19	1	0,203	10,050	10,861
20	1	0,228	7,068	8,007
21	1	0,445	9,423	11,254
22	1	0,326	11,636	12,978
23	1	0,292	10,026	11,230
24	1	0,291	8,583	9,740
25	1	0,310	11,620	12,895
26	1	0,336	10,920	11,573
27	1	0,189	9,130	9,910
28	1	0,409	10,225	11,910
29	1	0,484	11,099	13,095
30	1	0,253	11,3978	12,419
31	1	11,951	11,545	0,406

TAB. 4.7 – Tableau représente la prédiction (la borne inf et sup).

Conclusion :

On conclut que le prix et la consommation soient d'une certaine manière liées, on peut le comprendre. En revanche, imaginer que le prix influe directement sur la consommation paraît étrange. Cela voudrait dire qu'en diminuant artificiellement le prix d'un véhicule, on pourrait diminuer la consommation, concernant la cylindrée, la taille de moteur, on s'étonne comme même qu'elle ne joue aucun rôle sur la consommation. Cela voudrait dire qu'on peut augmenter la taille de moteur sans que cela ne soit préjudiciable à la consommation.

Conclusion générale

Nous nous intéressons dans ce mémoire au modèle linéaire général. Un intérêt particulier a été donné au modèle simple et multiple. Dans la partie pratique, on d'abord modéliser le problème donné par un modèle mathématique de régression linéaire multiple. Ensuite, on a passé à l'estimation du modèle par la méthode des moindres carrés, et on a calculé la matrice de corrélation. Par la suite, on a calculé les prédictions des observations, les résidus, les résidus standards, l'écart type, et les bornes inférieure et supérieure de chaque observation. On a fini par un tableau de L'ANOVA sur lequel on a tiré une conclusion. Dans les deux cas, régression linéaire et modèle linéaire, on a été amené à poser le même modèle : $y = X\beta + \epsilon$.

Cependant, les hypothèses sont différentes : dans le modèle linéaire X est un tableau de données certaines, alors qu'en régression X est aléatoire.

Le vecteur des résidus ϵ a une matrice variance quelconque Σ dans le modèle linéaire, alors qu'en régression le vecteur ϵ a pour une matrice de variance $\sigma^2 I$ car l'hypothèse d'échantillonnage suppose les observations indépendantes.

Les objectifs sont également différents, en régression, on veut ajuster au mieux y , dans le modèle linéaire, on cherche à estimer l'effet moyen des variables explicatives.

Si on considère dans le modèle de régression linéaire multiple les variables explicatives comme des constantes, ce qui revient à travailler conditionnellement aux X , il est clair que ceci revient au même que de poser le modèle linéaire $(y; X\beta; \sigma^2 I)$ si tous les individus ont le même poids.

En fait, la plupart des propriétés de la régression multiple s'obtiennent conditionnellement aux variables explicatives comme en régression simple, ce qui nous autorisera à ne plus parler que le modèle $(y; X\beta; \sigma^2 I)$.

Par ailleurs, l'utilisation complète de modèle linéaire suppose comme la matrice Σ . Or, en pratique, on ignore Σ et, faute de mieux, on fait couramment l'hypothèse simplificatrice

que Σ est diagonale (non corrélation des erreurs) et que tous les termes sont égaux (homoscédasticité) c'est-à-dire que $\Sigma = \sigma^2 I$, quitté à vérifier a posteriori sur les résultats de la validité de ces deux hypothèses.

Ceci explique la confusion entre modèle linéaire et régression multiple ; dans ce qui suit, nous ne ferons plus la distinction, car nous référons désormais à l'unique modèle simplificateur $(y; X\beta; \sigma^2 I)$, en supposant que les poids des observations sont égaux entre eux.

Remarquons pour finir que le terme de linéaire s'applique en fait au vecteur β , et aux variables explicatives ; ainsi la régression polynomiale $y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \dots + \beta_p x^p$ est un cas particulier du modèle général où l'on prend p variables explicatives x^1, x^2, \dots, x^p .

Bibliographie

- [1] V.ROUSSON, *Analyse de régression appliquée*, 2ème édition, DUNOD,Paris, 2004.
- [2] P.J., *Robust Statistics*,New York, John Wiley and Sons,1981.
- [3] W.J Gentle J.E., *Statistical Compunting*, New York, Marcel Dekker 1980.
- [4] P.,Nelder J. A.,*Generalised linear Models*,New York,John Wiley and Sons,1982.
- [5] F., TUKEY J.W., *Data Analysis and Regression*,Reading, Massachussets, 1977.
- [6] R.H., *Classical and Modern Regression with Application*, Boston Massachussets, 1990.
- [7] F.A., *Matrices with Application in Statistics*. Californie, Wad-sworth, 1983.
- [8] J., *Regression Diagnostics*,Newberry Park,Califirnie,Sage, 1991.
- [9] R.J.,WILSON W.J., *Regression Analysis*.London,Academic Press,1998.
- [10] Y., *Statistique Dictionnaire encyclopédique* .Päris, Springer, 2004.
- [11] R.D., WEISBERG S., Sporta, *In Introduction to Regression Graphics*.New York,John Wiley and Sons, 1993.
- [12] G.W., *Collineary and least Squares Regression*, Statistical Scienc,1987.
- [13] P.G., *Robust Regression,Asymptotics,Conjectures,and Monte Carlo*, Annals of Statistics,1973.
- [14] S.C.,WEDDERBURN J.F., *Prediction, Linear Regression and Minimum Sun of Relative Errors*, Technometrics,1977.
- [15] I., *Transformation and Influentiel Observation in Minimum Sum of Absolute Errors Regression*,Journal of the American Statistical,1988.
- [16] , *Résumé de Cours de Modèle de Régression*, 10 Janvier 2011.
- [17] S., *Régression linéaire simple*,École Polytechnique de Montréal 10 Janvier 2011.
- [18] K.Khaldi, *Méthodes statistiques et Probabilités* . Ed,Alger, 2000.