

RÉSUMÉ DU COURS D'ÉCONOMÉTRIE

Yves Tillé

16 décembre 2008

Avertissement

Ce document n'est pas un compte rendu exhaustif du cours d'Économétrie, mais un résumé. Il reprend les principaux développements, mais il est complété au cours par de nombreux graphiques, commentaires, et approfondissements. Nous remercions Jérôme Taillard pour la préparation de plusieurs exercices, Guido Pult pour nous avoir donné plusieurs exercices et Ines Pasini pour son aide à la dactylographie. Les étudiants sont invités à consulter les ouvrages de références suivants cités dans la bibliographie : Judge et al. (1985), Johnston (1988), Theil (1979), Maddala (1988), Gourieroux and Monfort (1989a), Gourieroux and Monfort (1989b), Greene (1990), Cohen and Pradel (1993), Bourbonnais (1993), Johnston (1997), Johnson (1999), Ruud (2000).

Yves Tillé

Chapitre 1

Éléments d'algèbre linéaire

1.1 Espace vectoriel

1.1.1 Vecteur

Un élément de \mathbb{R}^n est une suite ordonnée de n éléments de \mathbb{R} . On peut disposer cette suite, appelée vecteur soit en ligne, soit en colonne.

Exemple 1.1 Le vecteur $\mathbf{a} = [3 \ 0]$, est un vecteur ligne et le vecteur

$$\mathbf{b} = \begin{pmatrix} 3 \\ -2 \\ 0 \end{pmatrix}$$

est un vecteur colonne.

La transposition transforme un vecteur ligne en vecteur colonne et réciproquement.

Exemple 1.2 Si $\mathbf{a} = (3 \ 0)$, la transposée de \mathbf{a} est

$$\mathbf{a}' = \begin{pmatrix} 3 \\ 0 \end{pmatrix}.$$

1.1.2 Multiplication par un scalaire et addition

On peut multiplier un vecteur par un scalaire. Soit un scalaire $c \in \mathbb{R}$ et un vecteur colonne \mathbf{a} de \mathbb{R}^n , alors

$$c \times \mathbf{a} = c \times \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ca_1 \\ \vdots \\ ca_n \end{pmatrix}.$$

Deux vecteurs lignes (ou deux vecteurs colonnes) peuvent s'additionner s'ils sont de même dimension.

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix}.$$

En utilisant la multiplication par un scalaire et l'addition, on peut définir une combinaison linéaire de deux vecteurs \mathbf{a} et \mathbf{b} :

$$c_1 \mathbf{a} + c_2 \mathbf{b} = c_1 \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + c_2 \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} c_1 a_1 + c_2 b_1 \\ \vdots \\ c_1 a_n + c_2 b_n \end{pmatrix}.$$

où $c_1, c_2 \in \mathbb{R}$.

1.1.3 Définition d'un espace vectoriel

On se réfère à la définition suivante : la définition suivante :

Définition 1.1 Soit \mathbb{K} un corps commutatif d'élément unité noté 1. On nomme espace vectoriel sur \mathbb{K} , un ensemble E muni d'une loi de composition interne $(+)$ conférant à E la structure de groupe commutatif ou abélien, et d'une seconde loi dite externe, application de $E \times \mathbb{K}$ dans E notée (\times) , aussi appelée multiplication, faisant intervenir les éléments de \mathbb{K} , appelés scalaires. Cette loi externe doit vérifier les axiomes suivants, $x, y \in E$, $a, b \in \mathbb{K}$ désignant des scalaires :

1. $a \times (x + y) = a \times x + a \times y$ (distributivité)
2. $(a + b) \times x = a \times x + b \times x$ (distributivité)
3. $a \times (b \times x) = ab \times x$ (associativité)
4. $1 \times x = x$

Si on prend $\mathbb{K} = \mathbb{R}$, on vérifie que \mathbb{R}^n doté de la loi interne $+$ et de la loi externe \times est un espace vectoriel.

1.1.4 Vecteurs linéairement indépendants

Définition 1.2 Les vecteurs $\mathbf{u}_1, \dots, \mathbf{u}_j, \dots, \mathbf{u}_J$ sont dit linéairement indépendants, si

$$a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \dots + a_J \mathbf{u}_J = \mathbf{0}$$

implique que $a_1 = a_2 = \dots = a_J = 0$.

1.1.5 Sous-espace vectoriel

Définition 1.3 Un sous-ensemble non-vide V de \mathbb{R}^n est un sous-espace vectoriel, si pour tous $\mathbf{u}, \mathbf{v} \in V$,

1. $\mathbf{u} + \mathbf{v} \in V$,
2. $a\mathbf{u} \in V$ pour tout $a \in \mathbb{R}$.

1.1.6 Système générateur d'un sous-espace vectoriel

Définition 1.4 Un ensemble de p vecteurs $\mathbf{u}_1, \dots, \mathbf{u}_p$ du sous-espace vectoriel V forment un système générateur de V si et seulement si

1. $\mathbf{u}_1, \dots, \mathbf{u}_p$ sont tous différents de $\mathbf{0}$,
2. pour tout $\mathbf{v} \in V$, on peut écrire $\mathbf{v} = a_1 \mathbf{u}_1 + \dots + a_p \mathbf{u}_p$.

1.1.7 Base d'un sous-espace vectoriel

Définition 1.5 Un ensemble de p vecteurs $\mathbf{u}_1, \dots, \mathbf{u}_p$ du sous-espace vectoriel V forment une base de V si et seulement si

1. ils sont linéairement indépendants,
2. ils forment un système générateur de V .

Autrement dit, tout vecteur de V peut s'écrire comme une combinaison linéaire de $\mathbf{u}_1, \dots, \mathbf{u}_p$.

1.1.8 Base canonique de \mathbb{R}^n

La base canonique de \mathbb{R}^n est

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

1.1.9 Dimension d'un sous-espace vectoriel

Définition 1.6 La dimension d'un sous-espace vectoriel est le plus petit nombre de vecteurs suffisants pour l'engendrer.

Cette dimension correspond en particulier au nombre de vecteurs constituant une base quelconque de V .

1.2 Espace euclidien

1.2.1 Produit scalaire

On définit la multiplication d'un vecteur ligne \mathbf{a} par un vecteur colonne \mathbf{b} comme le résultat scalaire :

$$\mathbf{a} \times \mathbf{b} = (a_1 \dots a_n) \times \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^n a_i b_i.$$

Le produit scalaire de deux vecteurs colonnes \mathbf{u} et \mathbf{b} de même dimension est noté $\langle \mathbf{u}, \mathbf{b} \rangle$ et est défini par :

$$\langle \mathbf{u}, \mathbf{b} \rangle = \mathbf{u}'\mathbf{b} = (u_1 \dots u_n) \times \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^n u_i b_i.$$

Définition 1.7 Un espace euclidien est un espace vectoriel muni d'un produit scalaire.

1.2.2 Norme

Définition 1.8 La norme (ou longueur) d'un vecteur colonne \mathbf{u} est

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

vecteur de norme égale à 1 est dit normé.

1.2.3 Distance entre deux vecteurs

Définition 1.9 La distance entre les vecteurs \mathbf{u} et \mathbf{v} de \mathbb{R}^n est définie par

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{\sum_{i=1}^n (u_i - v_i)^2}.$$

Définition 1.10 La projection d'un vecteur \mathbf{u} sur un vecteur \mathbf{v} est définie par

$$p_v(\mathbf{u}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v}}{\|\mathbf{v}\|^2}. \quad (1.1)$$

1.2.4 Vecteurs orthogonaux

Définition 1.11 Deux vecteurs non-nuls \mathbf{u} et \mathbf{v} de \mathbb{R}^n sont orthogonaux si

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0.$$

On note alors $\mathbf{u} \perp \mathbf{v}$

Théorème 1.1 (de Pythagore) Si \mathbf{u} et \mathbf{v} sont orthogonaux, alors

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

1.2.5 Orthogonal d'un sous-espace vectoriel

Définition 1.12 Un vecteur \mathbf{u} est orthogonal à un sous-espace vectoriel V si et seulement si il est orthogonal à tous les vecteurs de V , on note alors

$$\mathbf{u} \perp V.$$

Définition 1.13 Les sous-espaces V et W sont dits orthogonaux, si tout vecteur de V est orthogonal à tout vecteur de W .

Définition 1.14 L'ensemble de tous les vecteurs orthogonaux à V est appelé l'orthogonal de V et est noté V^\perp .

Propriété 1.1

- $(V^\perp)^\perp = V$,
- $V \cap V^\perp = \{\mathbf{0}\}$.

1.3 Application linéaire et matrices

1.3.1 Application linéaire

Une application $f(\cdot)$ de \mathbb{R}^J dans \mathbb{R}^I est dite linéaire si pour tous \mathbf{u}, \mathbf{v} , de \mathbb{R}^J et tout $a \in \mathbb{R}$

- $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$,
- $f(a\mathbf{u}) = af(\mathbf{u})$.

1.3.2 Matrice

Une matrice est un tableau de nombres. Par exemple :

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1J} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ij} & \dots & a_{iJ} \\ \vdots & & \vdots & & \vdots \\ a_{I1} & \dots & a_{Ij} & \dots & a_{IJ} \end{pmatrix}$$

est une matrice de I lignes et de J colonnes.

En statistique, on manipule souvent des matrices. Par convention, les lignes représentent souvent les unités statistiques, et les colonnes des variables.

Comme les vecteurs, les matrices peuvent être multipliées par un scalaire. On peut également additionner deux matrices à condition qu'elles aient le même nombre de lignes et de colonnes. Sous cette même condition, on peut aussi définir une combinaison linéaire de deux matrices.

1.3.3 Produit d'une matrice et d'un vecteur

Soient une matrice \mathbf{A} de dimension $I \times J$ et un vecteur colonne \mathbf{u} de dimension J le produit \mathbf{Au} est donné par

$$\mathbf{Au} = \begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1J} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ij} & \dots & a_{iJ} \\ \vdots & & \vdots & & \vdots \\ a_{I1} & \dots & a_{Ij} & \dots & a_{IJ} \end{pmatrix} \times \begin{pmatrix} u_1 \\ \vdots \\ u_j \\ \vdots \\ u_J \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^J a_{1j}u_j \\ \vdots \\ \sum_{j=1}^J a_{ij}u_j \\ \vdots \\ \sum_{j=1}^J a_{Ij}u_j \end{pmatrix}.$$

Le produit d'un vecteur par une matrice est la représentation d'une application linéaire dans la base canonique.

1.3.4 Produit matriciel

Soient deux matrices \mathbf{A} de dimension $I \times J$ et \mathbf{B} de dimension $J \times K$, alors le produit de ces deux matrices est donné par

$$\begin{aligned} \mathbf{AB} &= \begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1J} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ij} & \dots & a_{iJ} \\ \vdots & & \vdots & & \vdots \\ a_{I1} & \dots & a_{Ij} & \dots & a_{IJ} \end{pmatrix} \times \begin{pmatrix} b_{11} & \dots & b_{1k} & \dots & b_{1K} \\ \vdots & & \vdots & & \vdots \\ b_{j1} & \dots & b_{jk} & \dots & b_{jK} \\ \vdots & & \vdots & & \vdots \\ b_{J1} & \dots & b_{Jk} & \dots & b_{JK} \end{pmatrix} \\ &= \begin{pmatrix} c_{11} & \dots & c_{1k} & \dots & c_{1K} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \dots & c_{ik} & \dots & c_{iK} \\ \vdots & & \vdots & & \vdots \\ c_{I1} & \dots & c_{Ik} & \dots & c_{IK} \end{pmatrix} \\ &= \mathbf{C}, \end{aligned}$$

où

$$c_{ik} = \sum_{j=1}^J a_{ij} b_{jk}.$$

C'est le produit des lignes par les colonnes. La matrice \mathbf{C} est de dimension $(I \times K)$.

1.3.5 Transposition

Transposer une matrice revient à remplacer les lignes par les colonnes et vice versa. Par exemple, si

$$\mathbf{A} = \begin{pmatrix} -1 & 2 \\ 4 & 3 \\ -2 & 5 \end{pmatrix} \text{ alors } \mathbf{A}' = \begin{pmatrix} -1 & 4 & -2 \\ 2 & 3 & 5 \end{pmatrix}.$$

Remarque 1.1 Soient $\mathbf{A}, \mathbf{B}, \mathbf{C}$ de dimension respectives $(I \times J), (J \times K)$ et $(K \times L)$, alors la transposée de \mathbf{ABC} vaut

$$(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'.$$

1.3.6 Matrices carrées, symétriques et diagonales

Définition 1.15 Une matrice est dite carrée si elle a le même nombre de lignes et de colonnes.

Si un vecteur de dimension n est prémultiplié par une matrice carrée $n \times n$, le résultat est donc aussi de dimension n . Une matrice carrée $n \times n$ est donc une application linéaire de \mathbb{R}^n dans \mathbb{R}^n .

Définition 1.16 Une matrice est dite symétrique si elle est égale à sa transposée.

Une matrice symétrique est donc toujours carrée.

Définition 1.17 Une matrice est dite diagonale, si elle est carrée et que tous ses éléments extradiagonaux sont nuls.

Par exemple,

$$\mathbf{D} = \begin{pmatrix} 6 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

est une matrice diagonale.

Définition 1.18 Une matrice identité \mathbf{I} est une matrice diagonale dont tous les éléments de la diagonale sont égaux à 1.

Par exemple,

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

est une matrice identité de dimension 3×3 .

1.3.7 Rang d'une matrice

Définition 1.19 Le rang d'une matrice est le nombre maximum de lignes (ou de colonnes) linéairement indépendantes.

Propriété 1.2 Le rang est toujours inférieur ou égal au minimum du nombre de lignes et du nombre de colonnes de la matrice.

Définition 1.20 Si le rang de la matrice est égal au minimum du nombre de lignes et du nombre de colonnes, la matrice est dite de plein rang (ou de rang maximal).

Propriété 1.3 Le rang d'un produit de matrices est inférieur ou égal au rang de chaque matrice.

1.3.8 Trace d'une matrice

Définition 1.21 La trace d'une matrice carrée est la somme des éléments de sa diagonale.

Propriété 1.4 1. $\text{trace}(\mathbf{A} + \mathbf{B}) = \text{trace}(\mathbf{A}) + \text{trace}(\mathbf{B})$.
2. $\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$ mais $\text{trace}(\mathbf{AB}) \neq \text{trace}(\mathbf{A})\text{trace}(\mathbf{B})$.

1.3.9 Matrices inversibles

Définition 1.22 Une matrice carrée \mathbf{A} est dite inversible, s'il existe une matrice \mathbf{A}^{-1} qui vérifie $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

Propriété 1.5 Si une matrice carrée est de plein rang, alors elle est inversible.

1.3.10 Inversion par parties

Soit une matrice \mathbf{F} composée de quatre sous-matrices :

$$\mathbf{F} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}.$$

Les matrices \mathbf{A} et \mathbf{D} sont carrées et inversibles.

La technique d'inversion par partie permet d'obtenir l'inverse de \mathbf{F} .

$$\mathbf{F}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{BQCA}^{-1} & -\mathbf{A}^{-1}\mathbf{BQ} \\ -\mathbf{QCA}^{-1} & \mathbf{Q} \end{pmatrix}$$

où

$$\mathbf{Q} = (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}$$

Ce résultat peut être démontré aisément en réalisant le produit $\mathbf{F}^{-1}\mathbf{F}$.

1.3.11 Déterminant

Définition 1.23 Le déterminant d'une matrice carrée \mathbf{A} ($J \times J$) est noté $|\mathbf{A}|$ et est défini par

- Si $J = 1$, $|\mathbf{A}| = \mathbf{A}$
- Si $J > 1$,

$$|\mathbf{A}| = \sum_{i=1}^J (-1)^{i+j} |\mathbf{M}_{ij}| a_{ij},$$

pour tout j fixé, où $|\mathbf{M}_{ij}|$ est le mineur de a_{ij} . Le mineur est le déterminant de la matrice $(J-1) \times (J-1)$ obtenue en enlevant la colonne i et la ligne j de la matrice \mathbf{A} .

Exemple 1.3 Soit \mathbf{A} une matrice (2×2) ,

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

en prenant $j = 1$, on a

$$|\mathbf{A}| = a \times d - c \times b = ad - cb.$$

On peut aussi calculer le déterminant de \mathbf{A} en prenant $j = 2$.

Exemple 1.4 Soit une matrice \mathbf{A} de dimension (3×3) , le calcul se fait en prenant $j = 1$

$$\mathbf{A} = \begin{pmatrix} 2 & 7 & 6 \\ 9 & 5 & 1 \\ 4 & 3 & 8 \end{pmatrix}$$

alors son déterminant vaut

$$\begin{aligned} |\mathbf{A}| &= \begin{vmatrix} 5 & 1 \\ 3 & 8 \end{vmatrix} \times 2 - \begin{vmatrix} 7 & 6 \\ 3 & 8 \end{vmatrix} \times 9 + \begin{vmatrix} 7 & 6 \\ 5 & 1 \end{vmatrix} \times 4 \\ &= (5 \times 8 - 1 \times 3) \times 2 - (7 \times 8 - 3 \times 6) \times 9 + (7 \times 1 - 6 \times 5) \times 4 \\ &= 37 \times 2 - 38 \times 9 - 23 \times 4 \\ &= -360. \end{aligned}$$

Propriété 1.6

1. $|\mathbf{A}| = |\mathbf{A}'|$,
2. $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$, en particulier $|\mathbf{A}^k| = |\mathbf{A}|^k$.
3. $|c\mathbf{A}| = c^J |\mathbf{A}|$, (où \mathbf{A} est de dimension $J \times J$),

1.3.12 Quelques propriétés

Propriété 1.7 En général, si \mathbf{A} , \mathbf{B} et \mathbf{C} sont des matrices carrées de même dimension, on a

1. $\mathbf{AB} \neq \mathbf{BA}$,
2. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$,
3. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$,
4. $\mathbf{AI} = \mathbf{A} = \mathbf{IA}$, où \mathbf{I} est une matrice identité,
5. $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$,
6. $\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$,
7. $\text{trace}(\mathbf{A} + \mathbf{B}) = \text{trace}(\mathbf{A}) + \text{trace}(\mathbf{B})$,
8. $\det \mathbf{A} = \det \mathbf{A}'$,
9. $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$.

1.3.13 Matrices orthogonales

Définition 1.24 Une matrice Γ est dite orthogonale si son inverse est égale à sa transposée :

$$\Gamma' = \Gamma^{-1}.$$

1.3.14 Valeurs propres et vecteurs propres

Définition 1.25 Soit \mathbf{A} une matrice $J \times J$. λ_i est une valeur propre de \mathbf{A} si λ_i est une solution de l'équation

$$|\mathbf{A} - \lambda \mathbf{I}| = 0.$$

Propriété 1.8

- Une matrice carrée symétrique de dimension $J \times J$ possède toujours J valeurs propres.
- La trace d'une matrice carrée est toujours égale à la somme des valeurs propres.
- Le déterminant d'une matrice carrée symétrique est toujours égal au produit de ses valeurs propres.

Définition 1.26 Le vecteur $\mathbf{u}_i \neq 0$ est un vecteur propre de \mathbf{A} associé à la valeur propre λ_i si

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i.$$

Propriété 1.9 Si \mathbf{A} est une matrice $J \times J$ réelle symétrique, il existe J vecteurs propres normés et orthogonaux.

Théorème 1.2 (de diagonalisation) Soient \mathbf{A} une matrice symétrique ($J \times J$), et $\mathbf{u}_i, \lambda_i, i = 1, \dots, J$, ses valeurs propres et vecteurs propres associés. Soient la matrice orthogonale Γ dont les colonnes sont les J vecteurs propres de \mathbf{A} , et la matrice diagonale Λ ayant sur sa diagonale principale les J valeurs propres. Alors

- $\Gamma' \mathbf{A} \Gamma = \Lambda$,
- $\mathbf{A} = \Gamma \Lambda \Gamma'$.

1.3.15 Formes et applications linéaires, formes quadratiques

Définition 1.27 Soient \mathbf{A} une matrice ($I \times I$), \mathbf{B} une matrice ($J \times I$), \mathbf{a} un vecteur colonne de \mathbb{R}^J et \mathbf{b} un vecteur colonne de dimension I . On appelle

- forme linéaire définie par le vecteur \mathbf{a}' , l'application de \mathbb{R}^I dans \mathbb{R}

$$\mathbf{a}'\mathbf{b},$$

- application linéaire de \mathbb{R}^I dans \mathbb{R}^J définie par la matrice \mathbf{B} ,

$$\mathbf{B}\mathbf{b},$$

- et forme quadratique définie par la matrice \mathbf{A} , l'expression

$$\mathbf{b}'\mathbf{A}\mathbf{b}.$$

Définition 1.28 Une matrice \mathbf{A} de dimension ($I \times I$) est dite définie positive si

$$\mathbf{b}'\mathbf{A}\mathbf{b} > 0,$$

pour tout $\mathbf{b} \in \mathbb{R}^I \setminus \{0\}$.

Définition 1.29 Une matrice \mathbf{A} de dimension ($I \times I$) est dite semi-définie positive si

$$\mathbf{b}'\mathbf{A}\mathbf{b} \geq 0,$$

pour tout $\mathbf{b} \in \mathbb{R}^I$.

Propriété 1.10 Une condition nécessaire et suffisante pour qu'une matrice soit définie positive (resp. semi-définie positive) est que toutes ses valeurs propres soient strictement positives (resp. positives ou nulles).

Propriété 1.11 Pour toute matrice \mathbf{D} , la matrice $\mathbf{D}'\mathbf{D}$ est semi-définie positive.

Démonstration

En posant $\mathbf{a} = \mathbf{D}\mathbf{b}$ la forme quadratique $\mathbf{b}'\mathbf{D}'\mathbf{D}\mathbf{b}$ peut s'écrire

$$\mathbf{b}'\mathbf{D}'\mathbf{D}\mathbf{b} = \mathbf{a}'\mathbf{a} = \sum_i a_i^2 \geq 0.$$

□

Propriété 1.12 Une matrice définie positive est toujours inversible.

1.3.16 Image et noyau d'une matrice

Définition 1.30 Le noyau d'une matrice \mathbf{A} de dimension $I \times J$ est le sous-espace de \mathbb{R}^J défini par

$$\text{Ker}(\mathbf{A}) = \{\mathbf{u} \in \mathbb{R}^J \mid \mathbf{A}\mathbf{u} = \mathbf{0}\}.$$

La définition implique que tous les vecteurs de $\text{Ker}(\mathbf{A})$ sont orthogonaux à tous les vecteurs lignes contenus dans la matrice \mathbf{A} .

Définition 1.31 L'image d'une matrice \mathbf{B} de dimension $I \times J$ est le sous-espace de \mathbb{R}^I défini par

$$\text{Im}(\mathbf{B}) = \{\mathbf{x} \in \mathbb{R}^I \mid \text{il existe } \mathbf{u} \in \mathbb{R}^J \text{ tel que } \mathbf{B}\mathbf{u} = \mathbf{x}\}.$$

Le sous-espace $\text{Im}(\mathbf{B})$ est l'ensemble des vecteurs qui peuvent s'écrire comme une combinaison linéaire des colonnes de \mathbf{B} . L'image de la matrice \mathbf{B} est souvent appelé sous-espace engendré par les colonnes de \mathbf{B} . La dimension de l'image de \mathbf{B} est égale au rang de \mathbf{B} .

Remarque 1.2 Le sous-espace $\text{Im}(\mathbf{B})$ est l'orthogonal de $\text{Ker}(\mathbf{B}')$.

Propriété 1.13 Si $\mathbf{u} \in \text{Im}(\mathbf{B})$ et $\mathbf{v} \in \text{Ker}(\mathbf{B}')$, alors \mathbf{u} et \mathbf{v} sont orthogonaux.

En statistique, on utilise souvent des matrices \mathbf{X} (individus-variables) de dimension $n \times p$ avec $n \geq p$. Le sous-espace engendré par les colonnes de \mathbf{X} est l'image de \mathbf{X} .

1.4 Projection et matrice idempotente

1.4.1 Projection

L'opération de projection se déduit du théorème suivant :

Théorème 1.3 Soit V un sous-espace vectoriel de \mathbb{R}^n , alors tout vecteur $\mathbf{u} \in \mathbb{R}^n$ se décompose de manière unique en une somme d'un vecteur de V et d'un vecteur de V^\perp .

1.4.2 Projection orthogonale

Définition 1.32 Soit V un sous-espace de \mathbb{R}^n , l'application linéaire qui à un vecteur \mathbf{u} fait correspondre un vecteur \mathbf{u}^* tel que $\mathbf{u} - \mathbf{u}^*$ soit orthogonal à V est appelé projection orthogonale ($\mathbf{u}^* \in V$).

1.4.3 Projection orthogonale dans l'image et le noyau d'une matrice

Le projecteur orthogonal dans l'image d'une matrice \mathbf{X} de plein rang de dimension $n \times p$ avec $n \geq p$ est donné par

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Le projecteur orthogonal dans le noyau d'une matrice \mathbf{X}' de plein rang de dimension $n \times p$ avec $n \geq p$ est donné par

$$\mathbf{P}_X^\perp = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I} - \mathbf{P}_X.$$

Remarque 1.3 Si $\mathbf{X} = \mathbf{v}$ est un vecteur, alors le projecteur est

$$\mathbf{P}_v = \mathbf{v}(\mathbf{v}'\mathbf{v})^{-1}\mathbf{v}' = \mathbf{v}\|\mathbf{v}\|^{-2}\mathbf{v}' = \frac{\mathbf{v}\mathbf{v}'}{\|\mathbf{v}\|^2},$$

et la projection de \mathbf{u} sur \mathbf{v}

$$p_v(\mathbf{u}) = \mathbf{P}_v\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v}'\mathbf{u} = \mathbf{v}\frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{v}\|^2},$$

ce qui correspond à la définition donnée en (1.10).

1.4.4 Matrice idempotente

Définition 1.33 Une matrice \mathbf{P} est dite idempotente si $\mathbf{P}\mathbf{P} = \mathbf{P}$.

Une matrice de projection est idempotente.

Remarque 1.4 Les matrices \mathbf{P}_X et \mathbf{P}_X^\perp sont évidemment idempotentes, en effet

$$\begin{aligned} \mathbf{P}_X\mathbf{P}_X &= \{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\} \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\underbrace{\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}_{=\mathbf{I}}\mathbf{X}' \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{P}_X. \end{aligned}$$

De plus

$$\mathbf{P}_X^\perp\mathbf{P}_X^\perp = (\mathbf{I} - \mathbf{P}_X)(\mathbf{I} - \mathbf{P}_X) = \mathbf{I} - 2\mathbf{P}_X + \underbrace{\mathbf{P}_X\mathbf{P}_X}_{=\mathbf{P}_X} = \mathbf{I} - \mathbf{P}_X = \mathbf{P}_X^\perp.$$

Le projecteur orthogonal dans le noyau d'une matrice \mathbf{X}' de plein rang de dimension $n \times p$ est donné par

$$\mathbf{P}_X^\perp = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I} - \mathbf{P}_X.$$

Théorème 1.4 Toutes les valeurs propres d'une matrice idempotente valent 1 ou 0.

Démonstration

Un vecteur propre non-nul \mathbf{u} d'une matrice \mathbf{P} doit satisfaire au système d'équation

$$\mathbf{P}\mathbf{u} = \lambda\mathbf{u}, \tag{1.2}$$

où λ est la valeur propre associée à \mathbf{u} . En multipliant (1.2) par \mathbf{P} , on obtient

$$\underbrace{\mathbf{P}\mathbf{P}}_{\mathbf{P}}\mathbf{u} = \mathbf{P}\lambda\mathbf{u},$$

et donc,

$$\lambda\mathbf{u} = \lambda^2\mathbf{u}.$$

En prémultipliant par \mathbf{u}' on a

$$\mathbf{u}'\lambda\mathbf{u} = \mathbf{u}'\lambda^2\mathbf{u},$$

on obtient donc

$$\lambda = \lambda^2,$$

ce qui n'est vérifié que si λ vaut 0 ou 1. \square

Comme la trace d'une matrice carrée est aussi la somme de ses valeurs propres, la trace d'une matrice idempotente est le nombre de valeurs propres non-nulles, ce qui donne la propriété suivante.

Propriété 1.14 *La trace d'une matrice idempotente est égale à son rang.*

Remarque 1.5 Le rang et la trace de $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ sont égaux au rang de la matrice $(\mathbf{X}'\mathbf{X})^{-1}$. Cette matrice est supposée de plein rang (sinon $\mathbf{X}'\mathbf{X}$ ne serait pas inversible). Le rang de $(\mathbf{X}'\mathbf{X})^{-1}$ et donc de $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ est donc égal au nombre de colonnes de \mathbf{X} . Le rang de \mathbf{P}_X est la dimension du sous-espace sur lequel projette \mathbf{P}_X .

1.4.5 Projecteurs obliques

Il existe des projecteurs non-orthogonaux. On parle alors de projecteurs obliques. Soit \mathbf{Z} une matrice ayant le même nombre de lignes et de colonnes que \mathbf{X} , alors

$$\mathbf{P}_O = \mathbf{X}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'$$

est un projecteur oblique. Il est facile de vérifier que la matrice \mathbf{P}_O est idempotente et qu'elle projette sur l'image de \mathbf{X} .

1.4.6 Théorème des trois perpendiculaires

Théorème 1.5 *Soit V et W deux sous-espaces vectoriels tels que $V \subset W$, alors*

$$\mathbf{P}_V\mathbf{P}_W = \mathbf{P}_W\mathbf{P}_V = \mathbf{P}_V.$$

1.5 Dérivée par rapport à un vecteur

1.5.1 Gradient

Soit une fonction $f(\cdot)$ de \mathbb{R}^p dans \mathbb{R} :

$$f(\mathbf{x}) = f(x_1, \dots, x_j, \dots, x_p).$$

On suppose en outre que toutes les dérivées partielles existent. On appelle gradient de $f(\cdot)$ le vecteur des dérivées partielles :

$$\text{grad } f = \frac{\partial f}{\partial \mathbf{x}'} = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_j}, \dots, \frac{\partial f}{\partial x_p} \right).$$

1.5.2 Derivation d'une forme linéaire

Soit \mathbf{a} un vecteur de \mathbb{R}^p , alors

$$\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}'} = \left(\frac{\partial \sum_{i=1}^p a_i x_i}{\partial x_1}, \dots, \frac{\partial \sum_{i=1}^p a_i x_i}{\partial x_j}, \dots, \frac{\partial \sum_{i=1}^p a_i x_i}{\partial x_p} \right) = (a_1, \dots, a_j, \dots, a_p) = \mathbf{a}'.$$

1.5.3 Derivation d'une application linéaire

Soit \mathbf{A} une matrice de dimension $q \times p$, alors

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} \sum_{j=1}^p a_{1j}x_j \\ \vdots \\ \sum_{j=1}^p a_{ij}x_j \\ \vdots \\ \sum_{j=1}^p a_{qj}x_j \end{pmatrix}.$$

On a

$$\frac{\partial \mathbf{Ax}}{\partial x_j} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{qj} \end{pmatrix}.$$

Donc,

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}'} = \left(\begin{pmatrix} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{q1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{qj} \end{pmatrix}, \dots, \begin{pmatrix} a_{1p} \\ \vdots \\ a_{ip} \\ \vdots \\ a_{qp} \end{pmatrix} \right) = \begin{pmatrix} a_{11} & \dots & a_{1i} & \dots & a_{1p} \\ \vdots & & \vdots & & \vdots \\ a_{j1} & \dots & a_{ji} & \dots & a_{jp} \\ \vdots & & \vdots & & \vdots \\ a_{q1} & \dots & a_{qi} & \dots & a_{qp} \end{pmatrix} = \mathbf{A}.$$

1.5.4 Dérivée d'une forme quadratique

Soit \mathbf{A} une matrice de dimension $p \times p$, alors

$$\mathbf{x}'\mathbf{Ax} = \sum_{i=1}^p \sum_{j=1}^p a_{ij}x_i x_j = \sum_{i=1}^p a_{ii}x_i^2 + \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p a_{ij}x_i x_j.$$

Donc,

$$\frac{\partial \mathbf{x}'\mathbf{Ax}}{\partial x_k} = 2a_{kk}x_k + \sum_{j \neq k} a_{kj}x_j + \sum_{i \neq k} a_{ik}x_i = \sum_{j=1}^p a_{kj}x_j + \sum_{i=1}^p a_{ik}x_i,$$

et

$$\frac{\partial \mathbf{x}'\mathbf{Ax}}{\partial \mathbf{x}} = \begin{pmatrix} \sum_{j=1}^p a_{1j}x_j + \sum_{i=1}^p a_{i1}x_i \\ \vdots \\ \sum_{j=1}^p a_{kj}x_j + \sum_{i=1}^p a_{ik}x_i \\ \vdots \\ \sum_{j=1}^p a_{pj}x_j + \sum_{i=1}^p a_{ip}x_i \end{pmatrix} = \mathbf{Ax} + \mathbf{A}'\mathbf{x}.$$

Si la matrice \mathbf{A} est symétrique, on a

$$\frac{\partial \mathbf{x}'\mathbf{Ax}}{\partial \mathbf{x}} = 2\mathbf{Ax}$$

Exercices

Exercice 1.1

Calculez

$$\frac{\partial (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})}{\partial \mathbf{b}},$$

où $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^n$, et \mathbf{X} est une matrice de dimension $n \times p$.

Exercice 1.2

1. Construisez des projecteurs orthogonaux $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$, sur des sous-espaces engendrés par les colonnes des matrices

$$\mathbf{X}_1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \mathbf{X}_2 = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}, \mathbf{X}_3 = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

2. Construisez les trois projecteurs qui projettent sur l'orthogonal des sous-espaces engendré par les colonnes de $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$.
3. Vérifiez que ces 6 projecteurs sont des matrices idempotentes.
4. Projetez le vecteur

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}$$

au moyen de ces 6 projecteurs.

Exercice 1.3 Soient les matrices

$$\mathbf{A} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}, \mathbf{C} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

Décomposez le vecteur $\mathbf{z} = (z_1, \dots, z_n)$ en fonction de ses projections sur respectivement

1. $\text{Ker}(\mathbf{A}')$ et $\text{Im}(\mathbf{A})$
2. $\text{Ker}(\mathbf{B}')$ et $\text{Im}(\mathbf{B})$
3. $\text{Ker}(\mathbf{C}')$ et $\text{Im}(\mathbf{C})$.

Exercice 1.4 Soient les matrices

$$\begin{aligned} &1. \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad 2. \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad 3. \begin{pmatrix} 1/5 & 3/5 \\ 3/5 & 9/5 \end{pmatrix} \\ &4. \begin{pmatrix} 1/9 & 4/9 \\ 4/9 & 16/9 \end{pmatrix}, \quad 5. \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad 6. \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}}. \end{aligned}$$

Calculez

1. leur rang
2. leur trace.

Quelles sont les matrices idempotentes et orthogonales ?

Avec les matrices idempotentes, projetez le vecteur $\begin{pmatrix} x_1 & x_2 \end{pmatrix}$ ou $\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}$ selon leur dimension.

Exercice 1.5 Soient \mathbf{X} et \mathbf{Z} , deux matrices de plein rang de dimension $n \times p$ définissant le même sous-espace vectoriel.

1. Donnez l'application linéaire (la matrice) permettant de passer de \mathbf{X} à \mathbf{Z} et réciproquement. Cette matrice est définie en fonction de \mathbf{X} et \mathbf{Z} .
2. Montrez que les projecteurs orthogonaux sur les sous-espaces engendrés par les colonnes de \mathbf{X} et \mathbf{Z} sont égaux.

Exercice 1.6 Soient les matrices

$$\mathbf{A} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

Construisez les projecteurs sur

$$\text{Im}(\mathbf{A}) \quad \text{Im}(\mathbf{B})$$

notés respectivement \mathbf{P}_A et \mathbf{P}_B .

Si $\text{Im}(\mathbf{A}) \subset \text{Im}(\mathbf{B})$ vérifier le théorème des 3 perpendiculaires.

Chapitre 2

Géométrie des moindres carrés

2.1 Série statistique bivariable

On s'intéresse à deux variables x et y . Ces deux variables sont mesurées sur les n unités d'observation. Pour chaque unité, on obtient donc deux mesures. La série statistique est alors une suite de n couples des valeurs prises par les deux variables sur chaque individu :

$$(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n).$$

Chacune des deux variables peut être soit quantitative, soit qualitative.

2.1.1 Représentation graphique de deux variables

Dans ce cas, chaque couple est composé de deux valeurs numériques. Un couple de nombres (entiers ou réels) peut toujours être représenté comme un point dans un plan

$$(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n).$$

Exemple 2.1 On mesure le poids Y et la taille X de 20 individus.

TAB. 2.1 – Taille et poids de 20 individus

y_i	x_i	y_i	x_i
60	155	75	180
61	162	76	175
64	157	78	173
67	170	80	175
68	164	85	179
69	162	90	175
70	169	96	180
70	170	96	185
72	178	98	189
73	173	101	187

2.1.2 Analyse des variables

Les variables x et y peuvent être analysées séparément. On peut calculer tous les paramètres dont les moyennes et les variances :

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

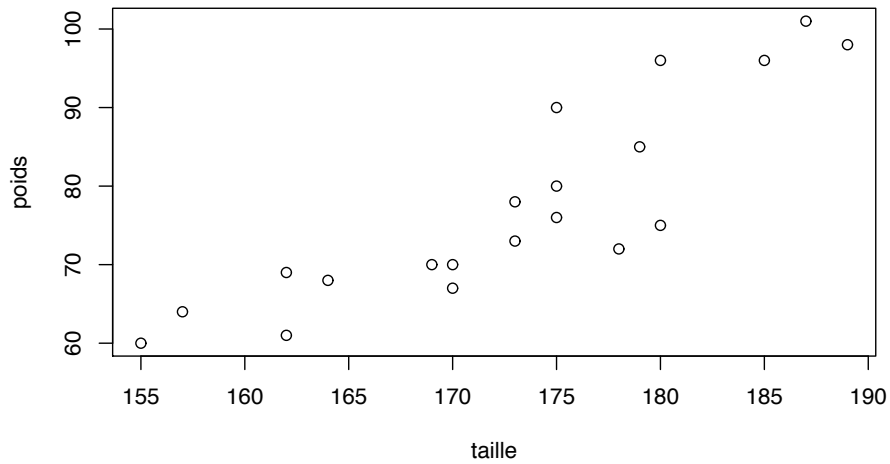


FIG. 2.1 – Le nuage de points

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Ces paramètres sont appelés *paramètres marginaux* : *variances marginales*, *moyennes marginales*, *écarts-types marginaux*, etc.

2.1.3 Covariance

La *covariance* est définie

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

Remarque 2.1

- La covariance peut prendre des valeurs positives, négatives ou nulles.
- Quand $x_i = y_i$, pour tout $i = 1, \dots, n$, la covariance est égale à la variance.
- La covariance peut également s'écrire

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}.$$

2.1.4 Corrélation

Le *coefficient de corrélation* est la covariance divisée par les deux écart-types marginaux

$$r_{xy} = \frac{s_{xy}}{s_x s_y}.$$

Le *coefficient de détermination* est le carré du coefficient de corrélation

$$r_{xy}^2 = \frac{s_{xy}^2}{s_x^2 s_y^2}.$$

Remarque 2.2

- Le coefficient de corrélation mesure la dépendance linéaire entre deux variables.

$$\begin{aligned} - & -1 \leq r_{xy} \leq 1 \\ - & 0 \leq r_{xy}^2 \leq 1 \end{aligned}$$

2.1.5 Droite de régression

La *droite de régression* est la droite qui ajuste au mieux un nuage de points au sens des moindres carrés.

On considère que la variable X est explicative et que la variable Y est dépendante. L'équation d'une droite est

$$y = a + bx.$$

Le coefficient a est appelé la constante, et le coefficient b la pente de la droite de régression. Le principe des *moindres carrés* consiste à chercher la droite qui minimise

$$M(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2.$$

Le minimum s'obtient en annulant les dérivées partielles par rapport à a et b .

$$\begin{cases} \frac{\partial M(a, b)}{\partial a} = - \sum_{i=1}^n 2(y_i - a - bx_i) = 0 \\ \frac{\partial M(a, b)}{\partial b} = - \sum_{i=1}^n 2(y_i - a - bx_i)x_i = 0. \end{cases}$$

On obtient un système de deux équations à deux inconnues, qui peuvent également s'écrire

$$\begin{cases} \bar{y} = a + b\bar{x} \\ \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i - b \sum_{i=1}^n x_i^2 = 0. \end{cases}$$

La première équation montre que la droite passe par le point (\bar{x}, \bar{y}) . De plus, on obtient

$$a = \bar{y} - b\bar{x}.$$

En remplaçant a par sa valeur dans la seconde équation divisée par n , on a

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n x_i y_i - (\bar{y} - b\bar{x})\bar{x} - b \frac{1}{n} \sum_{i=1}^n x_i^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x}\bar{y} - b \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right) \\ &= s_{xy} - bs_x^2 \\ &= 0, \end{aligned}$$

ce qui donne

$$\begin{cases} b = \frac{s_{xy}}{s_x^2} \\ a = \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x}. \end{cases}$$

La droite de régression est donc

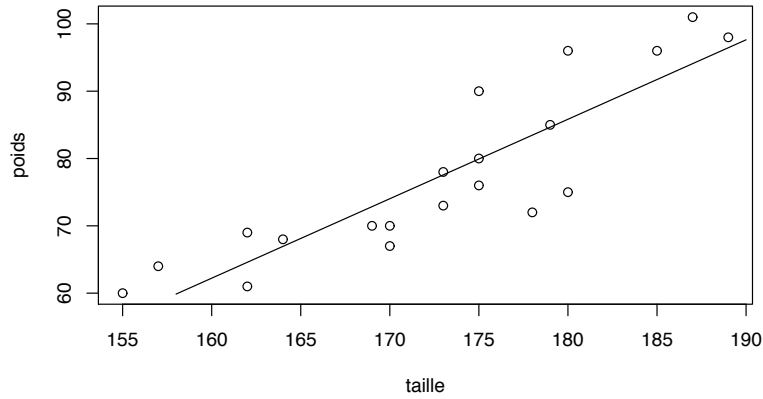
$$y = \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x} + \frac{s_{xy}}{s_x^2} x,$$

ce qui peut s'écrire aussi

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x}).$$

Remarque 2.3 La droite de régression de y en x n'est pas la même que la droite de régression de x en y .

FIG. 2.2 – La droite de régression



2.1.6 Résidus et valeurs ajustées

Les *valeurs ajustées* sont obtenues au moyen de la droite de régression :

$$y_i^* = a + bx_i.$$

Les valeurs ajustées sont les “prédictions” des y_i réalisées au moyen de la variable x et de la droite de régression de y en x .

Remarque 2.4 La moyenne des valeurs ajustées est \bar{y} .

Les *résidus* sont les différences entre les valeurs observées et les valeurs ajustées de la variable dépendante :

$$e_i = y_i - y_i^*.$$

Les résidus représentent la partie inexpliquée des y_i par la droite de régression.

Remarque 2.5

– La moyenne des résidus est nulle :

$$\sum_{i=1}^n e_i = 0.$$

– De plus,

$$\sum_{i=1}^n x_i e_i = 0.$$

2.1.7 Variance de régression et variance résiduelle

La *variance de régression* est la variance des valeurs ajustées.

$$s_Y^2 = \frac{1}{n} \sum_{i=1}^n (y_i^* - \bar{y})^2.$$

Théorème 2.1 La variance de régression peut également s’écrire

$$s_Y^2 = s_y^2 r^2,$$

où r^2 est le coefficient de détermination.

Démonstration

$$\begin{aligned}
s_Y^2 &= \frac{1}{n} \sum_{i=1}^n (y_i^* - \bar{y})^2 \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ \bar{y} + \frac{s_{xy}}{s_x^2} (x_i - \bar{x}) - \bar{y} \right\}^2 \\
&= \frac{s_{xy}^2}{s_x^4} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\
&= \frac{s_{xy}^2}{s_x^2} \\
&= s_y^2 \frac{s_{xy}^2}{s_x^2 s_y^2} \\
&= s_y^2 r^2.
\end{aligned}$$

□

La *variance résiduelle* est définie par :

$$s_e^2 = \frac{1}{n} \sum_{i=1}^n e_i^2.$$

Théorème 2.2 *La variance résiduelle peut également s'écrire*

$$s_e^2 = s_y^2 (1 - r^2),$$

où r^2 est le *coefficient de détermination*.

Démonstration

$$\begin{aligned}
s_e^2 &= \frac{1}{n} \sum_{i=1}^n e_i^2 \\
&= \frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2 \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ y_i - \bar{y} - \frac{s_{xy}}{s_x^2} (x_i - \bar{x}) \right\}^2 \\
&= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 + \frac{s_{xy}^2}{s_x^4} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 - 2 \frac{s_{xy}}{s_x^2} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\
&= s_y^2 + \frac{s_{xy}^2}{s_x^2} - 2 \frac{s_{xy}^2}{s_x^2} \\
&= s_y^2 \left(1 - \frac{s_{xy}^2}{s_x^2 s_y^2} \right).
\end{aligned}$$

□

Théorème 2.3 *La variance marginale est la somme de la variance de régression et de la variance résiduelle,*

$$s_y^2 = s_Y^2 + s_e^2.$$

La démonstration découle directement des deux théorèmes précédents.

2.2 La régression multivariée

2.2.1 Représentation matricielle des données

La matrice

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix}$$

peut représenter des données statistiques. Plus précisément, on suppose que x_{ij} représente la valeur prise par la variable explicative j sur l'unité statistique i . De même, le vecteur $\mathbf{y} = (y_1 \dots y_i \dots y_n)'$ représente les valeurs prises par la variable dépendante sur les n unités statistiques. Dans la plupart des applications, on supposera également que la première variable est la constante, c'est-à-dire que $x_{i1} = 1, i = 1, \dots, n$. (Néanmoins, il est intéressant dans certains cas particulier d'utiliser une régression sans constante.) On supposera alors que la matrice est de la forme :

$$\mathbf{X} = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix}.$$

Dans ce qui suit, on suppose toujours que la première variable est une constante. Si ce n'est pas le cas, nous le notifierons expressément.

2.2.2 Principe des moindres carrés

La régression de \mathbf{y} en \mathbf{X} au sens des moindres carrés consiste à chercher l'ajustement qui minimise en \mathbf{b} :

$$Q(\mathbf{b}) = \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}),$$

où $\mathbf{b} = (b_1 \dots b_p)'$. Pour obtenir le minimum, de $Q(\mathbf{b})$, on annule le vecteur des dérivées

$$\frac{\partial Q(\mathbf{b})}{\partial \mathbf{b}} = -2\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{0},$$

ce qui donne la valeur de \mathbf{b} :

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}.$$

En faisant l'hypothèse que $\mathbf{X}'\mathbf{X}$ est inversible, on peut déterminer \mathbf{b} :

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}.$$

2.2.3 Valeurs ajustées et résidus

Le vecteur des valeurs ajustées est le vecteur des prédictions de \mathbf{y} au moyen de \mathbf{X} et de \mathbf{b} , c'est-à-dire

$$\mathbf{y}^* = \mathbf{X}\mathbf{b} = \underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\mathbf{P}_X} \mathbf{y}.$$

Le vecteur des valeurs ajustées peut être interprété comme la projection de \mathbf{y} sur le sous-espace engendré par les colonnes de la matrice \mathbf{X} .

$$\mathbf{y}^* = \mathbf{P}_X \mathbf{y},$$

où \mathbf{P}_X est un projecteur (c'est-à-dire une matrice idempotente) sur le sous-espace engendré par les colonnes de \mathbf{X} .

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

Le vecteur des résidus est la différence entre \mathbf{y} et \mathbf{y}^* .

$$\mathbf{e} = \mathbf{y} - \mathbf{y}^* = \mathbf{y} - \mathbf{X}\mathbf{b} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}.$$

Le vecteur des valeurs ajustées peut également être interprété comme la projection de \mathbf{y} dans le noyau de \mathbf{X}' (ou l'orthogonal du sous-espace engendré par les colonnes de \mathbf{X}).

$$\mathbf{e} = \mathbf{P}_X^\perp \mathbf{y}, \quad (2.1)$$

où \mathbf{P}_X^\perp est un projecteur (c'est-à-dire une matrice idempotente) sur le noyau de \mathbf{X}' .

$$\mathbf{P}_X^\perp = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

Propriété 2.1

- $\mathbf{y} = \mathbf{y}^* + \mathbf{e}$,
- \mathbf{y}^* est une combinaison linéaire des colonnes de \mathbf{X} ,
- \mathbf{y}^* et \mathbf{e} sont orthogonaux,
- \mathbf{e} est orthogonal avec toutes les colonnes de \mathbf{X} , c'est-à-dire $\mathbf{e}'\mathbf{X} = \mathbf{0}$.

2.2.4 Variance de régression et variance résiduelle

Soit le vecteur de \mathbb{R}^n contenant n fois la moyenne de la variable y :

$$\bar{\mathbf{y}} = (\bar{y}, \dots, \bar{y})'.$$

La variance peut être définie simplement par :

$$s_y^2 = \frac{1}{n}(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

La variance de régression est la moyenne des valeurs ajustées :

$$s_Y^2 = \frac{1}{n}(\mathbf{y}^* - \bar{\mathbf{y}})'(\mathbf{y}^* - \bar{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^n (y_i^* - \bar{y})^2.$$

La variance résiduelle est la variance résiduelle :

$$s_e^2 = \frac{1}{n}\mathbf{e}'\mathbf{e} = \frac{1}{n}(\mathbf{y} - \mathbf{y}^*)'(\mathbf{y} - \mathbf{y}^*) = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2 = \frac{1}{n} \sum_{i=1}^n e_i^2.$$

2.2.5 Coefficient de détermination

Le coefficient de détermination vaut

$$R^2 = \frac{s_Y^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}.$$

Il est important de noter que le R^2 ne peut être calculé que si la régression inclut une constante. Si ce n'est pas le cas, le R^2 peut prendre des valeurs négatives. La racine carrée du coefficient de détermination est appelée le coefficient de corrélation multiple.

2.3 Matrice de variance-covariance et matrice de corrélation

Si la première colonne de la matrice \mathbf{X} contient uniquement des 1, alors on peut calculer les covariances entre les $p-1$ dernières variables. La matrice variance-covariance, de dimension $(p-1) \times (p-1)$, des variables explicatives est

$$\Sigma = \begin{pmatrix} s_2^2 & \cdots & s_{2j} & \cdots & s_{2p} \\ \vdots & & \vdots & & \vdots \\ s_{j2} & \cdots & s_j^2 & \cdots & s_{jp} \\ \vdots & & \vdots & & \vdots \\ s_{p2} & \cdots & s_{pj} & \cdots & s_p^2 \end{pmatrix}, \quad (2.2)$$

où

$$s_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij},$$

et

$$s_j^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2.$$

Si la première colonne de la matrice \mathbf{X} est une constante, alors la matrice variance-covariance est une matrice de dimension $(p-1) \times (p-1)$ correspondant aux $p-1$ dernières colonnes de \mathbf{X} .

On peut également construire la matrice diagonale des écart-types :

$$\mathbf{S} = \begin{pmatrix} s_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & s_j & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & s_p \end{pmatrix}.$$

La matrice des corrélations :

$$\mathbf{R} = \begin{pmatrix} 1 & \cdots & r_{2j} & \cdots & r_{2p} \\ \vdots & \ddots & \vdots & & \vdots \\ r_{j2} & \cdots & 1 & \cdots & r_{jp} \\ \vdots & & \vdots & \ddots & \vdots \\ r_{p2} & \cdots & r_{pj} & \cdots & 1 \end{pmatrix},$$

est obtenue par

$$\mathbf{R} = \mathbf{S}^{-1} \mathbf{\Sigma} \mathbf{S}^{-1}.$$

2.4 Corrélations partielles

Soit deux variables y et z et le vecteur de leurs valeurs \mathbf{y} et \mathbf{z} sur les n unités de l'échantillon. La matrice idempotente $\mathbf{P}_X^\perp = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ permet d'obtenir

- le vecteur des résidus de la régression de y en X

$$\mathbf{e}_{y|X} = \mathbf{P}_X^\perp \mathbf{y},$$

- le vecteur des résidus de la régression de z en X

$$\mathbf{e}_{z|X} = \mathbf{P}_X^\perp \mathbf{z}.$$

Le coefficient de corrélation partielle est le coefficient de corrélation entre $\mathbf{e}_{y|X}$ et $\mathbf{e}_{z|X}$. Si la première colonne de la matrice \mathbf{X} contient une colonne de constante, alors ce coefficient s'écrit

$$r_{yz|x_2, \dots, x_p} = \frac{\mathbf{e}_{y|X}' \mathbf{e}_{z|X}}{\sqrt{\mathbf{e}_{y|X}' \mathbf{e}_{y|X} \mathbf{e}_{z|X}' \mathbf{e}_{z|X}}} = \frac{\mathbf{y}' \mathbf{P}_X^\perp \mathbf{z}}{\sqrt{\mathbf{y}' \mathbf{P}_X^\perp \mathbf{y} \mathbf{z}' \mathbf{P}_X^\perp \mathbf{z}}}.$$

Le coefficient de corrélation partielle mesure la corrélation entre les variables y et z auxquelles on a enlevé la partie explicable par les variables de \mathbf{X} .

2.5 Condition pour que la somme des résidus soit nulle

La matrice \mathbf{X} peut contenir une variable constante de manière explicite, c'est-à-dire qu'une des colonnes de cette matrice contient une variable constante. La constante peut également être définie de manière implicite, ce qui signifie qu'il existe une combinaison linéaire des colonnes de \mathbf{X} qui permet d'obtenir une colonne de uns. Formellement, on suppose qu'il existe un vecteur $\boldsymbol{\lambda}$ de \mathbb{R}^p tel que $\mathbf{X}\boldsymbol{\lambda} = \mathbf{1}_n = (1 \ \cdots \ 1 \ \cdots \ 1)'$.

Théorème 2.4 *Si la matrice \mathbf{X} contient une variable constante définie de manière explicite ou implicite, alors la somme des résidus est nulle.*

Démonstration

On a

$$\sum_{i=1}^n e_i = \mathbf{1}_n' \mathbf{e}$$

Or, il existe un vecteur $\boldsymbol{\lambda}$ de \mathbb{R}^p tel que $\mathbf{X}\boldsymbol{\lambda} = \mathbf{1}_n$. On obtient donc

$$\begin{aligned} \sum_{i=1}^n e_i &= \boldsymbol{\lambda}' \mathbf{X}' \mathbf{e} \\ &= \boldsymbol{\lambda}' \mathbf{X}' \{ \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \} \mathbf{y} \\ &= \left\{ \boldsymbol{\lambda}' \mathbf{X}' - \underbrace{\boldsymbol{\lambda}' \mathbf{X}' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'}_{\mathbf{I}} \right\} \mathbf{y} \\ &= 0. \end{aligned}$$

□

Une autre manière d'aboutir à ce résultat, consiste à se remémorer que le vecteur de résidus est toujours orthogonal aux variables explicatives, c'est-à-dire

$$\mathbf{e}' \mathbf{X} = \mathbf{0}.$$

Or, s'il existe un vecteur $\boldsymbol{\lambda}$ de \mathbb{R}^p tel que $\mathbf{X}\boldsymbol{\lambda} = \mathbf{1}_n$, alors

$$\mathbf{e}' \mathbf{X} \boldsymbol{\lambda} = \mathbf{e}' \mathbf{1}_n = \mathbf{0}' \boldsymbol{\lambda} = 0.$$

Si la somme des résidus est nulle, la moyenne des valeurs ajustées est égale à la moyenne des valeurs observées, autrement dit

$$\frac{1}{n} \sum_{i=1}^n y_i^* = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}.$$

2.6 Décomposition en sommes de carrés

Théorème 2.5 *Soit une régression pour laquelle la constante est une variable explicative (éventuellement définie de manière implicite), alors la somme des carrés totale des écarts à la moyenne*

$$SC_{tot} = (\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) = \sum_{i=1}^n (y_i - \bar{y})^2$$

se décompose donc en une somme de deux termes :

- la somme des carrés expliquée par la régression,

$$SC_{regr} = (\mathbf{y}^* - \bar{\mathbf{y}})'(\mathbf{y}^* - \bar{\mathbf{y}}) = \sum_{i=1}^n (y_i^* - \bar{y})^2,$$

– la somme des carrés des résidus

$$SC_{res} = \mathbf{e}'\mathbf{e} = \sum_{i=1}^n (y_i - y_i^*)^2 = \sum_{i=1}^n e_i^2. \quad (2.3)$$

Démonstration En notant $\bar{\mathbf{y}}$ le vecteur de \mathbb{R}^n contenant n fois la moyenne \bar{y} , on a

$$\mathbf{y} - \bar{\mathbf{y}} = \mathbf{y}^* - \bar{\mathbf{y}} + \mathbf{e}.$$

Donc,

$$(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) = (\mathbf{y}^* - \bar{\mathbf{y}} + \mathbf{e})'(\mathbf{y}^* - \bar{\mathbf{y}} + \mathbf{e}) = (\mathbf{y}^* - \bar{\mathbf{y}})'(\mathbf{y}^* - \bar{\mathbf{y}}) + \mathbf{e}'\mathbf{e} + 2\mathbf{e}'(\mathbf{y}^* - \bar{\mathbf{y}})$$

or \mathbf{e} et $(\mathbf{y}^* - \bar{\mathbf{y}})$ sont orthogonaux. En effet \mathbf{e} est toujours orthogonal à \mathbf{y}^* et, $\mathbf{e}'\bar{\mathbf{y}} = \bar{y} \sum_{i=1}^n e_i$. Or la somme des résidus est nulle quand la constante est une variable explicative. Donc $\mathbf{e}'(\mathbf{y}^* - \bar{\mathbf{y}}) = 0$, ce qui donne finalement

$$(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) = (\mathbf{y}^* - \bar{\mathbf{y}})'(\mathbf{y}^* - \bar{\mathbf{y}}) + \mathbf{e}'\mathbf{e}.$$

□

2.7 Régression avec les données centrées

Supposons que la première colonne de la matrice \mathbf{X} soit composée de constantes :

$$\mathbf{X} = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix}.$$

Dans ce cas, la régression multiple s'écrit :

$$y_i = b_1 + x_{i2}b_2 + x_{i3}b_3 + \cdots + x_{ip}b_p + e_i. \quad (2.4)$$

On peut aussi travailler avec les données centrées. En sommant sur les i et en divisant par n l'équation (2.4), on obtient :

$$\bar{y} = b_1 + \bar{x}_2b_2 + \bar{x}_3b_3 + \cdots + \bar{x}_pb_p, \quad (2.5)$$

et donc en soustrayant (2.5) à (2.4), on a finalement :

$$y_i - \bar{y} = (x_{i2} - \bar{x}_2)b_2 + (x_{i3} - \bar{x}_3)b_3 + \cdots + (x_{ip} - \bar{x}_p)b_p + e_i. \quad (2.6)$$

Définissons maintenant

1. $\tilde{\mathbf{b}}$: le vecteur de \mathbb{R}^{p-1} composé des $p-1$ dernières composantes de \mathbf{b} , $\tilde{\mathbf{b}} = (b_2, b_3, \dots, b_p)'$,
2. $\tilde{\mathbf{X}}$: la matrice $n \times (p-1)$ composée des $p-1$ dernières colonnes de \mathbf{X} ,

$$\tilde{\mathbf{X}} = \begin{pmatrix} x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & & \vdots & & \vdots \\ x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix},$$

3. $\mathbf{1} = (1, 1, \dots, 1)'$: le vecteur colonne de n uns,
4. la matrice idempotente qui centre les valeurs :

$$\mathbf{P}_c = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n} = \begin{pmatrix} 1-1/n & -1/n & -1/n & \cdots & -1/n \\ -1/n & 1-1/n & -1/n & \cdots & -1/n \\ -1/n & -1/n & 1-1/n & \cdots & -1/n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & -1/n & \cdots & 1-1/n \end{pmatrix}, \quad (2.7)$$

5. $\mathbf{y}_c = \mathbf{P}_c \mathbf{y} = \mathbf{y} - \mathbf{1}\bar{y} = \mathbf{y} - \bar{\mathbf{y}} = (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})'$

6. $\mathbf{X}_c = \mathbf{P}_c \tilde{\mathbf{X}}$ la matrice $\tilde{\mathbf{X}}$ centrée

$$\mathbf{X}_c = \begin{pmatrix} x_{12} - \bar{x}_2 & \cdots & x_{1j} - \bar{x}_j & \cdots & x_{1p} - \bar{x}_p \\ \vdots & & \vdots & & \vdots \\ x_{i2} - \bar{x}_2 & \cdots & x_{ij} - \bar{x}_j & \cdots & x_{ip} - \bar{x}_p \\ \vdots & & \vdots & & \vdots \\ x_{n2} - \bar{x}_2 & \cdots & x_{nj} - \bar{x}_j & \cdots & x_{np} - \bar{x}_p \end{pmatrix}.$$

La régression multiple peut maintenant s'écrire :

$$\mathbf{y}_c = \mathbf{X}_c \tilde{\mathbf{b}} + \mathbf{e}.$$

Le vecteur $\tilde{\mathbf{b}}$ est évidemment défini par

$$\tilde{\mathbf{b}} = (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c' \mathbf{y}_c = \left(\frac{\mathbf{X}_c' \mathbf{X}_c}{n} \right)^{-1} \frac{\mathbf{X}_c' \mathbf{y}_c}{n}. \quad (2.8)$$

Cette présentation est intéressante à plus d'un titre. En effet $(\mathbf{X}_c' \mathbf{X}_c)/n$ n'est autre que la matrice variance-covariance Σ donnée en (2.2).

$$\Sigma = \frac{\mathbf{X}_c' \mathbf{X}_c}{n} = \begin{pmatrix} s_2^2 & \cdots & s_{2j} & \cdots & s_{2p} \\ \vdots & & \vdots & & \vdots \\ s_{j2} & \cdots & s_j^2 & \cdots & s_{jp} \\ \vdots & & \vdots & & \vdots \\ s_{p2} & \cdots & s_{pj} & \cdots & s_p^2 \end{pmatrix},$$

et $\mathbf{X}_c' \mathbf{y}_c/n$ est le vecteur des covariances entre les variables explicatives et la variable dépendante :

$$\frac{\mathbf{X}_c' \mathbf{y}_c}{n} = \begin{pmatrix} s_{2y} \\ \vdots \\ s_{jy} \\ \vdots \\ s_{py} \end{pmatrix}.$$

où

$$s_{jy} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y}),$$

pour $j = 2, \dots, n$.

Comme,

$$\mathbf{y}_c = \mathbf{X}_c \tilde{\mathbf{b}} + \mathbf{e},$$

la décomposition en somme de carrés vient directement :

$$\mathbf{y}_c' \mathbf{y}_c = (\mathbf{X}_c \tilde{\mathbf{b}} + \mathbf{e})' (\mathbf{X}_c \tilde{\mathbf{b}} + \mathbf{e}) = \tilde{\mathbf{b}}' \mathbf{X}_c' \mathbf{X}_c \tilde{\mathbf{b}} + \mathbf{e}' \mathbf{e} + \underbrace{2\mathbf{e}' \mathbf{X}_c \tilde{\mathbf{b}}}_0.$$

Le dernier terme s'annule, car les résidus observés sont orthogonaux aux colonnes de \mathbf{X} . On peut donc à nouveau décomposer la somme des carrés totales en une somme de deux termes :

$$SC_{tot} = SC_{regr} + SC_{res},$$

où

– la somme des carrés totales

$$SC_{tot} = \mathbf{y}_c' \mathbf{y}_c = \sum_{i=1}^n (y_i - \bar{y})^2, \quad (2.9)$$

- la somme des carrés expliquée par la régression,

$$SC_{regr} = \tilde{\mathbf{b}}' \mathbf{X}'_c \mathbf{X}_c \tilde{\mathbf{b}}, \quad (2.10)$$

car

$$\mathbf{X}_c \tilde{\mathbf{b}} = \begin{pmatrix} \sum_{j=2}^p b_j (x_{1j} - \bar{x}_j) \\ \vdots \\ \sum_{j=2}^p b_j (x_{ij} - \bar{x}_j) \\ \vdots \\ \sum_{j=2}^p b_j (x_{nj} - \bar{x}_j) \end{pmatrix} = \begin{pmatrix} y_1^* - \bar{y} \\ \vdots \\ y_i^* - \bar{y} \\ \vdots \\ y_n^* - \bar{y} \end{pmatrix}$$

et que donc

$$\tilde{\mathbf{b}}' \mathbf{X}'_c \mathbf{X}_c \tilde{\mathbf{b}} = (\mathbf{y}^* - \bar{\mathbf{y}})' (\mathbf{y}^* - \bar{\mathbf{y}}) = \sum_{i=1}^n (y_i^* - \bar{y})^2 = SC_{regr},$$

- la somme des carrés des résidus

$$SC_{res} = \mathbf{e}' \mathbf{e} = \sum_{i=1}^n e_i^2. \quad (2.11)$$

2.8 Retour au cas bivarié

2.8.1 Méthode 1

Le cas particulier le plus fréquemment étudié consiste à utiliser deux variables explicatives ($p = 2$) : une constante et une variable x_i . Dans ce cas,

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

On a alors

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

$$\begin{aligned} (\mathbf{X}' \mathbf{X})^{-1} &= \frac{1}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \\ &= \frac{1}{n^2 \left\{ \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 \right\}} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \\ &= \frac{1}{n^2 s_x^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \\ &= \frac{1}{n^2 s_x^2} \begin{pmatrix} ns_x^2 + n\bar{x}^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \\ &= \frac{1}{ns_x^2} \begin{pmatrix} s_x^2 + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \end{aligned}$$

où

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2.$$

De plus,

$$\mathbf{X}' \mathbf{y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix} = n \begin{pmatrix} \bar{y} \\ s_{xy} + \bar{x}\bar{y} \end{pmatrix},$$

ce qui permet de calculer \mathbf{b}

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \frac{1}{s_x^2} \begin{pmatrix} (s_x^2 + \bar{x}^2)\bar{y} - \bar{x}(s_{xy} + \bar{x}\bar{y}) \\ -\bar{x}\bar{y} + (s_{xy} + \bar{x}\bar{y}) \end{pmatrix} = \begin{pmatrix} \bar{y} - \bar{x} \frac{s_{xy}}{s_x^2} \\ \frac{s_{xy}}{s_x^2} \end{pmatrix}.$$

En général, on note

$$b_1 = \bar{y} - \bar{x} \frac{s_{xy}}{s_x^2},$$

et

$$b_2 = \frac{s_{xy}}{s_x^2}.$$

On a finalement le vecteur des valeurs ajustées

$$\mathbf{y}^* = (y_i^*) = \mathbf{X}\mathbf{b},$$

avec

$$y_i^* = 1 \times b_1 + x_i b_2 = \left(\bar{y} - \bar{x} \frac{s_{xy}}{s_x^2} \right) + x_i \frac{s_{xy}}{s_x^2} = \bar{y} + (x_i - \bar{x}) \frac{s_{xy}}{s_x^2}.$$

Le cas bivarié consiste donc à utiliser deux variables explicatives, la première est la constante et la seconde est la variable x .

2.8.2 Méthode 2

Une autre manière de traiter le même problème est de d'utiliser les données centrées. Dans ce cas, on a

$$\mathbf{y}_c = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_i - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}, \quad \mathbf{X}_c = \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_i - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}.$$

On obtient

$$\mathbf{X}_c' \mathbf{X}_c = n s_x^2, \quad \mathbf{X}_c' \mathbf{y}_c = n s_{xy} \quad \text{et} \quad \tilde{\mathbf{b}} = (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c' \mathbf{y}_c = \frac{s_{xy}}{s_x^2}.$$

Il reste à déduire b_1 de l'équation

$$\bar{y} = b_1 + \frac{s_{xy}}{s_x^2} \bar{x},$$

ce qui donne

$$b_1 = \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x}, \quad \text{et} \quad b_2 = \frac{s_{xy}}{s_x^2}.$$

Exercices

Exercice 2.1 Au moyen du tableau 2.1, calculez

1. tous les paramètres marginaux,
2. la covariance,
3. la droite de regression de la taille par le poids,
4. les résidus et les valeurs ajustées,
5. le coefficient de la régression, la variance résiduelle et la variance de régression.

Exercice 2.2 En quoi consiste la régression,

1. quand une seule variable x est utilisée,

2. quand seule la constante est utilisée,
3. quand l'échantillon est partitionné en p parties notées U_1, \dots, U_p et que $x_{ij} = 1$ si l'unité i est dans la partie j et 0 sinon ?

Représentez les deux droites de régression, pour les points 1 et 2.

Exercice 2.3 À partir du tableau 2.2, calculez les coefficients de corrélation et de régression a et b de la régression de y en x .

TAB. 2.2 – Données pour les variables x et y

t	y_t	x_t
1983	7389.99	8000
1984	8169.65	9000
1985	8831.71	9500
1986	8652.84	9500
1987	8788.08	9800
1988	9616.21	11000
1989	10593.45	12000
1990	11186.11	13000
1991	12758.09	15000
1992	13869.62	16000
Somme	99855.75	112800
Moyenne	9985.57	11280

Exercice 2.4 Application du principe des moindres carrés :

Soit

$$Q(\mathbf{b}) = \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2,$$

qui peut également s'écrire

$$Q(b_1, \dots, b_p) = \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} b_j \right)^2.$$

Annulez les dérivées partielles

$$\frac{\partial Q}{\partial b_j} = 0.$$

Écrivez ensuite ce système de p équations à p inconnues sous forme matricielle.

Exercice 2.5 (extrait de Cohen and Pradel, 1993) Parmi les relations suivantes donnant y en fonction de x et peut être z , quelles sont celles qui peuvent être déterminées à l'aide d'un modèle linéaire ?

1. $y = ax + b$
2. $y = ax^2 + b$
3. $y = ax^2 + bx + c$
4. $y = ax^3 + b$
5. $y = x^a z^b$
6. $y = \frac{1}{1 + a \exp^{-bx}}$

7. $y = \frac{c}{1 + a \exp^{-bx}}$
8. $y = x^2 + ax + b$
9. $y = a \log(x) + 5$
10. $y = ab^x + c^z$
11. $y = \frac{a}{x-1} + b$
12. $y = a \ln(x) + bz^5 + c$

Exercice 2.6 Dans un modèle où on cherche un ajustement linéaire de Y sur X et la constante, on dispose des résultats suivants portant sur 52 observations :

$$y_t^* = 1.286 - 0.43x_t,$$

$$\bar{x} = 1.063 \quad s_y^2 = 0.00137 \quad s_x^2 = 0.00686$$

Déterminez successivement les valeurs du coefficient de corrélation linéaire entre X et Y , le coefficient de détermination R^2 et les SC_{tot} , SC_{res} et SC_{regr} de la régression.

Exercice 2.7 Soit une matrice

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

Calculez le coefficient de corrélation partiel $r_{yz|x}$ et exprimez-le en fonction des coefficients de corrélation (non-partiels) r_{yz} , r_{xy} et r_{yx} .

Exercice 2.8 A partir des données du tableau 2.3, calculez le vecteur des coefficients de la régression des y_i en x_{i1} et x_{i2} (avec une constante). Les données sont les suivantes : Indication : travailler avec la matrice

TAB. 2.3 – Données sur le travail, le capital et la production

Entreprise(i)	Travail(x_i)	Capital(z_i)	Production(y_i)
1	7389.99	8000	60
2	8169.65	9000	120
3	8831.71	9500	190
4	8652.84	9500	250
5	8788.08	9800	300
6	9616.21	11000	360
7	10593.45	12000	380
8	11186.11	13000	430
9	12758.09	15000	440

variance-covariance permet de simplifier considérablement les calculs (voir calcul de $\tilde{\mathbf{b}}$ dans l'expression (2.8)).

Exercice 2.9 On procède à l'estimation d'un modèle linéaire avec une constante. Les informations disponibles sont :

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 250 & 0 & 0 \\ 0 & 200 & 100 \\ 0 & 100 & 100 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 500 \\ 140 \\ 100 \end{pmatrix}$$

$$\mathbf{y}'\mathbf{y} = 200$$

1. Calculez :
 - (a) La taille de l'échantillon
 - (b) $\sum_{i=1}^n x_{i1}$; $\sum_{i=1}^n x_{i1}^2$
 - (c) $\sum_{i=1}^n x_{i2}$; $\sum_{i=1}^n x_{i2}^2$
 - (d) $\sum_{i=1}^n x_{i1}x_{i2}$
2. Calculez la droite de régression des y_i en x_{i1} et x_{i2} (avec constante).
3. Calculez la matrice variance-covariance des variables explicatives.
4. Calculez la matrice des corrélations des variables explicatives.

Exercice 2.10 Retour au cas bivarié.

Calculez les droites de régression de

$$\begin{cases} y & \text{en } x \\ x & \text{en } y \end{cases}$$

1. Si sur un graphique on a x en abscisse et y en ordonnée, quelle est la droite ayant la plus grande pente ? (Attention la réponse dépend de la valeur du coefficient de corrélation)
2. Quelle est le point d'intersection des deux droites (faites les calculs) ?

Chapitre 3

Rappel sur le calcul des probabilités, les variables aléatoires, et l'inférence statistique

3.1 Probabilités

3.1.1 Événement

Une expérience est dite aléatoire si on ne peut pas prédire *a priori* son résultat. On note ω un résultat possible de cette expérience aléatoire. L'ensemble de tous les résultats possibles est noté Ω . Par exemple, si on jette deux pièces de monnaie, on peut obtenir les résultats

$$\Omega = \{(P, P), (F, P), (P, F), (F, F)\},$$

avec F pour “face” et P pour “pile”. Un événement est une assertion logique sur une expérience aléatoire. Formellement, un événement est un sous-ensemble de Ω .

Exemple 3.1 L'expérience peut consister à jeter un dé, alors

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

et un événement, noté A , est “obtenir un nombre pair”. On a alors

$$A = \{2, 4, 6\}.$$

Soient deux événements A et B , si $A \cap B = \emptyset$, alors on dit qu'ils sont mutuellement exclusifs.

Exemple 3.2 Par exemple, si on jette un dé, l'événement “obtenir un nombre pair” et l'événement “obtenir un nombre impair” ne peuvent pas être obtenus en même temps. Ils sont mutuellement exclusifs. D'autre part, si l'on jette un dé, les événements A : “obtenir un nombre pair” n'est pas mutuellement exclusif avec l'événement B : “obtenir un nombre inférieur ou égal à 3”. En effet, l'intersection de A et B est non-vide et consiste en l'événement “obtenir 2”.

On appelle complémentaire d'un événement

$$\overline{A} = \Omega \setminus A.$$

On va associer à Ω l'ensemble \mathcal{A} de toutes les parties (ou sous-ensembles) de Ω .

Exemple 3.3 Si on jette une pièce de monnaie alors $\Omega = \{P, F\}$, et

$$\mathcal{A} = \{\emptyset, \{F\}, \{P\}, \{P, F\}\}.$$

Définition 3.1 Les événements A_1, \dots, A_n forment un système complet d'événements, si ils constituent une partition de Ω , c'est-à-dire si

- tous les couples A_i, A_j sont mutuellement exclusifs quand $i \neq j$,
- $\bigcup_{i=1}^n A_i = \Omega$.

3.1.2 Axiomatique des Probabilités

Définition 3.2 Une probabilité $P(\cdot)$ est une application de \mathcal{A} dans $[0, 1]$, telle que :

- $\Pr(\Omega) = 1$,
- Pour tout ensemble dénombrable d'événements A_1, \dots, A_n tels que $A_i \cap A_j = \emptyset$, pour tout $i \neq j$,

$$\Pr\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \Pr(A_i).$$

A partir des axiomes, on peut déduire les propriétés suivantes :

- $\Pr(\emptyset) = 0$,
- $\Pr(\bar{A}) = 1 - \Pr(A)$,
- $\Pr(A) \leq \Pr(B)$ si $A \subset B$,
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$,
- $\Pr(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \Pr(A_i)$,
- Si A_1, \dots, A_n forment un système complet d'événements, alors

$$\sum_{i=1}^n \Pr(B \cap A_i) = \Pr(B).$$

3.1.3 Probabilités conditionnelles et indépendance

Définition 3.3 Soient deux événements A et B , si $\Pr(B) > 0$, alors

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Définition 3.4 Deux événements A et B sont dits indépendants si

$$\Pr(A|B) = \Pr(A).$$

On peut montrer facilement que si A et B sont indépendants, alors

$$\Pr(A \cap B) = \Pr(A)\Pr(B).$$

3.1.4 Théorème des probabilités totales et théorème de Bayes

Théorème 3.1 (des probabilités totales) Soit A_1, \dots, A_n un système complet d'événements, alors

$$\Pr(B) = \sum_{i=1}^n \Pr(A_i)\Pr(B|A_i).$$

En effet,

$$\sum_{i=1}^n \Pr(A_i)\Pr(B|A_i) = \sum_{i=1}^n \Pr(B \cap A_i).$$

Comme les événements $A_i \cap B$ sont mutuellement exclusifs,

$$\sum_{i=1}^n \Pr(B \cap A_i) = \Pr\bigcup_{i=1}^n (B \cap A_i) = \Pr(B).$$

Théorème 3.2 (de Bayès) Soit A_1, \dots, A_n un système complet d'événements, alors

$$\Pr(A_i|B) = \frac{\Pr(A_i)\Pr(B|A_i)}{\sum_{j=1}^n \Pr(A_j)\Pr(B|A_j)}.$$

En effet, par le théorème des probabilités totales,

$$\frac{\Pr(A_i)\Pr(B|A_i)}{\sum_{j=1}^n \Pr(A_j)\Pr(B|A_j)} = \frac{\Pr(B \cap A_i)}{\Pr(B)} = \Pr(A_i|B).$$

3.2 Variables aléatoires

3.2.1 Définition

La notion de variable aléatoire formalise l'association d'une valeur au résultat d'une expérience aléatoire.

Définition 3.5 Une variable aléatoire X est une application de l'ensemble fondamental Ω dans \mathbb{R} .

Exemple 3.4 On considère une expérience aléatoire consistant à lancer deux pièces de monnaie. L'ensemble des résultats possibles est

$$\Omega = \{(F, F), (F, P), (P, F), (P, P)\}.$$

Chacun des éléments de Ω a une probabilité $1/4$. Une variable aléatoire va associer une valeur à chacun des éléments de Ω . Considérons la variable aléatoire représentant le nombre de "Faces" obtenus :

$$X = \begin{cases} 0 & \text{avec une probabilité } 1/4 \\ 1 & \text{avec une probabilité } 1/2 \\ 2 & \text{avec une probabilité } 1/4. \end{cases}$$

3.2.2 Variables aléatoires discrètes

Définition, espérance et variance

Une variable aléatoire discrète prend uniquement des valeurs entières (de \mathbb{Z}).

Une distribution de probabilité $p_X(x)$ est une fonction qui associe à chaque valeur entière une probabilité.

$$p_X(x) = \Pr(X = x), x \in \mathbb{Z}.$$

La fonction de répartition est définie par

$$F_X(x) = \Pr(X \leq x) = \sum_{z \leq x} p_X(z).$$

L'espérance mathématique d'une variable aléatoire discrète est donné par

$$\mu = E(X) = \sum_{x \in \mathbb{Z}} xp_X(x),$$

et sa variance

$$\sigma^2 = \text{var}(X) = E\left(\{X - E(X)\}^2\right) = \sum_{x \in \mathbb{Z}} p_X(x)(x - \mu)^2.$$

Variable indicatrice ou bernoullienne

La variable indicatrice X de paramètre $p \in [0, 1]$ a la distribution de probabilité suivante :

$$X = \begin{cases} 1 & \text{avec une probabilité } p \\ 0 & \text{avec une probabilité } 1 - p. \end{cases}$$

L'espérance vaut

$$\mu = E(X) = 0 \times (1 - p) + 1 \times p = p,$$

et la variance vaut

$$\sigma^2 = \text{var}(X) = E(X - p)^2 = (1 - p)(0 - p)^2 + p(1 - p)^2 = p(1 - p).$$

Exemple 3.5 On tire au hasard une boule dans une urne contenant 18 boules rouges et 12 boules blanches. Si X vaut 1 si la boule est rouge et 0 sinon, alors X a une loi bernoullienne de paramètre $p = 18/(18+12) = 0.6$.

Variable binomiale

Une variable X suit une loi binomiale de paramètre $0 < p < 1$ et d'exposant n , si

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \dots, n-1, n,$$

où

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

La somme de ces probabilités vaut 1, en effet

$$\sum_{x=0}^n \Pr(X = x) = \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} = \{p + (1-p)\}^n = 1.$$

L'espérance et la variance sont données par

$$E(X) = np, \quad \text{var}(X) = np(1-p).$$

Exemple 3.6 On tire au hasard avec remise et de manière indépendante 5 boules dans une urne contenant

18 boules rouges et 12 boules blanches. Si X est le nombre de boules rouges obtenues, alors X a une loi binomiale de paramètre $p = 18/(18+12) = 0.6$, et d'exposant $n = 5$. Donc,

$$\Pr(X = x) = \binom{5}{x} 0.6^x 0.4^{5-x}, x = 0, 1, \dots, 4, 5,$$

ce qui donne

$$\begin{aligned} \Pr(X = 0) &= \frac{5!}{0!(5-0)!} 0.6^0 \times 0.4^{5-0} = 1 \times 0.4^5 = 0.01024 \\ \Pr(X = 1) &= \frac{5!}{1!(5-1)!} 0.6^1 \times 0.4^{5-1} = 5 \times 0.6^1 \times 0.4^4 = 0.0768 \\ \Pr(X = 2) &= \frac{5!}{2!(5-2)!} 0.6^2 \times 0.4^{5-2} = 10 \times 0.6^2 \times 0.4^3 = 0.2304 \\ \Pr(X = 3) &= \frac{5!}{3!(5-3)!} 0.6^3 \times 0.4^{5-3} = 10 \times 0.6^3 \times 0.4^2 = 0.3456 \\ \Pr(X = 4) &= \frac{5!}{4!(5-4)!} 0.6^4 \times 0.4^{5-4} = 5 \times 0.6^4 \times 0.4^1 = 0.2592 \\ \Pr(X = 5) &= \frac{5!}{5!(5-5)!} 0.6^5 \times 0.4^{5-5} = 1 \times 0.6^5 = 0.07776 \end{aligned}$$

Variable de Poisson

La variable X suit une loi de Poisson, de paramètre $\lambda \in \mathbb{R}^+$ si

$$\Pr(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, 3, \dots$$

L'espérance et la variance d'une loi de Poisson sont égales au paramètre λ

$$E(X) = \lambda, \quad \text{var}(X) = \lambda.$$

3.2.3 Variable aléatoire continue

Définition, espérance et variance

Une variable aléatoire continue prend des valeurs dans \mathbb{R} ou dans un intervalle de \mathbb{R} .

La probabilité qu'une variable aléatoire continue soit inférieure à une valeur particulière est donnée par sa fonction de répartition.

$$\Pr(X \leq x) = F(x).$$

La fonction de répartition d'une variable aléatoire continue est toujours :

- dérivable,
- positive : $F(x) \geq 0$, pour tout x ,
- croissante,
- $\lim_{x \rightarrow \infty} F(x) = 1$,
- $\lim_{x \rightarrow -\infty} F(x) = 0$.

On a

$$\Pr(a \leq X \leq b) = F(b) - F(a).$$

La fonction de densité d'une variable aléatoire continue est la dérivée de la fonction de répartition en un point

$$f(x) = \frac{dF(x)}{dx}.$$

Une fonction de densité est toujours :

- positive : $f(x) \geq 0$, pour tout x ,
- d'aire égale à un : $\int_{-\infty}^{\infty} f(x)dx = 1$.

On a évidemment la relation :

$$F(b) = \int_{-\infty}^b f(x)dx.$$

La probabilité que la variable aléatoire soit inférieure à une valeur quelconque vaut :

$$\Pr(X \leq a) = \int_{-\infty}^a f(x)dx = F(a)$$

La probabilité que la variable aléatoire prenne une valeur comprise entre a et b vaut

$$\Pr(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a).$$

Si la variable aléatoire est continue, la probabilité qu'elle prenne exactement une valeur quelconque est nulle :

$$\Pr(X = a) = 0.$$

L'espérance d'une variable aléatoire continue est définie par :

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$

et la variance

$$\text{var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx.$$

Variable uniforme

Une variable est dite uniforme dans un intervalle $[a, b]$, (avec $a < b$) si sa répartition est :

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ (x - a)/(b - a) & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b. \end{cases}$$

Sa densité est alors

$$f(x) = \begin{cases} 0 & \text{si } x < a \\ 1/(b - a) & \text{si } a \leq x \leq b \\ 0 & \text{si } x > b. \end{cases}$$

On peut montrer que

$$\mu = E(X) = \frac{b+a}{2}$$

et

$$\sigma^2 = \text{var}(X) = \frac{(b-a)^2}{12}.$$

Les logiciels gènèrent en général des variables aléatoires uniformes dans $[0,1]$.

Variable normale

Une variable aléatoire X est dite normale si sa densité vaut

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2. \quad (3.1)$$

De manière synthétique, pour noter que X a une distribution normale de moyenne μ et de variance σ^2 on écrit :

$$X \sim N(\mu, \sigma^2).$$

On peut montrer que

$$E(X) = \mu,$$

et

$$\text{var}(X) = \sigma^2.$$

La fonction de répartition vaut

$$F_{\mu,\sigma^2}(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2} \left(\frac{u-\mu}{\sigma} \right)^2 du.$$

3.2.4 Distribution bivariable

Deux variables aléatoires peuvent avoir une distribution jointe.

Cas continu

Soit deux variables aléatoires X et Y continues, leur distribution de densité $f(x,y)$ est une fonction continue, positive, et telle que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1.$$

La fonction de répartition jointe est définie par

$$F(x,y) = \Pr(X \leq x \text{ et } Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u,v) dv du.$$

On appelle densités marginales les fonctions

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy, \text{ et } f_Y(y) = \int_{-\infty}^{\infty} f(x,y) dx.$$

Avec les distributions marginales, on peut définir les moyennes marginales, et les variances marginales :

$$\begin{aligned} \mu_X &= \int_{-\infty}^{\infty} x f_X(x) dx, \text{ et } \mu_Y = \int_{-\infty}^{\infty} y f_Y(y) dy, \\ \sigma_X^2 &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx, \text{ et } \sigma_Y^2 = \int_{-\infty}^{\infty} (y - \mu_Y)^2 f_Y(y) dy. \end{aligned}$$

On appelle densités conditionnelles, les fonctions

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} \text{ et } f(y|x) = \frac{f(x,y)}{f_X(x)}.$$

Avec les distributions conditionnelles, on peut définir les moyennes conditionnelles, et les variances conditionnelles :

$$\mu_X(y) = \int_{-\infty}^{\infty} xf(x|y)dx, \text{ et } \mu_Y(x) = \int_{-\infty}^{\infty} yf(y|x)dy,$$

$$\sigma_X^2(y) = \int_{-\infty}^{\infty} \{x - \mu_X(y)\}^2 f(x|y)dx, \text{ et } \sigma_Y^2(x) = \int_{-\infty}^{\infty} \{y - \mu_Y(x)\}^2 f(y|x)dy.$$

Enfin, la covariance entre X et Y est définie par

$$\sigma_{xy} = \text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)dxdy.$$

3.2.5 Indépendance de deux variables aléatoires

Deux variables aléatoires X et Y sont dites indépendantes, si

$$\Pr(X \leq x \text{ et } Y \leq y) = \Pr(X \leq x)\Pr(Y \leq y), \text{ pour tout } x, y \in \mathbb{R}.$$

– Si X et Y sont discrètes, cela implique que

$$\Pr(X = x \text{ et } Y = y) = \Pr(X = x)\Pr(Y = y), \text{ pour tout } x, y \in \mathbb{Z}.$$

– Si X et Y sont continues, en notant $f_X(\cdot)$ et $f_Y(\cdot)$ les fonctions de densité marginales respectives de X et Y , et en notant $f_{XY}(x, y)$ la densité jointe des deux variables, alors X et Y sont indépendants si

$$f_{XY}(x, y) = f_X(x)f_Y(y), x, y \in \mathbb{R}.$$

3.2.6 Propriétés des espérances et des variances

De manière générale, pour des variables aléatoires X et Y , et avec a et b constants :

$$E(a + bX) = a + bE(X)$$

$$E(aY + bX) = aE(Y) + bE(X)$$

$$\text{var}(a + bX) = b^2\text{var}(X).$$

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y).$$

De plus, si X et Y sont indépendantes :

$$E(XY) = E(X)E(Y)$$

$$\text{cov}(X, Y) = 0,$$

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y).$$

Enfin, il est possible de calculer l'espérance et la variance d'une somme de variables aléatoires indépendantes, et identiquement distribuées.

Théorème 3.3 Soit X_1, \dots, X_n une suite de variables aléatoires, indépendantes et identiquement distribuées et dont la moyenne μ et la variance σ^2 existent et sont finies, alors si

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

on a

$$E(\bar{X}) = \mu, \text{ et } \text{var}(\bar{X}) = \frac{\sigma^2}{n}.$$

Démonstration

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu.$$

et

$$\text{var}(\bar{X}) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}.$$

□

3.2.7 Autres variables aléatoires

Variable khi-carrée

Soit une suite de variables aléatoires indépendantes, normales, centrées réduites, X_1, \dots, X_p , (c'est-à-dire de moyenne nulle et de variance égale à 1), alors la variable aléatoire

$$\chi_p^2 = \sum_{i=1}^p X_i^2,$$

est appelée variable aléatoire khi-carré à p degrés de liberté.

Il est possible de montrer que

$$E(\chi_p^2) = p,$$

et que

$$\text{var}(\chi_p^2) = 2p.$$

Variable de Student

Soit une variable aléatoire X normale centrée réduite, et une variable aléatoire khi-carré χ_p^2 à p degrés de liberté, indépendante de X , alors la variable aléatoire

$$t_p = \frac{X}{\sqrt{\chi_p^2/p}}$$

est appelée variable aléatoire de Student à p degrés de liberté.

Variable de Fisher

Soient deux variables aléatoires khi-carrés indépendantes χ_p^2, χ_q^2 , respectivement à p et q degrés de liberté, alors la variable aléatoire

$$F_{p,q} = \frac{\chi_p^2/p}{\chi_q^2/q}$$

est appelée variable aléatoire de Fisher à p et q degrés de liberté.

Remarque 3.1 Il est facile de montrer que le carré d'une variable de Student à q degrés de liberté est une variable de Fisher à 1 et q degrés de liberté.

3.2.8 Variable normale multivariée

Le vecteur de variables aléatoires $\mathbf{X} = (X_1, \dots, X_p)'$ a une distribution normale multivariée de moyenne $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$ et de matrice variance-covariance $\boldsymbol{\Sigma}$ (on suppose par simplicité que $\boldsymbol{\Sigma}$ est de plein rang), si sa fonction de densité est donnée par

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (3.2)$$

pour tout $\mathbf{x} \in \mathbb{R}^p$.

Remarque 3.2 Si $p = 1$, on retrouve l'expression (3.1).

Un cas particulier est important : supposons que la matrice variance-covariance peut s'écrire $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$, ce qui signifie que toutes les composantes du vecteur \mathbf{X} sont non-corrélées. Dans ce cas,

$$\begin{aligned}
f_X(\mathbf{x}) &= \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \\
&= \frac{1}{(2\pi)^{p/2} (\prod_{j=1}^p \sigma_j^2)^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \\
&= \frac{1}{(2\pi)^{p/2} (\prod_{j=1}^p \sigma_j)} \exp \left[-\sum_{j=1}^p \frac{(x_j - \mu_j)^2}{2\sigma_j^2} \right] \\
&= \frac{1}{(2\pi)^{p/2} (\prod_{j=1}^p \sigma_j)} \prod_{j=1}^p \exp \left[-\frac{(x_j - \mu_j)^2}{2\sigma_j^2} \right] \\
&= \prod_{j=1}^p \frac{1}{(2\pi)^{1/2} \sigma_j} \exp \left[-\frac{(x_j - \mu_j)^2}{2\sigma_j^2} \right] \\
&= \prod_{j=1}^p f_{X_j}(x_j),
\end{aligned}$$

où

$$f_{X_j}(x_j) = \frac{1}{(2\pi\sigma_j^2)^{1/2}} \exp \left[-\frac{(x_j - \mu_j)^2}{2\sigma_j^2} \right],$$

est la densité de la variable X_j . On constate que s'il y a absence de corrélation entre les variables normales, alors la densité du vecteur normal peut s'écrire comme un produit de densités. Dans le cas multinormal (et seulement dans ce cas), l'absence de corrélation implique donc l'indépendance des variables aléatoires.

De manière générale, si \mathbf{X} est un vecteur de variables aléatoires de moyenne $\boldsymbol{\mu}$ et de matrice variance-covariance Σ , et si \mathbf{A} est une matrice $q \times p$ de constantes, alors

$$\mathbb{E}(\mathbf{AX}) = \mathbf{A}\mathbb{E}(\mathbf{X}) = \mathbf{A}\boldsymbol{\mu},$$

et

$$\text{var}(\mathbf{AX}) = \mathbf{A}\text{var}(\mathbf{X})\mathbf{A}' = \mathbf{A}\Sigma\mathbf{A}'.$$

Dans le cas normal, on a en plus la propriété suivante :

Propriété 3.1 *Toute combinaison linéaire d'un vecteur de variables aléatoires normales est normal (Cependant sa matrice variance-covariance n'est pas nécessairement de plein rang).*

Donc, si \mathbf{X} est un vecteur multinormal de moyenne $\boldsymbol{\mu}$ et de matrice variance-covariance Σ et si \mathbf{A} est une matrice $q \times p$ de constantes, alors on écrit

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma),$$

et on a

$$\mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}').$$

Comme une projection est une combinaison linéaire, on a aussi que :

Propriété 3.2 *Toute projection d'un vecteur des variables aléatoires normales est normale.*

3.3 Inférence statistique

3.3.1 Modélisation

La modélisation est une approche qui consiste à approcher la réalité par un modèle plus simple. Le modèle ne pourra jamais représenter complètement la réalité dans toute sa complexité. Le modèle est une simplification. La maxime des modélisateurs dit que "tous les modèles sont faux, mais certains sont utiles". Comme le modèle ne peut tout décrire, il restera toujours une partie inexpliquée qui sera supposée aléatoire. Le calcul des probabilités est alors introduit pour prendre en compte la partie inexpliquée par le modèle. Dans la démarche de la modélisation, la randomization est donc introduite à titre d'hypothèse.

3.3.2 Intervalle de confiance

Pour ne pas donner sèchement la valeur d'un estimateur $\hat{\theta}$ d'un paramètre θ , on préfère produire un intervalle $[L^-, L^+]$ dans lequel pourrait se trouver le paramètre inconnu avec une certaine probabilité que l'on note $1 - \alpha$ (α est une probabilité petite). On relativise ainsi l'information donnée par l'estimateur $\hat{\theta}$. Pour pouvoir construire un intervalle de confiance, il faut connaître la distribution de probabilité de $\hat{\theta}$ (ou au moins une approximation de cette distribution de probabilité).

3.3.3 Tests d'hypothèses

Tests d'hypothèses simples

Le test d'hypothèses consiste à énoncer deux hypothèses sur un paramètre θ , dont une seule est vraie. Par exemple, on peut tester

- l'hypothèse nulle H_0 que $\theta = \theta_0$,
- l'hypothèse alternative H_1 que $\theta = \theta_1$.

L'objectif est de prendre une décision sur H_0 qui consistera à rejeter H_0 (RH_0) ou à ne pas rejeter H_0 ($\bar{R}H_0$). La décision est prise sur base des données observées, et peut donc conduire à deux types d'erreurs :

- Rejeter H_0 alors que H_0 est vraie, cette erreur est appelée erreur de première espèce.
- Ne pas rejeter H_0 alors que H_0 est fausse, cette erreur est appelée erreur de deuxième espèce.

TAB. 3.1 – Erreur de première et seconde espèce

	H_0 est vraie	H_0 est fausse
RH_0	Erreur de 1ère espèce	Décision correcte
$\bar{R}H_0$	Décision correcte	Erreur de 2ème espèce

La probabilité de commettre une erreur de première espèce est notée α , et la probabilité de commettre une erreur de deuxième espèce est notée β . Dans la théorie des tests d'hypothèses, on fixe α petit.

La décision prise sur base des données observées ne peut pas être exacte, on calcule donc les probabilités de commettre les erreurs.

TAB. 3.2 – Probabilité de commettre les erreurs

	H_0 est vraie	H_0 est fausse
RH_0	$\Pr(RH_0 H_0 \text{ vraie}) = \alpha$	$\Pr(RH_0 H_0 \text{ fausse}) = 1 - \beta$
$\bar{R}H_0$	$\Pr(\bar{R}H_0 H_0 \text{ vraie}) = 1 - \alpha$	$\Pr(\bar{R}H_0 H_0 \text{ fausse}) = \beta$

La quantité

$$\Pr(RH_0|H_0 \text{ fausse}) = \Pr(RH_0|H_1 \text{ vraie}) = 1 - \beta,$$

est appelée la puissance du test. Pour construire un test d'hypothèses, on fixe α petit (par ex : 0,05), et on cherche la règle de décision la plus puissante, c'est-à-dire, celle qui maximise $1 - \beta$.

Tests d'hypothèses composites

En pratique, on ne teste pas des hypothèses simples, mais des hypothèses composites. En effet, les questions que l'on se pose sur le paramètre sont du type "Le paramètre θ est-il strictement plus grand qu'une certaine valeur θ_0 ?" Ce type d'hypothèse composite amène à la construction de test du type :

$$1) \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} \quad 2) \begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} \quad 3) \begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

Remarque 3.3 L'égalité doit toujours être dans l'hypothèse nulle, donc si la question est : " θ est-il strictement plus grand que θ_0 ?" on posera l'hypothèse alternative $H_1 : \theta > \theta_0$ et donc $H_0 : \theta \leq \theta_0$.

Il existe des techniques statistiques qui permettent de construire des tests puissants. Le test aboutit à la construction d'une statistique de test notée T et d'un intervalle d'acceptation que l'on note IA et qui est construit pour un α particulier. Souvent la statistique de test est l'estimateur $\hat{\theta}$ de θ . La décision qui se prend en général en fonction d'un estimateur de T est du type :

- On rejette H_0 si $T \notin IA$
- On ne rejette pas H_0 si $T \in IA$

Exercices

Exercice 3.1 Soient \mathbf{X} , un vecteur de \mathbb{R}^p , de variables aléatoires de moyenne $\boldsymbol{\mu}$ et de matrice variance-covariance $\boldsymbol{\Sigma}$ et \mathbf{A} est une matrice $q \times p$ de constantes. Montrez que $E(\mathbf{AX}) = \mathbf{A}\boldsymbol{\mu}$ et que $\text{var}(\mathbf{AX}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$.

Exercice 3.2 Dans une ville, on évalue à 20% les individus qui approuvent la politique économique du président, les 80% restant s'y opposent.

1. Quelle est la probabilité que parmi 8 personnes choisies au hasard, 3 exactement approuvent la politique économique ?
2. Quelle est la probabilité que parmi 8 personnes choisies au hasard, un nombre inférieur ou égal à 3 personnes approuvent la politique économique ?
3. Un meeting organisé par les opposants a réuni 10% des opposants et 1% des individus favorables. Déterminez les probabilités qu'un participant au meeting, choisi au hasard, soit un opposant.
4. Donnez les expressions de l'espérance et de la variance de la loi de probabilité utilisée.
5. Calculez les valeurs de l'espérance et de la variance.

Chapitre 4

Le modèle linéaire général

4.1 Le modèle

4.1.1 Définition du modèle linéaire général

En économétrie, on ne considère pas simplement que les variables sont observées sur des unités statistiques. On postule l'existence d'un modèle qui régit les relations entre les variables. La relation la plus simple est une relation linéaire, entre les variables explicatives et la variable dépendante.

Le modèle linéaire général s'écrit

$$y_i = \sum_{j=1}^p x_{ij}\beta_j + \varepsilon_i,$$

où

- x_{ij} représente la valeur prise par la j ème variable sur l'individu i , les x_{ij} sont supposés non-aléatoires,
- β_j est la j ème composante du coefficient de régression,
- les ε_i sont des variables aléatoires telles que
 - $E(\varepsilon_i) = 0$ pour tout i ,
 - $E(\varepsilon_i \varepsilon_k) = 0$ pour tout $i \neq k$,
 - $E(\varepsilon_i^2) = \sigma_\varepsilon^2$ pour tout i .

4.1.2 Hypothèses du modèle linéaire général

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

- 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).

4.1.3 Données observées, et formulation matricielle

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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

où

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

Seuls \mathbf{y} et \mathbf{X} sont observés.

Les hypothèses du modèle linéaire général peuvent être reformulées :

- La matrice \mathbf{X} n'est pas aléatoire,

- La matrice \mathbf{X} est supposée de plein rang (Dans le cas contraire, on dit qu'il y a multicollinéarité, c'est-à-dire qu'au moins une des colonnes de la matrice peut s'exprimer comme une combinaison linéaire des autres colonnes),
- $E(\boldsymbol{\varepsilon}) = \mathbf{0}$,
- $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2$ (homoscédasticité).
- $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ (toutes les corrélations sont nulles).

Remarque 4.1 La somme des termes d'erreur $\sum_{i=1}^n \varepsilon_i$, n'est pas nécessairement nulle.

4.1.4 Autre présentation du modèle linéaire général

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

- $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ où \mathbf{X} est une matrice $n \times p$ et $\boldsymbol{\beta} \in \mathbb{R}^p$,
- $\text{var}(\mathbf{y}) = \mathbf{I}\sigma_\varepsilon^2$ où \mathbf{I} est une matrice identité $n \times n$ et σ_ε^2 est un scalaire.

Cette formulation est équivalente à la précédente.

4.2 Estimation du modèle

4.2.1 Estimation par les moindres carrés (ordinaires)

L'objectif est d'estimer $\boldsymbol{\beta}$ et σ_ε^2 . La méthode des moindres carrés consiste à minimiser en $\boldsymbol{\beta}$, l'expression

$$\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

La solution (voir section 2.2.2) fournit l'estimateur des moindres carrés (ordinaires) $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$, qui se note

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

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

Définition 4.1 Un estimateur est dit sans biais si son espérance mathématique est égale au paramètre à estimer, quelle que soit la valeur de ce paramètre.

Théorème 4.1 L'estimateur $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ est sans biais.

Démonstration

Comme

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

On a

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

□

Théorème 4.2 $\text{var}(\hat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}$.

Démonstration

Comme

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon},$$

on a

$$\begin{aligned}
\text{var}(\hat{\boldsymbol{\beta}}) &= \text{var} \{ (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varepsilon} \} \\
&= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \text{var} \{ \boldsymbol{\varepsilon} \} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{I} \sigma_{\varepsilon}^2 \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma_{\varepsilon}^2 \underbrace{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}}_{\mathbf{I}} \\
&= \sigma_{\varepsilon}^2 (\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}$$

□

Théorème 4.3 (de Gauss-Markov) *L'estimateur $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$ est le meilleur (au sens de la plus petite variance) estimateur linéaire en \mathbf{y} sans biais de $\boldsymbol{\beta}$.*

Démonstration

Soit $\boldsymbol{\beta}^* = \mathbf{C} \mathbf{y}$, un estimateur linéaire. En posant $\mathbf{B} = \mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$, on a $\boldsymbol{\beta}^* = (\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{y}$. Comme

$$\mathbb{E}(\boldsymbol{\beta}^*) = \mathbb{E} \{ (\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') (\mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}) \} = (\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{X} \boldsymbol{\beta} = \mathbf{B} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta},$$

pour que $\boldsymbol{\beta}^*$ soit sans biais, il faut que

$$\mathbf{B} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta} = \boldsymbol{\beta},$$

c'est-à-dire que

$$\mathbf{B} \mathbf{X} \boldsymbol{\beta} = \mathbf{0},$$

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

$$\mathbf{B} \mathbf{X} = \mathbf{0}. \quad (4.1)$$

Calculons maintenant la variance de $\boldsymbol{\beta}^*$:

$$\begin{aligned}
\text{var}(\boldsymbol{\beta}^*) &= (\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \text{var}(\mathbf{y}) (\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}')' \\
&= (\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{I} \sigma_{\varepsilon}^2 (\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}')' \\
&= \left\{ \mathbf{B} \mathbf{B}' + \underbrace{\mathbf{B} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}}_0 + \underbrace{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{B}'}_0 + (\mathbf{X}'\mathbf{X})^{-1} \right\} \sigma_{\varepsilon}^2.
\end{aligned}$$

Par (4.1), on a finalement

$$\text{var}(\boldsymbol{\beta}^*) = \{ \mathbf{B} \mathbf{B}' + (\mathbf{X}'\mathbf{X})^{-1} \} \sigma_{\varepsilon}^2. \quad (4.2)$$

La matrice $\mathbf{B} \mathbf{B}'$ est semi-définie positive. Tous les éléments de sa diagonale sont positifs. Donc, le meilleur estimateur est obtenu quand $\mathbf{B} = \mathbf{0}$.

□

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

$$\mathbf{e} = \hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = \mathbf{P}_X^{\perp} \mathbf{y}.$$

Notre objectif est de calculer $\mathbb{E}(\mathbf{e}'\mathbf{e})$. Pour obtenir le résultat, on utilisera le théorème général suivant.

Lemme 4.1 *Soit un vecteur \mathbf{u} composé de n variables aléatoires d'espérances nulles, et tel que $\text{var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}$, et \mathbf{A} une matrice symétrique non-aléatoire, alors*

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

Démonstration

$$\mathbb{E}(\mathbf{u}' \mathbf{A} \mathbf{u}) = \sum_{i=1}^n a_{ii} \underbrace{\mathbb{E}(u_i^2)}_{\sigma_u^2} + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \underbrace{\mathbb{E}(u_i u_j)}_0.$$

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

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

□

Grâce au lemme 4.1, on peut calculer l'espérance de $\mathbf{e}'\mathbf{e}$.

Théorème 4.4 Soit $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, alors

$$E(\mathbf{e}'\mathbf{e}) = (n - p)\sigma_\varepsilon^2$$

Démonstration

Nous avons vu en section 2.1 que \mathbf{e} peut également s'écrire

$$\mathbf{e} = (\mathbf{I} - \mathbf{P}_X) \mathbf{y}, \quad (4.3)$$

où \mathbf{P}_X est un projecteur (c'est-à-dire une matrice idempotente) sur le sous-espace engendré par les colonnes de \mathbf{X} :

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

Donc,

$$\mathbf{e} = (\mathbf{I} - \mathbf{P}_X) \mathbf{y} = (\mathbf{I} - \mathbf{P}_X) (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta} - \mathbf{P}_X \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{P}_X \boldsymbol{\varepsilon}.$$

Or $\mathbf{P}_X \mathbf{X} = \mathbf{X}$, ce qui donne

$$\mathbf{e} = \boldsymbol{\varepsilon} - \mathbf{P}_X \boldsymbol{\varepsilon} = (\mathbf{I} - \mathbf{P}_X) \boldsymbol{\varepsilon}.$$

On obtient

$$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{P}_X)'(\mathbf{I} - \mathbf{P}_X)\boldsymbol{\varepsilon},$$

et comme $(\mathbf{I} - \mathbf{P}_X)$ est symétrique et idempotente, on a

$$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{P}_X)\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}'\mathbf{I}\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}'\mathbf{P}_X\boldsymbol{\varepsilon}.$$

Par le lemme 4.1, on obtient

$$E(\mathbf{e}'\mathbf{e}) = \sigma_\varepsilon^2 \text{trace}(\mathbf{I}) - \sigma_\varepsilon^2 \text{trace}(\mathbf{P}_X).$$

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

$$E(\mathbf{e}'\mathbf{e}) = n\sigma_\varepsilon^2 - p\sigma_\varepsilon^2 = (n - p)\sigma_\varepsilon^2.$$

□

Le théorème 4.4 nous permet de construire un estimateur sans biais pour σ_ε^2 qui est :

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

La quantité $n - p$ est appelée nombre de degrés de liberté, et est le rang de $(\mathbf{I} - \mathbf{P}_X)$.

TAB. 4.1 – Tableau récapitulatif

Paramètre	Estimateur	Variance	Variance estimée
$\boldsymbol{\beta}$	$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$	$(\mathbf{X}'\mathbf{X})^{-1} \sigma_\varepsilon^2$	$(\mathbf{X}'\mathbf{X})^{-1} \hat{\sigma}_\varepsilon^2$
σ_ε^2	$\hat{\sigma}_\varepsilon^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - p}$	—	—

4.2.2 Estimateurs du maximum de vraisemblance

Une autre approche consiste à faire une hypothèse sur la distribution de probabilité de $\boldsymbol{\varepsilon}$. 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 $\boldsymbol{\varepsilon}$ a une distribution multinormale :

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2),$$

et, comme $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$,

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}\sigma_\varepsilon^2),$$

et donc

$$\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2).$$

De (3.2), on a

$$\begin{aligned} f_y(\mathbf{u}) &= \frac{1}{(2\pi)^{n/2} |\mathbf{I}\sigma_\varepsilon^2|^{1/2}} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{u} - \mathbf{X}\boldsymbol{\beta})' \mathbf{I}^{-1} (\mathbf{u} - \mathbf{X}\boldsymbol{\beta}) \right] \\ &= \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2}} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{u} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{u} - \mathbf{X}\boldsymbol{\beta}) \right], \text{ pour tout } \mathbf{u} \in \mathbb{R}^n. \end{aligned}$$

On se trouve dans un problème paramétrique classique. Comme \mathbf{y} et \mathbf{X} sont observés, on va estimer les paramètres $\boldsymbol{\beta}$ 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 :

$$L(\boldsymbol{\beta}, \sigma_\varepsilon^2) = f_y(\mathbf{y}) = \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2}} \exp - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma_\varepsilon^2}.$$

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 :

$$\ell(\boldsymbol{\beta}, \sigma_\varepsilon^2) = \log L(\boldsymbol{\beta}, \sigma_\varepsilon^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_\varepsilon^2) - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma_\varepsilon^2}.$$

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

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma_\varepsilon^2)}{\partial \boldsymbol{\beta}} = \frac{\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta}}{\sigma_\varepsilon^2} = \mathbf{0},$$

et

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma_\varepsilon^2)}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0.$$

La solution du maximum de vraisemblance pour $\boldsymbol{\beta}$ est donc la même que la solution des moindres carrés, et vaut :

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

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

$$\hat{\sigma}_{\varepsilon MV}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{\mathbf{e}'\mathbf{e}}{n}.$$

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

4.2.3 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

$$\lim_{n \rightarrow \infty} \Pr(|\hat{\theta} - \theta| > \varepsilon) = 0,$$

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 θ alors $f(\hat{\theta})$ est l'estimateur du maximum de vraisemblance de $f(\theta)$.
- Si l'estimateur du maximum de vraisemblance admet une solution unique, alors cet estimateur est convergent et asymptotiquement efficace du paramètre. De plus, cet estimateur converge en loi vers une normale.

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.

4.2.4 Distribution de probabilité des estimateurs

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

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

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

$$\hat{\beta} \sim \mathcal{N}(\beta, (\mathbf{X}'\mathbf{X})^{-1} \sigma_\varepsilon^2). \quad (4.4)$$

Lemme 4.2 Soient \mathbf{u} un vecteur aléatoire de distribution normale de \mathbb{R}^n , de moyennes nulles et de variance \mathbf{I} , et $\mathbf{\Gamma}$ une matrice orthogonale de dimension $n \times n$, alors

$$\mathbf{\Gamma}\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \text{ et } \mathbf{\Gamma}'\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Démonstration

On a $\mathbf{\Gamma}\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}\mathbf{\Gamma}')$, et $\mathbf{\Gamma}'\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}'\mathbf{\Gamma})$. Or, $\mathbf{\Gamma}' = \mathbf{\Gamma}^{-1}$, donc $\mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{I}$. □

L'inférence sur paramètres est basée sur le résultat général suivant.

Théorème 4.5 Soit un vecteur aléatoire \mathbf{u} de distribution normale, de moyennes nulles et de variance \mathbf{I} . Si \mathbf{P} est symétrique, idempotente et de rang p , alors $\mathbf{u}'\mathbf{P}\mathbf{u}$ est une variable χ_p^2 à p degrés de liberté.

Démonstration

La matrice \mathbf{P} admet une décomposition en valeurs propres et vecteurs propres. En vertu du théorème 1.2, si $\mathbf{\Lambda}$ représente la matrice diagonale ayant les valeurs propres λ_i de \mathbf{P} sur sa diagonale, et $\mathbf{\Gamma}$ est une matrice orthogonale contenant les n vecteurs propres de \mathbf{P} , alors on peut écrire :

$$\mathbf{P} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}'.$$

La forme quadratique peut s'écrire

$$\mathbf{u}'\mathbf{P}\mathbf{u} = \mathbf{u}'\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}'\mathbf{u} = \mathbf{v}'\mathbf{\Lambda}\mathbf{v},$$

où $\mathbf{v} = \mathbf{\Gamma}'\mathbf{u}$. En vertu du lemme 4.2, $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. En vertu du théorème 1.4, comme \mathbf{P} est idempotente et de rang p , \mathbf{P} a p valeurs propres égales à 1 et $n - p$ valeurs propres égales à 0. La forme quadratique

$$\mathbf{v}'\mathbf{\Lambda}\mathbf{v} = \sum_{i=1}^n v_i^2 \lambda_i = \sum_{i=1| \lambda_i=1}^n v_i^2$$

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

Corrolaire 4.1 Dans le modèle linéaire général avec des termes d'erreur normaux,

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

En effet,

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

donc

$$(\hat{\beta} - \beta)' \frac{\mathbf{X}'\mathbf{X}}{\sigma_\varepsilon^2} (\hat{\beta} - \beta) = \boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \frac{\mathbf{X}'\mathbf{X}}{\sigma_\varepsilon^2} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} = \frac{\boldsymbol{\varepsilon}'}{\sigma_\varepsilon} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \frac{\boldsymbol{\varepsilon}}{\sigma_\varepsilon}.$$

Comme la matrice $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ est symétrique idempotente et de rang p et que $\boldsymbol{\varepsilon}'/\sigma_\varepsilon$ est un vecteur multinormal non-corrélé, le corollaire s'obtient directement par le théorème 4.5. \square

Corrolaire 4.2 Dans le modèle linéaire général avec des termes d'erreur normaux,

$$\frac{\mathbf{e}'\mathbf{e}}{\sigma_\varepsilon^2} \sim \chi_{n-p}^2.$$

En effet,

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \mathbf{P}_X^\perp \boldsymbol{\varepsilon}$$

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

$$\frac{\mathbf{e}'\mathbf{e}}{\sigma_\varepsilon^2} = \frac{\boldsymbol{\varepsilon}'}{\sigma_\varepsilon} \mathbf{P}_X^\perp \mathbf{P}_X^\perp \frac{\boldsymbol{\varepsilon}}{\sigma_\varepsilon} = \frac{\boldsymbol{\varepsilon}'}{\sigma_\varepsilon} \mathbf{P}_X^\perp \frac{\boldsymbol{\varepsilon}}{\sigma_\varepsilon} \sim \chi_{n-p}^2.$$

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

Théorème 4.6 Soient les matrices \mathbf{B} ($p \times n$) et \mathbf{A} ($n \times n$) et un vecteur aléatoire $\mathbf{u} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma_u^2 \mathbf{I})$, alors les p formes linéaires $\mathbf{B}\mathbf{u}$ sont indépendantes de la forme quadratique $\mathbf{u}'\mathbf{A}\mathbf{u}$ si $\mathbf{B}\mathbf{A} = \mathbf{0}$.

Corrolaire 4.3 Dans le modèle linéaire avec des termes d'erreur normaux,

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

En effet, $\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{P}_X^\perp \boldsymbol{\varepsilon}$ où $\mathbf{P}_X^\perp = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ et $\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon}$ or $(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{P}_X^\perp = \mathbf{0}$, ce qui implique directement le corollaire.

Théorème 4.7 Soient deux matrices symétriques \mathbf{C} ($n \times n$) et \mathbf{A} ($n \times n$) et un vecteur aléatoire $\mathbf{u} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma_u^2 \mathbf{I})$, alors les deux formes quadratiques $\mathbf{u}'\mathbf{C}\mathbf{u}$ et $\mathbf{u}'\mathbf{A}\mathbf{u}$ sont indépendantes si $\mathbf{C}\mathbf{A} = \mathbf{0}$.

4.2.5 Synthèse des résultats

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

- $\hat{\beta}$ et $\hat{\sigma}_\varepsilon^2$ sont convergents, exhaustifs, efficaces et sans biais,
- $\hat{\beta}$ et $\hat{\sigma}_\varepsilon^2$ sont indépendants,
- $\hat{\beta} = \mathcal{N}(\beta, (\mathbf{X}'\mathbf{X})^{-1} \sigma_\varepsilon^2)$
- $\frac{(n-p)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} = \frac{\mathbf{e}'\mathbf{e}}{\sigma_\varepsilon^2} \sim \chi_{n-p}^2$,
- $(\hat{\beta} - \beta)' \frac{\mathbf{X}'\mathbf{X}}{\sigma_\varepsilon^2} (\hat{\beta} - \beta) \sim \chi_p^2$.

Exercices

Exercice 4.1 Soit une suite de variables aléatoires (v.a.) indépendantes et identiquement distribuées (i.i.d.) de loi $\mathcal{N}(\mu, \sigma^2)$

1. On considère que σ^2 est connue.
Estimez μ par la méthode du maximum de vraisemblance.
2. On considère que μ est connue.
Estimez σ^2 par la méthode du maximum de vraisemblance.
3. On considère que σ^2 et μ sont inconnues.
Estimez μ et σ^2 par la méthode du maximum de vraisemblance.

Exercice 4.2 On se place dans le cadre du modèle linéaire général (MLG) avec la normalité des erreurs.

1. Écrivez la fonction de vraisemblance quand

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

2. Écrivez la de manière scalaire (et non sous la forme matricielle).
3. Annulez les dérivées partielles par rapport à β_1 , β_2 et σ^2 .

Exercice 4.3 Soit une suite de v.a. X_1, \dots, X_n i.i.d. dont la densité d'un X_i est donné par

$$f_{\mathbf{x}_i}(x_i) = \begin{cases} \frac{1}{\theta} & \text{si } 0 \leq x_i \leq \theta, \\ 0 & \text{sinon.} \end{cases} \quad (4.5)$$

1. Dessinez la fonction de densité et la fonction de répartition de X_i .
2. Quelle est la densité jointe du vecteur $\mathbf{X} = (X_1, \dots, X_i, \dots, X_n)$?
3. Donnez la fonction de vraisemblance.
4. Estimez θ par maximum de vraisemblance.
5. Donnez les fonctions de densité et de répartition de l'estimateur du maximum de vraisemblance.
6. Calculez l'espérance de l'estimateur du maximum de vraisemblance.
7. Si l'estimateur est biaisé, faites une correction de non biais.
8. Soit deux estimateurs de l'espérance des X_i : la moyenne des X_i sur l'échantillon et l'estimateur du maximum de vraisemblance de θ débiaisé et divisé par deux. Quel est le plus efficace ?

Chapitre 5

Inférence dans le modèle linéaire

5.1 Intervalle de confiance sur un coefficient de régression

Dans le chapitre précédent nous avons vu que

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

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

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

On a également que

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

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

La quantité

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

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

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

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

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

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

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

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

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

5.2 Test d'un seul coefficient de régression

5.2.1 Construction du test

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

$$\begin{cases} H_0 : \beta_j = \beta_{j0} \\ H_1 : \beta_j \neq \beta_{j0}. \end{cases}$$

Sous H_0 , $\hat{\beta}_j \sim \mathcal{N}(\beta_{j0}, \sigma^2(\hat{\beta}_j))$ où

$$\sigma^2(\hat{\beta}_j) = [(\mathbf{X}'\mathbf{X})^{-1} \sigma_\varepsilon^2]_{jj}$$

est simplement la composante correspondante à la j ème ligne et la j ème colonne de $\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1} \sigma_\varepsilon^2$. On peut donc estimer simplement $\sigma^2(\hat{\beta}_j)$ par

$$\hat{\sigma}^2(\hat{\beta}_j) = [(\mathbf{X}'\mathbf{X})^{-1} \hat{\sigma}_\varepsilon^2]_{jj}.$$

Rappelons que $\hat{\sigma}_\varepsilon^2$ et $\hat{\beta}_j$ sont indépendants, et que

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

Donc

$$\frac{(n-p)\hat{\sigma}^2(\hat{\beta}_j)}{\sigma^2(\hat{\beta}_j)} = \frac{(n-p) [(\mathbf{X}'\mathbf{X})^{-1} \hat{\sigma}_\varepsilon^2]_{jj}}{[(\mathbf{X}'\mathbf{X})^{-1} \sigma_\varepsilon^2]_{jj}} = \frac{(n-p)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{n-p}^2.$$

De plus,

$$\frac{\hat{\beta}_j - \beta_{j0}}{\sigma(\hat{\beta}_j)} \sim \mathcal{N}(0, 1)$$

Sous H_0 , la statistique

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

a donc, sous H_0 , 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.

5.2.2 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 $y_i \sim \mathcal{N}(\mu, \sigma^2)$, ce qui peut s'écrire sous la forme d'un modèle linéaire

$$y_i = \mu + \varepsilon_i, i = 1, \dots, n,$$

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

$$\mathbf{y} = \mathbf{1}\mu + \boldsymbol{\varepsilon},$$

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

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

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

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

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

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

Par le corollaire 4.3, $\hat{\mu}$ et $\hat{\sigma}^2$ sont indépendants. De plus on a, par l'expression (4.4) :

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

Donc,

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

En outre, on peut écrire

$$K = \frac{(n-1)\hat{\sigma}^2}{\sigma^2} = \frac{\boldsymbol{\varepsilon}'\mathbf{P}_c\boldsymbol{\varepsilon}}{\sigma^2},$$

où \mathbf{P}_c la matrice idempotente de rang $n-1$ qui centre les valeurs :

$$\mathbf{P}_c = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n} = \begin{pmatrix} 1-1/n & -1/n & -1/n & \dots & -1/n \\ -1/n & 1-1/n & -1/n & \dots & -1/n \\ -1/n & -1/n & 1-1/n & \dots & -1/n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & -1/n & \dots & 1-1/n \end{pmatrix}. \quad (5.1)$$

Les variables aléatoires d et K sont indépendantes. De plus, par le théorème 4.5, $K \sim \chi_{n-1}^2$. 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.

5.3 Tests de Wald sur les coefficients de régression

5.3.1 Test général d'une contrainte linéaire

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

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}, \quad (5.2)$$

contre l'hypothèse alternative

$$H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}, \quad (5.3)$$

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

Exemple 5.1

- Le test $H_0 : \beta_j = c$ s'obtient en prenant $\mathbf{R} = (0 \dots 0 \underbrace{1}_{j\text{ième}} 0 \dots 0)$ et $\mathbf{r} = c$.
- Le test $H_0 : \beta_j = 0$ pour tout j s'obtient en prenant $\mathbf{R} = \mathbf{I}_p$ (matrice identité de dimension p) et \mathbf{r} est un vecteur de 0 de dimension p .

Sous l'hypothèse H_0 ,

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

De plus,

$$\text{var}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \text{var}(\mathbf{R}\hat{\boldsymbol{\beta}}) = \mathbf{R}\text{var}(\hat{\boldsymbol{\beta}})\mathbf{R}' = \sigma_\varepsilon^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'.$$

Examinons maintenant la forme quadratique :

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \text{var}(\mathbf{R}\hat{\boldsymbol{\beta}})^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \frac{1}{\sigma_\varepsilon^2} \boldsymbol{\varepsilon}' \mathbf{W} \boldsymbol{\varepsilon}, \quad (5.4)$$

où

$$\mathbf{W} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \left\{ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right\}^{-1} \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

On vérifie facilement que \mathbf{W} est une matrice idempotente, symétrique de rang q . Par le théorème 4.5, on obtient donc que

$$\frac{1}{\sigma_\varepsilon^2} \boldsymbol{\varepsilon}' \mathbf{W} \boldsymbol{\varepsilon} \sim \chi_q^2,$$

et donc

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \text{var}(\mathbf{R}\hat{\boldsymbol{\beta}})^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \frac{1}{\sigma_\varepsilon^2} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left\{ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right\}^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \sim \chi_q^2. \quad (5.5)$$

Si la forme quadratique (5.4) est grande, on soupçonne H_0 d'être faux. Cependant, on ne peut réaliser directement un test χ^2 car l'expression (5.5) dépend de σ_ε^2 qui est inconnu. On sait par ailleurs que

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

De plus, comme

$$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{P}_X)\boldsymbol{\varepsilon},$$

et que $(\mathbf{I} - \mathbf{P}_X)\mathbf{W} = \mathbf{0}$, par le théorème (4.7), on a l'indépendance de $\mathbf{e}'\mathbf{e}/\sigma_\varepsilon^2$ et de $\boldsymbol{\varepsilon}'\mathbf{W}\boldsymbol{\varepsilon}$.

On peut construire une statistique de test

$$F_c = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left\{ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right\}^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \frac{1}{q}}{\mathbf{e}'\mathbf{e} \frac{1}{n-p}}. \quad (5.6)$$

Sous H_0 , le numérateur et le dénominateur de F_c sont indépendants, et ont, à une constante près, une distribution χ^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 5.2, si

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

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

5.3.2 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 \mathbf{X} est composée de uns, c'est-à-dire

que $x_{i1} = 1$ pour tout $i = 1, \dots, n$. La matrice \mathbf{R} est de dimension $(p-1) \times p$ et vaut :

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

Alors

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

et

$$\mathbf{r} = \mathbf{0} \in \mathbb{R}^{p-1}.$$

Le test devient alors :

$$\begin{cases} H_0 : & \beta_j = 0, \text{ pour tout } j = 2, \dots, p, \\ H_1 : & \text{au moins un des } \beta_j \neq 0, \end{cases}$$

ce qui peut aussi s'écrire

$$\begin{cases} H_0 : & \mathbf{R}\boldsymbol{\beta} = \mathbf{0}, \\ H_1 : & \mathbf{R}\boldsymbol{\beta} \neq \mathbf{0}. \end{cases}$$

ou encore

$$\begin{cases} H_0 : & \tilde{\boldsymbol{\beta}} = \mathbf{0}, \\ H_1 : & \tilde{\boldsymbol{\beta}} \neq \mathbf{0}. \end{cases}$$

Théorème 5.1

$$\left\{ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right\}^{-1} = \tilde{\mathbf{X}}' \mathbf{P}_c \tilde{\mathbf{X}} = n\boldsymbol{\Sigma}, \quad (5.7)$$

où \mathbf{P}_c est l'opérateur qui centre les données déjà présenté dans l'expression (2.7)

$$\mathbf{P}_c = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n},$$

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

Démonstration

On peut écrire

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \mathbf{u}' \\ \mathbf{u} & \mathbf{Z} \end{pmatrix} = \begin{pmatrix} n & \sum_i x_{i2} & \sum_i x_{i3} & \dots & \sum_i x_{ip} \\ \sum_i x_{i2} & \sum_i x_{i2}^2 & \sum_i x_{i2}x_{i3} & \dots & \sum_i x_{i2}x_{ip} \\ \sum_i x_{i3} & \sum_i x_{i2}x_{i3} & \sum_i x_{i3}^2 & \dots & \sum_i x_{i3}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_i x_{ip} & \sum_i x_{i2}x_{ip} & \sum_i x_{i3}x_{ip} & \dots & \sum_i x_{ip}^2 \end{pmatrix},$$

où

$$\mathbf{u} = (\sum_i x_{i2} \quad \sum_i x_{i3} \quad \dots \quad \sum_i x_{ip})',$$

et

$$\mathbf{Z} = \begin{pmatrix} \sum_i x_{i2}^2 & \sum_i x_{i2}x_{i3} & \dots & \sum_i x_{i2}x_{ip} \\ \sum_i x_{i2}x_{i3} & \sum_i x_{i3}^2 & \dots & \sum_i x_{i3}x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i x_{i2}x_{ip} & \sum_i x_{i3}x_{ip} & \dots & \sum_i x_{ip}^2 \end{pmatrix}.$$

Par la méthode d'inversion par partie, on a

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} n & \mathbf{u}' \\ \mathbf{u} & \mathbf{Z} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{n} + \frac{1}{n^2} \mathbf{u}'\mathbf{Q}\mathbf{u} & -\frac{1}{n} \mathbf{u}'\mathbf{Q} \\ -\frac{1}{n} \mathbf{Q}\mathbf{u} & \mathbf{Q} \end{pmatrix},$$

où

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

De plus,

$$(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1} = \mathbf{Q}^{-1} = \mathbf{Z} - \frac{1}{n} \mathbf{u} \mathbf{u}' = n\mathbf{\Sigma},$$

où $\mathbf{\Sigma}$ est la matrice variance-covariance définie en (2.2). □

L'expression (5.5) est alors la somme des carrés de la régression (voir expression (2.10)) :

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \text{var}(\mathbf{R}\hat{\boldsymbol{\beta}})^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \hat{\boldsymbol{\beta}}' \tilde{\mathbf{X}}' \mathbf{P}_c \tilde{\mathbf{X}} \hat{\boldsymbol{\beta}} = \text{SC}_{regr}.$$

En considérant l'expression (2.3), la statistique de test (5.6) devient :

$$F_c = \frac{\text{SC}_{regr}/(p-1)}{\text{SC}_{res}/(n-p)}, \quad (5.8)$$

ce qui peut également s'écrire

$$F_c = \frac{(\text{SC}_{tot} - \text{SC}_{res})/(p-1)}{\text{SC}_{res}/(n-p)}.$$

Ce test est généralement résumé au moyen du tableau d'analyse de la variance (voir tableau 5.1).

TAB. 5.1 – Tableau d'analyse de la variance

Source de variation	Sommes des carrés	Degrés de liberté	Carrés moyens	F_c
Régression	SC_{regr}	$p-1$	$\text{CM}_{regr} = \frac{\text{SC}_{regr}}{p-1}$	$F_c = \text{CM}_{regr}/\text{CM}_{res}$
Résiduelle	SC_{res}	$n-p$	$\text{CM}_{res} = \frac{\text{SC}_{res}}{n-p}$	
Totale	SC_{tot}	$n-1$	$\text{CM}_{tot} = \frac{\text{SC}_{tot}}{n-1}$	

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.

5.3.3 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 :

$$\begin{cases} H_0 : \beta_j = \beta_{j0} \\ H_1 : \beta_j \neq \beta_{j0}. \end{cases}$$

Pour ce faire, on prend

- $q = 1$,
- $\mathbf{R} = (0 \dots \underbrace{1}_{\text{unité } j} \dots 0)$,
- $\mathbf{r} = \beta_{j0}$.

On obtient

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

L'expression (5.6) devient

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

Sous H_0 , F_c suit une distribution de Fisher 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. Ce test n'est autre que le test de Student développé en section 5.2.1. 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 (voir section 3.2.7).

5.4 Analyse de la variance à un facteur

5.4.1 Le problème

L'analyse de la variance à un facteur est un cas particulier du modèle linéaire général. On suppose que les observations sont réparties dans H groupes. Les H groupes correspondent souvent à un traitement spécifique ou à une caractéristique des unités d'observation. L'objectif est de tester d'hypothèse nulle que les moyennes de tous les groupes sont égales. Si on note y_{ih} la valeur prise par l'observation i du groupe h , et n_h Le nombre d'observations du groupe h , avec

$$\sum_{h=1}^H n_h = n.$$

Le modèle s'écrit :

$$y_{ih} = \mu_h + \varepsilon_{ih}, \quad (5.9)$$

pour tout $h = 1, \dots, H$, et $i = 1, \dots, n_h$, où les μ_h sont H constantes et les ε_i sont des termes d'erreur indépendants, identiquement distribués ayant une distribution normale de moyenne nulle et de variance σ_ε^2 . Le modèle (5.9) est un cas particulier du modèle linéaire général. Nous allons examiner deux méthodes permettant de tester l'hypothèse d'égalité des moyennes des groupes, ce qui s'écrit

$$\begin{cases} H_0 & \mu_1 = \mu_2 = \dots = \mu_H \\ H_1 & \text{au moins un des } \mu_h \text{ est différent des autres.} \end{cases} \quad (5.10)$$

5.4.2 Méthode 1

La première méthode consiste à écrire le modèle (5.9) sous la forme d'un modèle linéaire général où :

- \mathbf{y} est le vecteur des n observations de y_{ih}
- $\boldsymbol{\beta} = (\mu_1 \dots \mu_h \dots \mu_H)'$ est le paramètre du modèle,
- $\boldsymbol{\varepsilon}$ est le vecteur des termes d'erreur,
- \mathbf{X} est la matrice $(n \times H)$ des variables explicatives qui est définie par :

$$x_{ih} = \begin{cases} 1 & \text{si l'observation } i \text{ est dans le groupe } h \\ 0 & \text{sinon} \end{cases}$$

ce qui donne, quand les unités sont rangées selon leurs groupes,

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \hline \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \hline 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \quad (5.11)$$

On peut dès lors écrire le modèle (5.9) sous la forme matricielle habituelle

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

La matrice $\mathbf{X}'\mathbf{X}$ est une matrice diagonale qui vaut

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & n_h & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & n_H \end{pmatrix},$$

et son inverse vaut

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{n_1} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \frac{1}{n_h} & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \frac{1}{n_H} \end{pmatrix}.$$

On a également le produit

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} \sum_{i=1}^{n_1} y_{i1} \\ \vdots \\ \sum_{i=1}^{n_h} y_{iH} \end{pmatrix}.$$

Enfin, l'estimateur de β est donné par

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_h \\ \vdots \\ \bar{y}_H \end{pmatrix},$$

où \bar{y}_h est la moyenne du groupe h et l'estimateur de μ_h :

$$\hat{\mu}_h = \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{ih},$$

pour $h = 1, \dots, H$. On a alors directement

– les valeurs ajustées

$$y_{ih}^* = \bar{y}_h,$$

– et les résidus

$$e_{ih} = y_{ih} - \bar{y}_h.$$

Pour réaliser le test donné en (5.10), on va utiliser la méthode de Wald développée dans la section (5.3). Le test (5.10) est un cas particulier du test (5.2) en prenant la matrice de contraintes \mathbf{R} de dimension $(H-1) \times H$ suivante :

$$\begin{aligned} \mathbf{R} &= \begin{pmatrix} 1 - \frac{n_1}{n} & -\frac{n_2}{n} & \dots & -\frac{n_{H-1}}{n} & -\frac{n_H}{n} \\ -\frac{n_1}{n} & 1 - \frac{n_2}{n} & \dots & -\frac{n_{H-1}}{n} & -\frac{n_H}{n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{n_1}{n} & -\frac{n_2}{n} & \dots & 1 - \frac{n_{H-1}}{n} & -\frac{n_H}{n} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} - \begin{pmatrix} \frac{n_1}{n} & \frac{n_2}{n} & \dots & \frac{n_{H-1}}{n} & \frac{n_H}{n} \\ \frac{n_1}{n} & \frac{n_2}{n} & \dots & \frac{n_{H-1}}{n} & \frac{n_H}{n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{n_1}{n} & \frac{n_2}{n} & \dots & \frac{n_{H-1}}{n} & \frac{n_H}{n} \end{pmatrix}, \end{aligned}$$

et \mathbf{r} est un vecteur de zéros de dimension $H-1$. On obtient après quelques calculs :

$$\mathbf{R}\beta = \begin{pmatrix} \mu_1 - \frac{1}{n} \sum_{h=1}^H n_h \mu_h \\ \vdots \\ \mu_h - \frac{1}{n} \sum_{h=1}^H n_h \mu_h \\ \vdots \\ \mu_{H-1} - \frac{1}{n} \sum_{h=1}^H n_h \mu_h \end{pmatrix} = \begin{pmatrix} \mu_1 - \mu \\ \vdots \\ \mu_h - \mu \\ \vdots \\ \mu_{H-1} - \mu \end{pmatrix},$$

et, de la même manière,

$$\mathbf{R}\hat{\beta} = \begin{pmatrix} \bar{y}_1 - \bar{y} \\ \vdots \\ \bar{y}_h - \bar{y} \\ \vdots \\ \bar{y}_{H-1} - \bar{y} \end{pmatrix},$$

où

$$\mu = \frac{1}{n} \sum_{h=1}^H n_h \mu_h,$$

et \bar{y} est la moyenne des observations :

$$\bar{y} = \frac{1}{n} \sum_{h=1}^H \sum_{i=1}^{n_h} y_{ih} = \frac{1}{n} \sum_{h=1}^H n_h \bar{y}_h.$$

Tester $\mathbf{R}\beta = \mathbf{r}$ équivaut, dans ce cas, à tester l'hypothèse nulle de (5.10). Pour calculer la statistique du test donné en (5.6), on doit calculer $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$. Après quelques calculs, on obtient :

$$\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' = \frac{1}{n} \begin{pmatrix} \frac{n}{n_1} - 1 & -1 & \cdots & -1 \\ -1 & \frac{n}{n_2} - 1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & \frac{n}{n_{H-1}} - 1 \end{pmatrix},$$

qui est une matrice de dimension $(H-1) \times (H-1)$. On peut vérifier par une simple multiplication que l'inverse de cette matrice vaut

$$\{\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\}^{-1} = \begin{pmatrix} n_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & n_h & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & n_{H-1} \end{pmatrix} + \frac{\mathbf{nn}'}{n_H},$$

ou $\mathbf{n}' = (n_1 \ n_2 \ \cdots \ n_{H-1})$. Enfin, après quelques calculs, on obtient

$$(\mathbf{R}\hat{\beta} - \mathbf{r})' \{\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\}^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) = \sum_{h=1}^H n_h (\bar{y}_h - \bar{y})^2,$$

qui n'est autre que la somme de carrés de la régression. Cette somme de carrés est souvent appelée pour ce cas particulier : somme des carrés inter-groupes (SC_{INTER}).

Au dénominateur de l'expression (5.6), on a

$$\mathbf{e}'\mathbf{e} = \sum_{h=1}^H \sum_{i=1}^{n_h} (y_{ih} - \bar{y}_h)^2,$$

c'est la somme des carrés des résidus qui est appelée pour ce cas particulier : somme des carrés intra-groupes (SC_{INTRA}).

Si l'on considère la somme des carrés totale,

$$SC_{TOT} = \sum_{h=1}^H \sum_{i=1}^{n_h} (y_{ih} - \bar{y})^2,$$

on a la décomposition classique des sommes de carrés

$$SC_{TOT} = SC_{INTRA} + SC_{INTER}.$$

On peut enfin construire la statistique de test de l'expression (5.6). Comme $q = H-1$, on a

$$F_c = \frac{SC_{INTER}/(H-1)}{SC_{INTRA}/(n-H)}. \quad (5.12)$$

On construit le tableau 5.2 d'analyse de la variance.

La règle de décision consiste à rejeter H_0 si $F_c > F_{1-\alpha, H-1, n-H}$ où $F_{1-\alpha, H-1, n-H}$ est le quantile d'ordre $1-\alpha$ d'une variable aléatoire de Fischer à $H-1$ et $n-H$ degrés de liberté.

TAB. 5.2 – Tableau d’analyse de la variance à un facteur

Source de variation	Sommes de carrés	Degrés de liberté	Carrés moyens	F_c
INTER	SC_{INTER}	$H - 1$	$CM_{INTER} = \frac{SC_{INTER}}{H-1}$	$F_c = \frac{CM_{INTER}}{CM_{INTRA}}$
INTRA	SC_{INTRA}	$n - H$	$CM_{INTRA} = \frac{SC_{INTRA}}{n-H}$	
TOTALE	SC_{TOT}	$n - 1$	$CM_{TOT} = \frac{SC_{TOT}}{n-1}$	

5.4.3 Méthode 2

Une autre manière d’écrire le modèle (5.9) sous la forme d’un modèle linéaire consiste à poser

$$\alpha_h = \mu_h - \mu, h = 1, \dots, H,$$

où

$$\mu = \frac{1}{n} \sum_{h=1}^H n_h \mu_h.$$

Le modèle s’écrit alors

$$y_{ih} = \mu + \alpha_h + \varepsilon_{ih}, \quad (5.13)$$

avec la contrainte que

$$\sum_{h=1}^H n_h \alpha_h = 0. \quad (5.14)$$

Le modèle (5.13) a maintenant $H + 1$ paramètres, et une contrainte sur les paramètres du modèle. Afin de pouvoir écrire ce modèle sous la forme d’un modèle linéaire, on intègre la contrainte dans le modèle, sachant que

$$\alpha_H = -\frac{1}{n_H} \sum_{h=1}^{H-1} n_h \alpha_h, \quad (5.15)$$

ce qui donne

$$\begin{cases} y_{ih} = \mu + \alpha_h + \varepsilon_{ih} & \text{si } 1 \leq h \leq H - 1 \\ y_{iH} = \mu - \frac{1}{n_H} \sum_{h=1}^{H-1} n_h \alpha_h + \varepsilon_{iH} & \text{sinon.} \end{cases} \quad (5.16)$$

Pour tester l’égalité des moyennes, on peut réaliser le test

$$\begin{cases} H_0 : \alpha_h = 0, \text{ pour tout } h = 1, \dots, H - 1 \\ H_1 : \text{au moins un des } \alpha_h \text{ est différent de } 0. \end{cases}$$

On remarque qu’un test sur les $H - 1$ premiers coefficients α_h suffit, en vertu de l’expression (5.15). Le modèle (5.13) s’écrit comme un modèle linéaire général

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où

$$\boldsymbol{\beta}' = (\mu \ \alpha_1 \ \alpha_2 \cdots \alpha_{H-1}),$$

et la matrice \mathbf{X} est de dimension $n \times H$ et est donnée par

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \hline 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \hline \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \hline 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \\ \hline 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \\ 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \\ 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \end{pmatrix}. \quad (5.17)$$

La première colonne de la matrice est donc une constante. Comme l'objectif est de tester la nullité des coefficients de regression à l'exception de la constante, on se retrouve dans le cas de la section (5.3.2). Estimons les paramètres du modèle. On a

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & 0 & 0 & \cdots & 0 \\ 0 & n_1(1 + \frac{n_1}{n_H}) & \frac{n_1 n_2}{n_H} & \cdots & \frac{n_1 n_{H-1}}{n_H} \\ 0 & \frac{n_1 n_2}{n_H} & n_2(1 + \frac{n_2}{n_H}) & \cdots & \frac{n_2 n_{H-1}}{n_H} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{n_1 n_{H-1}}{n_H} & \frac{n_2 n_{H-1}}{n_H} & \cdots & n_{H-1}(1 + \frac{n_{H-1}}{n_H}) \end{pmatrix}.$$

Son inverse est

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 1/n & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{n_1} - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ 0 & -\frac{1}{n} & \frac{1}{n_2} - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{1}{n_{H-1}} - \frac{1}{n} \end{pmatrix}.$$

Le vecteur $\mathbf{X}'\mathbf{y}$ vaut

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} n\bar{y} \\ n_1(\bar{y}_1 - \bar{y}_H) \\ \vdots \\ n_{H-1}(\bar{y}_{H-1} - \bar{y}_H) \end{pmatrix}.$$

On peut donc calculer l'estimateur de β .

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} \bar{y} \\ \bar{y}_1 - \bar{y} \\ \vdots \\ \bar{y}_{H-1} - \bar{y} \end{pmatrix}.$$

L'estimateur de μ est donc \bar{y} et les estimateurs α_h sont

$$\hat{\alpha}_h = \bar{y}_h - \bar{y}, h = 1, \dots, H-1.$$

Les valeurs ajustées valent $\mathbf{X}\hat{\beta}$, ce qui donne, si $h \leq H-1$

$$y_{ih}^* = \hat{\mu} + \hat{\alpha}_h = \bar{y}_h,$$

et si $h = H$,

$$y_{iH}^* = \hat{\mu} - \sum_{h=1}^{H-1} \frac{\hat{\alpha}_h n_h}{n_H} = \bar{y}_H.$$

Les résidus valent

$$e_{ih} = y_{ih} - y_{ih}^* = y_{ih} - \bar{y}_h, h = 1, \dots, H,$$

On a donc la somme de carrés des résidus qui vaut à nouveau la somme des carrés intra-groupes

$$SC_{INTRA} = \sum_{i=1}^n e_i^2 = \sum_{h=1}^H \sum_{i=1}^{n_h} (y_{ih} - \bar{y}_h)^2,$$

et la somme des carrés de la régression qui vaut à nouveau la somme des carrés inter-groupes

$$SC_{INTER} = \sum_{h=1}^H \sum_{i=1}^{n_h} (y_{ih}^* - \bar{y})^2 = \sum_{h=1}^H n_h (\bar{y}_h - \bar{y})^2.$$

La statistique de test s'obtient directement à partir de l'expression (5.8) et vaut

$$F_c = \frac{SC_{INTER}/(H-1)}{SC_{INTRA}/(n-H)}$$

et est exactement la même que (5.12). En posant le modèle différemment, on estime d'autres paramètres, mais les résidus, les valeurs ajustées, et le test sont identiques.

5.5 Prédiction ponctuelle d'une valeur

5.5.1 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

$$\mathbf{x}_j = (x_{j1} \cdots x_{jp}).$$

La prédiction vient simplement et vaut :

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

Le prédicteur peut également s'écrire

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

Comme la vraie valeur vaut

$$y_j = \mathbf{x}_j \boldsymbol{\beta} + \varepsilon_j,$$

l'erreur de prévision est

$$\hat{y}_j - y_j = \mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} - \varepsilon_j.$$

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

$$\mathbb{E}(\hat{y}_j - y_j) = \mathbb{E}\{\mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} - \varepsilon_j\} = \mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbb{E}(\boldsymbol{\varepsilon}) - \mathbb{E}(\varepsilon_j) = 0.$$

Comme la valeur prédite se réfère à une nouvelle observation,

$$\mathbb{E}(\varepsilon_j \boldsymbol{\varepsilon}) = \mathbf{0},$$

et donc

$$\begin{aligned} \text{var}(\hat{y}_j - y_j) &= \text{var}\{\mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon}\} + \text{var}\{\varepsilon_j\} \\ &= \mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \sigma_\varepsilon^2 \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_j' + \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 \{\mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_j' + 1\}. \end{aligned}$$

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 $\hat{\boldsymbol{\beta}}$, et la seconde partie est due à l'erreur inconnue ε_j .

On estime la variance simplement par

$$\widehat{\text{var}}(\hat{y}_j - y_j) = \hat{\sigma}_\varepsilon^2 \{\mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_j' + 1\},$$

où $\hat{\sigma}_\varepsilon^2 = \mathbf{e}'\mathbf{e}/(n-p)$. Enfin, il est possible de construire un intervalle de confiance pour la prévision :

$$\text{IC}(1-\alpha) = \left[\hat{y}_j - t_{1-\alpha/2, n-p} \sqrt{\widehat{\text{var}}(\hat{y}_j - y_j)}; \hat{y}_j + t_{1-\alpha/2, n-p} \sqrt{\widehat{\text{var}}(\hat{y}_j - y_j)} \right].$$

5.5.2 Cas bivarié

Dans le cas où une seule variable explicative x et une constante sont utilisées, on a

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \begin{pmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}, \\ (\mathbf{X}'\mathbf{X})^{-1} &= \frac{1}{ns_x^2} \begin{pmatrix} s_x^2 + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}. \end{aligned}$$

De plus, on a $\mathbf{x}_j = (1, x_j)$. La variance de l'erreur de prévision devient alors

$$\begin{aligned} \text{var}(\hat{y}_j - y_j) &= \sigma_\varepsilon^2 \{\mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_j' + 1\} \\ &= \sigma_\varepsilon^2 \left[\frac{1}{ns_x^2} \{(s_x^2 + \bar{x}^2) \times 1 - \bar{x}x_j - x_j\bar{x} + x_j^2\} + 1 \right] \\ &= \frac{\sigma_\varepsilon^2}{n} \left\{ n + 1 + \frac{(x_j - \bar{x})^2}{s_x^2} \right\}. \end{aligned}$$

Plus x_j est éloigné de la moyenne \bar{x} , plus la variance augmente. Faire une prévision pour des valeurs extrêmes de la variable x est donc plus hasardeux.

On estime la variance simplement par

$$\widehat{\text{var}}(\hat{y}_j - y_j) = \frac{\hat{\sigma}_\varepsilon^2}{n} \left\{ n + 1 + \frac{(x_j - \bar{x})^2}{s_x^2} \right\},$$

où $\hat{\sigma}_\varepsilon^2 = \mathbf{e}'\mathbf{e}/(n-p)$.

5.6 Exemple d'analyse de la variance à un facteur

5.6.1 Les données

Un ensemble de magazines a été classé selon trois groupes selon qu'ils s'adressent à un public d'un niveau d'instruction élevé (groupe 1) moyen (groupe 2) ou bas (groupe 3). Dix-huit publicités ont été sélectionnées au hasard dans chaque type de magazines. On s'intéresse au nombre de mots dans ces publicités. On cherche à savoir si le nombre de mots dépend du type de public visé. Les données sont présentées dans le tableau 5.3.

TAB. 5.3 – Nombre de mots selon les groupes

Groupe 1	Groupe 2	Groupe 3	Groupe 1	Groupe 2	Groupe 3
205	191	162	80	94	68
203	219	31	208	206	32
229	205	85	89	197	50
208	57	111	49	68	208
146	105	88	93	44	81
230	109	60	46	203	83
215	82	97	34	139	195
153	88	169	39	72	111
205	39	78	88	67	208

5.6.2 Les résultats

Le traitement statistique nous donne les résultats présentés dans les tableaux 5.4 et 5.5.

TAB. 5.4 – Moyennes selon les groupes

Groupe	Moyennes	N	Écart-type
1	140.00	18	74.0374
2	121.39	18	64.2698
3	106.50	18	57.6299
Total	122.63	54	65.8770

TAB. 5.5 – Tableau d'analyse de la variance

	sommes de carrés	degrés de liberté	carrés moyens	F	Sign.
Inter Groupes	10141.815	2	5070.907	1.176	0.317
Intra Groupes	219866.778	51	4311.113		
Total	230008.593	53			

Le test n'est pas significatif. En effet $F = 1.176$ et la valeur du quantile d'ordre 0.95 d'une Fisher à 2 et 51 degrés de liberté vaut 3.2. Donc on ne peut pas rejeter l'hypothèse d'égalité des moyennes, malgré d'importants écarts des moyennes des groupes pour les valeurs observées.

Exercices

Exercice 5.1 En reprenant les calculs de l'exercice 2.6, et en supposant que l'on se trouve dans le cadre du MLG avec normalité des erreurs, estimez σ_ε^2 et faites les tests suivants avec $\alpha = 0.05$ et 0.01 :

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}$$

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

$$\begin{cases} H_0 : \beta_0 = 1 \\ H_1 : \beta_0 \neq 1. \end{cases}$$

Exercice 5.2 Construisez un test pour tester les hypothèses suivantes :

1. $H_0 : \beta_j = 0$ (uniquement pour un coefficient),
2. $H_0 : \beta_j = 0$, pour tout $j = 2, \dots, p$, (c'est-à-dire pour tous les coefficients sauf la constante),
3. $H_0 : \beta_i = \beta_j$ pour deux coefficients i et j donnés,
4. $H_0 : \mathbf{c}'\boldsymbol{\beta} = \gamma$ (test sur une combinaison linéaire des coefficients).

Proposez au moins deux solutions pour \mathbf{R} pour chaque test.

Exercice 5.3 On considère le modèle :

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t, t = 1, \dots, 10.$$

Les données de l'échantillon sont résumées de la façon suivante :

$$\sum_{i=1}^n y_t^2 = 177, \quad \sum_{i=1}^n y_t = 10, \quad \sum_{i=1}^n y_t x_{2t} = 20, \quad \sum_{i=1}^n y_t x_{3t} = 40,$$

$$\sum_{i=1}^n x_{2t}^2 = 5, \quad \sum_{i=1}^n x_{3t}^2 = 20, \quad \sum_{i=1}^n x_{2t} = \sum_{i=1}^n x_{3t} = \sum_{i=1}^n x_{2t} x_{3t} = 0.$$

1. Construisez le tableau d'analyse de la variance,
2. Calculer le R^2
3. Testez, au seuil de signification $\alpha = 0.05$, les hypothèses suivantes :

$$\begin{cases} H_0 : \beta_2 = \beta_3 = 0 \\ H_1 : \beta_2 \neq 0 \text{ ou } \beta_3 \neq 0 \end{cases}, \begin{cases} H_0 : \beta_3 = 3 \\ H_1 : \beta_3 \neq 3 \end{cases}$$

$$\begin{cases} H_0 : \beta_3 \leq 1,5 \\ H_1 : \beta_3 > 1,5 \end{cases}, \begin{cases} H_0 : \beta_2 + \beta_3 = 8 \\ H_1 : \beta_2 + \beta_3 \neq 8 \end{cases}$$

Exercice 5.4 En utilisant la technique d'inversion matricielle par parties, montrez l'égalité données en (5.7) :

$$\left\{ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right\}^{-1} = \tilde{\mathbf{X}}' \mathbf{P}_c \tilde{\mathbf{X}}$$

Indication : Soit une matrice

$$\mathbf{F} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}.$$

On a

$$\mathbf{F}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1} \mathbf{B} \mathbf{Q} \mathbf{C} \mathbf{A}^{-1} & -\mathbf{A}^{-1} \mathbf{B} \mathbf{Q} \\ -\mathbf{Q} \mathbf{C} \mathbf{A}^{-1} & \mathbf{Q} \end{pmatrix}$$

où

$$\mathbf{Q} = [\mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B}]^{-1}$$

Exercice 5.5 Reprenez les résultats de l'exercice 2.6 et 2.1, calculez et dessinez des intervalles de confiance pour la prévision de la variable expliquée (en choisissant quelques valeurs pour x).

TAB. 5.6 – Consommation de crème glacée

consommation y	revenu x_1	température x_2
386	78	41
374	79	56
393	81	63
425	80	68
406	76	69
344	78	65
327	82	61
288	79	47
269	76	32
256	79	24
286	82	28
298	85	26
329	86	32
318	83	40
381	84	55
381	82	63
470	80	72
443	78	72
386	84	67
342	86	60
319	85	44
307	87	40
284	94	32
326	92	27
309	95	28
359	96	33
376	94	41
416	96	52
437	91	64
548	90	71

Exercice 5.6 La consommation de crème glacée d'individus a été mesurée pendant 30 périodes. L'objectif est de déterminer si la consommation dépend du revenu et de la température. Les données sont dans le tableau 1. On sait en outre que

$$\begin{aligned}
\sum_{i=1}^n y_i &= 10783, & \sum_{i=1}^n x_{i1} &= 2538, & \sum_{i=1}^n x_{i2} &= 1473, \\
\sum_{i=1}^n y_i^2 &= 4001293, & \sum_{i=1}^n x_{i1}^2 &= 215846, & \sum_{i=1}^n x_{i2}^2 &= 80145, \\
\sum_{i=1}^n x_{i1}y_i &= 912813, & \sum_{i=1}^n x_{i2}y_i &= 553747, & \sum_{i=1}^n x_{i1}x_{i2} &= 123650,
\end{aligned}$$

et que

$$\begin{pmatrix} 215846 & 123650 \\ 123650 & 80145 \end{pmatrix}^{-1} = \begin{pmatrix} 3.987998 & -6.152797 \\ -6.152797 & 10.740450 \end{pmatrix} \times \frac{1}{100000}.$$

Considérons le modèle de régression

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i,$$

où les ε_i sont des termes d'erreur normaux indépendants et équidistribués. Attention ! Ce modèle n'a pas de constante.

1. Estimez β_1 et β_2 par la méthode des moindres carrés ordinaires.
2. Sachant que la somme des carrés des résidus vaut 38912.310, estimez la variance des erreurs.
3. Donnez la valeur ajustée et le résidu pour la première observation du tableau 1.
4. Estimez la matrice variance-covariance du vecteur $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$.
5. La somme des résidus de ce modèle est-elle nulle (réponse sans calcul)? Justifiez en deux lignes (et toujours sans calcul).
6. Testez (au niveau de 95%) la nullité du coefficient de régression de la variable “température”.

Exercice 5.7 En considérant le même modèle que dans l'exercice 5.6, on veut tester l'hypothèse que $2\beta_1 = \beta_2$.

1. Donnez une matrice de contrainte \mathbf{R} et le vecteur \mathbf{r} à utiliser pour construire ce test. (La notation est celle utilisée au cours).
2. Donnez l'expression théorique et simplifiée de la statistique de test.
3. Faites le test. Peut-on admettre au niveau de 95% l'hypothèse que $2\beta_1 = \beta_2$?

Exercice 5.8 Calculez l'estimateur de la variance des coefficients de régression dans le cas d'un modèle à une constante et une variable explicative. Écrivez ces variances de manière scalaire.

Exercice 5.9 Les matrices définies en (5.11) et (5.17) définissent le même sous-espace linéaire. Donnez les applications linéaires (les matrices) permettant de passer de la matrice (5.11) à la matrice (5.17) et réciproquement. Ensuite, faites le produit des deux matrices obtenues.

Exercice 5.10 Question préliminaire : soit $\mathbf{P}_c = \mathbf{I} - \frac{11'}{n}$, le projecteur qui centre les données, \mathbf{I} la matrice identité, et \mathbf{P}_X le projecteur sur le sous-espace engendré par les colonnes de la matrice \mathbf{X} . La première colonne de \mathbf{X} est constituée de 1. Montrez que

$$\mathbf{P}_c(\mathbf{I} - \mathbf{P}_X) = (\mathbf{I} - \mathbf{P}_X).$$

(Inutile de se lancer dans des calculs compliqués, un argument simple se référant à des résultats donnés au cours suffit).

Calculez ensuite les espérances des trois sommes des carrés pour le tableau d'analyse de la variance correspondant au test :

$$\begin{cases} H_0 : & \beta_j = 0, \text{ pour tout } j = 2, \dots, p, \\ H_1 : & \text{au moins un des } \beta_j \neq 0, \end{cases}$$

où β_1 est le coefficient de régression se rapportant à la constante,

1. dans le cas général où H_0 n'est pas supposé vrai,
2. dans le cas où H_0 est vrai.

Sous H_0 , que valent les espérances des trois carrés moyens?

Indications :

1. les calculs sont plus simples en utilisant le projecteur qui centre les données,
2. l'espérance d'une variable aléatoire khi-carré est égale à son nombre de degrés de liberté.

Exercice 5.11 Pour étudier le comportement maternel de rats de laboratoire, nous éloignons le bébé rat de sa mère d'une distance fixée et enregistrons le temps nécessaire à la mère (en secondes) pour ramener son bébé au nid. Nous réalisons cette expérience avec des bébés rats de 5, 20 et 35 jours. Les données figurent ci-dessous pour six bébés par groupe. On donne le tableau d'analyse de la variance suivant :

TAB. 5.7 – Temps selon les groupes

5 jours	15	10	25	15	20	18
20 jours	30	15	20	25	23	20
35 jours	40	35	50	43	45	40

TAB. 5.8 – Tableau d'analyse de la variance

	Sommes des carrés	degrés de liberté	Carrés moyens	F	Sig.
Inter-groupes	2100.000	2	1050.000	40.127	0.000
Intra-groupe	392.500	15	26.167		
Total	2492.500	17			

1. Peut-on dire au niveau de probabilité 0.05 que le temps nécessaire pour ramener le bébé dépend de l'âge ? Justifiez votre réponse.
2. Donnez le quantile d'ordre 0.95 de la variable de Fisher correspondant à l'analyse de la variance ?
3. À partir du tableau d'analyse de la variance donnez la variance (marginale) de la variable "secondes".

Exercice 5.12 Une autre partie de l'étude d'Eysenck (1974) mentionnée précédemment comparait les sujets plus jeunes et plus âgés quand à leur aptitude à se rappeler le matériel alors qu'on les avait prévenus qu'ils devaient mémoriser les données de manière à s'en souvenir ultérieurement (cette tâche exigeait vraisemblablement un niveau élevé de traitement.) Les données figurent dans le tableau 5.9 (la variable dépendante étant le nombre d'éléments rappelés).

TAB. 5.9 – Nombre d'éléments rappelés selon l'âge

Sujets plus jeunes :	21	19	17	15	22	16	22	22	18	21
Sujets plus âgés :	10	19	14	5	10	11	14	15	11	11

1. Effectuez une analyse de variance afin de comparer les moyennes de ces deux groupes.

Exercice 5.13 Une autre approche des données d'Eysenck (1974) consiste à comparer quatre groupes de sujets. L'un des groupes se composait de jeunes sujets à qui l'on présentait les mots dans une condition qui suscitait un niveau peu élevé de traitement. Un deuxième groupe se composait des sujets jeunes à qui l'on donnait des tâches requérant un niveau de traitement plus élevé. Les deux autres groupes comprenaient des sujets plus âgés à qui l'on donnait des tâches requérant un niveau de traitement soit peu élevé, soit élevé. Les données sont les suivantes :

1. Effectuez une analyse de variance à un critère de classification sur ces données.
2. Effectuez à présent une analyse de variance à un critère de classification en opposant les traitements 1 et 3 combinés ($n = 2$) aux traitements 2 et 4 combinés. A quelle question répondez-vous ?

Exercice 5.14 Cet exercice est une étude hypothétique similaire à une expérience importante réalisée par Siegel (1975) sur la tolérance à la morphine. La morphine est un médicament souvent utilisé pour atténuer la douleur. Cependant, des administrations répétées de morphine provoquent un phénomène de tolérance : la morphine a de moins en moins d'effet (la réduction de la douleur est de moins en moins forte) au fil du

TAB. 5.10 – Nombre d'éléments rappelés selon l'âge et le niveau

Jeunes/Peu élevé	8	6	4	6	7	6	5	7	9	7
Jeunes/Elevé	21	19	17	15	22	16	22	22	18	21
Agés/Peu élevé	9	8	6	8	10	4	6	5	7	7
Agés/Elevé	10	19	14	5	10	11	14	15	11	11

temps. Pour mettre en évidence la tolérance à la morphine, on a souvent recours à une expérience qui consiste à placer un rat sur une surface trop chaude. Lorsque la chaleur devient insupportable, le rat va se mettre à se lécher les pattes ; le temps de latence qui précède le moment où le rat commence à se lécher les pattes est utilisé comme mesure de sa sensibilité à la douleur. Un rat qui vient de recevoir une injection de morphine montre en général un temps de latence plus long, ce qui montre que sa sensibilité à la douleur est réduite. Le développement de la tolérance à la morphine est indiqué par le fait que les latences se raccourcissent progressivement (signe d'une sensibilité accrue) sous l'effet des injections répétées de morphine.

Prenons une expérience impliquant cinq groupes de rats. Chaque groupe participe à quatre essais, mais les données d'analyse sont uniquement prélevées lors du dernier essai critique (test). On désigne les groupes en indiquant le traitement appliqué lors des trois premiers essais puis du quatrième. Nous avons les cinq groupes suivant :

1. Le premier groupe (M-M) a reçu des injections de morphine lors des trois premiers essais dans l'environnement de test, puis de nouveau lors du quatrième essai, dans le même environnement ;
2. Le deuxième groupe (M-S) a reçu une injection de morphine (dans l'environnement de test) lors des trois premiers essais puis une solution saline lors du quatrième ;
3. Les animaux du troisième groupe (Mc-M) ont reçu une injection de morphine lors des trois premiers essais, effectués dans leur cage habituelle, puis la même injection lors du quatrième essai, mais dans l'environnement de test standard, qu'ils ne connaissaient pas ;
4. Le quatrième groupe (S-M) a reçu une injection de solution saline durant les trois premiers essais (dans l'environnement de test) et de morphine lors du quatrième ;
5. Enfin, le cinquième groupe (S-S) a reçu une injection de solution saline lors des quatre essais.

Les temps de latence (en secondes) selon les groupes sont présentés dans le tableau 5.11. Peut-on affirmer

TAB. 5.11 – Temps de latence selon les groupes

M-S	M-M	S-S	S-M	Mc-M
3	2	14	29	24
5	12	6	20	26
1	13	12	36	40
8	6	4	21	32
1	10	19	25	20
1	7	3	18	33
4	11	9	26	27
9	19	21	17	30

que :

1. Les cinq groupes ont une perception de la douleur identique malgré les différents traitements (à 99%) ;
Un tableau de l'analyse de la variance a déjà été partiellement calculé ;
2. Le groupe M-M et S-S ont une sensibilité différente à la douleur (à 99%).
Un tableau de l'analyse de la variance a déjà été partiellement calculé.

Pour répondre aux questions a. et b. (à traiter séparément) :

- Posez les hypothèses ;
- Complétez les tableaux de l'analyse de la variance ;

TAB. 5.12 – Tableau incomplet d'analyse de la variance

	sommes de carrés	degrés de liberté	carrés moyens	F
Inter Groupes	3497,6	?	?	?
Intra Groupes	?	?	32	
Total	?	?		

TAB. 5.13 – Tableau incomplet d'analyse de la variance

	sommes de carrés	degrés de liberté	carrés moyens	F
Inter Groupes	4	?	?	?
Intra Groupes	?	?	?	
Total	504	?		

– Testez les hypothèses.

Exercice 5.15 Les données suivantes représentent les tailles et poids réels pour des étudiants américains de sexe masculin. Les mesures sont exprimées en pouces et en livres.

1. Estimez les coefficients du modèle

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, i = 1, \dots, n,$$

où les ε_i sont des termes d'erreur normaux, non corrélés de moyenne nulle et homoscedastiques.

2. Donnez un estimateur sans biais de la variance des erreurs.

3. Que vaut la valeur ajustée pour un individu mesurant 70 pouces ?

4. Peut-on affirmer au niveau de probabilité de 0.95 pour-cents, que la pente de la droite de régression vaut 5 (test bilatéral) ?

On a déjà réalisé les calculs suivants :

$$\bar{x} = 70.7544, \quad \bar{y} = 158.26, \quad \frac{1}{n} \sum_{i=1}^n x_i^2 = 5012.7368, \quad \frac{1}{n} \sum_{i=1}^n y_i^2 = 25388.4386, \quad \frac{1}{n} \sum_{i=1}^n x_i y_i = 11226.33596.$$

Exercice 5.16 Une autre étude sur le même sujet nous donne la droite de régression suivante :

$$\text{Poids} = -155 + 4.5 \text{ taille} + \text{résidus}. \quad (5.18)$$

On se demande si il n'est pas possible d'invalider cette hypothèse au moyen des données précédentes.

1. Construisez un test permettant de tester l'hypothèse

$$\begin{cases} H_0 & : \beta_1 = -155 \text{ et } \beta_2 = 4.5 \\ H_1 & : \text{au moins un des deux coefficients est différent de ces valeurs} \end{cases}$$

le coefficient β_1 est la constante et β_2 est le coefficient de régression se rapportant à la variable taille. Construisez d'abord le test de manière théorique en simplifiant au maximum l'expression, et en l'écrivant de manière scalaire (et non plus matricielle).

2. Testez l'hypothèse au moyen des données de l'exercice précédent ($\alpha = 0.05$).

3. Les données permettent-elles d'infirmer le modèle (5.18) ?

TAB. 5.14 – Tailles (en pouces) et poids (en livres) des étudiants

Taille x	Poids y	Taille x	Poids y
70	150	73	170
67	140	74	180
72	180	66	135
75	190	71	170
68	145	70	157
69	150	70	130
71.5	164	75	185
71	140	74	190
72	142	71	155
69	136	69	170
67	123	70	155
68	155	72	215
66	140	67	150
72	145	69	145
73.5	160	73	155
73	190	73	155
69	155	71	150
73	165	68	155
72	150	69.5	150
74	190	73	180
72	195	75	160
71	138	66	135
74	160	69	160
72	155	66	130
70	153	73	155
67	145	68	150
71	170	74	148
72	175	73.5	155
69	175		

Exercice 5.17 Soit le modèle à 5 paramètres suivant (dit d’analyse de la variance à deux facteurs) :

$$y_{ijk} = \mu + \alpha_j + \gamma_k + \varepsilon_{ijk},$$

où $j = 1, 2$, $k = 1, 2$, $\alpha_1 + \alpha_2 = 0$, $\gamma_1 + \gamma_2 = 0$, et les ε_{ijk} sont homoscédastiques et non-corrélés. L’objectif est d’estimer ces 5 paramètres avec un programme de régression multivariée classique.

- Écrivez le modèle sous forme matricielle.
- Quel est le rang de la matrice \mathbf{X} des variables explicatives ?
- Réécrivez le modèle et la matrice \mathbf{X} de manière à obtenir une matrice de plein rang. Quels sont les liens entre les nouveaux et les anciens paramètres ?
- Comment peut-on procéder pour estimer les paramètres avec un programme de régression multivarié ?

Chapitre 6

Multicolinéarité et choix des variables

6.1 La multicolinéarité

Parfois, dans le modèle linéaire général,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

la matrice \mathbf{X} n'est pas de plein rang. La matrice $\mathbf{X}'\mathbf{X}$ n'est alors pas inversible. Cependant il est encore possible de réaliser une régression, au sens où l'on peut toujours définir le sous-espace engendré par les colonnes de \mathbf{X} et projeter le vecteur \mathbf{y} sur ce sous-espace. Pour réaliser cette projection on utilisera l'inverse généralisée d'une matrice.

Définition 6.1 La matrice \mathbf{A}^- est une inverse généralisée (ou pseudo-inverse) de la matrice \mathbf{A} si et seulement si

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}.$$

L'inverse généralisée n'est pas unique, il existe donc une multitude de solutions, mais il n'existe qu'une seule inverse généralisée dite de Moore-Penrose \mathbf{A}^+ qui vérifie

$$\begin{aligned}\mathbf{A}\mathbf{A}^+\mathbf{A} &= \mathbf{A}, \\ \mathbf{A}^+\mathbf{A}\mathbf{A}^+ &= \mathbf{A}^+, \\ \mathbf{A}\mathbf{A}^+ &= (\mathbf{A}\mathbf{A}^+)', \\ \mathbf{A}^+\mathbf{A} &= (\mathbf{A}^+\mathbf{A})' .\end{aligned}$$

Exemple 6.1 Soit

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

On a

$$\mathbf{A}^+ = \frac{1}{25} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix},$$

et

$$\mathbf{A}^+\mathbf{A} = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

Exemple 6.2 Soit

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}.$$

On a

$$\mathbf{B}^+ = \frac{1}{20} \begin{pmatrix} 2 & 0 & 0 \\ -10 & 5 & 5 \\ 6 & 0 & 0 \end{pmatrix},$$

$$\mathbf{B}^+\mathbf{B} = \frac{1}{2} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

et

$$\mathbf{B}\mathbf{B}^+ = \frac{1}{2} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Exemple 6.3 On peut calculer l'inverse généralisée de Moore-Penrose d'une matrice non-carrée. Soit \mathbf{X} une matrice $n \times p$ de plein rang, alors

$$\mathbf{X}^+ = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

On peut vérifier qu'on a bien les propriétés de l'inverse de Moore-Penrose.

Il est possible de montrer que le projecteur

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}',$$

ne dépend pas de la pseudo-inverse utilisée.

Il est donc possible de faire une régression même avec une matrice \mathbf{X} qui n'est pas de plein rang, car le projecteur \mathbf{P}_X est unique, et il est donc possible de calculer la valeur ajustée

$$\mathbf{y}^* = \mathbf{P}_X\mathbf{y},$$

et le résidu.

$$\mathbf{e} = \mathbf{y} - \mathbf{y}^* = (\mathbf{I} - \mathbf{P}_X)\mathbf{y}.$$

Cependant, si la matrice n'est pas de plein rang, il existe une indétermination sur les coefficients de régression. En effet

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

n'est pas unique. On pourrait choisir le coefficient donné par l'inverse de Moore-Penrose, il n'est alors pas possible de réaliser une inférence sur $\hat{\boldsymbol{\beta}}$ car la variance de certains coefficients de régression n'existe pas.

Si la matrice \mathbf{X} n'est pas de plein rang, il est toujours possible de réaliser une régression, c'est-à-dire de construire un projecteur sur le sous-espace engendré par les colonnes de la matrice \mathbf{X} . Cependant, il y aura une indétermination sur les coefficients de régression et la non-existence de certaines variances. Pour ces raisons, on préfère aborder le problème en supprimant la ou les variables redondantes, ce qui nous ramènera à un problème standard.

6.2 Détection de la multicolinéarité

6.2.1 Le problème

La multicolinéarité se définit donc comme le fait que la matrice n'est pas de plein rang. Le logiciel signale alors le problème, car il ne peut pas procéder à l'estimation des paramètres. En pratique, on est plutôt confronté à des cas un peu limite, quand la matrice est "presque" de rang maximum, ce qui se traduira par un déterminant très petit ou par une valeur propre très petite.

Dans des cas simples, on constate parfois que deux variables explicatives ont un coefficient de corrélation très proche de 1 ou -1, mais la multicolinéarité est parfois difficile à détecter, car la relation entre les variables explicatives peut être complexe.

Pour détecter la multicolinéarité, nous utiliserons les indicateurs suivants :

- R^2 , le coefficient de détermination pour le modèle de régression de la variable y par les variables

$$x_1, \dots, x_j, \dots, x_p.$$

- R_{-j}^2 le coefficient de détermination pour le modèle de régression de la variable y par les variables

$$x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p.$$

- R_j^2 , le coefficient de détermination pour le modèle de régression de la variable x_j par les variables

$$x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p.$$

- D le déterminant de la matrice de corrélation des variables $x_1, \dots, x_j, \dots, x_p$.

Si la variable x_j est une combinaison linéaire des autres variables explicatives, alors :

- R_{-j}^2 sera proche de R^2 ,
- R_j^2 sera proche de 1.

6.2.2 Méthode de Klein

La méthode de Klein consiste à examiner les carrés des coefficients de corrélation par paires r_{jk}^2 entre les variables explicatives x_j et x_k , avec $j \neq k$. Si l'un de ces coefficients est plus grand que R^2 , alors on peut soupçonner la multicollinéarité.

6.2.3 Test de Farrar et Glauber

Le test de Farrar et Glauber teste que le déterminant D de la matrice de corrélation est égal à 1. Le coefficient ne peut être égal à 1 que si les variables explicatives sont orthogonales, le test est donc

H_0 : Les variables explicatives sont des variables aléatoires non-corrélées

H_1 : Les variables explicatives sont des variables aléatoires corrélées.

Le test est basé sur le fait que sous H_0 et avec une hypothèse de normalité, la statistique

$$\chi_{obs}^2 = - \left\{ n - 1 - \frac{1}{6}(2p + 5) \right\} \log D$$

a une distribution khi-carré à $p(p-1)/2$ degrés de liberté.

6.2.4 Facteur d'inflation

Le facteur d'inflation est la quantité

$$FI = \frac{1}{1 - R_j^2}.$$

Si cette quantité est très grande pour une variable, alors on peut soupçonner la multicollinéarité, et que cette multicollinéarité soit due à la variable x_j . Si il y a multicollinéarité, au moins deux variables sont impliquées.

6.2.5 Coefficient de Theil

Le coefficient de Theil est

$$m = R^2 - \sum_{j=1}^p (R^2 - R_{-j}^2).$$

Le coefficient de Theil est égal à 0 si toutes les paires de variables ont des coefficients de corrélation nulles, il n'a alors pas de multicollinéarité. Si ce n'est pas le cas, le coefficient de Theil peut être positif ou négatif.

6.2.6 Résolution du problème

Si une variable explicative est manifestement une combinaison linéaire des autres variables, alors, on peut supprimer cette variable du modèle. Une autre optique consiste à utiliser une technique itérative pour la construction du modèle.

6.3 Méthodes de choix de variables

Afin de tenter de contrôler le problème de la multicollinéarité, plusieurs méthodes itératives de construction de modèles ont été proposées.

6.3.1 Méthode Backward

La technique consiste à introduire toutes les variables dans le modèle, à condition bien sûr que l'on puisse calculer sans indétermination les coefficients de régression. Tant que certains t de Student des coefficients de régression sont en dessous du seuil critique, à chaque étape, on élimine une variable. On élimine la variable ayant le t de Student le plus proche de 0.

6.3.2 Méthode Forward

La méthode Forward consiste à ajouter une variable dans le modèle à chaque étape. A la première étape, on introduit dans le modèle la variable x_j la plus corrélée avec y . Ensuite, à chaque étape, on calcule pour toutes les variables qui ne sont pas dans le modèle, les carrés des coefficients de corrélation partielle relativement aux variables qui sont déjà dans le modèle, afin de mesurer comment chacune des variables peut "améliorer" le modèle. On introduit dans le modèle la variable dont le carré du coefficient de corrélation partielle est le plus élevé. Enfin, on arrête la procédure, dès que l'on obtient un t de Student non-significatif.

6.3.3 Méthode Stepwise

La méthode Stepwise alterne une étape où l'on ajoute une variable dans le modèle avec une étape où l'on enlève éventuellement une variable du modèle. La règle d'introduction d'une variable est la même que dans la méthode Forward. Ensuite on fait un test de Student sur chacun des coefficients, et éventuellement, on enlève une variable du modèle.

6.3.4 Mise en garde

Il faut cependant être prudent en utilisant des méthodes itératives pour la construction de modèles. En effet, ces techniques sont basées sur la répétition de tests d'hypothèses. Ces tests sont susceptibles de conduire à des décisions fausses, et donc la multiplication de tests augmente les probabilités de prendre des décisions fausses. Plus le nombre de variables susceptibles d'entrer dans le modèle est élevé, plus on risque d'introduire à tort des variables non-pertinentes.

Chapitre 7

Méthode des moindres carrés généralisés

7.1 Les hypothèses du modèle linéaire général

Défini sous forme matricielle, le modèle linéaire s'écrit.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

où

- \mathbf{X} est une matrice de constantes (non-aléatoire) de plein rang de dimension $n \times p$ des x_{ij} .
- $\boldsymbol{\beta}$ est un vecteur (inconnu) de \mathbb{R}^p .
- $\boldsymbol{\varepsilon}$ est un vecteur (inconnu) de dimension n de variables aléatoires ε_i .
- $E(\boldsymbol{\varepsilon}) = \mathbf{0}$,
- $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2$ (homoscédasticité).
- $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ (toutes les corrélations sont nulles).

Seuls \mathbf{y} et \mathbf{X} sont observés.

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

- $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ où \mathbf{X} est une matrice $n \times p$ et $\boldsymbol{\beta} \in \mathbb{R}^p$,
- $\text{var}(\mathbf{y}) = \mathbf{I}\sigma_\varepsilon^2$ où \mathbf{I} est une matrice identité $n \times n$ et σ_ε^2 est un scalaire.

La condition $\text{var}(\mathbf{y}) = \mathbf{I}\sigma_\varepsilon^2$ inclut en fait deux hypothèses :

1. absence de corrélation entre les termes d'erreur (les éléments extradiagonaux de la matrice $\text{var}(\mathbf{y})$ sont nuls).
2. absence d'hétéroscédasticité (tous les éléments diagonaux de la matrice $\text{var}(\mathbf{y})$ sont égaux).

Dans beaucoup d'applications ces deux hypothèses ne sont pas réalistes. Dans des séries temporelles, les termes d'erreur sont souvent corrélés. De même, si l'on analyse des unités statistiques régies par un effet de taille, alors les variances, et donc les termes d'erreur, sont aussi régies par un effet de taille, il y a donc hétéroscédasticité. Dans un premier temps, nous allons lever ces deux hypothèses et proposer une méthode d'estimation : la méthode des moindres carrés généralisés. Ensuite, nous appliquerons cette méthode aux problèmes d'hétéroscédasticité et de corrélation des termes d'erreur.

7.2 La méthode des moindres carrés généralisés

Soit le modèle général :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (7.1)$$

où \mathbf{X} est une matrice de constantes,

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}, \text{var}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \boldsymbol{\Omega},$$

et $\boldsymbol{\Omega}$ est une matrice $n \times n$, symétrique, définie positive, de plein rang. La matrice $\boldsymbol{\Omega} = [\omega_{ij}]$ n'est pas nécessairement diagonale. Les hypothèses d'homoscédasticité et de non-corrélation des termes d'erreur sont donc levées.

Une première approche pour estimer β consiste à utiliser la méthode des moindres carrés généralisés. On minimise le critère :

$$Q_G(\beta) = (\mathbf{y} - \mathbf{X}\beta)' \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta).$$

En annulant la dérivée par rapport à β , on obtient l'estimateur par les moindres carrés généralisés (MCG) :

$$\frac{\partial Q_G(\beta)}{\partial \beta} = 2\mathbf{X}'\Omega^{-1} (\mathbf{y} - \mathbf{X}\beta) = \mathbf{0},$$

et finalement, en supposant que $(\mathbf{X}'\Omega^{-1}\mathbf{X})$ est inversible :

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

L'estimateur des moindres carrés généralisés est sans biais. En effet,

$$\mathbb{E}(\hat{\beta}_{MCG}) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1}\mathbb{E}(\mathbf{X}\beta + \varepsilon) = \beta.$$

Notons que l'estimateur des moindres carrés ordinaires (MCO) est également sans biais même pour le modèle (7.1)

$$\mathbb{E}(\hat{\beta}_{MCO}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbb{E}(\mathbf{X}\beta + \varepsilon) = \beta.$$

Le fait d'avoir des termes d'erreur corrélés et de l'hétéroscédasticité ne fait pas perdre la propriété d'absence de biais de $\hat{\beta}_{MCO}$. Cependant, l'estimateur des moindres carrés ordinaires n'est plus l'estimateur optimal pour le modèle (7.1).

Théorème 7.1 (*généralisé de Gauss-Markov*) Sous le modèle (7.1), l'estimateur des moindres carrés généralisés $\hat{\beta}_{MCG} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y}$ est le meilleur (au sens de la plus petite variance) estimateur linéaire en \mathbf{y} sans biais de β .

La démonstration est une généralisation du théorème de Gauss-Markov développée sous les hypothèses d'absence d'autocorrélation et d'hétéroscédasticité.

La variance de $\hat{\beta}_{MCG}$ se calcule assez facilement

$$\begin{aligned} \text{var}(\hat{\beta}_{MCG}) &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1} \text{var}(\mathbf{y}) \Omega^{-1} \mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1} \sigma_\varepsilon^2 \Omega \Omega^{-1} \mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \sigma_\varepsilon^2. \end{aligned}$$

et peut être estimée sans biais par

$$\widehat{\text{var}}(\hat{\beta}_{MCG}) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \hat{\sigma}_\varepsilon^2.$$

où

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-p} (\mathbf{y} - \mathbf{X}\hat{\beta}_{MCG})' \Omega^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}_{MCG}).$$

7.3 Estimateur des moindres carrés généralisés et projection oblique

L'estimateur des moindres carrés généralisés permet de définir des valeurs ajustées

$$\mathbf{y}^* = \mathbf{X}\hat{\beta}_{MCG} = \mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1}\mathbf{y} = \mathbf{P}_{XG}\mathbf{y}.$$

où \mathbf{P}_{XG} est une matrice idempotente représentant un projecteur oblique sur le sous-espace engendré par les colonnes de la matrice \mathbf{X} (l'image de \mathbf{X} ou $\text{Im}(\mathbf{X})$) :

$$\mathbf{P}_{XG} = \mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}'\Omega^{-1}.$$

On peut également calculer les résidus. En effet,

$$\hat{\varepsilon} = \mathbf{e} = \mathbf{y} - \mathbf{y}^* = (\mathbf{I} - \mathbf{P}_{XG})\mathbf{y} = \mathbf{P}_{XG}^\perp \mathbf{y} = \mathbf{P}_{XG}^\perp \varepsilon$$

où

$$\mathbf{P}_{XG}^\perp = \mathbf{I} - \mathbf{P}_{XG}.$$

La matrice \mathbf{P}_{XG}^\perp est également idempotente, et est aussi un projecteur oblique sur l'orthogonal du sous-espace engendré par les colonnes de la matrice \mathbf{X} .

7.4 Retour au moindres carrés ordinaires

Supposons que nous identifions une matrice \mathbf{M} de dimension $n \times n$ et de plein rang tel que le vecteur

$$\mathbf{u} = [u_i] = \mathbf{M}\boldsymbol{\varepsilon},$$

soit non-corrélés et homoscedastique, autrement dit, $E[u_i] = 0$, $\text{cov}[u_i, u_j] = 0, i \neq j$, et $\text{var}[u_i] = \sigma_u^2$.

Remarquons que

$$\text{var}[\mathbf{u}] = \mathbf{I}\sigma_u^2 = \text{var}[\mathbf{M}\boldsymbol{\varepsilon}] = \mathbf{M}\text{var}[\boldsymbol{\varepsilon}]\mathbf{M}' = \mathbf{M}\boldsymbol{\Omega}\sigma_\varepsilon^2\mathbf{M}'.$$

En prémultipliant cette dernière équation par \mathbf{M}' et en la postmultipliant par \mathbf{M} , on a

$$\mathbf{M}'\mathbf{M}\sigma_u^2 = \mathbf{M}'\mathbf{M}\boldsymbol{\Omega}\sigma_\varepsilon^2\mathbf{M}'\mathbf{M}.$$

Comme \mathbf{M} est de plein rang, $\mathbf{M}'\mathbf{M}$ est inversible. En inversant les deux membres de cette égalité, il vient

$$\frac{(\mathbf{M}'\mathbf{M})^{-1}}{\sigma_u^2} = \frac{(\mathbf{M}'\mathbf{M})^{-1}\boldsymbol{\Omega}^{-1}(\mathbf{M}'\mathbf{M})^{-1}}{\sigma_\varepsilon^2},$$

ce qui permet d'obtenir, en prémultipliant et en postmultipliant par $\mathbf{M}'\mathbf{M}$:

$$\frac{\mathbf{M}'\mathbf{M}}{\sigma_u^2} = \frac{\boldsymbol{\Omega}^{-1}}{\sigma_\varepsilon^2}.$$

Alors, en prémultipliant le modèle général (7.1) par \mathbf{M} , on obtient

$$\mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\varepsilon} = \mathbf{M}\mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

En posant $\check{\mathbf{y}} = \mathbf{M}\mathbf{y}$ et $\check{\mathbf{X}} = \mathbf{M}\mathbf{X}$, on peut réécrire le modèle

$$\check{\mathbf{y}} = \check{\mathbf{X}}\boldsymbol{\beta} + \mathbf{u}.$$

On retrouve alors un modèle classique avec des termes d'erreur d'autocorrélation nulle et homoscedastique. La matrice des variables explicatives est $\check{\mathbf{X}} = \mathbf{M}\mathbf{X}$ et le vecteur des variables expliquées est $\check{\mathbf{y}} = \mathbf{M}\mathbf{y}$. L'estimateur des moindres carrés ordinaires est alors

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

Comme $\mathbf{M}'\mathbf{M} = \boldsymbol{\Omega}^{-1}\sigma_u^2/\sigma_\varepsilon^2$, on retrouve l'estimateur par les moindres carrés généralisés :

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

7.5 Méthode du maximum de vraisemblance

Supposons que le modèle général ait des termes d'erreur multinormaux, autrement dit

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (7.2)$$

où \mathbf{X} est une matrice non-aléatoire et

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2\boldsymbol{\Omega}).$$

Si $\boldsymbol{\Omega}$ est de plein rang, la fonction de densité de $\boldsymbol{\varepsilon}$ vaut

$$f_{\boldsymbol{\varepsilon}}(\mathbf{u}) = \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2} |\boldsymbol{\Omega}|^{1/2}} \exp\left(-\frac{\mathbf{u}'\boldsymbol{\Omega}^{-1}\mathbf{u}}{2\sigma_\varepsilon^2}\right),$$

La fonction de vraisemblance s'écrit :

$$L(\boldsymbol{\beta}, \sigma_\varepsilon^2) = \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2} |\boldsymbol{\Omega}|^{1/2}} \exp - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma_\varepsilon^2}.$$

Le logarithme de la fonction de vraisemblance vaut :

$$\begin{aligned}\ell(\boldsymbol{\beta}, \sigma_\varepsilon^2) &= \log L(\boldsymbol{\beta}, \sigma_\varepsilon^2) \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_\varepsilon^2) - \frac{1}{2} \log |\boldsymbol{\Omega}| - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma_\varepsilon^2}.\end{aligned}$$

En annulant les dérivées partielles par rapport aux paramètres, on obtient

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma_\varepsilon^2)}{\partial \boldsymbol{\beta}} = \frac{\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{y} - \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X} \boldsymbol{\beta}}{\sigma_\varepsilon^2} = \mathbf{0},$$

et

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma_\varepsilon^2)}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0.$$

La solution du maximum de vraisemblance pour $\boldsymbol{\beta}$ est à nouveau la même que la solution des moindres carrés généralisés, et vaut :

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

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

$$\hat{\sigma}_{\varepsilon MV}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

L'estimateur $\hat{\sigma}_{\varepsilon MV}^2$ est biaisé, mais il est possible de construire un estimateur sans biais

$$\hat{\sigma}^2 = \frac{1}{n-p} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

7.6 Intérêt des moindres carrés généralisés

Le problème de ce résultat est que la matrice $\boldsymbol{\Omega}$ n'est pas toujours connue. Il faudra alors estimer $\boldsymbol{\Omega}$. Cependant $\boldsymbol{\Omega}$ est constitué de n termes diagonaux et de $n(n-1)/2$ termes extra-diagonaux. Il est donc impossible de considérer la matrice $\boldsymbol{\Omega}$ comme un ensemble de paramètres à estimer. En effet, on aurait $n(n+1)/2$ paramètres, pour seulement n valeurs observées. On ne peut donc pas se passer de la formulation d'hypothèses afin d'estimer $\boldsymbol{\Omega}$.

On peut dès à présent formuler deux hypothèses qui ne sont pas suffisantes pour pouvoir estimer $\boldsymbol{\Omega}$, mais qui réduisent considérablement le nombre de paramètres à estimer.

- La matrice $\boldsymbol{\Omega}$ est diagonale. Autrement dit, il y a seulement de l'hétéroscédasticité et pas d'auto-corrélation, c'est-à-dire que les éléments de la diagonale ne sont pas tous égaux.
- Tous les éléments diagonaux de $\boldsymbol{\Omega}$ sont égaux. Autrement dit, il y a homoscedasticité, et de l'auto-corrélation. Il reste alors $n(n-1)/2$ paramètres à estimer.

Ces hypothèses sont réalistes car l'autocorrélation et l'hétéroscédasticité ne ne présentent pas souvent conjointement. L'autocorrélation est spécifique des modèles de séries temporelles, et l'hétéroscédasticité est typique des modèles en coupe (pour un instant particulier).

7.7 Détection de l'hétéroscédasticité

7.7.1 Le problème

L'hétéroscédasticité apparaît quand on traite des unités pour lesquelles il peut exister un effet de taille. Par exemple, si les unités statistiques sont des entreprises, les variances liées aux grandes entreprises seront beaucoup plus élevées que les variances des petites entreprises pour toutes les variables liées à cet effet de taille : nombre de travailleurs, investissement, chiffre d'affaires.

On suppose alors que le modèle s'écrit

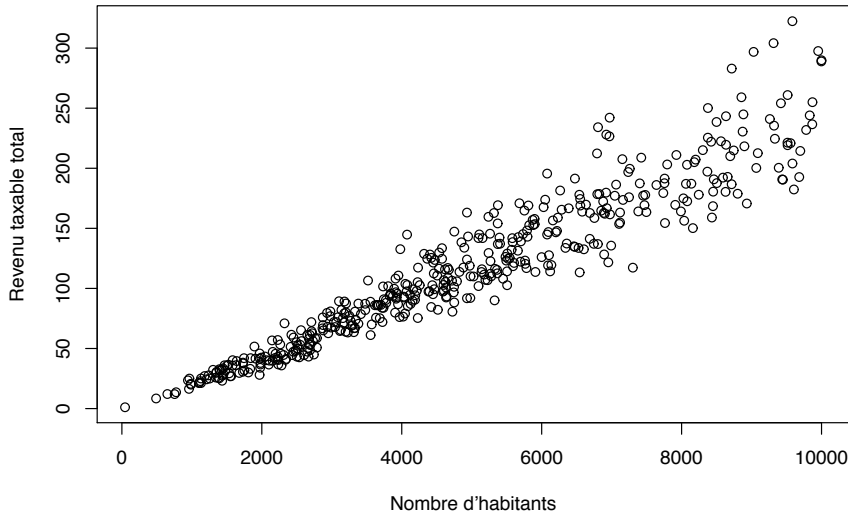
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

avec $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, et $\text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega}$, où

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_{\varepsilon 1}^2 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_{\varepsilon 2}^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{\varepsilon i}^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \sigma_{\varepsilon n}^2 \end{pmatrix}.$$

Exemple 7.1 Pour les 454 communes belges de moins de 20 000 habitants en 2004, on dispose de deux variables : le nombre habitants dans la communes en 2004, et le revenu total imposable de tous les habitants de la commune en millions d’euros. La Figure 7.1 montre le nuage de points pour le croisement de ces deux variables. La relation entre les deux variables est bien linéaire, mais la dispersion augmente avec la taille de la commune. C’est un cas typique d’hétéroscédasticité.

FIG. 7.1 – Nombre d’habitants et revenus total pour les 454 communes belges de moins de 20 000 habitants en 2004



7.7.2 Graphique des résidus

Il est toujours intéressant de représenter le nuage de points des résidus en fonction des différentes variables explicatives.

Exemple 7.2 Le Tableau 7.1 reprend la population en milliers d’habitants et les revenus totaux du cantons en millions de francs.

Le nuage de points de ces deux variables est présenté dans la Figure 7.2. La Figure 7.2 ne montre pas clairement l’hétéroscédasticité. Cependant, si l’on estime un modèle de régression simple par la méthode des moindres carrés ordinaires, on obtient

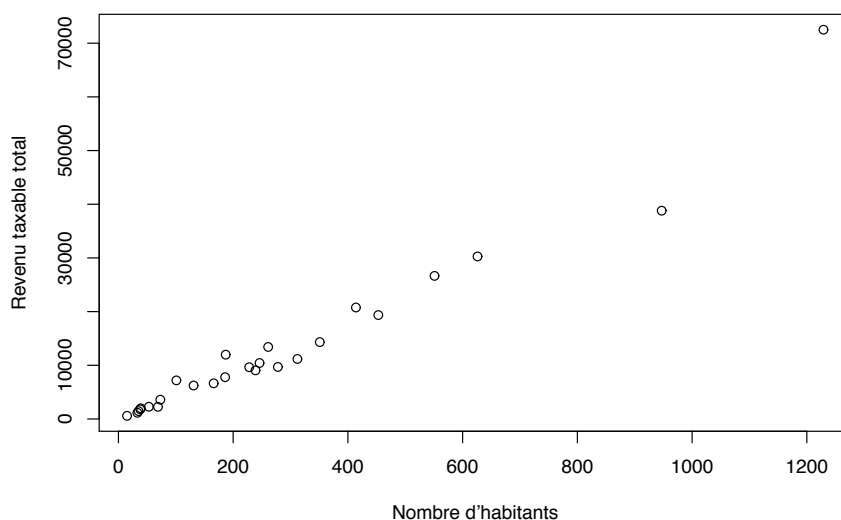
$$\text{Revenu} = -1353.66 + 51.81 \times \text{Population} + \text{Résidu}.$$

Dans la Figure 7.3, on a calculé les résidus de cette régression, et l’on a ordonné les cantons selon leur taille (en terme de population). Le graphique des résidus de la Figure 7.3 met bien en évidence l’hétéroscédasticité.

TAB. 7.1 – Population des cantons suisses en milliers d’habitants en 2001 et revenus des cantons

Canton	Population en milliers d’habitants	Revenu du canton en millions de fr.
Appenzell Rh.-Int	15	588
Obwald	33	1151
Uri	35	1468
Glaris	38	1796
Nidwald	39	1989
Appenzell Rh.-Ext	53	2273
Jura	69	2263
Schaffhouse	73	3593
Zoug	101	7191
Schwytz	131	6235
Neuchâtel	166	6645
Grisons	186	7782
Bâle-Ville	187	11978
Thurgovie	228	9639
Fribourg	239	9055
Soleure	246	10425
Bâle-Campagne	261	13415
Valais	278	9692
Tessin	312	11181
Lucerne	351	14319
Genève	414	20763
Saint-Gall	453	19356
Argovie	551	26655
Vaud	626	30272
Berne	947	38796
Zurich	1’229	72504

FIG. 7.2 – Nombre d’habitants et revenus total pour les cantons suisses



Exemple 7.3 Avec les données relatives aux communes belges de moins de 20000 habitants, les résidus sont présentés dans la Figure 7.4.

L’hétéroscédasticité apparaît en général sur les graphiques de résidus. Cependant, il est évidemment plus

FIG. 7.3 – Résidus de la régression en fonction des cantons classés par ordre croissant de population

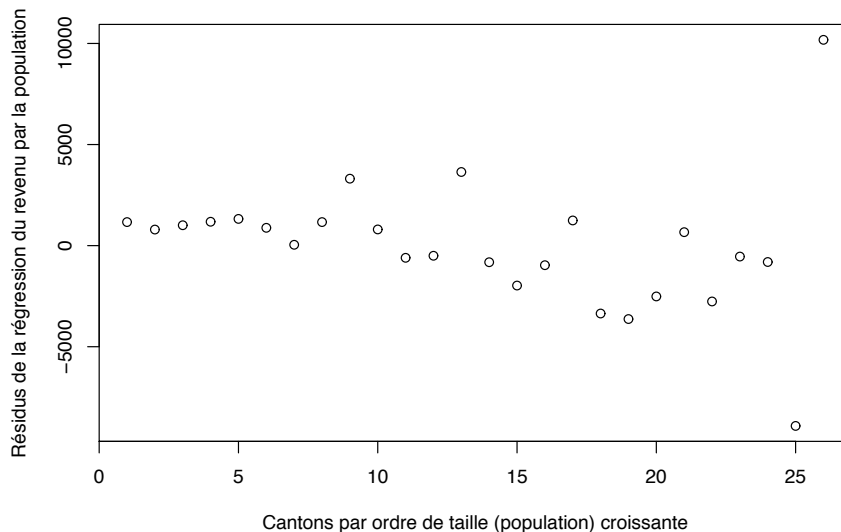
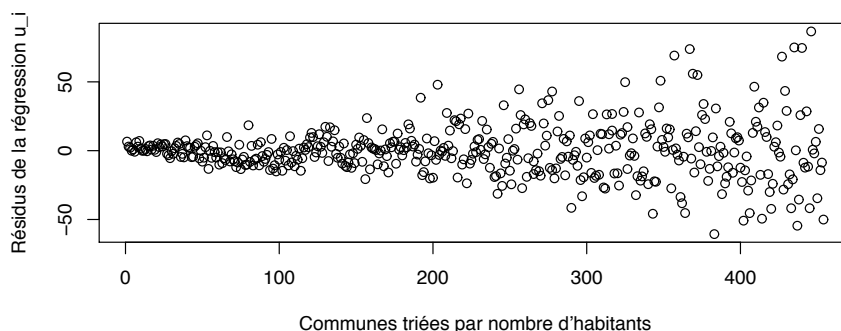


FIG. 7.4 – Résidus de la régression des revenus par la population en fonction des communes belges classés par ordre croissant de population



rigoureux de recourir à un test d'hypothèses.

7.7.3 Test de White

Le test de White sert à déterminer si les carrés des résidus sont liés aux variables explicatives. On estime d'abord les coefficients de la régression de la variable y par les variables explicatives x au moyen de la méthode des moindres carrés ordinaires. Ensuite, on effectue une seconde régression où la variable dépendante est le carré du résidu e_i de la première régression et les variables explicatives sont les variables explicatives de la première régression auxquelles on ajoute les carrés de ces variables et leurs produits.

Par exemple si le modèle pour lequel on soupçonne de l'hétéroscédasticité est

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i.$$

On estime par les MCO les coefficients β_0 , β_1 , β_2 et β_3 , ce qui permet d'estimer les résidus $e_i = \hat{\varepsilon}_i$. On considère ensuite la régression

$$e_i^2 = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i3} + \gamma_4 x_{i1}^2 + \gamma_5 x_{i2}^2 + \gamma_6 x_{i3}^2 + \gamma_7 x_{i1} x_{i2} + \gamma_8 x_{i1} x_{i3} + \gamma_9 x_{i2} x_{i3} + u_i.$$

Si on note R_e^2 le coefficient de détermination estimé pour ce nouveau modèle, il est possible de montrer que sous l'hypothèse nulle (H_0) d'homoscédasticité,

$$nR_e^2 \sim \chi_{q-1}^2,$$

où q est le nombre de variables utilisées dans ce nouveau modèle. On rejette H_0 si $nR_e^2 > \chi_{q-1, 1-\alpha}^2$. Un inconvénient de ce type de test est que le nombre de variables peut devenir très important au regard de la taille de l'échantillon.

Exemple 7.4 Avec les données sur les cantons suisses présentées dans la Tableau 7.1, on a d'abord estimé un simple modèle linéaire donné par :

$$\text{Revenu} = -1353.66 + 51.81 \times \text{Population} + \text{Résidus}.$$

On a ensuite estimé le modèle

$$\text{Résidus}^2 = \gamma_1 + \gamma_2 \times \text{Population} + \gamma_3 \times \text{Population}^2 + \text{Nouveaux Résidus},$$

et on a obtenu

$$\text{Résidus}^2 = 4959954.70 + -39391.24 \times \text{Population} + 101.21 \times \text{Population}^2 + \text{Nouveaux Résidus}.$$

Les coefficients sont présentés dans le Tableau 7.2 où l'on constate que les coefficient γ_2 et γ_3 sont significativement différents de 0. De plus $R^2 = 0.8963$.

TAB. 7.2 – Coefficients du modèle sur le carré des résidus

Coefficients	Estimate	Std. Error	statistique t	$\Pr(> t)$
constante	4959954.70	3077865.80	1.611	0.1207
Population	-39391.24	16919.21	-2.328	0.0291
Population ²	101.21	14.79	6.842	5.61e-07

On peut donc tester l'homoscédasticité. Comme $R^2 = 0.8963$, $nR^2 = 26 \times 0.8963 = 23.30$. De plus le quantile d'ordre 95% d'une variable aléatoire chi-carré à $q - 1 = 3 - 1 = 2$ degrés de liberté vaut $\chi_{2, 0.95}^2 = 5.991$, on rejette donc l'hypothèse d'homoscédasticité.

7.7.4 Test de Goldfeld-Quant

Le test de Goldfeld-Quant s'applique quand on soupçonne que l'hétéroscédasticité est liée à une variable particulière. Souvent cette variable est liée à un effet de taille. On soupçonne donc une relation positive entre une variable particulière et l'hétéroscédasticité. On commence par trier les données selon cette variable.

Ensuite on calcule deux régressions séparées sur les k premières et les k dernières observations. Il faut évidemment que $2k \leq n$, et que $k > p$ où p est le nombre de paramètres du modèle. On omet donc les $c = n - 2k$ observations centrales. On calcule ensuite SC_{res1} , SC_{res2} , SC_{regr1} , SC_{regr2} qui sont respectivement les sommes des carrés résiduelles et de la régression pour les k premières et les k dernières unités. La statistique de test est

$$\frac{SC_{res2}/(k-p)}{SC_{res1}/(k-p)},$$

et suit, sous l'hypothèse nulle (H_0) d'homoscédasticité, une loi de Fisher à $(k-p)$ et $(k-p)$ degrés de liberté.

Exemple 7.5 On utilise les données sur les cantons suisses présentées dans le Tableau 7.1. Les données ont été scindées en trois groupes : les 9 plus petit cantons, les 8 cantons intermédiaires et les 9 plus grands cantons.

- Sur les 9 plus petits cantons le modèle de régression estimé est donné par :

$$\text{Revenu} = -1009.73 + 68.86 \times \text{Population} + \text{Résidus}.$$

Les résidus valent

564.85232, -111.60586, 67.67657, 189.10020, 313.24141, -366.78162, -1478.52222, -423.95737, 1245.9

La somme des carrés des résidus vaut

$$SC_{res1} = 4522777.$$

- Sur les 9 plus grands cantons le modèle de régression estimé est donné par :

$$\text{Revenu} = -6887.5 + 59.2 \times \text{Population} + \text{Résidus}.$$

Les résidus valent

$$122.2070, \quad -401.5554, \quad 427.6876, \quad 3142.1572, \quad -573.5997, \quad 923.9085, \quad 100.9914, \quad -10377.8541, \quad 6636.0574.$$

La somme des carrés des résidus vaut

$$\text{SC}_{res2} = 163162186.$$

On peut dès lors calculer la statistique de test

$$F_{obs} = \frac{\text{SC}_{res2}/(k-p)}{\text{SC}_{res1}/(k-p)} = \frac{\sum_{i=18}^{26} \text{Résidus}_i^2}{\sum_{i=1}^9 \text{Résidus}_i^2} = \frac{163162186}{4522777} = 36.07566.$$

Comme $k = 9$, $p = 2$ et $k - p = 7$, sous H_0 , F_{obs} suit une loi de Fisher à $k - p = 7$ et $k - p = 7$ degrés de liberté. Comme $F_{7,7,0.95} = 3.787$, $F_{obs} > F_{7,7,0.95} = 3.787$, on rejette l'hypothèse d'homoscédasticité.

7.8 Estimation avec hétéroscédasticité

7.8.1 Si la variance est connue

Méthode 1

Dans certains cas, la variance peut être connue, par exemple, si les unités statistiques sont des entreprises, la variance peut être liée à un effet de taille notée z , et la taille de l'entreprise (par ex. le nombre de travailleurs peut être une variable connue). Dans ce cas, on est face au modèle :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

avec $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, et $\text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega}$, où

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_{\varepsilon 1}^2 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_{\varepsilon 2}^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{\varepsilon i}^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \sigma_{\varepsilon n}^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} z_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & z_i & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & z_n \end{pmatrix} = \sigma^2 \mathbf{Z},$$

où

$$\mathbf{Z} = \text{diag}(z_1, \dots, z_n).$$

Les valeurs z_i sont supposées strictement positives.

L'estimateur des moindres carrés généralisés peut être construit sans qu'il soit nécessaire de se poser des questions sur la valeur de σ^2 , en effet

$$\hat{\boldsymbol{\beta}}_{MCG} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = (\mathbf{X}'(\sigma^2\mathbf{Z})^{-1}\mathbf{X})^{-1} \mathbf{X}'(\sigma^2\mathbf{Z})^{-1}\mathbf{y} = (\mathbf{X}'\mathbf{Z}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}^{-1}\mathbf{y}. \quad (7.3)$$

Méthode 2

Il est également possible de traiter ce problème, avec les moindres carrés ordinaires, en effet, en notant

$$\mathbf{M} = \begin{pmatrix} 1/\sqrt{z_1} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1/\sqrt{z_2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/\sqrt{z_i} & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1/\sqrt{z_n} \end{pmatrix},$$

on a

$$\mathbf{M}'\mathbf{M} = \mathbf{Z}^{-1} = \mathbf{\Omega}^{-1}\sigma^2,$$

et donc le modèle

$$\mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\varepsilon},$$

avec $E(\mathbf{M}\boldsymbol{\varepsilon}) = \mathbf{0}$, et

$$\text{var}(\mathbf{M}\boldsymbol{\varepsilon}) = \mathbf{M}\mathbf{\Omega}\mathbf{M} = \mathbf{M}\mathbf{Z}\sigma^2\mathbf{M} = \sigma^2\mathbf{I}.$$

En pratique, on va simplement transformer les variables

$$\tilde{\mathbf{y}} = \mathbf{M}\mathbf{y} = \begin{pmatrix} \frac{y_1}{\sqrt{z_1}} & \cdots & \frac{y_i}{\sqrt{z_i}} & \cdots & \frac{y_n}{\sqrt{z_n}} \end{pmatrix},$$

$$\mathbf{u} = \mathbf{M}\boldsymbol{\varepsilon} = \begin{pmatrix} \frac{\varepsilon_1}{\sqrt{z_1}} & \cdots & \frac{\varepsilon_i}{\sqrt{z_i}} & \cdots & \frac{\varepsilon_n}{\sqrt{z_n}} \end{pmatrix},$$

et $\tilde{\mathbf{X}} = \mathbf{M}\mathbf{X}$ où $\tilde{\mathbf{X}} = (\tilde{x}_{ij})$ et $\tilde{x}_{ij} = x_{ij}/\sqrt{z_i}$. Le modèle s'écrit alors simplement

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{u},$$

et comme $\text{var}(\mathbf{u}) = \mathbf{I}\sigma^2$, on peut utiliser la méthode des moindres carrés ordinaire pour estimer $\boldsymbol{\beta}$.

Donc, avec $\tilde{\mathbf{y}} = \mathbf{M}\mathbf{y}$ comme vecteur de variables dépendantes et $\tilde{\mathbf{X}} = \mathbf{M}\mathbf{X}$ comme variables explicatives, on a à nouveau l'homoscédasticité. Dans ce cas, on peut utiliser l'estimateur par les moindres carrés ordinaires qui est

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}} = (\mathbf{X}'\mathbf{M}'\mathbf{M}\mathbf{X})^{-1} \mathbf{X}'\mathbf{M}'\mathbf{M}\mathbf{y} = (\mathbf{X}'\mathbf{Z}\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}\mathbf{y}$$

et qui n'est autre que (7.3).

7.8.2 Exemple de variance connue

Dans les données relatives aux communes belges de moins de 20000 habitants, la dispersion semble directement liée à la variable explicative (nombre d'habitants). La variance est donc proportionnelle au carré de la variable explicative. Nous pouvons considérer plusieurs modèles.

Modèle avec constante et hétéroscédasticité

Dans ce premier modèle, on utilise une régression avec une constante et des termes d'erreur hétéroscédastique.

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \text{ où } \text{var}(\varepsilon_i) = \sigma^2 x_i^2. \quad (7.4)$$

Pour se ramener à un modèle homoscédastique, on peut simplement diviser chacun des modèles par x_i , ce qui donne :

$$\frac{y_i}{x_i} = \frac{\beta_1}{x_i} + \beta_2 + \frac{\varepsilon_i}{x_i}, \text{ où } \text{var}(\varepsilon_i) = \sigma^2 x_i^2.$$

En posant $\tilde{y}_i = y_i/x_i$, $u_i = \varepsilon_i/x_i$ et $z_i = 1/x_i$, on obtient

$$\text{var}(u_i) = \text{var}\left(\frac{\varepsilon_i}{x_i}\right) = \frac{\text{var}(\varepsilon_i)}{x_i^2} = \frac{\sigma^2 x_i^2}{x_i^2} = \sigma^2.$$

Les nouveaux termes d'erreur sont maintenant homoscédastiques. Le nouveau modèle peut alors s'écrire

$$\tilde{y}_i = \beta_1 z_i + \beta_2 + u_i, \text{ où } \text{var}(u_i) = \sigma^2.$$

Le modèle se ramène à un modèle avec constante dont la variable dépendante est $\tilde{y}_i = y_i/x_i$ et la variable indépendante est $z_i = 1/x_i$. Les résultats sont données, dans le Tableau 7.3, qui montre que la coefficient β_1 n'est pas significativement différent de 0. On imaginera donc un modèle plus simple en supprimant le coefficient β_2 .

TAB. 7.3 – Estimation de paramètre de la régression avec constante de y/x par $1/x$

Coefficients	Estimations	Std. Error	statistique t	$\Pr(> t)$
constante (β_2)	1.143e-02	8.756e-05	130.556	$< 2e - 16$
$1/x$ (β_1)	-1.099e-01	1.521e-01	-0.722	0.47

Modèle sans constante et hétéroscédasticité

Au vu du résultat obtenu dans le Tableau 7.3, on peut supprimer du modèle (7.4) la constante. On obtient ainsi un modèle sans constante et avec hétéroscédasticité.

$$y_i = \beta x_i + \varepsilon_i, \text{ où } \text{var}(\varepsilon_i) = \sigma^2 x_i^2.$$

Pour se ramener à un modèle homoscédastique, on peut simplement diviser chacun des modèles par x_i , ce qui donne :

$$\frac{y_i}{x_i} = \beta + \frac{\varepsilon_i}{x_i}, \text{ où } \text{var}(\varepsilon_i) = \sigma^2 x_i^2.$$

En posant $\tilde{y}_i = y_i/x_i$ et $u_i = \varepsilon_i/x_i$, on obtient

$$\tilde{y}_i = \beta + u_i, \text{ où } \text{var}(u_i) = \sigma^2.$$

Le modèle réduit est donc extrêmement simple puisqu'on obtient une régression de $\tilde{y}_i = y_i/x_i$ par une constante comme variable explicative et que les termes d'erreur sont maintenant homoscédastiques.

En estimant le paramètre par la méthode des moindres carrés ordinaires, on obtient

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n \tilde{y}_i = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i},$$

ce qui donne

$$\tilde{y}_i = 0.01141 + \hat{u}_i,$$

et en multipliant par x_i , on revient au modèle de départ pour obtenir finalement :

$$y_i = 0.01141 \times x_i + \hat{\varepsilon}_i.$$

La Figure 7.5 montre, en outre, que l'hétéroscédasticité n'est presque plus présente dans les résidus \hat{u}_i .

FIG. 7.5 – Résidus \hat{u}_i de la régression sans constante du revenu par la population en fonction des communes classées par ordre croissant de population

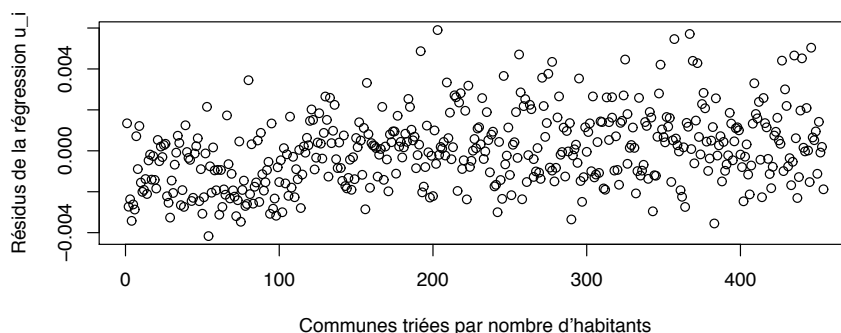
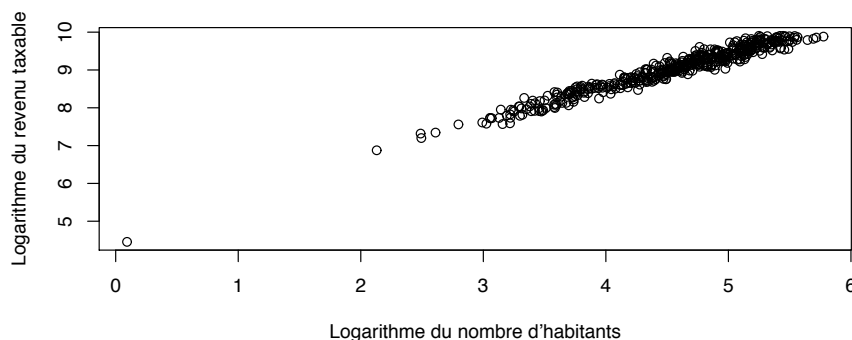


FIG. 7.6 – Nuage de points : logarithme du revenu par le logarithme du nombre d’habitants



Passage par le logarithme des variables

Quand on est en présence d’hétéroscédasticité, il est parfois intéressant d’utiliser le logarithme des variables. En utilisant les données des communes belges présentées dans la Figure 7.1, le nuage de points obtenu en croisant les logarithmes des variables “population” et “revenu” est présenté dans la Figure 7.6. On y constate que l’essentiel de l’hétéroscédasticité a disparu.

On pourrait donc concevoir le modèle suivant

$$\log(y_i) = \beta_1 + \beta_2 \log(x_i) + \varepsilon_i, \text{ où } \text{var}(\varepsilon_i) = \sigma^2. \quad (7.5)$$

Notons que ce modèle peut également s’écrire

$$\log\left(\frac{y_i}{x_i^{\beta_2}}\right) = \beta_1 + \varepsilon_i,$$

ou encore, en prenant l’exponentielle,

$$\left(\frac{y_i}{x_i^{\beta_2}}\right) = \exp \beta_1 \exp \varepsilon_i. \quad (7.6)$$

On peut estimer directement le modèle (7.5) par la méthode des moindres carrés ordinaires. Le Tableau 7.4 montre que les deux coefficients sont significativement différents de zéro.

TAB. 7.4 – Estimation de paramètre du modèle logarithmique

Coefficients	Estimations	Std. Error	statistique t	$\text{Pr}(> t)$
Constante	-5.21913	0.09921	-52.61	$< 2e - 16$
$\log(x)$	1.08139	0.01097	98.56	$< 2e - 16$

On obtient donc le modèle estimé

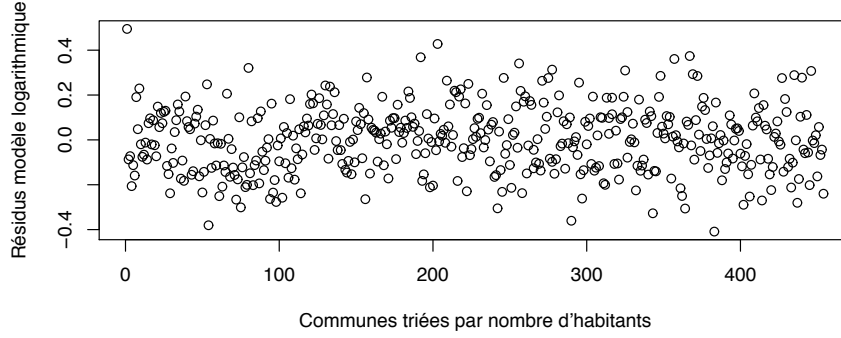
$$\log(y_i) = -5.21913 + 1.08139 \log(x_i) + \hat{\varepsilon}_i,$$

ou en écrivant sous la forme du modèle (7.6), on obtient

$$\left(\frac{y_i}{x_i^{1.08139}}\right) = \exp -5.21913 \exp \hat{\varepsilon}_i = 0.005412036 \cdot \exp \hat{\varepsilon}_i.$$

La Figure 7.7 montre que les résidus de la régression logarithmique sont homoscedastiques.

FIG. 7.7 – Résidus de la régression du modèle logarithmique classés par ordre croissant de population



7.8.3 Si la variance est inconnue

Dans la plupart des cas, on ne dispose pas d'une variable auxiliaire proportionnelle à la variance. Il est également exclu d'estimer cette variance, car la matrice

$$\mathbf{\Omega} = \begin{pmatrix} \sigma_{\varepsilon 1}^2 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_{\varepsilon 2}^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{\varepsilon i}^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \sigma_{\varepsilon n}^2 \end{pmatrix}$$

dépend de n paramètres. Le nombre de paramètres à estimer serait donc de $n + p$ et donc supérieur au nombre d'observations, ce qui est impossible.

Cependant, quand aucune hypothèse ne peut être faite sur la forme de l'hétéroscédasticité, White propose d'estimer la matrice $\mathbf{\Omega}$ par

$$\hat{\mathbf{\Omega}} = \begin{pmatrix} e_1^2 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & e_2^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & e_i^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & e_n^2 \end{pmatrix}, \quad (7.7)$$

où les e_i sont les résidus au moyen de la méthode des moindres carrés ordinaires. Notons que les e_i^2 sont des estimateurs biaisés de σ_i^2 , mais on peut montrer que ce sont des estimateurs convergents. On obtient alors l'estimateur de White

$$\hat{\beta}_{MCG} = \left(\mathbf{X}' \hat{\mathbf{\Omega}}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \hat{\mathbf{\Omega}}^{-1} \mathbf{y}. \quad (7.8)$$

Exemple 7.6 Reprenons les données sur les cantons suisses. Nous pouvons estimer par les MCO le modèle sans constante :

$$\text{Revenu} = \beta \times \text{Population} + \varepsilon,$$

on obtient l'estimation MCO :

$$\text{Revenu} = 49.450 \times \text{Population} + \hat{\varepsilon}.$$

Le Tableau 7.5 donne les résidus et les carrés des résidus. La Figure 7.8 montre le lien entre les carrés des résidus et la variable "population". Ensuite, on utilise la méthode des moindres carrés généralisés, en utilisant la matrice $\hat{\mathbf{\Omega}}$ définie en (7.7). On obtient alors l'estimation de White :

$$\text{Revenu} = 49.1982 \times \text{Population} + \hat{\varepsilon}.$$

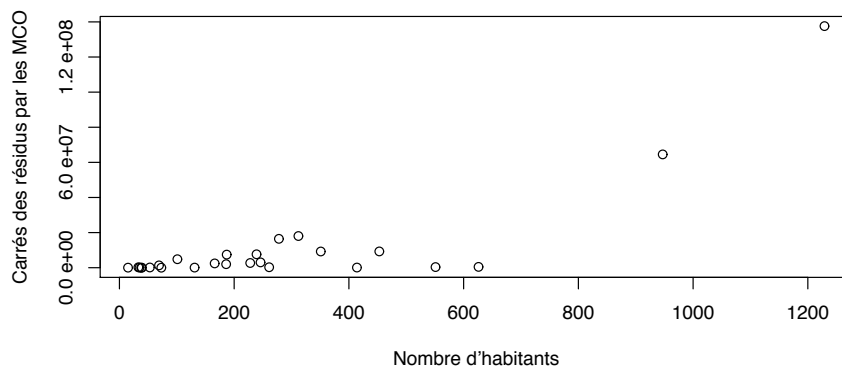
Notons que si l'on considère que l'hétéroscédasticité est donnée par $\text{var}(\varepsilon_k) = x_k^2 \sigma^2$, alors l'estimation par les moindres carrés généralisés (MCG) donne

$$\text{Revenu} = 45.343 \times \text{Population} + \hat{\varepsilon}.$$

TAB. 7.5 – Population des cantons suisses en milliers d'habitants, revenus des cantons, résidus de la régression et carrés des résidus

Canton	Population	Revenu	résidus	résidus ²
AppenzellRh.-Int	15	588	-153.75	23637.73
Obwald	33	1151	-480.84	231207.56
Uri	35	1468	-262.74	69032.25
Glaris	38	1796	-83.09	6903.79
Nidwald	39	1989	60.46	3655.57
AppenzellRh.-Ext	53	2273	-347.83	120988.97
Jura	69	2263	-1149.03	1320270.10
Schaffhouse	73	3593	-16.83	283.21
Zoug	101	7191	2196.58	4824960.09
Schwytz	131	6235	-242.91	59006.32
Neuchâtel	166	6645	-1563.65	2445007.73
Grisons	186	7782	-1415.65	2004054.36
Bâle-Ville	187	11978	2730.90	7457836.77
Thurgovie	228	9639	-1635.53	2674971.92
Fribourg	239	9055	-2763.48	7636827.02
Soleure	246	10425	-1739.63	3026308.85
Basel-Campagne	261	13415	508.63	258699.79
Valais	278	9692	-4055.02	16443184.69
Tessin	312	11181	-4247.31	18039641.13
Lucerne	351	14319	-3037.85	9228524.18
Genève	414	20763	290.82	84576.03
Saint-Gall	453	19356	-3044.72	9270314.64
Argovie	551	26655	-591.79	350216.39
Vaud	626	30272	-683.52	467198.46
Berne	947	38796	-8032.88	64527103.90
Zurich	1229	72504	11730.31	137600055.86

FIG. 7.8 – Données suisses, carrés des résidus par nombre d'habitants



7.9 L'autocorrélation des résidus

7.10 Un exemple d'autocorrélation

Le Tableau 7.6 contient les températures journalières du mois de janvier 2006 de la station de mesure d'Adelboden de Météosuisse. Ces données sont également présentées dans la Figure 7.9. qui montre directement un phénomène bien connu. La température ne change en général pas de manière brusque d'un jour à l'autre. La température d'un jour particulier ressemble donc souvent à la température du jour précédent.

TAB. 7.6 – Températures journalière du mois de janvier 2006

1	2	3	4	5	6	7	8	9	10	
0.6	1.5	-4.2	0.9	2.7	2.5	3.4	7.8	4.8	3.9	
11	12	13	14	15	16	17	18	19	20	
0.2	1.1	-1.3	-3	-0.2	1.5	1	1.3	-4.6	1.6	
21	22	23	24	25	26	27	28	29	30	31
0.9	-3.7	-5.4	-8.5	-11.1	-12	-13.3	-12.1	-13.2	-11.6	-6.9

Si T_t représente la température au jour t et T_{t-1} la température au jour précédent, la Figure 7.10 présente le nuage de points obtenu en croisant la température et la température du jour précédent.

FIG. 7.9 – Températures journalière du mois de janvier 2006

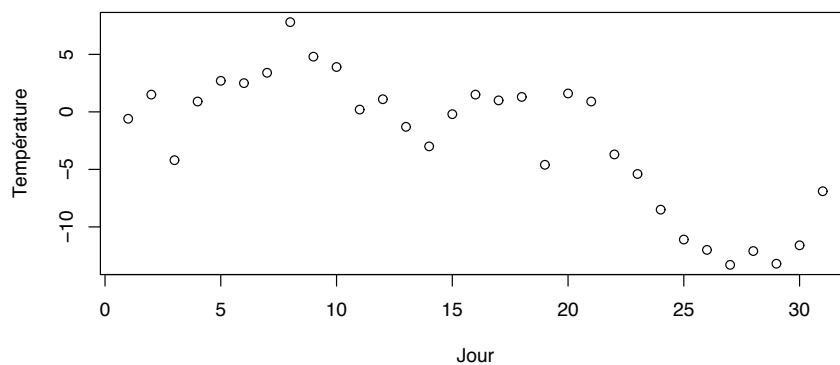
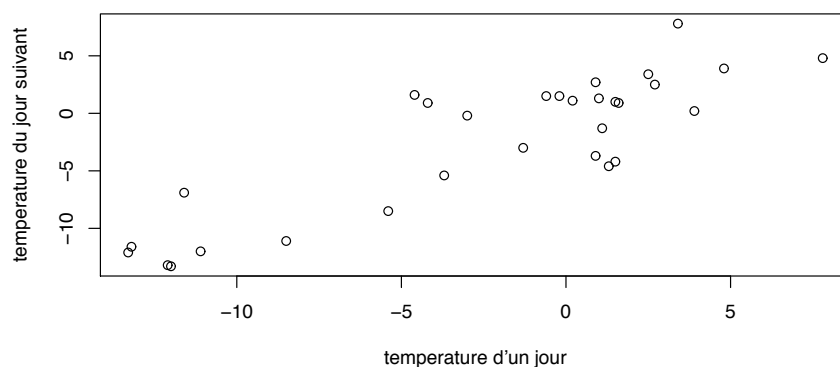


FIG. 7.10 – Températures journalière vs températures du jour suivant



On observe, en examinant la Figure 7.10, que les points semblent aligner le long d'une droite croissante. Pour prédire la température à un jour particulier, on pourrait dans un premier modèle simple utiliser la température du jours précédent. En utilisant un simple modèle linéaire, on obtient

$$T_t = \beta_1 + \beta_2 T_{t-1} + \varepsilon_t. \quad (7.9)$$

L'estimation des paramètres par les moindres carrés ordinaires donne

$$T_t = -0.49659 + 0.87665 \times T_{t-1} + \varepsilon_t,$$

et le R^2 est égal à 0.7564.

Cependant un simple test de Student nous montre que le coefficient β_1 n'est pas significativement différent de zéro. Il est donc plus judicieux d'estimer un modèle sans constante

$$T_t = \beta T_{t-1} + \varepsilon_t. \quad (7.10)$$

L'estimation du paramètre par les moindres carrés ordinaires donne

$$\hat{\beta} = \frac{\sum_{t=2}^{31} T_t T_{t-1}}{\sum_{t=2}^{31} T_{t-1}^2} = 0.9055,$$

ce qui donne

$$T_t = 0.9055 \times T_{t-1} + \hat{\varepsilon}_t.$$

Les modèles (7.9) et (7.10) où l'on tente d'expliquer une variable par ses valeurs passées s'appellent des modèles *autorégressif*. Ces modèles se justifient dans le traitement des séries temporelles, car généralement les valeurs des variables ne changent pas radicalement d'un temps à l'autre. Dans les séries temporelles, on utilisera souvent des modèles autorégressifs sur les termes d'erreur d'une régression pour prendre en compte cette "ressemblance".

7.10.1 La modélisation

Quand les données sont issues d'observations temporelles, on peut soupçonner les termes d'erreur du modèle linéaire d'être autocorrélés. Le modèle est alors

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

avec $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega}\sigma_\varepsilon^2$, et

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-3} & \rho_{n-2} & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-4} & \rho_{n-3} & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-5} & \rho_{n-4} & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \rho_{n-3} & \rho_{n-4} & \rho_{n-5} & \cdots & 1 & \rho_1 & \rho_2 \\ \rho_{n-2} & \rho_{n-3} & \rho_{n-4} & \cdots & \rho_1 & 1 & \rho_1 \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & \rho_2 & \rho_1 & 1 \end{pmatrix}.$$

Les coefficients $-1 < \rho_j < 1$ sont appelés coefficients d'autocorrélation. Cependant ce modèle est trop complexe pour être estimé directement, car il faudrait estimer $n - 1$ coefficients d'autocorrélation, ce qui est impossible avec seulement n paramètres. On aura donc recours à des modèles plus simple comme les modèles autorégressifs d'ordre 1.

7.10.2 Définition du processus autorégressif d'ordre un

Le processus autorégressif d'ordre un est un cas simple de série statistique dont les termes d'erreur sont autocorrélés. Considérons la série temporelle des ε_i définie pour toute valeur i de \mathbb{Z} , et régie par le modèle suivant :

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i, i \in \mathbb{Z}.$$

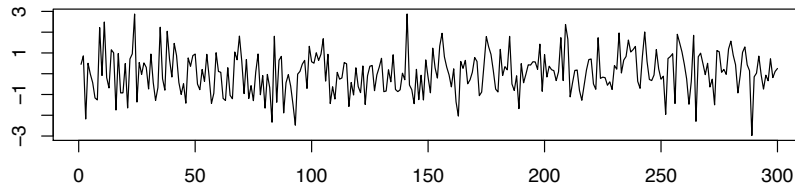
où

- les u_i sont de moyennes nulles, homoscedastiques, de variance σ_u^2 et non-corrélés, pour tout $i \in \mathbb{Z}$,
- $|\rho| < 1$,
- $\text{cov}(\varepsilon_{i-j}, u_i) = 0$, si j est positif.

7.10.3 Exemples de processus autorégressifs

Il est intéressant de générer des processus autorégressifs. Dans la Figure 7.11 une suite de variables aléatoires normales indépendantes de moyennes nulles et de variances égales à 1 ont été générées. Ce processus est appelé un bruit blanc gaussien.

FIG. 7.11 – Bruit blanc : Suite de variables normales centrées réduites



Dans la Figure 7.12, on a généré un processus autoregressif avec $\rho = 0.9$. La valeur du processus au temps t est très similaire à la valeur temps précédent. Dans la Figure 7.13, on a généré un processus autoregressif avec $\rho = 0.5$. La valeur du processus au temps t est similaire à la valeur temps précédent, mais cette similarité est moins forte qu'avec $\rho = 0.9$.

FIG. 7.12 – Processus autorégressif avec $\rho = 0.9$

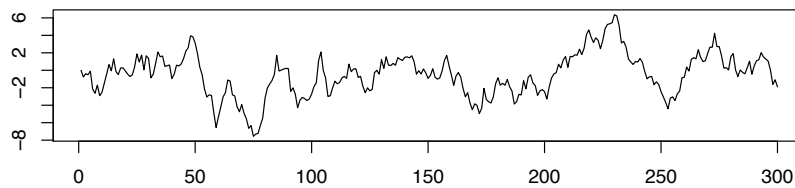
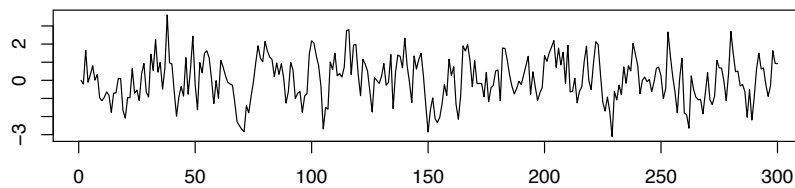


FIG. 7.13 – Processus autorégressif avec $\rho = 0.5$



Il est également possible de générer des processus autorégressif avec une valeur négative pour ρ . Dans la Figure 7.14, on a généré un processus autoregressif avec $\rho = -0.5$. La valeur du processus n'a en général pas le même signe au temps t et au temps $t - 1$. Ensuite, dans la Figure 7.15, on a généré un processus autoregressif avec $\rho = -0.9$. La valeur du processus n'est presque jamais le même signe au temps t et au temps $t - 1$.

Enfin, on a généré, dans la Figure 7.16 un processus avec $\rho = 1$. Ce processus est appelé une promenade aléatoire. Finalement, dans la Figure 7.17, on a généré un processus avec $\rho = 1.01$, qui n'est plus du tout stationnaire. A partir des deux dernières figures, on peut comprendre intuitivement l'importance de la

FIG. 7.14 – Processus autorégressif avec $\rho = -0.5$

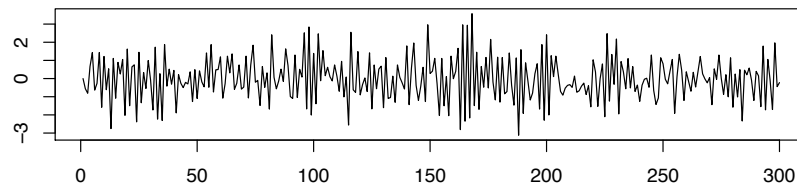
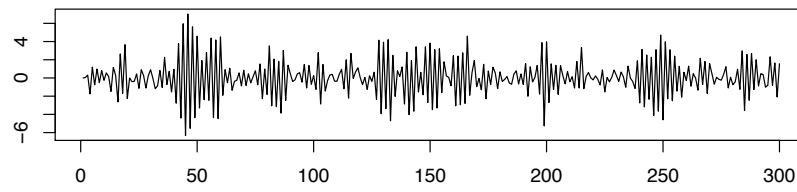


FIG. 7.15 – Processus autorégressif avec $\rho = -0.9$



condition $|\rho| < 1$, qui sert, en quelque sorte, à ramener le processus aux alentours de zero, ce qui garantit la stationnarité.

FIG. 7.16 – Promenade aléatoire : Processus autorégressif avec $\rho = 1$

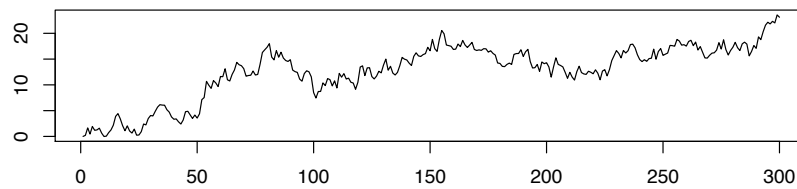
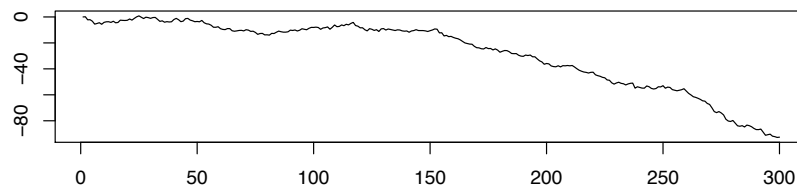


FIG. 7.17 – Processus non stationnaire $\rho = 1.01$



7.10.4 Espérance et variance du processus autorégressif d'ordre 1

Le caractère récursif de la définition de ε_i permet de réaliser le développement suivant :

$$\begin{aligned}
\varepsilon_i &= \rho\varepsilon_{i-1} + u_i \\
&= \rho(\rho\varepsilon_{i-2} + u_{i-1}) + u_i \\
&= \rho^2\varepsilon_{i-2} + \rho u_{i-1} + u_i \\
&= \rho^2(\rho\varepsilon_{i-3} + u_{i-2}) + \rho u_{i-1} + u_i \\
&= \rho^3\varepsilon_{i-3} + \rho^2 u_{i-2} + \rho u_{i-1} + u_i \\
&\vdots \\
&= \rho^j \varepsilon_{i-j} + \sum_{k=0}^{j-1} \rho^k u_{i-k}, \text{ avec } j > 0 \\
&\vdots \\
&= \sum_{k=0}^{\infty} \rho^k u_{i-k}.
\end{aligned} \tag{7.11}$$

On peut alors calculer l'espérance

$$\begin{aligned}
E(\varepsilon_i) &= E\left(\rho^j \varepsilon_{i-j} + \sum_{k=0}^{j-1} \rho^k u_{i-k}\right), \text{ avec } j > 0 \\
&= \rho^j E(\varepsilon_{i-j}) + \sum_{k=0}^{j-1} \rho^k E(u_{i-k}) \\
&= \rho^j E(\varepsilon_{i-j}) + \sum_{k=0}^{j-1} \rho^k \times 0 \\
&= \rho^j E(\varepsilon_{i-j}).
\end{aligned}$$

Si $|\rho| < 1$, alors en faisant tendre j vers l'infini, on obtient

$$E(\varepsilon_i) = \lim_{j \rightarrow \infty} \rho^j E(\varepsilon_{i-j}) = 0.$$

On peut également calculer la variance :

$$\begin{aligned}
\sigma_\varepsilon^2 = \text{var}(\varepsilon_i) &= \text{var}\left(\rho^j \varepsilon_{i-j} + \sum_{k=0}^{j-1} \rho^k u_{i-k}\right), \text{ avec } j > 0 \\
&= \rho^{2j} \text{var}(\varepsilon_{i-j}) + \sum_{k=0}^{j-1} \rho^{2k} \text{var}(u_{i-k}) \\
&= \rho^{2j} \text{var}(\varepsilon_{i-j}) + \sigma_u^2 \sum_{k=0}^{j-1} \rho^{2k} \\
&= \rho^{2j} \text{var}(\varepsilon_{i-j}) + \sigma_u^2 \frac{1 - \rho^{2j}}{1 - \rho^2}.
\end{aligned}$$

Si $|\rho| < 1$, alors en faisant tendre j vers l'infini, on obtient

$$\text{var}(\varepsilon_i) = \lim_{j \rightarrow \infty} \left(\rho^{2j} \text{var}(\varepsilon_{i-j}) + \sigma_u^2 \frac{1 - \rho^{2j}}{1 - \rho^2} \right) = \frac{\sigma_u^2}{1 - \rho^2}.$$

Ensuite, on peut calculer l'autocovariance en repartant de l'expression (7.11) :

$$\begin{aligned}
\text{cov}(\varepsilon_i, \varepsilon_{i-j}) &= \text{cov}\left(\rho^j \varepsilon_{i-j} + \sum_{k=0}^{j-1} \rho^k u_{i-k}, \varepsilon_{i-j}\right) \\
&= \text{cov}(\rho^j \varepsilon_{i-j}, \varepsilon_{i-j}) + \sum_{k=0}^{j-1} \text{cov}(\rho^k u_{i-k}, \varepsilon_{i-j}) \\
&= \rho^j \text{var}(\varepsilon_{i-j}) + \sum_{k=0}^{j-1} \rho^k \underbrace{\text{cov}(u_{i-k}, \varepsilon_{i-j})}_0 \\
&= \rho^j \sigma_\varepsilon^2 \\
&= \frac{\rho^j \sigma_u^2}{1 - \rho^2}, \text{ pour tout } j > 0.
\end{aligned}$$

Enfin, on calcule l'autocorrélation :

$$\text{corr}(\varepsilon_i, \varepsilon_{i-j}) = \frac{\text{cov}(\varepsilon_i, \varepsilon_{i-j})}{\sqrt{\text{var}(\varepsilon_i) \text{var}(\varepsilon_{i-j})}} = \frac{\rho^j \frac{1}{1-\rho^2} \sigma_u^2}{\sqrt{\sigma_u^2 \frac{1}{1-\rho^2} \sigma_u^2 \frac{1}{1-\rho^2}}} = \rho^j.$$

La série temporelle ε_i est donc homoscédastique. En effet, sa variance ne dépend pas de i . De plus, l'autocorrélation entre ε_i et ε_j ne dépend que la différence entre i et j .

7.10.5 Processus sur un intervalle de temps fini

En pratique, la série temporelle ne peut être observée que sur un intervalle de temps limité (de l'instant 1 à l'instant n). Sur cet intervalle, la série est régie par le même modèle :

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i, \text{ pour } i = 2, \dots, n.$$

Il est cependant nécessaire de faire des hypothèses supplémentaires sur le “démarrage” de la série (sur ε_1). Les hypothèses deviennent :

- les u_i sont de moyennes nulles, homoscédastiques, de variance σ_u^2 et non-corrélés,
- $|\rho| < 1$,
- $E(\varepsilon_1) = 0$,
- $\text{var}(\varepsilon_1) = \frac{1}{1-\rho^2} \sigma_u^2$,
- $E(\varepsilon_1 u_i) = 0$, pour $i = 2, \dots, n$.

On constate à nouveau que $E(\varepsilon_i) = 0$, et que

$$\begin{aligned}
\varepsilon_i &= \rho \varepsilon_{i-1} + u_i \\
&= \rho(\rho \varepsilon_{i-2} + u_{i-1}) + u_i \\
&= \rho^2 \varepsilon_{i-2} + \rho u_{i-1} + u_i \\
&= \rho^2(\rho \varepsilon_{i-3} + u_{i-2}) + \rho u_{i-1} + u_i \\
&= \rho^3 \varepsilon_{i-3} + \rho^2 u_{i-2} + \rho u_{i-1} + u_i \\
&\vdots \\
&= \rho^j \varepsilon_{i-j} + \sum_{k=0}^{j-1} \rho^k u_{i-k}, j < i \\
&= \rho^{i-1} \varepsilon_1 + \sum_{k=0}^{i-2} \rho^k u_{i-k},
\end{aligned}$$

ce qui permet de calculer la variance

$$\begin{aligned}
\text{var}(\varepsilon_i) &= \text{var}\left(\rho^{i-1}\varepsilon_1 + \sum_{k=0}^{i-2} \rho^k u_{i-k}\right) \\
&= \rho^{2(i-1)} \frac{\sigma_u^2}{1-\rho^2} + \sum_{k=0}^{i-2} \rho^{2k} \sigma_u^2 \\
&= \left\{ \rho^{2(i-1)} \frac{1}{1-\rho^2} + \sum_{k=0}^{i-2} \rho^{2k} \right\} \sigma_u^2 \\
&= \left\{ \rho^{2(i-1)} \frac{1}{1-\rho^2} + \frac{1-\rho^{2(i-1)}}{1-\rho^2} \right\} \sigma_u^2 \\
&= \frac{\sigma_u^2}{1-\rho^2},
\end{aligned}$$

et les covariances se calculent de la même manière

$$\text{cov}(\varepsilon_i, \varepsilon_{i-j}) = \frac{\sigma_u^2 \rho^j}{1-\rho^2}, \text{ pour tout } j > 0.$$

On peut donc construire la matrice variance-covariance du vecteur $\boldsymbol{\varepsilon} = (\varepsilon_1 \dots \varepsilon_i \dots \varepsilon_n)'$:

$$\text{var}(\boldsymbol{\varepsilon}) = \sigma_u^2 \boldsymbol{\Omega},$$

où

$$\boldsymbol{\Omega} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-3} & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-4} & \rho^{n-3} & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-5} & \rho^{n-4} & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \rho^{n-3} & \rho^{n-4} & \rho^{n-5} & \dots & 1 & \rho & \rho^2 \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \dots & \rho & 1 & \rho \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & \rho^2 & \rho & 1 \end{pmatrix}. \quad (7.12)$$

Cette matrice est inversible et l'on peut vérifier par une simple multiplication que son inverse est :

$$\boldsymbol{\Omega}^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho & 0 \\ 0 & 0 & 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}.$$

Le processus autorégressif d'ordre un ne dépend que d'un seul paramètre ρ . Ce paramètre peut être estimé par la méthode des moindres carrés qui consiste à minimiser la quantité :

$$Q(\rho) = \sum_{i=2}^n (\varepsilon_i - \rho \varepsilon_{i-1})^2.$$

On obtient :

$$\hat{\rho} = \frac{\sum_{i=2}^n \varepsilon_i \varepsilon_{i-1}}{\sum_{i=2}^n \varepsilon_{i-1}^2}.$$

Exemple 7.7 Avec les séries de températures données dans le Tableau 7.6, on obtient

$$\hat{\rho} = \frac{\sum_{i=2}^n \varepsilon_i \varepsilon_{i-1}}{\sum_{i=2}^n \varepsilon_{i-1}^2} = 0.9055.$$

7.10.6 Le test de Durbin-Watson

Considérons un modèle du type

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2$, et $\text{cov}(\varepsilon_i, \varepsilon_j) = \rho_{i-j}\sigma_\varepsilon^2$. On peut estimer $\boldsymbol{\beta}$ au moyen de l'estimateur des moindres carrés ordinaires $\hat{\boldsymbol{\beta}}_{MCO}$, ce qui ne procure pas un estimateur optimal, mais cet estimateur est sans biais. On peut dès lors calculer les résidus

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

Le test de Durbin-Watson consiste à tester l'hypothèse nulle

$$H_0 : \rho_1 = 0,$$

contre l'hypothèse alternative

$$H_1 : \rho_1 \neq 0.$$

Ce test utilise la statistique de test de Durbin-Watson

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \approx 2 - 2 \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2} \approx 2(1 - \hat{\rho})$$

où e_i est le résidu estimé au moyen d'une régression par les moindres carrés ordinaires.

On constate que :

- quand ρ est proche de 0, la statistique de Durbin-Watson est proche de 2,
- quand ρ est proche de 1, la statistique de Durbin-Watson est proche de 0,
- quand ρ est proche de -1, la statistique de Durbin-Watson est proche de 4.

La règle de décision pour un test de niveau α consiste à rejeter H_0 si

$$DW \notin [A_\alpha, 4 - A_\alpha],$$

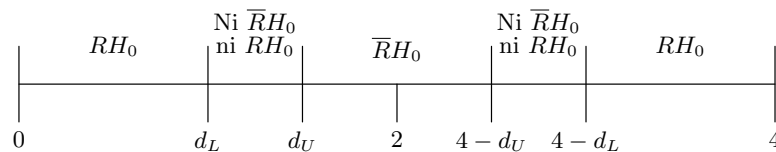
où A_α est la valeur critique. Durbin et Watson ont cependant montré que A_α dépend de la matrice \mathbf{X} . Pour chaque matrice \mathbf{X} , les valeurs critiques sont différentes. Durbin et Watson ont calculé des tables statistiques qui encadrent les valeurs A_α pour toutes valeurs de \mathbf{X} , ces bornes sont notées d_L et d_U .

En pratique, la règle de décision est donc

- on rejette H_0 si $DW < d_L$ ou si $DW > 4 - d_L$,
- on ne rejette pas H_0 si $DW \in [d_U, 4 - d_U]$,
- on ne peut pas conclure au sujet de H_0 si $DW \in [d_L, d_U]$ ou si $DW \in [4 - d_U, 4 - d_L]$.

Cette règle de décision est illustrée dans la Figure 7.18

FIG. 7.18 – Règle de décision pour le test de Durbin-Watson



Exemple 7.8 Le tableau 7.7 contient la consommation de boeuf et le prix du boeuf aux Etats-Unis de 1925 à 1941. On a fait la régression de la consommation par le prix et on a obtenu

$$\text{Consommation} = 85.239 - 0.466 \times \text{Prix} + e_i.$$

Les résidus ont également été ajoutés dans le tableau. Les Figures 7.19 et 7.19 présentent respectivement les graphiques des résidus soit en fonction des années soit en fonction du résidu de l'année précédente. L'autocorrélation apparaît clairement. On obtient :

TAB. 7.7 – Consommation et prix du boeuf aux Etats-Unis

Année	Prix du boeuf	Consommation de boeuf	Résidus Consommation par prix
1925	59.7	58.6	1.15471
1926	59.7	59.4	1.95471
1927	63	53.7	-2.20896
1928	71	48.1	-4.08451
1929	71	49	-3.18451
1930	74.2	48.2	-2.49473
1931	72.1	47.9	-3.7724
1932	79	46	-2.46006
1933	73.1	50.8	-0.40684
1934	70.2	55.2	2.64305
1935	82.2	52.2	5.22972
1936	68.4	57.3	3.90505
1937	73	54.4	3.1466
1938	70.2	53.6	1.04305
1939	67.8	53.9	0.22571
1940	63.4	54.2	-1.52274
1941	56	60	0.83215

FIG. 7.19 – Résidus selon les années

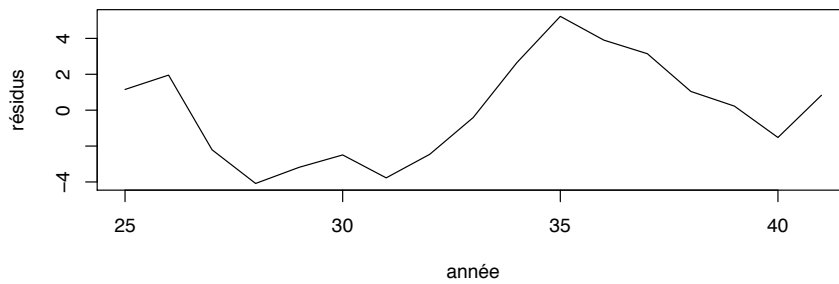
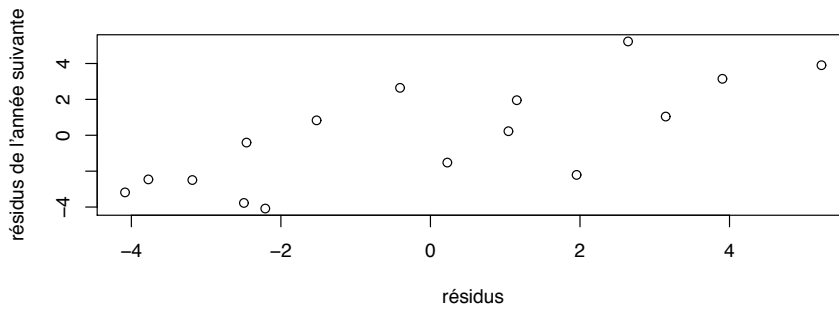


FIG. 7.20 – Résidus d'une année vs résidus de l'année suivante



$$\hat{\rho} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=2}^n e_{i-1}^2} = 0.7547252.$$

La statistique de Durbin-Watson vaut

$$DW \approx 2(1 - \hat{\rho}) = 0.4905496.$$

En pratique la règle de décision est

- on rejette H_0 si $DW < d_L$ ou si $DW > 4 - d_L$,
- on ne rejette pas H_0 si $DW \in [d_U, 4 - d_U]$,
- on ne peut pas conclure au sujet de H_0 si $DW \in [d_L, d_U]$ ou si $DW \in [4 - d_U, 4 - d_L]$.

Avec une seule variable explicative dans le modèle et $n = 17$, on obtient dans la table $d_L = 1.13$ et $d_u = 1.38$. Donc on rejette H_0 car $DW = 0.4905496 < d_L = 1.13$.

7.11 Estimation avec des termes d'erreur autocorrélés

7.11.1 Le modèle et estimation par les MCG

Méthode 1

On suppose que le modèle linéaire s'écrit

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

et le vecteur $\boldsymbol{\varepsilon}$ des ε_i est un processus autorégressif d'ordre 1.

Si ρ était connu, on pourrait donc directement estimer $\boldsymbol{\beta}$ par les moindres carrés généralisés :

$$\hat{\boldsymbol{\beta}}_{MCG} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}, \quad (7.13)$$

où $\boldsymbol{\Omega}$ est la matrice variance-covariance donnée en (7.12).

Méthode 2

Pour se ramener à la méthode des moindres carrés ordinaires, on peut vérifier par simple multiplication que $\boldsymbol{\Omega}^{-1} = \mathbf{M}'\mathbf{M}$, où

$$\mathbf{M} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{pmatrix}.$$

De plus,

$$\mathbf{M}\boldsymbol{\varepsilon} = \begin{pmatrix} \sqrt{1-\rho^2}\varepsilon_1 \\ -\rho\varepsilon_1 + \varepsilon_2 \\ \vdots \\ -\rho\varepsilon_{i-1} + \varepsilon_i \\ \vdots \\ -\rho\varepsilon_{n-1} + \varepsilon_n \end{pmatrix}. \quad (7.14)$$

En remplaçant, dans (7.14) les ε_i par $\rho\varepsilon_{i-1} + u_i$, on obtient

$$\mathbf{M}\boldsymbol{\varepsilon} = \begin{pmatrix} \sqrt{1-\rho^2}\varepsilon_1 \\ u_2 \\ \vdots \\ u_i \\ \vdots \\ u_n \end{pmatrix}. \quad (7.15)$$

On a donc $E(\mathbf{M}\boldsymbol{\varepsilon}) = 0$, et $\text{var}(\mathbf{M}\boldsymbol{\varepsilon}) = \mathbf{I}_{\sigma_u^2}$. Le modèle

$$\mathbf{My} = \mathbf{MX} + \underbrace{\mathbf{M}\boldsymbol{\varepsilon}}_{\mathbf{u}},$$

est donc un modèle linéaire général avec des termes d'erreur homoscédastiques et non-corrélés. L'estimateur linéaire optimal est alors l'estimateur des moindres carrés ordinaires qui s'écrit :

$$\hat{\beta}_{MCO} = (\mathbf{X}'\mathbf{M}'\mathbf{M}\mathbf{X})^{-1} \mathbf{X}'\mathbf{M}'\mathbf{M}\mathbf{y}, \quad (7.16)$$

et qui est le même que l'estimateur par les moindres carrés généralisés.

7.11.2 Cas où ρ est inconnu

En pratique, ρ est toujours inconnu. Cochrane et Orcutt suggèrent d'utiliser une procédure itérative. On commence d'abord par effectuer une régression classique par les MCO. On obtient ainsi des résidus \mathbf{e} , ce qui permet d'obtenir une première estimation approximative de ρ

$$\hat{\rho} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=2}^n e_{i-1}^2}.$$

Ensuite, on répète les deux opérations suivantes.

1. Connaissant une approximation de ρ , on peut estimer le coefficient de régression au moyen de l'expression (7.13) ou (7.16). On obtient ainsi une estimation de β qui permet d'obtenir une nouvelle estimation des résidus.
2. À partir de ces nouveaux résidus, on recalcule une estimation de ρ .

En répétant ces deux opérations plusieurs fois, on aboutit à une solution, qui n'est pas nécessairement optimale.

Il est également possible d'obtenir une solution du maximum de vraisemblance, en supposant que les u_i ont une distribution normale. Cette méthode, plus complexe, permet d'estimer en même temps β et ρ .

Exercices

Exercice 7.1 Soit le modèle à trois variables explicatives :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + \varepsilon_t$$

Nous disposons des observations annuelles de 1971 à 1990 :

Année	y_t	x_{t1}	x_{t2}	x_{t3}
1971	87.4	98.6	99.1	108.5
1972	97.6	101.2	99.1	110.1
1973	96.7	102.4	98.9	110.4
1974	98.2	100.9	110.8	104.3
1975	99.8	102.3	108.2	107.2
1976	100.5	101.5	105.6	105.8
1977	103.2	101.6	109.8	107.8
1978	107.8	101.6	108.7	103.4
1979	96.6	99.8	100.6	102.7
1980	88.9	100.3	81.0	104.1
1981	75.1	97.6	68.6	99.2
1982	76.9	97.2	70.9	99.7
1983	84.6	97.3	81.4	102.0
1984	90.6	96.0	102.3	94.3
1985	103.1	99.2	105.0	97.7
1986	105.1	100.3	110.5	101.1
1987	96.4	100.3	92.5	102.3
1988	104.4	104.1	89.3	104.4
1989	110.7	105.3	93.0	108.5
1990	127.1	107.6	106.6	111.3

Le but de cet exercice est de détecter une éventuelle autocorrélation d'ordre 1 des erreurs.

Pour ce faire :

1. estimez les coefficients du modèle par moindres carrés ordinaires,
2. calculez la statistique de Durbin-Watson, effectuez le test et commentez le résultat.

Exercice 7.2 Reprenez les données et le modèle de l'exercice précédent et essayez de corriger les effets liés à l'autocorrélation des erreurs. Pour cela, il y a plusieurs techniques possibles.

En voici deux :

1. estimez le ρ à partir de la statistique de Durbin-Watson et avec cet estimateur, transformez les données pour obtenir un modèle qui satisfasse aux hypothèses des moindres carrés ordinaires,
2. estimez le modèle en spécifiant au logiciel qu'il y a de l'autocorrélation d'ordre 1.

Commentez les résultats obtenus.

Remarque : Eviews estimera le ρ par maximum de vraisemblance et l'utilisera pour estimer le modèle par moindres carrés généralisés.

Exercice 7.3 Le but de cet exercice est de gagner de l'expérience dans les méthodes pour tester la présence d'auto-corrélation de 1er ordre dans les modèles contenant des variables endogènes retardées comme régresseurs. Vous utiliserez les statistiques m et h de Durbin et vous ferez des régressions utilisant les MCO et la technique de Hildreth-Lu. Les données nécessaires sont sur le site de l'Université de Neuchâtel division ecopo. Le nom du fichier est Kopcke1. Ce sont des données trimestrielles allant de 1952 : 1 à 1986 : 4. Pour cet exercice, vous aurez uniquement besoin des séries IE (investissement en équipement), IS (Investissement en construction) et Y (PIB).

1. En utilisant les MCO, estimez, pour les deux types d'investissements, l'équation suivante :

$$I_t = \mu\lambda Y_t - (1 - \delta)\mu\lambda Y_{t-1} + (1 - \lambda)I_{t-1} + \varepsilon_t$$

À partir de ces résultats, donnez les estimations implicites de μ le coefficient capital/output, δ le taux de dépréciation du capital et λ le coefficient d'ajustement. Est-ce que le taux de dépréciation δ correspond bien au 0.15 pour les équipements et 0.05 pour la construction obtenue par Kopcke ? En utilisant la statistique de Durbin-Watson donnée par Eviews, testez avec $\alpha = 0.05$ l'hypothèse nulle de l'absence d'auto-corrélation. Puisque DW est approximativement égale à $2(1 - \rho)$, calculez l'estimation implicite de ρ . Pourquoi est-ce que cette estimation peut être biaisé vers zéro ?

2. James Durbin (1970) développa deux statistiques pour ce contexte qui sont strictement valide asymptotiquement mais qui sont aussi fréquemment utilisées en petits échantillons. Il y a la statistique suivante, le h de Durbin :

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T(\text{var}\beta_3)}}$$

où $\hat{\rho}$ est l'estimation de ρ faite au premier point, T est la taille de l'échantillon (ici 124) et $\text{var}\beta_3$ est l'estimation faite au premier point de la variance du coefficient lié à la variable dépendante retardée. h est asymptotiquement normalement distribuée (centrée réduite). Calculez le h , et en utilisant la table de la normale centrée réduite, testez avec $\alpha = 0.05$ l'hypothèse nulle $h = 0$ (et donc $\rho = 0$). Dans certains cas, si $T(\text{var}\beta_3) > 1$, la racine est négative. Dans ce cas, cette statistique est inutilisable. On utilisera donc plutôt la statistique m de Durbin. Plus précisément, du premier point, récupérez les 124 résidus. Ensuite, estimez par MCO l'équation suivante (avec les mêmes variables explicatives que ci-dessus et avec en plus les résidus retardés d'une période) :

$$e_t = c + \beta_1 Y_t + \beta_2 Y_{t-1} + \beta_3 I_{t-1} + \rho^* e_{t-1} + \varepsilon_t$$

sur la période d'observation 1952 : 2-1986 : 4. Donnez l'estimation de ρ dans ce cas (ρ^*) et testez la nullité de ρ^* avec $\alpha = 0.05$.

- Maintenant, estimez l'équation du premier point en spécifiant de l'auto-corrélation de premier ordre. D'une part, en utilisant la technique intégrée dans le logiciel Eviews et d'autre part, en utilisant la technique de Hildreth-Lu. Cette technique consiste à "balayer" tout le spectre possible de ρ avec dans un premier temps un pas de 0.05. Sur la base des résultats initiaux, affinez votre balayage avec un pas de 0.01. Le critère pour la sélection du ρ dans ce cas est la minimisation de la somme des carrés des résidus (SSR) des équations transformées qui sont estimées par MCO (voir exercice 5.2).
- Comparez tous vos résultats et faites un commentaire.

Exercice 7.4 Estimez de manière optimale \hat{b}_0 et \hat{b}_1 du modèle

$$y_t = b_0 t + b_1 t^2 + \varepsilon_t$$

où : Les ε_t sont hétéroscédastiques avec $\varepsilon_t \sim \mathcal{N}(0, \sigma^2 t^2)$

TAB. 7.8 – Données selon le temps et le carré du temps

y_t	t	t^2
7	1	1
8	2	4
10	3	9
12	4	16
15	5	25

(Indication : Trouvez l'astuce !)

Exercice 7.5 Exercice sur l'hétéroscédasticité à faire avec Eviews.

On suppose que l'analyse théorique permet de conclure au bien-fondé d'une estimation de la relation suivante entre les dépenses publiques pour l'enseignement et le revenu :

$$\text{Dépenses}_t = a + b \times \text{Revenu}_t$$

La relation inverse (effet des dépenses d'enseignement sur le revenu) et les autres facteurs influençant la dépense sont ainsi ignorés. On utilise les données du tableau ci-après se référant aux cantons suisses et pour l'année 1991 (millions de CHF).

- Transférez les données ci-dessous sur un fichier Excel et, à partir de cet emplacement, transférez-les sur Eviews.
- Estimer par MCO ladite relation et commenter le résultat obtenu.
- Vérifier si le calcul ci-dessus est affecté par le phénomène d'hétéroscédasticité et cela à l'aide :
 - du graphique des résidus en rapport avec la variable explicative ;
 - du test de White ;
 Commenter les résultats obtenus.
- Dans le but, le cas échéant, d'éviter l'hétéroscédasticité et aussi afin d'améliorer l'intérêt économique des résultats, effectuez l'estimation en logarithmes de la fonction ci-dessus. Donnez intuitivement la raison de prendre les logarithmes afin d'éviter l'hétéroscédasticité. Commentez les résultats en utilisant le test de White également.
- Reprenez les données originales et estimer le modèle à l'aide des moindres carrés pondérés. Pour cela dans la boîte *Equation Specification*, allez dans *Options* et sélectionnez *Heteroscedasticity* en haut à gauche. Commentez le résultat en utilisant le test de White.

Indication : Eviews effectue automatiquement le test de White. Pour cela, il faut changer de vue lorsque le résultat de la régression est affiché sur l'écran. Il suffit de cliquer sur

View/ResidualTest/White Heteroscedasticity(crossterms).

TAB. 7.9 – Dépenses d'enseignement et revenus selon les cantons

Cantons	Enseignement	Revenu
Zurich	2252	65574
Bern	1937	36886
Luzern	399	11719
Uri	44	1196
Schwyz	101	4194
Obwalden	21	984
Nidwalden	22	1400
Glarus	44	1749
Zug	116	6037
Fribourg	438	7859
Solothurn	256	8857
Bale-ville	541	11655
Bale-campagne	349	10005
Schaffouse	77	2703
Appenzell A.	41	1869
Appenzell I.	12	456
St-Gallen	406	15857
Grison	178	7058
Aarau	740	20318
Thurgovie	190	7125
Tessin	444	9922
Vaud	1319	24103
Valais	434	8068
Neuchâtel	280	5834
Genève	1464	22034
Jura	117	2128

Exercice 7.6 Exercice sur la saisonnalité par variables muettes.

Une entreprise cherche à appréhender une relation entre ses ventes et ses dépenses publicitaires. Le directeur marketing dispose des données sur les ventes et les dépenses publicitaires entre 1988 et 1992 se trouvant ci-dessous.

1. Vous êtes son conseiller et vous lui recommandez de faire la régression classique :

$$\text{VENTES}_t = a + b \times \text{PUB}_t + \varepsilon_t$$

Commentez les résultats obtenus.

2. Tracer le graphique comprenant la série des ventes et celle de la publicité du Tableau 7.10. Que pouvez-vous en conclure ?
3. Spécifier et estimer le modèle adéquat.
4. Tracer les ventes réalisées et les ventes estimées par les régressions respectives (un graphe pour chaque régression).
5. Expliquez la fonction `@seas(n)` de Eviews et mettez-la en oeuvre ici.

Indication : Eviews calcule automatiquement tous ces graphes. Pour afficher des séries, il faut sélectionner les séries voulues, allez dans *View/Show*. Puis, lorsque le groupe contenant les séries apparaît, retournez dans *View/Graph/line*. Pour afficher graphiquement les résultats de la régression, c'est plus simple. Lorsque le résultat de la régression apparaît, allez dans *View/Actual, Fitted.../Actual, Fitted, ...Graph*.

TAB. 7.10 – Ventes et dépenses publicitaires

Années		T_1	T_2	T_3	T_4
1988	Ventes	164	198	85	179
	Pub.	34	36	32	29
1989	Ventes	168	201	98	197
	Pub.	45	67	76	75
1990	Ventes	197	209	100	216
	Pub.	75	78	72	75
1991	Ventes	223	245	119	260
	Pub.	78	81	84	83
1992	Ventes	298	309	124	267
	Pub.	89	82	81	83

Exercice 7.7 Exercice sur les séries temporelles.

Soient les processus :

1. $X_t = \rho X_{t-1} + u_t$ où $|\rho| < 1$, $t \in \mathbb{Z}$
C'est un processus autorégressif d'ordre 1 symbolisé par le terme AR(1).
2. $Y_t = \phi u_{t-1} + u_t$ où $|\phi| < 1$, $t \in \mathbb{Z}$
C'est un processus à moyenne mobile d'ordre 1 symbolisé par le terme MA(1).

Partie théorique :

Calculez d'une part :

1. $\text{var}(X_t)$
2. $\text{cov}(X_t, X_{t-j})$
3. $\text{corr}(X_t, X_{t-j})$

et d'autre part :

1. $\text{var}(Y_t)$
2. $\text{cov}(Y_t, Y_{t-j})$
3. $\text{corr}(Y_t, Y_{t-j})$

pour $j = 1, \dots, \infty$.

Partie pratique :

Générez sur EvIEWS des $u_t \sim \mathcal{N}(0, 1)$ et avec, générez une réalisation de X_t et de Y_t . Ensuite, estimez (visualisez) la fonction d'autocorrélation grâce au corrélogramme.

Utilisez les valeurs suivantes pour ρ et ϕ :

$$\begin{cases} \rho = 0.8 & \rho = 0.3 \\ \rho = -0.8 & \rho = -0.3 \\ \phi = 0.8 & \phi = 0.3 \\ \phi = -0.8 & \phi = -0.3 \end{cases}$$

Comparez avec les calculs qui ont été faits dans la partie théorique.

Exercice 7.8 On considère la relation suivante entre l'épargne du ménage E_i et son revenu R_i :

$$E_i = \beta R_i + \varepsilon_i, \quad \text{avec } i = 1, \dots, n.$$

où ε_i est un terme d'erreur vérifiant

- $E(\varepsilon_i) = 0$, pour tout i ,
- $E(\varepsilon_i \varepsilon_j) = 0$, pour tout $i \neq j$,

On considère en outre 3 modèles de variances sur les termes d'erreur ε_i :

- Modèle 1 : $E(\varepsilon_i^2) = \sigma^2$, pour tout i ,
- Modèle 2 : $E(\varepsilon_i^2) = \sigma^2 R_i$, pour tout i ,
- Modèle 3 : $E(\varepsilon_i^2) = \sigma^2 R_i^2$, pour tout i .

1. Donnez une signification du modèle et interpréter le paramètre β .
2. Donnez les meilleurs estimateurs de β sous ces trois modèles. Commentez les résultats obtenus.
3. Que vaut la variance l'estimateur des MCG et des MCO sous le modèle 3 ($E(\varepsilon_i^2) = \sigma^2 R_i^2$) ? Comparez sa variance avec celle de l'estimateur par les MCG. Le résultat est-il conforme au théorème de Gauss-Markov ?

Exercice 7.9 On considère le modèle linéaire sans constante suivant :

$$y_i = \beta x_i + \varepsilon_i, \quad \text{avec } i = 1, \dots, n \quad \text{et} \quad \varepsilon_i \sim N(0, \sigma^2). \quad (7.17)$$

Les résidus sont non-corrélés.

1. Recherche du meilleur estimateur de β :

- (a) On pose $\beta^* = \sum_{i=1}^n c_i y_i$ un estimateur linéaire de β ; les c_i sont des constantes. Sous quelles conditions β^* est-il sans biais ?
- (b) Donner l'expression de la variance de β^* .
- (c) Le problème est donc de déterminer les c_i tels que la variance de β^* soit minimale. Écrire le programme d'optimisation sous contrainte et déterminer l'expression de β^* .

2. Estimation de l'équation 7.17 par MCO : déterminer $\hat{\beta}$ l'estimateur de β par moindres carrés ordinaires.
3. Estimation par la méthode du maximum de vraisemblance :
 - (a) Écrire la fonction de vraisemblance de l'échantillon (y_1, \dots, y_n) .
 - (b) Estimer β et σ^2 par la méthode du maximum de vraisemblance.
4. Montrer que σ_{MV}^2 est biaisé et calculer son biais.

Exercice 7.10 (d'après J. Krishnakumar) Une entreprise possède n points de vente. Elle enregistre pour le mois de mars 2001 les ventes de chaque magasin. Soit y_{1i} les ventes du magasin i au mois de mars. On postule

$$y_{1i} = a + \varepsilon_{1i}, i = 1, \dots, n.$$

En avril, l'entreprise fait une grosse campagne de publicité à la télévision. A la fin avril, elle enregistre, dans le même ordre, les ventes des n magasins notées y_{2i} . On postule

$$y_{2i} = a + b + \varepsilon_{2i}, i = 1, \dots, n.$$

Le paramètre a représente la vente moyenne avant la campagne et b mesure l'effet de la campagne sur la vente moyenne. Les hypothèses sont

$$\begin{aligned} E(\varepsilon_{1i}) &= E(\varepsilon_{2i}) = 0, \text{ pour tout } i = 1, \dots, n, \\ E(\varepsilon_{1i}^2) &= E(\varepsilon_{2i}^2) = \sigma^2, \text{ pour tout } i = 1, \dots, n, \\ E(\varepsilon_{1i}\varepsilon_{2i}) &= \rho\sigma^2, \text{ pour tout } i = 1, \dots, n, \\ E(\varepsilon_{si}\varepsilon_{tj}) &= 0, \text{ pour tout } i \neq j = 1, \dots, n, \quad s, t = 1, 2. \end{aligned}$$

1. Écrire ce modèle sous la forme

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

où $\boldsymbol{\gamma} = (\alpha \ \beta)'$. Précisez la notation utilisée.

2. Donnez les estimateurs de α et β par les moindres carrés ordinaires et par les moindres carrés généralisés.
3. Comparez les deux méthodes d'estimation.
4. En déduire les estimateurs de a et b .
5. Donner la matrice variance-covariance des estimateurs obtenus.
6. Proposez des estimateurs pour σ^2 et ρ .
7. Construisez un test pour tester l'hypothèse nulle $b = 0$. Traitez les cas ρ connu et ρ inconnu séparément, en supposant les termes d'erreur normaux.

Chapitre 8

Variables instrumentales et équations simultanées

8.1 Erreurs sur la variable explicative

Considérons le modèle linéaire général

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

avec $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, et $\text{var}(\boldsymbol{\varepsilon}) = \mathbf{I}\sigma_{\varepsilon}^2$. En économie, il n'est pas toujours possible de faire l'hypothèse que la matrice \mathbf{X} est constante. Dans certains cas, on doit admettre que la matrice \mathbf{X} est aléatoire et est corrélée avec le vecteur des résidus $\boldsymbol{\varepsilon}$.

L'estimateur des moindres carrés est alors biaisé. En effet,

$$E(\hat{\boldsymbol{\beta}}) = E\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right\} = E\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right\} = \boldsymbol{\beta} + \underbrace{E\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right\}}_{\neq 0}.$$

Si on suppose que

$$\frac{1}{n}\mathbf{X}'\mathbf{X} \xrightarrow{P} \boldsymbol{\Sigma}_{XX}, \quad (8.1)$$

et que

$$\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon} \xrightarrow{P} \boldsymbol{\Sigma}_{X\varepsilon}, \quad (8.2)$$

alors

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}} &= \text{plim}_{n \rightarrow \infty} \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right\} \\ &= \text{plim}_{n \rightarrow \infty} \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right\} \\ &= \boldsymbol{\beta} + \text{plim}_{n \rightarrow \infty} \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right\} \\ &= \boldsymbol{\beta} + \text{plim}_{n \rightarrow \infty} \left\{\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} \frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon}\right\} \\ &= \boldsymbol{\beta} + \boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{X\varepsilon}. \end{aligned}$$

L'estimateur n'est donc pas convergent.

8.2 Variables instrumentales

Pour obtenir un estimateur convergent, on utilise q (avec $q \geq p$) autres variables dont la matrice des valeurs prises est notée \mathbf{Z} et est de dimension $n \times q$. Ces variables sont appelées *variables instrumentales*. Si

en outre on suppose que la matrice \mathbf{Z} n'est pas aléatoire, les variables instrumentales sont non-corrélées au vecteur de résidus $\boldsymbol{\varepsilon}$, et donc

$$\frac{1}{n}(\mathbf{Z}'\boldsymbol{\varepsilon}) \xrightarrow{P} \mathbf{0}. \quad (8.3)$$

En prémultipliant les deux membres du modèle linéaire par \mathbf{Z}' , on obtient

$$\mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}'\boldsymbol{\varepsilon}. \quad (8.4)$$

Comme \mathbf{Z} n'est pas aléatoire, on a

$$\text{var}(\mathbf{Z}'\boldsymbol{\varepsilon}) = \mathbf{Z}'\text{var}(\boldsymbol{\varepsilon})\mathbf{Z} = \mathbf{Z}'\sigma_{\varepsilon}^2\mathbf{Z}.$$

L'équation (8.4) peut être vue comme un nouveau modèle linéaire généralisé. On pourrait estimer $\boldsymbol{\beta}$ en utilisant la méthode des moindres carrés généralisés, ce qui donne

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{VI} &= \{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\sigma_{\varepsilon}^2\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\}^{-1} \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\sigma_{\varepsilon}^2\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \\ &= \{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\}^{-1} \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1} \mathbf{X}'\mathbf{P}_Z\mathbf{y}. \end{aligned}$$

où \mathbf{P}_Z est une matrice idempotente, qui projette sur le sous-espace engendré par les colonnes de \mathbf{Z} :

$$\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'.$$

Cet estimateur peut également s'écrire :

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{VI} &= \{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\}^{-1} \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \\ &= \{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\}^{-1} \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + \{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\}^{-1} \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}, \\ &= \boldsymbol{\beta} + \left\{ \frac{1}{n}\mathbf{X}'\mathbf{Z} \left(\frac{1}{n}\mathbf{Z}'\mathbf{Z} \right)^{-1} \frac{1}{n}\mathbf{Z}'\mathbf{X} \right\}^{-1} \frac{1}{n}\mathbf{X}'\mathbf{Z} \left(\frac{1}{n}\mathbf{Z}'\mathbf{Z} \right)^{-1} \frac{1}{n}\mathbf{Z}'\boldsymbol{\varepsilon}. \end{aligned}$$

Sous l'hypothèse (8.3) et en supposant que

$$\frac{1}{n}\mathbf{X}'\mathbf{Z} \xrightarrow{P} \boldsymbol{\Sigma}_{XZ},$$

et

$$\frac{1}{n}\mathbf{Z}'\mathbf{Z} \xrightarrow{P} \boldsymbol{\Sigma}_{ZZ},$$

on a

$$\hat{\boldsymbol{\beta}}_{VI} \xrightarrow{P} \boldsymbol{\beta} + \{\boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\Sigma}_{XZ}'\}^{-1} \boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_{ZZ}^{-1} \times \mathbf{0} = \boldsymbol{\beta}. \quad (8.5)$$

L'estimateur par les variables instrumentales est donc convergent.

8.3 Doubles moindres carrés

La méthode des variables instrumentales peut aussi être présentée comme une double application de la méthode des moindres carrés. À la première étape, on projette les variables explicatives sur le sous-espace engendré par les variables instrumentales :

$$\hat{\mathbf{X}} = \mathbf{P}_Z\mathbf{X},$$

où $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

Ensuite, on utilise $\hat{\mathbf{X}}$ comme variable explicative pour faire la régression sur \mathbf{y} , ce qui donne le coefficient de régression

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{VI} &= \{\hat{\mathbf{X}}'\hat{\mathbf{X}}\}^{-1} \hat{\mathbf{X}}'\mathbf{y} \\ &= \{\mathbf{X}'\mathbf{P}_Z\mathbf{X}\}^{-1} \mathbf{X}'\mathbf{P}_Z\mathbf{y}. \end{aligned}$$

8.4 Cas où $q = p$

Un cas particulier est intéressant quand $q = p$ et que la matrice \mathbf{Z} est de plein rang. La matrice $\mathbf{Z}'\mathbf{X}$ est alors inversible.

$$\begin{aligned}\hat{\beta}_{VI} &= \{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\}^{-1}\mathbf{X}'\mathbf{Z}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}\mathbf{X}'\mathbf{Z}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}.\end{aligned}$$

L'estimateur de $\hat{\beta}_{VI}$ est alors beaucoup plus simple.

8.5 Application à l'économie

8.5.1 Un exemple : le modèle keynesien

Considérons le modèle keynesien élémentaire dans lequel interviennent trois variables :

- la consommation C ,
- le revenu national Y ,
- l'investissement I .

Le modèle est régi par deux équations.

- La première équation est une équation de consommation, régie par une relation linéaire stochastique :

$$C_i = a + bY_i + u_i,$$

où les u_i sont des variables aléatoires non-corrélées homoscédastiques de variance σ_u^2 . Le paramètre b est la propension marginale à consommer. Dans la théorie keynesienne, $0 < b < 1$, dans la théorie monétariste $b = 1$.

- La seconde équation non-stochastique, c'est une identité comptable

$$Y_i = C_i + I_i. \quad (8.6)$$

Il n'y a pas de résidu, car elle exprime une relation comptable.

Le modèle économétrique dans sa forme structurelle est donc :

$$\begin{cases} C_i = a + bY_i + u_i \\ Y_i = C_i + I_i. \end{cases} \quad (8.7)$$

Une variable est dite exogène si elle est non-corrélée aux résidus. Une variable est endogène si elle est corrélée aux résidus. Il ne faut pas confondre variables exogènes et variables explicatives. La variable Y_i est explicative pour l'équation (8.6), mais nous allons voir qu'elle ne peut pas être exogène. La variable I_i est supposée exogène.

Il est possible de résoudre ce système d'équation. Après quelques calculs, on obtient la forme dite réduite du modèle, c'est-à-dire que les variables endogènes sont exprimées seulement en fonction des variables exogènes :

$$\begin{cases} C_i = \frac{a}{1-b} + \frac{b}{1-b}I_i + \frac{u_i}{1-b} \\ Y_i = \frac{a}{1-b} + \frac{1}{1-b}I_i + \frac{u_i}{1-b}. \end{cases}$$

La seconde équation nous montre que Y_i est endogène, car cette variable est forcément corrélée avec les résidus u_i . Il est possible de calculer la covariance :

$$\text{cov}(Y_i, u_i) = \text{cov}\left(\frac{a}{1-b} + \frac{b}{1-b}I_i + \frac{u_i}{1-b}, u_i\right) = \frac{1}{1-b}\sigma_u^2.$$

Avec la première équation du modèle structurel donné en (8.7), on se trouve donc dans le problème embarrassant où la variable explicative Y_i (qui est endogène) est corrélée aux résidus.

8.5.2 Estimation par la méthode des moindres carrés indirects

Il est possible estimer les paramètres de la forme réduite par la méthode des moindres. En effet, en posant

$$\pi_1 = \frac{a}{1-b}, \quad \pi_2 = \frac{b}{1-b}, \quad \pi_3 = \frac{a}{1-b}, \quad \pi_4 = \frac{1}{1-b},$$

on obtient

$$\begin{cases} C_i = \pi_1 + \pi_2 I_i + \frac{u_i}{1-b} \\ Y_i = \pi_3 + \pi_4 I_i + \frac{u_i}{1-b}. \end{cases}$$

Pour ces deux équations, la variable explicative est exogène. On peut donc estimer les paramètres de la forme réduite par les moindres carrés ordinaires :

$$\hat{\pi}_2 = \frac{\sum_{i=1}^n (C_i - \bar{C})(I_i - \bar{I})}{\sum_{i=1}^n (I_i - \bar{I})^2},$$

$$\hat{\pi}_1 = \bar{C} - \hat{\pi}_2 \bar{I},$$

$$\hat{\pi}_4 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(I_i - \bar{I})}{\sum_{i=1}^n (I_i - \bar{I})^2},$$

$$\hat{\pi}_3 = \bar{Y} - \hat{\pi}_4 \bar{I},$$

où

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i, \quad \bar{I} = \frac{1}{n} \sum_{i=1}^n I_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Remarquons au passage que, comme $Y_i = C_i + I_i$,

$$\begin{aligned} \hat{\pi}_4 &= \frac{\sum_{i=1}^n (Y_i - \bar{Y})(I_i - \bar{I})}{\sum_{i=1}^n (I_i - \bar{I})^2} \\ &= \frac{\sum_{i=1}^n (C_i - \bar{C} + I_i - \bar{I})(I_i - \bar{I})}{\sum_{i=1}^n (I_i - \bar{I})^2} \\ &= \frac{\sum_{i=1}^n (C_i - \bar{C})(I_i - \bar{I})}{\sum_{i=1}^n (I_i - \bar{I})^2} + \frac{\sum_{i=1}^n (I_i - \bar{I})(I_i - \bar{I})}{\sum_{i=1}^n (I_i - \bar{I})^2} \\ &= \hat{\pi}_2 + 1. \end{aligned}$$

De plus

$$\hat{\pi}_3 = \bar{Y} - \hat{\pi}_4 \bar{I} = \bar{C} + \bar{I} - (\hat{\pi}_2 + 1) \bar{I} = \bar{C} - \hat{\pi}_2 \bar{I} = \hat{\pi}_1.$$

Maintenant que l'on dispose d'estimateurs sans biais de π_1 , π_2 , π_3 et π_4 , et que l'on sait en outre que

$$b = \frac{\pi_2}{\pi_4},$$

on pourrait estimer b , par

$$\hat{b} = \frac{\hat{\pi}_2}{\hat{\pi}_4} = \frac{\sum_{i=1}^n (I_i - \bar{I})(C_i - \bar{C})}{\sum_{i=1}^n (I_i - \bar{I})(Y_i - \bar{Y})}. \quad (8.8)$$

Cet estimateur est biaisé, car le ratio deux estimateurs sans biais ne fournit pas un estimateur sans biais. Ensuite, on peut estimer a par

$$\frac{\hat{\pi}_1}{\hat{\pi}_4} \text{ ou } \frac{\hat{\pi}_3}{\hat{\pi}_4},$$

ce qui donne le même estimateur, car $\hat{\pi}_1 = \hat{\pi}_3$.

8.5.3 Estimation par la méthode des variables instrumentales

Nous allons montrer que l'estimateur (8.8) n'est autre que l'estimateur par les variables instrumentales où I_i est la variable instrumentale. Ainsi, l'estimateur est biaisé, mais comme on l'a vu ci-dessus il est convergent. Pour utiliser les variables instrumentales, on considère d'abord le modèle de régression de la variable explicative par la variable instrumentale, qu'on note

$$Y_i = c + dI_i + \varepsilon_i,$$

où les ε_i sont non-corrélés et identiquement distribués. On estime les paramètres c et d par les moindres carrés ordinaires, ce qui donne

$$\hat{d} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(I_i - \bar{I})}{\sum_{i=1}^n (I_i - \bar{I})^2},$$

et

$$\hat{c} = \bar{Y} - \hat{d}\bar{I}.$$

On peut alors construire les valeurs ajustées de Y en I , qui valent

$$Y_i^* = \hat{c} + \hat{d}I_i.$$

Ensuite, on estime le coefficient de régression b de la première équation de la forme structurelle du modèle, mais on utilise comme variable explicative Y_i^* à la place de Y_i

$$C_i = a + bY_i^* + u_i^*.$$

On obtient l'estimateur par les variables instrumentales :

$$\hat{b}_{VI} = \frac{\sum_{i=1}^n (Y_i^* - \bar{Y})(C_i - \bar{C})}{\sum_{i=1}^n (Y_i^* - \bar{Y})^2}.$$

On peut simplifier cette expression. En effet, comme

$$Y_i^* - \bar{Y} = \hat{c} + \hat{d}I_i - (\hat{c} + \hat{d}\bar{I}) = \hat{d}(I_i - \bar{I}),$$

on obtient

$$\begin{aligned} \hat{b}_{VI} &= \frac{\sum_{i=1}^n \hat{d}(I_i - \bar{I})(C_i - \bar{C})}{\sum_{i=1}^n \hat{d}^2(I_i - \bar{I})^2} \\ &= \frac{1}{\hat{d}} \frac{\sum_{i=1}^n (I_i - \bar{I})(C_i - \bar{C})}{\sum_{i=1}^n (I_i - \bar{I})^2} \\ &= \frac{\sum_{i=1}^n (I_i - \bar{I})^2}{\sum_{i=1}^n (Y_i - \bar{Y})(I_i - \bar{I})} \frac{\sum_{i=1}^n (I_i - \bar{I})(C_i - \bar{C})}{\sum_{i=1}^n (I_i - \bar{I})^2} \\ &= \frac{\sum_{i=1}^n (I_i - \bar{I})(C_i - \bar{C})}{\sum_{i=1}^n (Y_i - \bar{Y})(I_i - \bar{I})}, \end{aligned}$$

ce qui est le même estimateur que par la méthode des moindres carrés indirects (8.8).

Exemples d'équations simultanées

Exemple 8.1 Soit le modèle d'équilibre où q^o est la quantité offerte, q^d la quantité demandée, p le prix du bien, et z le climat

Equation d'offre :

$$q_t^o = a + bp_t + cz_t + u_t$$

Equation de demande :

$$q_t^d = a' + b'p_t + v_t$$

Hypothèse d'équilibre :

$$q_t^o = q_t^d = q_t$$

La forme structurelle est donc la suivante :

$$\begin{cases} q_t = a + bp_t + cz_t + u_t \\ q_t = a' + b'p_t + v_t \end{cases}$$

où p_t et q_t sont endogènes et z_t exogène.

Comme

$$a + bp_t + cz_t + u_t = a' + b'p_t + v_t,$$

on obtient

$$p_t = \frac{a - a'}{b' - b} + \frac{cz_t}{b' - b} + \frac{u_t - v_t}{b' - b}. \quad (8.9)$$

De plus,

$$b'q_t - bq_t = b'(a + bp_t + cz_t + u_t) - b(a' + b'p_t + v_t) = ab' - a'b + b'cz_t + b'u_t - bv_t,$$

ce qui donne

$$q_t = \frac{ab' - a'b}{b' - b} + \frac{b'cz_t}{b' - b} + \frac{b'u_t - bv_t}{b' - b}. \quad (8.10)$$

Les équation (8.9) et (8.10) permettent d'écrire la forme réduite :

$$\begin{cases} q_t = \pi_1 + \pi_2 z_t + \varepsilon_t \\ p_t = \pi_3 + \pi_4 z_t + \eta_t, \end{cases}$$

avec

$$\begin{aligned} \pi_1 &= \frac{ab' + a'b}{b' - b} & \pi_2 &= \frac{cb'}{b' - b} & \pi_3 &= \frac{a - a'}{b' - b} & \pi_4 &= \frac{c}{b' - b} \\ \varepsilon_t &= \frac{b'u_t - bv_t}{b' - b} & \eta_t &= \frac{u_t - v_t}{b' - b} \end{aligned}$$

Il est possible d'estimer π_1, π_2, π_3 et π_4 (paramètres réduits) par les moindres carrés ordinaires, mais il y a un problème d'identification pour remonter aux paramètres structurels (a, b, c, a', b') . En effet, le nombre de paramètres structurels (5) est plus grand que le nombre de paramètres réduits (4). Toutefois, les paramètres a' et b' sont identifiables, en effet :

$$\frac{\pi_2}{\pi_4} = \frac{\frac{cb'}{b' - b}}{\frac{c}{b' - b}} = b'$$

$$\pi_1 - b'\pi_3 = \frac{ab' + a'b}{b' - b} - b' \left(\frac{a - a'}{b' - b} \right) = a'$$

alors que les paramètres a, b, c ne sont pas identifiables.

Exemple 8.2 Modèle d'équilibre avec une variable supplémentaire

Avec : q^o = quantité offerte ; q^d = quantité demandée ; p = prix du bien ; z = climat et ; x = revenu des ménages

Equation d'offre :

$$q_t^o = a + bp_t + cz_t + u_t$$

Equation de demande :

$$q_t^d = a' + b'p_t + dx_t + v_t$$

Hypothèse d'équilibre :

$$q_t^o = q_t^d = q_t$$

Forme structurelle :

$$\begin{cases} q_t = a + bp_t + cz_t + u_t \\ q_t = a' + b'p_t + dx_t + v_t \end{cases}$$

Avec p_t et q_t comme variables endogènes et z_t et x_t comme variable exogène.

Forme réduite :

$$\begin{cases} p_t = \pi_1 + \pi_2 z_t + \pi_3 x_t + \varepsilon_t \\ q_t = \pi_4 + \pi_5 z_t + \pi_6 x_t + \eta_t \end{cases}$$

Avec

$$\begin{aligned} \pi_1 &= \frac{a - a'}{b' - b} & \pi_2 &= \frac{c}{b' - b} \\ \pi_3 &= \frac{-d}{b' - b} & \pi_4 &= \frac{ab' - a'b}{b' - b} & \pi_5 &= \frac{b'c}{b' - b} & \pi_6 &= \frac{-bd}{b' - b} \end{aligned}$$

Nous avons donc 6 paramètres dans la forme réduite et 6 paramètres dans la forme structurelle. C'est une situation favorable à l'identification, mais cela ne la garanti pas. Dans notre cas, il n'y a pas de problème d'identification :

$$\frac{\pi_5}{\pi_2} = b'$$

$$\frac{\pi_6}{\pi_3} = b$$

$$\pi_2(b' - b) = c$$

$$-\pi_3(b' - b) = d$$

$$\pi_4 - b'\pi_1 = a'$$

$$\pi_4 - b\pi_1 = a$$

Le modèle est donc identifiable à la suite de l'ajout d'un paramètre exogène dans la forme structurelle qui se traduit par deux paramètres en plus dans la forme réduite.

Exemple 8.3

Equation d'offre :

$$q_t^o = a + bp_t + u_t$$

Equation de demande :

$$q_t^d = a' + b'p_t + c'x_t + d'z_t + v_t$$

Hypothèse d'équilibre :

$$q_t^o = q_t^d = q_t$$

Forme structurelle :

$$\begin{cases} q_t = a + bp_t + u_t \\ q_t = a' + b'p_t + c'x_t + d'z_t + v_t \end{cases}$$

Avec p_t et q_t comme variables endogènes et z_t et x_t comme variable exogène.

On a directement

$$a + bp_t + u_t = a' + b'p_t + c'x_t + d'z_t + v_t,$$

et donc

$$(b' - b)p_t = a - a' - c'x_t - d'z_t + u_t - v_t.$$

et donc

$$p_t = \frac{a - a'}{b' - b} - \frac{c'x_t}{b' - b} - \frac{d'z_t}{b' - b} + \frac{u_t - v_t}{b' - b}. \quad (8.11)$$

D'autre part, on a

$$\frac{q_t}{b} - \frac{q_t}{b'} = \frac{a + bp_t + u_t}{b} - \frac{a' + b'p_t + c'x_t + d'z_t + v_t}{b'},$$

ce qui donne

$$(b' - b)q_t = b'(a + bp_t + u_t) - b(a' + b'p_t + c'x_t + d'z_t + v_t) = ab' - a'b - bc'x_t - bd'z_t + b'u_t - bv_t. \quad (8.12)$$

Les equations (8.11) et (8.12) permettent d'écrire la forme réduite :

$$\begin{cases} p_t = \frac{a - a'}{b' - b} - \frac{c'x_t}{b' - b} - \frac{d'z_t}{b' - b} + \frac{u_t - v_t}{b' - b} \\ q_t = \frac{ab' - a'b}{b' - b} - \frac{bc'x_t}{b' - b} - \frac{bd'z_t}{b' - b} + \frac{b'u_t - bv_t}{b' - b} \end{cases}$$

On redéfinit les paramètres :

$$\begin{cases} p_t = \pi_1 + \pi_2 x_t + \pi_3 z_t + \varepsilon_t \\ q_t = \pi_4 + \pi_5 x_t + \pi_6 z_t + \eta_t \end{cases}$$

Avec

$$\begin{aligned} \pi_1 &= \frac{a - a'}{b' - b} & \pi_2 &= \frac{-c'}{b' - b} & \pi_3 &= \frac{-d'}{b' - b} \\ \pi_4 &= \frac{ab' - a'b}{b' - b} & \pi_5 &= \frac{-bc'}{b' - b} & \pi_6 &= \frac{-bd'}{b' - b} \end{aligned}$$

Nous avons donc 6 paramètres dans la forme réduite et 6 paramètres dans la forme structurelle. Cette situation est favorable à l'identification, mais cela ne la garantit pas. En effet, dans notre cas :

$$\frac{\pi_6}{\pi_3} = b \quad \frac{\pi_5}{\pi_2} = b$$

On dit dans ce cas que le paramètre b est suridentifié, c'est-à-dire qu'il est défini par deux estimateurs distincts. De ce fait, le paramètre a est aussi sur-identifié (car il dépend du paramètre b) :

$$\pi_4 - b\pi_1 = a$$

L'équation d'offre est donc sur-identifiée et l'équation de demande non-identifiable.

8.6 Méthodes d'estimation

8.6.1 Moindres carrés indirects (MCI)

1. On écrit la forme réduite du modèle, c'est-à-dire qu'on reformule le modèle pour que seules les variables exogènes soient explicatives ;
2. On estime les paramètres de la forme réduite par les moindres carrés ordinaires (MCO) ;
3. On estime les paramètres de la forme structurelle en utilisant les relations algébriques entre les paramètres de la forme réduite et de la forme structurelle.

Une condition nécessaire (mais pas suffisante) pour que les paramètres de la forme structurelle soient identifiables est qu'il y ait au moins autant de paramètres dans la forme réduite que dans la forme structurelle.

8.6.2 Doubles moindres carrés (2MC)

On estime directement les paramètres de la forme structurelle en utilisant toutes les variables exogènes comme variables instrumentales.

Propriétés :

- Si le modèle est juste identifié, la méthode des 2MC donne le même résultat que la méthode des MCI
- Si le modèle est sous-identifié, la méthode des 2MC ne donne pas de résultats.

Condition d'ordre : Une condition nécessaire pour qu'une équation d'un modèle à équations simultanées soit identifiable est que le nombre de variables explicatives de l'équation soit inférieur ou égal au nombre de variables exogènes de l'ensemble du modèle.

La méthode des 2MC a l'avantage, par rapport à la méthode des MCI, de ne donner qu'un seul estimateur en cas de sur-identification.

Exemple 8.4

$$\begin{cases} C_t = \alpha + \beta Y_t + u_t \\ Y_t = C_t + I_t \end{cases}$$

Avec deux variables exogènes (I_t et la constante α) et 2 variables explicatives dans la première équation (la deuxième étant une identité), le modèle est juste identifiable.

Exemple 8.5

$$\begin{cases} q_t = a + bp_t + cz_t + u_t \\ q_t = a' + b'p_t + u_t \end{cases}$$

Avec 2 variables exogènes (z_t et les constantes a, a'), la première équation (avec 3 variables explicatives) n'est pas identifiable, mais la deuxième équation (avec 2 variables explicatives) est identifiable.

Exemple 8.6

$$\begin{cases} q_t = a + bp_t + cz_t + u_t \\ q_t = a' + b'p_t + c'x_t + v_t \end{cases}$$

Avec 3 variables exogènes (z_t, x_t et les constantes a, a') et 3 variables explicatives dans chaque équation, le modèle est juste identifiable.

Exemple 8.7

$$\begin{cases} q_t = a + bp_t + u_t \\ q_t = a' + b'p_t + c'x_t + d'z_t + v_t \end{cases}$$

Avec 3 variables exogènes (z_t , x_t et les constantes a, a'), la première équation (avec 2 variables explicatives) est sur-identifiée et la deuxième équation (avec 4 variables explicatives) est sous-identifiée.

Exemple 8.8 Soit la forme structurelle

$$\begin{cases} C_t = a + bY_t + u_t \\ I_t = c + dY_{t-1} + v_t \\ Y_t = C_t + I_t. \end{cases}$$

Avec C_t, I_t, Y_t comme variables endogènes et Y_{t-1} et la constante comme variables exogènes. On a

$$I_t = Y_t - C_t.$$

Avec la deuxième équation, on obtient

$$Y_t = C_t + c + dY_{t-1} + v_t,$$

ce qui donne

$$\begin{cases} C_t = a + bY_t + u_t \\ C_t = Y_t - c - dY_{t-1} + v_t. \end{cases}$$

En soustrayant ces deux équations, on a

$$0 = a + c + (b - 1)Y_t + dY_{t-1} + u_t - v_t,$$

ce qui donne

$$Y_t = \frac{a + c}{1 - b} + \frac{d}{1 - b}Y_{t-1} + \frac{u_t - v_t}{1 - b}.$$

En soustrayant à la première équation, la seconde multipliée par b , on a

$$C_t - bC_t = a + bc + 0 + bdY_t + u_t - bv_t,$$

soit

$$C_t = \frac{a + bc}{1 - b} + \frac{bd}{1 - b}Y_{t-1} + \frac{u_t - bv_t}{1 - b}.$$

On obtient ainsi la forme réduite :

$$\begin{cases} I_t = c + dY_{t-1} + v_t \\ C_t = \frac{a+bc}{1-b} + \frac{bd}{1-b}Y_{t-1} + \varepsilon_t \\ Y_t = \frac{a+c}{1-b} + \frac{d}{1-b}Y_{t-1} + \eta_t. \end{cases}$$

Avec deux variables exogènes dans le modèle et 2 variables explicatives par équation, le modèle est juste identifiable.

Exemple 8.9 Modèle macro-économique de Klein (extrait de Bourbonnais, 1993)

Forme structurelle :

$$\begin{cases} C_t = a_0 + a_1P_t + a_2P_{t-1} + a_3(W_t + W'_t) + \eta_{1t} \\ I_t = b_0 + b_1P_t + b_2P_{t-1} + b_3K_{t-1} + \eta_{2t} \\ W_t = c_0 + c_1X_t + c_2X_{t-1} + c_3t + \eta_{3t} \\ X_t = C_t + I_t + G_t \\ P_t = X_t - W_t - Tax_t \\ K_t = I_t + K_{t-1} \end{cases}$$

Avec : C_t = consommation ; P_t = profit ; W_t et W'_t = salaires dans l'industrie et l'administration ; I_t =

investissements ; K_t = stock ; X_t = production industrielle ; G_t = dépenses publiques et Tax_t = impôts.

Les trois premières équations comportent chacune 4 variables explicatives et les trois dernières équations sont des identités comptables. Étant donné qu'il y a 8 variables exogènes :

$$P_{t-1}, X_{t-1}, K_{t-1}, t, G_t, Tax_t, W'_t$$

et les constantes, cela implique une sur-identification du modèle ; il faut donc utiliser la méthode des 2MC dans ce cas.

Exercices

Exercice 8.1 Soit la forme structurelle d'un modèle d'équations simultanées,

$$\begin{cases} y_t = a + bx_t + u_t \\ y_t = c + v_t \end{cases}$$

où x_t et y_t sont endogènes, $t = 1, \dots, n$ les u_t et les v_t sont des résidus homoscedastiques tels que $E(u_t) = 0$, $\text{var}(u_t) = \sigma_u^2$, $E(u_t, u_j) = 0, t \neq j$, $E(v_t) = 0$, $\text{var}(v_t) = \sigma_v^2$, $E(v_t, v_j) = 0, t \neq j$, $E(u_t, v_t) = 0$, pour tout t .

1. Écrivez la forme réduite du modèle.
2. Quelle est la covariance entre x_t et u_t , (en fonction de σ_u^2 et de σ_v^2) ?
3. Quelle est la corrélation entre x_t et u_t ?
4. Donner les estimateurs des paramètres de la forme réduite du modèle.
5. Les paramètres de la forme structurelle du modèle sont-ils identifiables, tous ou seulement certains d'entre eux ? Donnez l'expression des estimateurs par les moindres carrés indirects pour les paramètres identifiables.
6. Si le paramètre a était connu, quel serait l'estimateur par les moindres carrés indirects de b ?

Chapitre 9

Modèles à choix discret

9.1 Modèles probit, logit et à probabilité linéaire

9.1.1 Le modèle

Souvent, la variable dépendante y_i mesure une qualité ou une caractéristique. Dans ce cas, y_i sera codée de la manière suivante :

$$y_i = \begin{cases} 1 & \text{si l'unité } i \text{ possède la caractéristique} \\ 0 & \text{sinon} \end{cases}$$

On dispose de p variables explicatives $x_1, \dots, x_j, \dots, x_p$, et on note x_{ij} la valeur prise par la j ème variable explicative sur l'unité i de la population. De plus, on note

$$\mathbf{x}_i = (x_{i1}, \dots, x_{ij}, \dots, x_{ip})',$$

le vecteur colonne des p valeurs prises par les variables explicatives sur l'unité statistique i . Les x_{ij} sont supposés constants (non-aléatoires). Considérons également une fonction $F(\cdot)$ de \mathbb{R} dans $[0, 1]$, croissante, et dérivable, telle que $\lim_{z \rightarrow -\infty} F(z) = 0$, et $\lim_{z \rightarrow \infty} F(z) = 1$. Le modèle à choix discret s'écrit

$$y_i = \begin{cases} 1 & \text{avec une probabilité } F(\mathbf{x}'_i \boldsymbol{\beta}) \\ 0 & \text{avec une probabilité } 1 - F(\mathbf{x}'_i \boldsymbol{\beta}). \end{cases} \quad (9.1)$$

avec y_i indépendant de y_j pour tout $i \neq j$. Le modèle à choix discret peut également s'écrire :

$$y_i = F(\mathbf{x}'_i \boldsymbol{\beta}) + \varepsilon_i,$$

où

- $\boldsymbol{\beta}$ est un vecteur de p coefficients de régression,

- ε_i est une variable aléatoire telle que $E(\varepsilon_i) = 0$, et $E(\varepsilon_i \varepsilon_j) = 0$.

Comme y_i vaut 1 ou 0, on peut déterminer la distribution de probabilité exacte de ε_i

$$\varepsilon_i = \begin{cases} -F(\mathbf{x}'_i \boldsymbol{\beta}) & \text{avec une probabilité } 1 - F(\mathbf{x}'_i \boldsymbol{\beta}) \\ 1 - F(\mathbf{x}'_i \boldsymbol{\beta}) & \text{avec une probabilité } F(\mathbf{x}'_i \boldsymbol{\beta}). \end{cases}$$

Les résidus ne sont donc pas homoscédastiques, car ils dépendent des \mathbf{x}_i . Leur variance vaut

$$\begin{aligned} \text{var}(\varepsilon_i) &= [1 - F(\mathbf{x}'_i \boldsymbol{\beta})][F(\mathbf{x}'_i \boldsymbol{\beta})]^2 + F(\mathbf{x}'_i \boldsymbol{\beta})[1 - F(\mathbf{x}'_i \boldsymbol{\beta})]^2 \\ &= [1 - F(\mathbf{x}'_i \boldsymbol{\beta})]F(\mathbf{x}'_i \boldsymbol{\beta}). \end{aligned}$$

Le modèle peut également s'écrire :

$$\begin{aligned} y_k &\in \{0, 1\}, \\ E(y_i) &= F(\mathbf{x}'_i \boldsymbol{\beta}), \\ \text{var}(y_i) &= [1 - F(\mathbf{x}'_i \boldsymbol{\beta})]F(\mathbf{x}'_i \boldsymbol{\beta}) \\ \text{cov}(y_i, y_j) &= 0, i \neq j. \end{aligned}$$

9.1.2 Choix de la fonction $F(\cdot)$

Le choix de la fonction $F(\cdot)$ détermine le modèle. Les trois modèles les plus utilisés sont les modèles logit, probit et en probabilité linéaire.

Le modèle logit

Le modèle logit consiste à utiliser une fonction logistique,

$$F(z) = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z}.$$

Le modèle probit

Le modèle probit consiste à utiliser la fonction de répartition d'une variable normale centrée réduite,

$$F(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Le modèle de probabilité linéaire

Le modèle de probabilité linéaire consiste à utiliser une fonction linéaire tronquée,

$$F(z) = \begin{cases} 0 & \text{si } z < -0,5 \\ z + 0,5 & \text{si } -0,5 \leq z \leq 0,5 \\ 1 & \text{si } 0,5 < z. \end{cases}$$

9.1.3 Remarques sur la fonction $F(\cdot)$

La fonction $F(\cdot)$ est supposée croissante, dérivable, telle que $F(0) = 0,5$, $\lim_{z \rightarrow -\infty} F(z) = 0$, et $\lim_{z \rightarrow \infty} F(z) = 1$. Théoriquement, n'importe quelle fonction de répartition d'une variable aléatoire continue pourrait être utilisée pour construire un modèle à choix discret chacune des fonctions $F(\cdot)$ possède évidemment une fonction de densité associée :

$$f(z) = F'(z) = \frac{dF(z)}{dz}.$$

Modèle logit

La densité est

$$\begin{aligned} f(z) &= \frac{e^z}{1 + e^z} - \frac{e^{2z}}{(1 + e^z)^2} \\ &= \frac{e^z}{1 + e^z} \left[1 - \frac{e^z}{1 + e^z} \right] \\ &= F(z) [1 - F(z)]. \end{aligned}$$

Modèle probit

La fonction $f(\cdot)$ est simplement la fonction de densité d'une variable aléatoire normale centrée réduite.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Modèle en probabilité linéaire

La fonction $F(\cdot)$ est

$$f(z) = \begin{cases} 0 & \text{si } z < -0,5 \\ 1 & \text{si } -0,5 \leq z < 0,5 \\ 0 & \text{si } 0,5 \leq z. \end{cases}$$

9.1.4 Estimation par les moindres carrés

L'estimation par les moindres carrés (ordinaires) consiste à minimiser en β

$$Q(\beta) = \sum_{i=1}^n [y_i - F(\mathbf{x}'_i \beta)]^2.$$

Si on note

$$f(z) = F'(z) = \frac{dF(z)}{dz},$$

alors on a

$$\frac{\partial F(\mathbf{x}'_i \beta)}{\partial \beta} = f(\mathbf{x}'_i \beta) \mathbf{x}_i.$$

Pour trouver le minimum en β de $Q(\beta)$, on annule le vecteur des dérivées partielles de $Q(\beta)$ en β :

$$\frac{\partial Q(\beta)}{\partial \beta} = 2 \sum_{i=1}^n [y_i - F(\mathbf{x}'_i \beta)] f(\mathbf{x}'_i \beta) \mathbf{x}_i = \mathbf{0},$$

ce qui donne

$$\sum_{i=1}^n y_i f(\mathbf{x}'_i \beta) \mathbf{x}_i = \sum_{i=1}^n F(\mathbf{x}'_i \beta) f(\mathbf{x}'_i \beta) \mathbf{x}_i. \quad (9.2)$$

L'expression (9.2) est un système non linéaire de p équations à p inconnues. Ce système ne peut être résolu qu'au moyen d'un algorithme (méthode de Newton).

9.1.5 Méthode du maximum de vraisemblance

Une autre méthode d'estimation est la méthode du maximum de vraisemblance. Comme

$$\begin{aligned} \Pr(y_1 \cdots y_n) &= \left\{ \prod_{i|y_i=1} F(\mathbf{x}'_i \beta) \right\} \left\{ \prod_{i|y_i=0} [1 - F(\mathbf{x}'_i \beta)] \right\}, \\ &= \prod_{i=1}^n \left\{ F(\mathbf{x}'_i \beta)^{y_i} [1 - F(\mathbf{x}'_i \beta)]^{1-y_i} \right\}, \end{aligned}$$

la fonction de vraisemblance est

$$L(\beta; y_1, \dots, y_n) = \prod_{i=1}^n \left\{ F(\mathbf{x}'_i \beta)^{y_i} [1 - F(\mathbf{x}'_i \beta)]^{1-y_i} \right\}.$$

Le logarithme de la fonction de vraisemblance est donc

$$\begin{aligned} \ell(\beta; y_1, \dots, y_n) &= \log L(\beta; y_1, \dots, y_n) \\ &= \sum_{i=1}^n \{ y_i \log F(\mathbf{x}'_i \beta) + (1 - y_i) \log [1 - F(\mathbf{x}'_i \beta)] \}. \end{aligned}$$

Pour déterminer l'estimateur du maximum de vraisemblance, on annule la dérivée de $\ell(\beta; y_1, \dots, y_n)$, ce qui donne

$$\begin{aligned} &\frac{\partial \ell(\beta; y_1, \dots, y_n)}{\partial \beta} \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{F(\mathbf{x}'_i \beta)} f(\mathbf{x}'_i \beta) \mathbf{x}_i - \frac{1 - y_i}{1 - F(\mathbf{x}'_i \beta)} f(\mathbf{x}'_i \beta) \mathbf{x}_i \right\} \\ &= \sum_{i=1}^n \frac{f(\mathbf{x}'_i \beta) \mathbf{x}_i [y_i - F(\mathbf{x}'_i \beta)]}{F(\mathbf{x}'_i \beta) [1 - F(\mathbf{x}'_i \beta)]} \\ &= \mathbf{0}. \end{aligned} \quad (9.3)$$

On obtient à nouveau un système de p équation à p inconnus. Cependant ce système n'est pas le même que celui obtenu par la méthode des moindres carrés. Les deux méthodes donnent donc des estimateurs distincts.

Modèle logit

Dans le cas du modèle logit on a

$$f(z) = \frac{dF(z)}{dz} = F(z)[1 - F(z)],$$

et donc l'égalité (9.3) devient

$$\sum_{i=1}^n \mathbf{x}_i y_i = \sum_{i=1}^n \mathbf{x}_i F(\mathbf{x}'_i \boldsymbol{\beta}).$$

9.1.6 Interprétation selon une variable latente

Souvent les modèles à choix discret sont présentés au moyen d'une variable aléatoire latente (non-observable) qui régit les y_i :

$$y_i = \begin{cases} 1 & \text{si } z_i > 0 \\ 0 & \text{si } z_i \leq 0. \end{cases}$$

La variable latente peut être régie par un modèle linéaire général :

$$z_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i,$$

avec $\text{cov}(\mathbf{u}) = \mathbf{I}\sigma_u^2$, et $\mathbf{u} = (u_1 \cdots u_n)'$. De plus les résidus u_i sont supposés avoir une fonction de répartition $F_u(\cdot)$. On peut alors écrire :

$$\begin{aligned} \Pr(y_i = 1) &= \Pr(\mathbf{x}'_i \boldsymbol{\beta} + u_i > 0) \\ &= \Pr(u_i > -\mathbf{x}'_i \boldsymbol{\beta}) \\ &= 1 - \Pr(u_i \leq -\mathbf{x}'_i \boldsymbol{\beta}) \\ &= 1 - F_u(-\mathbf{x}'_i \boldsymbol{\beta}). \end{aligned}$$

Si la densité est symétrique et de moyenne nulle, alors $F_u(-z) = 1 - F_u(z)$ et

$$\Pr(y_i = 1) = F_u(\mathbf{x}'_i \boldsymbol{\beta}),$$

et donc

$$\Pr(y_i = 0) = 1 - F_u(\mathbf{x}'_i \boldsymbol{\beta}).$$

On retrouve donc exactement la formulation donnée en (9.1). En introduisant une variable latente, la fonction $F_u(\cdot)$ peut alors s'interpréter directement comme la fonction de répartition des résidus.

9.1.7 Évaluation de la qualité du modèle

Considérons $\hat{\boldsymbol{\beta}}$ un estimateur de $\boldsymbol{\beta}$, qui peut être défini pour chacune des trois modèles, soit par la méthode du maximum de vraisemblance, soit par la méthode des moindres carrés. Il est possible d'estimer la probabilité $F(\mathbf{x}'_i \boldsymbol{\beta})$ par

$$\hat{y}_i = F(\mathbf{x}'_i \hat{\boldsymbol{\beta}}).$$

De plus, on peut réaliser une prédiction de y_i en utilisant

$$\hat{y}_i^* = \begin{cases} 1 & \text{si } \hat{y}_i \geq 0,5 \\ 0 & \text{si } \hat{y}_i < 0,5, \end{cases}$$

ce qui peut également s'écrire

$$\hat{y}_i^* = \begin{cases} 1 & \text{si } \mathbf{x}'_i \hat{\boldsymbol{\beta}} \geq 0 \\ 0 & \text{si } \mathbf{x}'_i \hat{\boldsymbol{\beta}} < 0, \end{cases}$$

Les deux indices principaux permettant d'évaluer la qualité du modèle sont :

- Le coefficient de détermination calculé entre les y_i et les \hat{y}_i .
- La proportion d'unités bien classées, définie par

$$P = \frac{1}{n} \left(n - \sum_{i=1}^n |y_i - \hat{y}_i^*| \right)$$

9.2 Analyse discriminante

9.2.1 Le modèle

Supposons que les données soient partitionnées en deux groupes notés G_1 et G_2 selon les valeurs de la variable dépendante y qui prend uniquement les valeurs 0 et 1 :

- l'unité $i \in G_1$ si $y_i = 1$,
- l'unité $i \in G_2$ si $y_i = 0$.

Les variables explicatives \mathbf{x}_i sont supposées aléatoires, continues, et indépendantes et sont régies par une fonction de densité qui dépend du groupe

- \mathbf{x}_i a une densité $f_1(\mathbf{x})$ si $i \in G_1$
- \mathbf{x}_i a une densité $f_2(\mathbf{x})$ si $i \in G_2$.

En outre, on suppose qu'une proportion p_1 d'unités appartient à G_1 et une proportion p_2 appartient à G_2 .

9.2.2 La règle bayésienne

Pour une unité particulière, le théorème de Bayes permet de déduire la probabilité d'appartenance à un groupe :

$$\Pr(\text{unité } i \in G_1 | \mathbf{x}_i) = \frac{p_1 f_1(\mathbf{x}_i)}{p_1 f_1(\mathbf{x}_i) + p_2 f_2(\mathbf{x}_i)}, \quad (9.4)$$

$$\Pr(\text{unité } i \in G_2 | \mathbf{x}_i) = \frac{p_2 f_2(\mathbf{x}_i)}{p_1 f_1(\mathbf{x}_i) + p_2 f_2(\mathbf{x}_i)}. \quad (9.5)$$

La règle de décision consiste à classer l'unité i dans G_1 si

$$\Pr(\text{unité } i \in G_1 | \mathbf{x}_i) \geq \Pr(\text{unité } i \in G_2 | \mathbf{x}_i),$$

et à classer l'unité i dans G_2 dans le cas contraire. Comme les dénominateurs de (9.4) et (9.5) sont égaux, on classe l'unité i dans G_1 si

$$p_1 f_1(\mathbf{x}_i) \geq p_2 f_2(\mathbf{x}_i),$$

ou si

$$\frac{f_1(\mathbf{x}_i)}{f_2(\mathbf{x}_i)} \geq \frac{p_2}{p_1}.$$

9.2.3 Le modèle multinormal

Un cas relativement simple s'obtient quand $f_1(\mathbf{x})$ et $f_2(\mathbf{x})$ ont une distribution multinormale de même matrice variance-covariance Σ , mais dont les moyennes dépendent du groupe, autrement dit

$$f_j(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_j)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_j) \right]$$

avec $j = 1, 2$. Dans ce cas,

$$\begin{aligned} \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &= \frac{\exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_1)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \right]}{\exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_2)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_2) \right]} \\ &= \exp \frac{1}{2} [(\mathbf{x} - \boldsymbol{\mu}_2)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_2) - (\mathbf{x} - \boldsymbol{\mu}_1)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_1)] \\ &= \exp \left[\mathbf{x}' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \frac{1}{2} \boldsymbol{\mu}_2' \Sigma^{-1} \boldsymbol{\mu}_2 - \frac{1}{2} \boldsymbol{\mu}_1' \Sigma^{-1} \boldsymbol{\mu}_1 \right] \end{aligned}$$

La règle bayésienne devient : on classe l'unité i dans G_1 si

$$\begin{aligned} \frac{f_1(\mathbf{x}_i)}{f_2(\mathbf{x}_i)} &= \exp \left[\mathbf{x}_i' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \frac{1}{2} \boldsymbol{\mu}_2' \Sigma^{-1} \boldsymbol{\mu}_2 - \frac{1}{2} \boldsymbol{\mu}_1' \Sigma^{-1} \boldsymbol{\mu}_1 \right] \\ &\geq \frac{p_2}{p_1}, \end{aligned}$$

ce qui s'écrit aussi

$$S(\mathbf{x}_i) \geq 0,$$

où

$$S(\mathbf{x}) = \mathbf{x}'_i \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \frac{1}{2} \boldsymbol{\mu}'_2 \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 - \frac{1}{2} \boldsymbol{\mu}'_1 \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \log \frac{p_2}{p_1}.$$

La fonction $S(\mathbf{x})$ est appelée fonction de score ou statistique d'Anderson. La fonction $S(\mathbf{x})$ est estimée simplement en prenant

$$\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{x}}_1, \hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{x}}_2$$

et

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-2} \left[\sum_{i \in G_1} (\mathbf{x}_i - \bar{\mathbf{x}}_1)(\mathbf{x}_i - \bar{\mathbf{x}}_1)' + \sum_{i \in G_2} (\mathbf{x}_i - \bar{\mathbf{x}}_2)(\mathbf{x}_i - \bar{\mathbf{x}}_2)' \right],$$

où

$$\bar{\mathbf{x}}_1 = \frac{1}{n_1} \sum_{i \in G_1} \mathbf{x}_i \text{ et } \bar{\mathbf{x}}_2 = \frac{1}{n_2} \sum_{i \in G_2} \mathbf{x}_i.$$

Exercices

Exercice 9.1 Soit une variable dépendante binaire y_i prenant les valeurs 0 et 1. Supposons que l'on dispose de deux variables explicatives définies de la manière suivante :

$$\begin{aligned} x_{i1} &= 1 \text{ pour tout } i = 1, \dots, n \\ x_{i2} &= \begin{cases} 1 & \text{si } i \text{ présente une caractéristique} \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

Si on note

$$\begin{aligned} n_{x_2} &= \sum_{i=1}^n x_{i2}, \\ n_y &= \sum_{i=1}^n y_i, \\ n_{x_2 y} &= \sum_{i=1}^n x_{i2} y_i, \end{aligned}$$

estimer le coefficient de régression pour toutes les méthodes proposées.

Chapitre 10

Exercices récapitulatifs

Exercice 10.1 Exercice concernant chapitre 1.

1. Donnez le rang de la matrice suivante :

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 2 & 1 \\ 6 & 3 & 8 \end{pmatrix}$$

2. Effectuez le produit \mathbf{Ax} , où :

$$\mathbf{x} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$$

3. Donnez en quelques phrases une interprétation géométrique de l'estimation données par moindres carrés ordinaires.

Exercice 10.2 Exercice concernant le chapitre 2.

1. Montrez, dans le cas bivarié, que :

$$\sum_{i=1}^n e_i = 0$$

2. Expliquez en une phrase le principe des moindres carrés et donner l'estimation de \mathbf{b} (en fonction de la matrice \mathbf{X} et du vecteur \mathbf{y}) qui en découle.

Exercice 10.3 Exercice concernant le chapitre 3.

1. Donnez la définition mathématique de l'espérance d'une variable aléatoire discrète et de même pour une variable aléatoire continue et donnez une interprétation de l'espérance en une phrase.
2. Montrez que le carré d'une variable de Student à q degrés de liberté est une variable de Fisher à 1 et q degrés de liberté.

Exercice 10.4 Exercice concernant le chapitre 4.

1. Citez les hypothèses du modèle linéaire général sous forme mathématique avec explications en français.
2. Citez l'hypothèse faite concernant les erreurs dans l'estimation par maximum de vraisemblance et expliquez en quelques phrases la technique de l'estimation par maximum de vraisemblance.
3. Donnez la définition d'un estimateur sans biais et donnez un exemple (avec calculs!).

Exercice 10.5 Soit le modèle :

$$y_t = a + bt + \varepsilon_t \text{ avec } t = 1, \dots, n$$

1. Donnez les estimations par MCO de a et b en les simplifiant autant que possible.
2. Calculer la variance de $\hat{\beta}$ (où $\beta = (a, b)$) à nouveau en la simplifiant autant que possible.

Exercice 10.6 Définissez la notion de variable exogène (au sens statistique). La notion de variable exogène est-elle la même que celle de variable explicative ? (réponse sans calcul et en 3 lignes maximum)

Exercice 10.7 En utilisant les variables instrumentales, on utilise la matrice

$$\mathbf{X} \{ \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} \}^{-1} \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$$

1. Montrez que cette matrice est idempotente (et est donc un projecteur). (réponse en 3 lignes)
2. Sur quel sous-espace cette matrice projette-t-elle ? (réponse en 1 ligne).

Exercice 10.8 La régression peut s'écrire comme une projection sur un sous-espace. Quelle est la signification géométrique du nombre de degrés de liberté $n - p$ par lequel on divise la somme des carrés des résidus ? (réponse sans calcul et 2 lignes maximum).

Exercice 10.9 Exercice basé sur le chapitre 2 du livre de Ernst E. Berndt, The practice of Econometrics-classic and contemporary.

Le CAPM est à la base de la théorie moderne du portefeuille. C'est un modèle d'évaluation pour les actifs financiers qui fut développé dans les années 60. Ce modèle met en relation la rentabilité d'un titre financier avec la rentabilité du marché et cela d'une manière très simple. L'idée de base est la suivante. Les investisseurs sont rémunérés pour le risque qu'ils prennent. Lorsqu'ils investissent dans un titre, ils prennent d'une part un risque spécifique (risque lié à l'entreprise ou à son secteur d'activité) et d'autre part un risque systématique ou risque de marché (risque lié aux conditions macro-économique du pays par exemple). En diversifiant son portefeuille, l'investisseur pourra éliminer une bonne partie du risque spécifique ; par contre, le risque systématique ne pourra être éliminé puisque toutes les entreprises y sont confrontées. Par conséquent, l'investisseur ne sera rémunéré que pour le risque systématique qu'il prendra. Cette exposition au risque de marché s'appelle β ; elle correspond à la covariance entre le rendement du titre (ou du portefeuille) et le rendement du marché divisé par la variance du marché. Ainsi selon ce modèle très simple la prime d'un actif i (défini comme le rendement du titre i moins le taux sans risque) est donnée par l'exposition au risque du marché (β multiplié par la prime de risque du marché (défini comme la rentabilité du marché moins le taux sans risque)). Sous sa forme mathématique, on a :

$$R_i - R_f = \beta \times [R_m - R_f]$$

Le but sera de tester ce modèle. Pour se faire nous allons prendre la spécification suivante :

$$R_i - R_f = \alpha + \beta * [R_m - R_f] + \epsilon_i$$

où

R_m est la rentabilité mensuelle du marché

R_f est le taux sans risque

$\epsilon_i \sim \mathcal{N}(0, \alpha^2)$

Fait très intéressant : l'estimation du paramètre β par MCO est donné par $\widehat{\text{cov}}(R_i, R_m) \div \widehat{\text{var}}(R_m)$

Ainsi l'estimation du β par MCO rejoint la définition du β donnée ci-dessus. Il est donc tout à fait approprié d'estimer le β par MCO.

1. Avant de commencer, réfléchissez aux tests d'hypothèses que vous pourriez mettre en oeuvre. Quelle devrait être la valeur de α selon le modèle théorique ? Que pouvez-vous dire concernant le β d'une entreprise plutôt risquée ? De celui d'une entreprise plutôt stable (nommée souvent "blue chip") ? Et d'un portefeuille essayant de répliquer le marché ?

2. Comme toujours, la 1ère étape sur Eviews consiste à créer un espace de travail. Les données sont mensuelles (**Monthly**) et la période d'observation va du mois de janvier 1976 (notation :**1976 : 1**) au mois de décembre 1987 (notation :**1987 : 12**).
3. Maintenant, importez les séries de rentabilité CONED (Consolidated Edison), DEC (Digital Equipment Company), MARKET (marché), RKFREE (taux sans risque). Attention, les observations de ces séries vont de 1978 :01 à 1987 :12. Par défaut Eviews choisit la période d'observation (**sample**) égale à la période spécifiée lors de la création de l'espace de travail (**workfile range**). Ainsi, il faudra adapter la période d'observation à celle de nos séries. Pour cela, il suffit de sélectionner la case **Sample** dans la barre des menus et de spécifier la période voulue. Une fois cela fait, procédez comme à l'exercice 1 pour importer les données. Cette fois-ci les séries à importer sont du type ASCII (fichier texte). La fenêtre **ASCIITextImport** qui apparaîtra en suivant la même démarche qu'à l'exercice 1 est légèrement différente de la fenêtre **ExcelSpreadsheetImport** que vous avez rencontré à l'exercice 1. Dans la 1ère case, il faut spécifier le nom de chaque série s'il n'est pas déjà dans le fichier ou le nombre de séries à importer si les séries sont déjà nommées dans le fichier. Pour vous aider, il y a une case en bas de la fenêtre pour visualiser le fichier. Attention aux délimiteurs. Pour le reste des options sur cette fenêtre, je vous laisse faire la traduction, c'est straightforward!
4. Nous avons les séries pour les rentabilités, mais rappelez-vous, nous voulons faire des régressions sur les primes. Il faut donc créer de nouvelles séries en prenant les différences entre les rentabilités des titres ou du marché et le rendement sans risque. Pour se faire, allez dans **Quick/Generateseries** et entrez une équation du type :

$$P_MARKET = MARKET - RKFREE.$$

Faites-le pour toutes les séries importées. Une fois cela fait, profitez-en pour visualiser les primes sous une autre vue.

5. Il est temps de faire des régressions. Prenez comme variable expliquée la prime de CONED (entreprise produisant de l'électricité) et ensuite prenez la prime de DEC (entreprise évoluant dans le secteur informatique). Que constatez-vous par rapport aux β et α estimés ? Sont-ils comme vous les imaginiez ? Étant donné que le R^2 de la régression indique la proportion de la variation de la variable expliquée (risque total) qui est expliquée par la variation de la variable explicative (risque systématique), le R^2 dans le CAPM nous donne la part du risque de marché (systématique) dans le risque total du titre. Que constatez-vous ? Donnez la part du risque spécifique et systématique par rapport au risque total.
6. Pour les deux régressions, testez $\alpha = 0$ contre $\alpha \neq 0$, $\beta = 0$ contre $\beta \neq 0$, $\beta = 1$ contre $\beta \neq 1$ et le test joint : $\alpha = 0, \beta = 1$ contre $\alpha \neq 0, \beta \neq 1$. Pour cela sélectionnez

View/Coefficient tests/Wald – Coefficient Restrictions

et spécifiez la contrainte linéaire que vous voulez tester. Attention, les coefficients de la régression sont stockés dans Eviews dans le vecteur c. Pour accéder au premier coefficient estimé (très souvent la constante), il faut taper c(1), pour le deuxième coefficient estimé c(2), etc ...

7. Le CAPM suppose que tous les investisseurs accèdent simultanément à l'information et l'utilisent rationnellement dans leur décision d'investissement. Nous allons tester cette hypothèse en faisant de l'analyse d'événement. Le but est de savoir si un événement générant de l'information se répercute significativement au niveau du marché. Pour cela vous avez le fichier EVENTS qui contient plusieurs séries. Attention, la période d'observation n'est pas la même qu'avant. Vous savez ce qu'il faut faire dans ce cas maintenant ! La série GPU contient les rentabilités observées de la General Public Utilities. Cette entreprise est propriétaire de la Three Mile Island plant. Le but est d'analyser l'effet de l'incident nucléaire qui se produisit dans cette station nucléaire le 28 mars 1979.
8. Faites la régression comme précédemment. Trouvez le graphe des résidus et regardez le résidu en avril 1979. Que constatez-vous ? Pourrait-on améliorer notre modèle pour prendre en compte cet événement ?
9. Oui, on peut. Pour ce faire nous allons ajouter à notre modèle une variable muette qui agira comme un détecteur d'événement. Cette variable prendra la valeur un pour le mois d'avril 1979 et zéro partout ailleurs. Pour créer cette série simplement, allez dans **Quick/GenerateSeries** Et vous entrez TMI-DUM=0 (TMIDUM sera le nom de votre variable muette). Puis vous changez le période d'observation pour n'avoir que la période 1979 :4 (pour cela, il suffit d'aller dans Sample et de spécifier la période voulue) et vous refaites la même procédure que ci-dessus, mais cette fois-ci en spécifiant TMIDUM=1.

Maintenant remettez la période d'observation que vous aviez avant (1976 :1 à 1985 :12). Et votre variable muette est là ! Souvent, les variables muettes sont des variables saisonnières. Dans ce cas, EvIEWS a déjà des fonctions préprogrammées. Veuillez vous référer à la fonction @seas(n) pour plus de détails.

- Maintenant, il ne reste plus qu'à faire la régression en n'oubliant pas d'ajouter la variable muette comme variable explicative pour prendre en compte l'incident nucléaire. Regardez le graphe des résidus. Que constatez-vous ? Regardez l'output de la régression. Est-ce que la variable muette est statistiquement significative ? Que peut-on conclure sur l'importance de cet événement et sur l'efficacité du marché dans ce cas ?

Exercice 10.10 Exercice d'introduction au logiciel EvIEWS 3.0.

L'entreprise Nambe Mills basée à Santa Fe au Nouveau Mexique fabrique de la vaisselle et d'autres produits ménagers. Pour mieux maîtriser ses coûts et optimiser sa production, elle désire connaître précisément le rapport entre le temps qu'il faut pour polir un de leur produit et le diamètre de ce dernier.

- Pour se faire, vous disposez de 59 données en coupe dans le fichier **don.polish.xls**. En premier lieu, il faut créer un espace de travail (**workfile**) qui prend en compte des données en coupe allant de 1 à 59. Pour cela, allez dans **File/New/workfile**. La fenêtre **WorkfileRange** apparaît ; il suffit de choisir Undated or Irregular et de spécifier dans les deux cases en-dessous la première observation (**1**) et la dernière (**59**).
- Importez les 3 séries qui se trouvent dans le fichier Excel **don.polish.xls**. Pour se faire, allez dans **Procs/Import/ReadText – Lotus – Excel** ; le fichier se trouve sur un des disques partagés de l'Université (**Etu commun** sur Noir). Puisque les séries sont déjà nommées, il suffira de spécifier le nombre de séries dans la première case de la fenêtre **ExcelSpreadsheetImport** (il y en a trois qui nous intéressent). Attention les données qui nous intéressent commencent dans la case B2.
- Vous avez fait le plus dur. Il est maintenant temps d'enregistrer tout ça sur votre disquette. Cliquez sur la case **Save**. Et enregistrez votre workfile (.wfl) sur **Disquette3 $\frac{1}{4}$ (A :)** en choisissant un nom approprié.
- Avant de manipuler vos séries, il est utile de vérifier si l'importation des données s'est faite correctement. Pour cela, vous allez créer un groupe (**Group**) en sélectionnant les trois séries importées. Plus précisément vous sélectionnez une des trois séries puis vous maintenez pressé **CTRL** et cliquez sur les deux autres. Vous avez mis en évidence les trois séries ; maintenant double cliquez (avec le bouton de gauche) sur une des trois et vous voyez l'option pour ouvrir un groupe ; faites-le en cliquant sur **OpenGroup**. Sauvez ce groupe en le nommant. Allez sur **Name** et choisissez un nom approprié. Attention, ayez toujours en tête qu'EvIEWS ne permet pas de choisir des noms de plus de 16 caractères (ou 8 dépendant de l'objet). So keep it short ! Et évitez les accents.
- Vous avez vérifié et les données sont en ordre. Maintenant vous pouvez commencer le travail d'économétrie à proprement parler. La première étape consiste toujours à "prendre connaissance des données", c'est-à-dire à se familiariser avec les données en les visualisant par exemple. Cette étape est très importante car elle nous permet de déceler très rapidement des tendances, de la non-stationnarité, des retournements ou tout simplement des données aberrantes dans nos séries. Pour une vue simultanée des trois séries sélectionnez **View/MultipleGraphs/Line**. Il apparaîtra les trois petits graphiques à l'écran. Vous avez sûrement remarqué que le tableau contenant les données des trois séries a disparu. Ceci est normale. En effet, EvIEWS travaille avec des objets (**Series, Group, Equation, Matrix**, etc.) et ces objets peuvent être "visualisés" de différentes manières. Pour voir le choix qui vous est proposé allez dans **View**. Comme vous le constatez, le choix est impressionnant. Si vous voulez à nouveau visualiser le groupe sous l'angle des données (c.-à-d. de visualiser toutes les observations de chaque série) il vous faut sélectionner **SpreadSheet**. C'est un bon moment pour sélectionner différents points de vue de l'objet et ainsi découvrir une partie de toutes les potentialités que recèlent ce logiciel. Par exemple, en sélectionnant **Correlations** vous allez voir apparaître la matrice de corrélations entre les variables. Cette matrice peut vous permettre d'évaluer assez rapidement les risques de multi-collinéarité entre les différentes variables explicatives.
- Maintenant que vous vous êtes familiarisés avec les données, il est temps de faire des régressions. Vous allez estimer un modèle de régression pour estimer le temps de polissage d'un objet (**time**) en fonction de son diamètre (**diam**) et en ajoutant une variable muette (**dum_{cass}**) qui prend en compte le fait

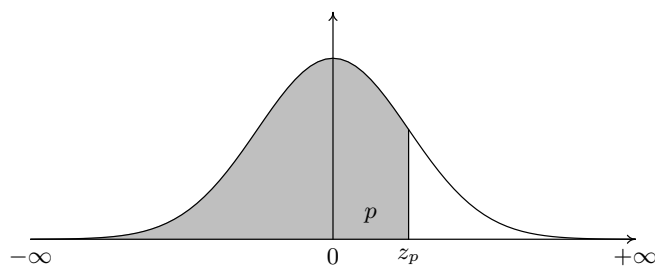
que les casseroles (plus complexe) prennent plus de temps à polir que la vaisselle. Avant de faire la régression, demandez-vous si il est utile ou non de mettre une constante dans le modèle. Pour confirmer vos soupçons (!), faites la régression avec et sans la constante.

7. Pour estimer le modèle, sélectionnez **Quick** du menu principal et choisissez **EstimateEquation...** Cela ouvrira la fenêtre **EquationSpecification**. Tapez dans la première case, en premier lieu, le nom de la variable expliquée, ensuite si vous voulez une constante, tapez **C** pour l'inclure dans le modèle (le logiciel reconnaîtra ce **C** comme la constante à inclure ; n'appellez donc jamais une de vos séries **C** , cela risquerait de poser problème !) et ensuite tapez le nom de chaque variable explicative du modèle. Il est important de laisser un espace entre chaque variable. En économétrie, il arrive fréquemment de devoir prendre des différences premières, secondes, etc. Le logiciel a pris ce fait en compte et permet de simplifier la tâche de l'utilisateur ; pour des différences premières du logarithme d'une série nommée ABC, il suffit de taper dlog(ABC) ; pour les différences premières d'une série ABC, il suffit de taper d(ABC) . Et si l'on veut une fonction de la variable comme le logarithme du ABC, il suffit de taper log(ABC) . Par défaut la méthode d'estimation est celle des moindres carrées (**LS – LeastSquares**). C'est ce que vous voulez. Reste à spécifier sur quel échantillon la régression se fera. Par défaut Eviews spécifie la taille de l'échantillon que vous avez donnée lors de la création du workfile au point 1. L'échantillon est toujours le même, il ne reste plus qu'à presser **OK** et la régression s'effectuera. Ce n'était pas si dur !
8. Les résultats de la régression sont apparus à l'écran. Vous remarquerez que la qualité de la régression est bonne. Pour des données en coupe, un R^2 de 40% peut déjà être considéré comme bon. Ces résultats sont intéressants, mais ils ne sont valables que si les hypothèses du modèle linéaire générale sont satisfaites. Or, il est bien connu que lorsque l'on a des données en coupe, un effet taille peut apparaître et l'hypothèse d'homoscédasticité n'est plus satisfaite dans ce cas. Une première approche pour observer cela est de changer de vue. Et oui, l'estimation LS est un objet **Equation** et comme tout objet sur Eviews il y a plusieurs vues possible pour étudier l'objet. Je vous conseille de sélectionner **View/Actual, Fitted, Residual/Graph**. Ce graphe vous donne les valeurs estimées et observées par le modèle de la variable expliquée et en-dessous les résidus. Regardez les résidus. Vous observez des piques parfois très grand qui peuvent signaler une présence d'hétéroscédasticité. Mais pour en être sûr, créez un groupe comprenant la série **diam** et la série **resid** (qui contient par défaut les résidus de la dernière régression effectuée par le logiciel). Maintenant le but est de produire un graphe qui met en rapport les résidus avec le diamètre des produits. Pour cela, il faut à nouveau changer de vue, sélectionnez **View/Graph/Scatter/SimpleScatter**. L'hétéroscédasticité est maintenant flagrante. Pour conserver à part ce graphe, sélectionnez l'option **Freeze**. Cette option "gèle" l'image à l'écran et l'intègre dans un nouvel objet qu'il faudra nommer. Les résultats de la régression précédente sont donc inutilisable, il faut corriger cet effet.
9. Avant de corriger l'effet, il faut s'assurer qu'on est bien en présence d'hétéroscédasticité. Pour cela, il existe un test statistique, c'est le test de White. Ce test peut se faire sur Eviews. Reprenez la fenêtre contenant votre régression et changez de vue (**View/ResidualTest/White/Heteroskedasticity**). Le test indique clairement que l'hypothèse nulle d'homoscédasticité est rejetée.
10. Dans le but d'éviter l'hétéroscédasticité et également afin de faciliter l'interprétation économique, on effectuera la même régression, mais cette fois-ci en prenant le logarithme des séries **time** et **diam**. Donnez intuitivement la raison de prendre les logarithmes. Commenter vos résultats, discutez du comportement des erreurs dans ce cas-ci et faites le test de White. Félicitations, vous venez de terminer votre premier exercice sur Eviews !

Chapitre 11

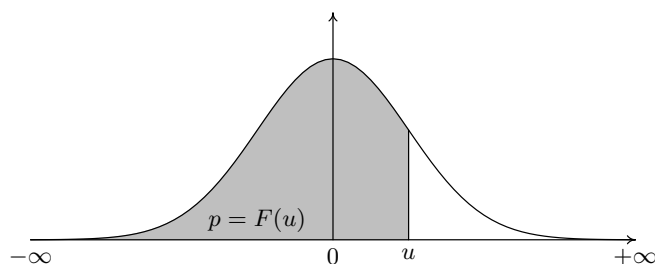
Tables statistiques

TAB. 11.1 – Table des quantiles d’une variable normale centrée réduite



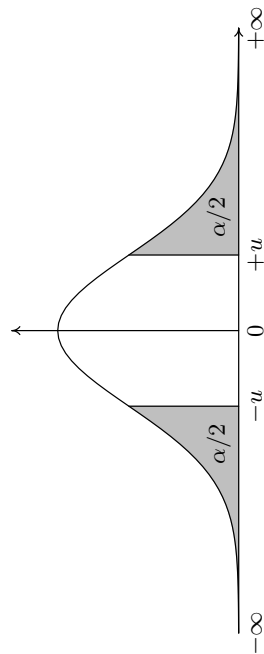
Ordre du quantile (p)	Quantile (z_p)	Ordre du quantile (p)	Quantile (z_p)
0.500	0.0000	0.975	1.9600
0.550	0.1257	0.976	1.9774
0.600	0.2533	0.977	1.9954
0.650	0.3853	0.978	2.0141
0.700	0.5244	0.979	2.0335
0.750	0.6745	0.990	2.3263
0.800	0.8416	0.991	2.3656
0.850	1.0364	0.992	2.4089
0.900	1.2816	0.993	2.4573
0.950	1.6449	0.994	2.5121
0.970	1.8808	0.995	2.5758
0.971	1.8957	0.996	2.6521
0.972	1.9110	0.997	2.7478
0.973	1.9268	0.998	2.8782
0.974	1.9431	0.999	3.0902

TAB. 11.2 – Fonction de répartition de la loi normale centrée réduite
(Probabilité de trouver une valeur inférieur à u)



u	0.0	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

TAB. 11.3 – Quantiles de la loi normale centrée réduite
(u : valeur ayant la probabilité α d'être dépassé en valeur absolue)



α	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	∞	2.5758	2.3263	2.1701	2.0537	1.9600	1.8808	1.8119	1.7507	1.6954
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.4395	1.4051	1.3722	1.3408	1.3106
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.1503	1.1264	1.1031	1.0803	1.0581
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.9346	0.9154	0.8965	0.8779	0.8596
0.4	0.8416	0.8239	0.8064	0.7892	0.7722	0.7554	0.7388	0.7225	0.7063	0.6903
0.5	0.6745	0.6588	0.6433	0.6280	0.6128	0.5978	0.5828	0.5681	0.5534	0.5388
0.6	0.5244	0.5101	0.4958	0.4817	0.4677	0.4538	0.4399	0.4261	0.4125	0.3989
0.7	0.3853	0.3719	0.3585	0.3451	0.3319	0.3186	0.3055	0.2924	0.2793	0.2663
0.8	0.2533	0.2404	0.2275	0.2147	0.2019	0.1891	0.1764	0.1637	0.1510	0.1383
0.9	0.1257	0.1130	0.1004	0.0878	0.0753	0.0627	0.0502	0.0376	0.0251	0.0125

TAB. 11.4 – Table des quantiles d'une variable χ^2 à n degrés de liberté

	ordre du quantile					
	0.01	0.025	0.05	0.95	0.975	0.99
$n=1$	0.000157	0.000982	0.003932	3.841	5.024	6.635
2	0.02010	0.05064	0.103	5.991	7.378	9.210
3	0.115	0.216	0.352	7.815	9.348	11.34
4	0.297	0.484	0.711	9.488	11.14	13.28
5	0.554	0.831	1.145	11.07	12.83	15.09
6	0.872	1.237	1.635	12.59	14.45	16.81
7	1.239	1.690	2.167	14.07	16.01	18.48
8	1.646	2.180	2.733	15.51	17.53	20.09
9	2.088	2.700	3.325	16.92	19.02	21.67
10	2.558	3.247	3.940	18.31	20.48	23.21
11	3.053	3.816	4.575	19.68	21.92	24.72
12	3.571	4.404	5.226	21.03	23.34	26.22
13	4.107	5.009	5.892	22.36	24.74	27.69
14	4.660	5.629	6.571	23.68	26.12	29.14
15	5.229	6.262	7.261	25.00	27.49	30.58
16	5.812	6.908	7.962	26.30	28.85	32.00
17	6.408	7.564	8.672	27.59	30.19	33.41
18	7.015	8.231	9.390	28.87	31.53	34.81
19	7.633	8.907	10.12	30.14	32.85	36.19
20	8.260	9.591	10.85	31.41	34.17	37.57
21	8.897	10.28	11.59	32.67	35.48	38.93
22	9.542	10.98	12.34	33.92	36.78	40.29
23	10.20	11.69	13.09	35.17	38.08	41.64
24	10.86	12.40	13.85	36.42	39.36	42.98
25	11.52	13.12	14.61	37.65	40.65	44.31
26	12.20	13.84	15.38	38.89	41.92	45.64
27	12.88	14.57	16.15	40.11	43.19	46.96
28	13.56	15.31	16.93	41.34	44.46	48.28
29	14.26	16.05	17.71	42.56	45.72	49.59
30	14.95	16.79	18.49	43.77	46.98	50.89
31	15.66	17.54	19.28	44.99	48.23	52.19
32	16.36	18.29	20.07	46.19	49.48	53.49
33	17.07	19.05	20.87	47.40	50.73	54.78
34	17.79	19.81	21.66	48.60	51.97	56.06
35	18.51	20.57	22.47	49.80	53.20	57.34
36	19.23	21.34	23.27	51.00	54.44	58.62
37	19.96	22.11	24.07	52.19	55.67	59.89
38	20.69	22.88	24.88	53.38	56.90	61.16
39	21.43	23.65	25.70	54.57	58.12	62.43
40	22.16	24.43	26.51	55.76	59.34	63.69
42	23.65	26.00	28.14	58.12	61.78	66.21
44	25.15	27.57	29.79	60.48	64.20	68.71
46	26.66	29.16	31.44	62.83	66.62	71.20
48	28.18	30.75	33.10	65.17	69.02	73.68
50	29.71	32.36	34.76	67.50	71.42	76.15
60	37.48	40.48	43.19	79.08	83.30	88.38
70	45.44	48.76	51.74	90.53	95.02	100.43
80	53.54	57.15	60.39	101.88	106.63	112.33
90	61.75	65.65	69.13	113.15	118.14	124.12
100	70.06	74.22	77.93	124.34	129.56	135.81
110	78.46	82.87	86.79	135.48	140.92	147.41
120	86.92	91.57	95.70	146.57	152.21	158.95

TAB. 11.5 – Table des quantiles d’une variable de Student à n degrés de liberté

	ordre du quantile			
	0.95	0.975	0.99	0.995
$n=1$	6.314	12.71	31.82	63.66
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
31	1.696	2.040	2.453	2.744
32	1.694	2.037	2.449	2.738
33	1.692	2.035	2.445	2.733
34	1.691	2.032	2.441	2.728
35	1.690	2.030	2.438	2.724
36	1.688	2.028	2.434	2.719
37	1.687	2.026	2.431	2.715
38	1.686	2.024	2.429	2.712
39	1.685	2.023	2.426	2.708
40	1.684	2.021	2.423	2.704
50	1.676	2.009	2.403	2.678
60	1.671	2.000	2.390	2.660
70	1.667	1.994	2.381	2.648
80	1.664	1.990	2.374	2.639
90	1.662	1.987	2.368	2.632
100	1.660	1.984	2.364	2.626
120	1.658	1.980	2.358	2.617
∞	1.645	1.960	2.327	2.576

TAB. 11.6 – Table des quantiles d'ordre 0.95 d'une variable de Fisher à n_1 et n_2 degrés de liberté

	$n_1=1$	2	3	4	5	6	7	8	9	10	12	14	16	20	30	∞
$n_2=1$	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.4	246.5	248.0	250.1	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.45	19.46	19.50
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.715	8.692	8.660	8.617	8.526
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.873	5.844	5.803	5.746	5.628
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.636	4.604	4.558	4.496	4.365
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.956	3.922	3.874	3.808	3.669
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.529	3.494	3.445	3.376	3.230
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.237	3.202	3.150	3.079	2.928
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.025	2.989	2.936	2.864	2.707
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.865	2.828	2.774	2.700	2.538
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.739	2.701	2.646	2.570	2.404
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.637	2.599	2.544	2.466	2.296
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.554	2.515	2.459	2.380	2.206
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.484	2.445	2.388	2.308	2.131
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.424	2.385	2.328	2.247	2.066
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.373	2.333	2.276	2.194	2.010
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381	2.329	2.289	2.230	2.148	1.960
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.290	2.250	2.191	2.107	1.917
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.256	2.215	2.155	2.071	1.878
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.225	2.184	2.124	2.039	1.843
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.250	2.197	2.156	2.096	2.010	1.812
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.226	2.173	2.131	2.071	1.984	1.783
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.204	2.150	2.109	2.048	1.961	1.757
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.183	2.130	2.088	2.027	1.939	1.733
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.165	2.111	2.069	2.007	1.919	1.711
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220	2.148	2.094	2.052	1.990	1.901	1.691
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.132	2.078	2.036	1.974	1.884	1.672
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.118	2.064	2.021	1.959	1.869	1.654
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.104	2.050	2.007	1.945	1.854	1.638
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.092	2.037	1.995	1.932	1.841	1.622
32	4.149	3.295	2.901	2.668	2.512	2.399	2.313	2.244	2.189	2.142	2.070	2.015	1.972	1.908	1.817	1.594
34	4.130	3.276	2.883	2.650	2.494	2.380	2.294	2.225	2.170	2.123	2.050	1.995	1.952	1.888	1.795	1.569
36	4.113	3.259	2.866	2.634	2.477	2.364	2.277	2.209	2.153	2.106	2.033	1.977	1.934	1.870	1.776	1.547
38	4.098	3.245	2.852	2.619	2.463	2.349	2.262	2.194	2.138	2.091	2.017	1.962	1.918	1.853	1.760	1.527
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.003	1.948	1.904	1.839	1.744	1.509
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.952	1.895	1.850	1.784	1.687	1.438
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.917	1.860	1.815	1.748	1.649	1.389
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.834	1.775	1.728	1.659	1.554	1.254
∞	3.841	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831	1.752	1.692	1.644	1.571	1.459	1.000

TAB. 11.7 – Table des quantiles d'ordre 0.99 d'une variable de Fisher à n_1 et n_2 degrés de liberté

	$n_1=1$	2	3	4	5	6	7	8	9	10	12	14	16	20	30	∞
$n_2=1$	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6143	6170	6209	6261	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.44	99.45	99.47	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.92	26.83	26.69	26.51	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.25	14.15	14.02	13.84	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.888	9.770	9.680	9.553	9.379	9.020
6	13.75	10.93	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.605	7.519	7.396	7.229	6.880
7	12.25	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.359	6.275	6.155	5.992	5.650
8	11.26	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.559	5.477	5.359	5.198	4.859
9	10.56	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	5.005	4.924	4.808	4.649	4.311
10	10.04	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.601	4.520	4.405	4.247	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.397	4.293	4.213	4.099	3.941	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.052	3.972	3.858	3.701	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.857	3.778	3.665	3.507	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.698	3.619	3.505	3.348	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.564	3.485	3.372	3.214	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.451	3.372	3.259	3.101	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.353	3.275	3.162	3.003	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.269	3.190	3.077	2.919	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.195	3.116	3.003	2.844	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.130	3.051	2.938	2.778	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.072	2.993	2.880	2.720	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	3.019	2.941	2.827	2.667	2.305
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.973	2.894	2.781	2.620	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.930	2.852	2.738	2.577	2.211
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.892	2.813	2.699	2.538	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.857	2.778	2.664	2.503	2.131
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.824	2.746	2.632	2.470	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.795	2.716	2.602	2.440	2.064
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.868	2.767	2.689	2.574	2.412	2.034
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.742	2.663	2.549	2.386	2.006
32	7.499	5.336	4.459	3.969	3.652	3.427	3.258	3.127	3.021	2.934	2.798	2.696	2.618	2.503	2.340	1.956
34	7.444	5.289	4.416	3.927	3.611	3.386	3.218	3.087	2.981	2.894	2.758	2.657	2.578	2.463	2.299	1.911
36	7.396	5.248	4.377	3.890	3.574	3.351	3.183	3.052	2.946	2.859	2.723	2.622	2.543	2.428	2.263	1.872
38	7.353	5.211	4.343	3.858	3.542	3.319	3.152	3.021	2.915	2.828	2.692	2.591	2.512	2.397	2.232	1.837
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.563	2.484	2.369	2.203	1.805
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698	2.562	2.461	2.382	2.265	2.098	1.683
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.394	2.315	2.198	2.028	1.601
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.234	2.154	2.035	1.860	1.381
∞	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.082	2.000	1.878	1.696	1.000

TAB. 11.8 – Valeur critique du test de Durbin-Watson au seuil de 5%

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

k est le nombre de variables explicatives (constante exclue).

n est la taille de l'échantillon.

TAB. 11.9 – Quantiles du coefficient de corrélation de Pearson d’une variable aléatoire normale bivariée sous l’hypothèse que $\rho = 0$

n	ordre du quantile			
	0.05	0.025	0.005	0.0005
4	0.900	0.950	0.990	0.999
5	0.805	0.878	0.959	0.991
6	0.729	0.811	0.917	0.974
7	0.669	0.754	0.875	0.951
8	0.621	0.707	0.834	0.925
9	0.582	0.666	0.798	0.898
10	0.549	0.632	0.765	0.872
11	0.521	0.602	0.735	0.847
12	0.497	0.576	0.708	0.823
13	0.476	0.553	0.684	0.801
14	0.458	0.532	0.661	0.780
15	0.441	0.514	0.641	0.760
16	0.426	0.497	0.623	0.742
17	0.412	0.482	0.606	0.725
18	0.400	0.468	0.590	0.708
19	0.389	0.456	0.575	0.693
20	0.378	0.444	0.561	0.679
21	0.369	0.433	0.549	0.665
22	0.360	0.423	0.537	0.652
23	0.352	0.413	0.526	0.640
24	0.344	0.404	0.515	0.629
25	0.337	0.396	0.505	0.618
26	0.330	0.388	0.496	0.607
27	0.323	0.381	0.487	0.597
28	0.317	0.374	0.479	0.588
29	0.311	0.367	0.471	0.579
30	0.306	0.361	0.463	0.570
35	0.283	0.334	0.430	0.532
40	0.264	0.312	0.403	0.501
45	0.248	0.294	0.380	0.474
50	0.235	0.279	0.361	0.451
60	0.214	0.254	0.330	0.414
70	0.198	0.235	0.306	0.385
80	0.185	0.220	0.286	0.361
90	0.174	0.207	0.270	0.341
100	0.165	0.197	0.256	0.324
200	0.117	0.139	0.182	0.231
300	0.095	0.113	0.149	0.189
400	0.082	0.098	0.129	0.164
500	0.074	0.088	0.115	0.147
1000	0.052	0.062	0.081	0.104

n est la taille de l’échantillon.

Bibliographie

- Bourbonnais, R. (1993). *Econométrie*. Dunod, Paris.
- Cohen, M. and Pradel, J. (1993). *Econométrie*. Litec, Paris.
- Gourieroux, C. and Monfort, A. (1989a). *Statistics and Econometric Models*, volume 1. Press Syndicate of the University of Cambridge, Cambridge.
- Gourieroux, C. and Monfort, A. (1989b). *Statistics and Econometric Models*, volume 2. Press Syndicate of the University of Cambridge, Cambridge.
- Greene, W. (1990). *Econometric Analysis*. Macmillan Publishing Company, New York.
- Johnson, J. et DiNardo, J. (1999). *Méthodes Économétriques*. Economica, Paris, 4 edition.
- Johnston, J. (1988). *Econometrics Methods*. McGraw-Hill, Singapore, 4 edition.
- Johnston, J. et DiNardo, J. (1997). *Méthodes économétriques*. Economica, Paris, 4ème edition.
- Judge, G., Griffiths, W., Carter Hill, R., Lütkepohl, H., and Lee, T. (1985). *The Theory and Practice of Econometrics*. Wiley, USA, 2 edition.
- Maddala, G. (1988). *Introduction to Econometrics*. Macmillan Publishing company, New York.
- Ruud, P. (2000). *An Introduction to classical Econometric Theory*. Oxford University Press, New York, Oxford.
- Theil, H. (1979). *Principles of Econometrics*. Wiley Hamilton publication, Canada.

Liste des tableaux

2.1	Taille et poids de 20 individus	16
2.2	Données pour les variables x et y	29
2.3	Données sur le travail, le capital et la production	30
3.1	Erreur de première et seconde espèce	41
3.2	Probabilité de commettre les erreurs	41
4.1	Tableau récapitulatif	46
5.1	Tableau d'analyse de la variance	56
5.2	Tableau d'analyse de la variance à un facteur	61
5.3	Nombre de mots selon les groupes	65
5.4	Moyennes selon les groupes	65
5.5	Tableau d'analyse de la variance	65
5.6	Consommation de crème glacée	67
5.7	Temps selon les groupes	69
5.8	Tableau d'analyse de la variance	69
5.9	Nombre d'éléments rappelés selon l'âge	69
5.10	Nombre d'éléments rappelés selon l'âge et le niveau	70
5.11	Temps de latence selon les groupes	70
5.12	Tableau incomplet d'analyse de la variance	71
5.13	Tableau incomplet d'analyse de la variance	71
5.14	Tailles (en pouces) et poids (en livres) des étudiants	72
7.1	Population des cantons suisses en milliers d'habitants en 2001 et revenus des cantons	82
7.2	Coefficients du modèle sur le carré des résidus	84
7.3	Estimation de paramètre de la régression avec constante de y/x par $1/x$	87
7.4	Estimation de paramètre du modèle logarithmique	88
7.5	Population des cantons suisses en milliers d'habitants, revenus des cantons, résidus de la régression et carrés des résidus	90
7.6	Températures journalière du mois de janvier 2006	91
7.7	Consommation et prix du boeuf aux Etats-Unis	99
7.8	Données selon le temps et le carré du temps	103
7.9	Dépenses d'enseignement et revenus selon les cantons	104
7.10	Ventes et dépenses publicitaires	105
11.1	Table des quantiles d'une variable normale centrée réduite	130
11.2	Fonction de répartition de la loi normale centrée réduite	131
11.3	Quantiles de la loi normale centrée réduite	132
11.4	Table des quantiles d'une variable χ^2 à n degrés de liberté	133
11.5	Table des quantiles d'une variable de Student à n degrés de liberté	134
11.6	Table des quantiles d'ordre 0.95 d'une variable de Fisher à n_1 et n_2 degrés de liberté	135
11.7	Table des quantiles d'ordre 0.99 d'une variable de Fisher à n_1 et n_2 degrés de liberté	136
11.8	Valeur critique du test de Durbin-Watson au seuil de 5%	137
11.9	Quantiles du coefficient de corrélation de Pearson d'une variable aléatoire normale bivariée sous l'hypothèse que $\rho = 0$	138

Table des figures

2.1	Le nuage de points	17
2.2	La droite de régression	19
7.1	Nombre d'habitants et revenus total pour les 454 communes belges de moins de 20 000 habitants en 2004	81
7.2	Nombre d'habitants et revenus total pour les cantons suisses	82
7.3	Résidus de la régression en fonction des cantons classés par ordre croissant de population	83
7.4	Résidus de la régression des revenus par la population en fonction des communes belges classés par ordre croissant de population	83
7.5	Résidus \hat{u}_i de la régression sans constante du revenu par la population en fonction des communes classées par ordre croissant de population	87
7.6	Nuage de points : logarithme du revenu par le logarithme du nombre d'habitants	88
7.7	Résidus de la régression du modèle logarithmique classés par ordre croissant de population	89
7.8	Données suisses, carrés des résidus par nombre d'habitants	90
7.9	Températures journalière du mois de janvier 2006	91
7.10	Températures journalière vs températures du jour suivant	91
7.11	Bruit blanc : Suite de variables normales centrées réduites	93
7.12	Processus autorégressif avec $\rho = 0.9$	93
7.13	Processus autorégressif avec $\rho = 0.5$	93
7.14	Processus autorégressif avec $\rho = -0.5$	94
7.15	Processus autorégressif avec $\rho = -0.9$	94
7.16	Promenade aléatoire : Processus autorégressif avec $\rho = 1$	94
7.17	Processus non stationnaire $\rho = 1.01$	94
7.18	Règle de décision pour le test de Durbin-Watson	98
7.19	Résidus selon les années	99
7.20	Résidus d'une année vs résidus de l'année suivante	99

Table des matières

1	Éléments d'algèbre linéaire	2
1.1	Espace vectoriel	2
1.1.1	Vecteur	2
1.1.2	Multiplication par un scalaire et addition	2
1.1.3	Définition d'un espace vectoriel	3
1.1.4	Vecteurs linéairement indépendants	3
1.1.5	Sous-espace vectoriel	3
1.1.6	Système générateur d'un sous-espace vectoriel	3
1.1.7	Base d'un sous-espace vectoriel	3
1.1.8	Base canonique de \mathbb{R}^n	3
1.1.9	Dimension d'un sous-espace vectoriel	4
1.2	Espace euclidien	4
1.2.1	Produit scalaire	4
1.2.2	Norme	4
1.2.3	Distance entre deux vecteurs	4
1.2.4	Vecteurs orthogonaux	4
1.2.5	Orthogonal d'un sous-espace vectoriel	5
1.3	Application linéaire et matrices	5
1.3.1	Application linéaire	5
1.3.2	Matrice	5
1.3.3	Produit d'une matrice et d'un vecteur	5
1.3.4	Produit matriciel	6
1.3.5	Transposition	6
1.3.6	Matrices carrées, symétriques et diagonales	6
1.3.7	Rang d'une matrice	7
1.3.8	Trace d'une matrice	7
1.3.9	Matrices inversibles	7
1.3.10	Inversion par parties	7
1.3.11	Déterminant	8
1.3.12	Quelques propriétés	8
1.3.13	Matrices orthogonales	9
1.3.14	Valeurs propres et vecteurs propres	9
1.3.15	Formes et applications linéaires, formes quadratiques	9
1.3.16	Image et noyau d'une matrice	10
1.4	Projection et matrice idempotente	10
1.4.1	Projection	10
1.4.2	Projection orthogonale	10
1.4.3	Projection orthogonale dans l'image et le noyau d'une matrice	11
1.4.4	Matrice idempotente	11
1.4.5	Projecteurs obliques	12
1.4.6	Théorème des trois perpendiculaires	12
1.5	Dérivée par rapport à un vecteur	12
1.5.1	Gradient	12
1.5.2	Dérivation d'une forme linéaire	12
1.5.3	Dérivation d'une application linéaire	12

1.5.4	Dérivée d'une forme quadratique	13
2	Géométrie des moindres carrés	16
2.1	Série statistique bivariable	16
2.1.1	Représentation graphique de deux variables	16
2.1.2	Analyse des variables	16
2.1.3	Covariance	17
2.1.4	Corrélation	17
2.1.5	Droite de régression	18
2.1.6	Résidus et valeurs ajustées	19
2.1.7	Varianse de régression et variance résiduelle	19
2.2	La régression multivariée	21
2.2.1	Représentation matricielle des données	21
2.2.2	Principe des moindres carrés	21
2.2.3	Valeurs ajustées et résidus	21
2.2.4	Varianse de régression et variance résiduelle	22
2.2.5	Coefficient de détermination	22
2.3	Matrice de variance-covariance et matrice de corrélation	22
2.4	Corrélations partielles	23
2.5	Condition pour que la somme des résidus soit nulle	24
2.6	Décomposition en sommes de carrés	24
2.7	Régression avec les données centrées	25
2.8	Retour au cas bivarié	27
2.8.1	Méthode 1	27
2.8.2	Méthode 2	28
3	Rappel sur le calcul des probabilités, les variables aléatoires, et l'inférence statistique	32
3.1	Probabilités	32
3.1.1	Événement	32
3.1.2	Axiomatique des Probabilités	33
3.1.3	Probabilités conditionnelles et indépendance	33
3.1.4	Théorème des probabilités totales et théorème de Bayes	33
3.2	Variables aléatoires	34
3.2.1	Définition	34
3.2.2	Variables aléatoires discrètes	34
3.2.3	Variable aléatoire continue	36
3.2.4	Distribution bivariable	37
3.2.5	Indépendance de deux variables aléatoires	38
3.2.6	Propriétés des espérances et des variances	38
3.2.7	Autres variables aléatoires	39
3.2.8	Variable normale multivariée	39
3.3	Inférence statistique	40
3.3.1	Modélisation	40
3.3.2	Intervalle de confiance	41
3.3.3	Tests d'hypothèses	41
4	Le modèle linéaire général	43
4.1	Le modèle	43
4.1.1	Définition du modèle linéaire général	43
4.1.2	Hypothèses du modèle linéaire général	43
4.1.3	Données observées, et formulation matricielle	43
4.1.4	Autre présentation du modèle linéaire général	44
4.2	Estimation du modèle	44
4.2.1	Estimation par les moindres carrés (ordinaires)	44
4.2.2	Estimateurs du maximum de vraisemblance	47
4.2.3	Propriétés des estimateurs du maximum de vraisemblance	47
4.2.4	Distribution de probabilité des estimateurs	48

4.2.5	Synthèse des résultats	49
5	Inférence dans le modèle linéaire	51
5.1	Intervalle de confiance sur un coefficient de régression	51
5.2	Test d'un seul coefficient de régression	52
5.2.1	Construction du test	52
5.2.2	Modèle linéaire avec uniquement une constante	52
5.3	Tests de Wald sur les coefficients de régression	53
5.3.1	Test général d'une contrainte linéaire	53
5.3.2	Test global des coefficients de régression	54
5.3.3	Test de Fisher sur un coefficient de régression	56
5.4	Analyse de la variance à un facteur	57
5.4.1	Le problème	57
5.4.2	Méthode 1	57
5.4.3	Méthode 2	61
5.5	Prévision ponctuelle d'une valeur	63
5.5.1	Cas général	63
5.5.2	Cas bivarié	64
5.6	Exemple d'analyse de la variance à un facteur	65
5.6.1	Les données	65
5.6.2	Les résultats	65
6	Multicolinéarité et choix des variables	73
6.1	La multicolinéarité	73
6.2	Détection de la multicolinéarité	74
6.2.1	Le problème	74
6.2.2	Méthode de Klein	75
6.2.3	Test de Farrar et Glauber	75
6.2.4	Facteur d'inflation	75
6.2.5	Coefficient de Theil	75
6.2.6	Résolution du problème	75
6.3	Méthodes de choix de variables	76
6.3.1	Méthode Backward	76
6.3.2	Méthode Forward	76
6.3.3	Méthode Stepwise	76
6.3.4	Mise en garde	76
7	Méthode des moindres carrés généralisés	77
7.1	Les hypothèses du modèle linéaire général	77
7.2	La méthode des moindres carrés généralisés	77
7.3	Estimateur des moindres carrés généralisés et projection oblique	78
7.4	Retour au moindres carrés ordinaires	79
7.5	Méthode du maximum de vraisemblance	79
7.6	Intérêt des moindres carrés généralisés	80
7.7	Détection de l'hétéroscédasticité	80
7.7.1	Le problème	80
7.7.2	Graphique des résidus	81
7.7.3	Test de White	83
7.7.4	Test de Goldfeld-Quant	84
7.8	Estimation avec hétéroscédasticité	85
7.8.1	Si la variance est connue	85
7.8.2	Exemple de variance connue	86
7.8.3	Si la variance est inconnue	89
7.9	L'autocorrélation des résidus	91
7.10	Un exemple d'autocorrélation	91
7.10.1	La modélisation	92
7.10.2	Définition du processus autorégressif d'ordre un	92

7.10.3	Exemples de processus autorégressifs	93
7.10.4	Espérance et variance du processus autorégressif d'ordre 1	95
7.10.5	Processus sur un intervalle de temps fini	96
7.10.6	Le test de Durbin-Watson	98
7.11	Estimation avec des termes d'erreur autocorrélés	100
7.11.1	Le modèle et estimation par les MCG	100
7.11.2	Cas où ρ est inconnu	101
8	Variables instrumentales et équations simultanées	108
8.1	Erreurs sur la variable explicative	108
8.2	Variables instrumentales	108
8.3	Doubles moindres carrés	109
8.4	Cas où $q = p$	110
8.5	Application à l'économie	110
8.5.1	Un exemple : le modèle keynesien	110
8.5.2	Estimation par la méthode des moindres carrés indirects	111
8.5.3	Estimation par la méthode des variables instrumentales	112
8.6	Méthodes d'estimation	116
8.6.1	Moindres carrés indirects (MCI)	116
8.6.2	Doubles moindres carrés (2MC)	116
9	Modèles à choix discret	119
9.1	Modèles probit, logit et à probabilité linéaire	119
9.1.1	Le modèle	119
9.1.2	Choix de la fonction $F(\cdot)$	120
9.1.3	Remarques sur la fonction $F(\cdot)$	120
9.1.4	Estimation par les moindres carrés	121
9.1.5	Méthode du maximum de vraisemblance	121
9.1.6	Interprétation selon une variable latente	122
9.1.7	Évaluation de la qualité du modèle	122
9.2	Analyse discriminante	123
9.2.1	Le modèle	123
9.2.2	La règle bayésienne	123
9.2.3	Le modèle multinormal	123
10	Exercices récapitulatifs	125
11	Tables statistiques	130

Index

- analyse
 - de la variance à un facteur, 57
 - des variables, 16
 - discriminante, 123
- application linéaire, 5, 9
- autocorrélation des résidus, 91
- autocovariance, 96
- base
 - canonique, 3
 - d'un sous-espace vectoriel, 3
- bruit blanc gaussien, 93
- coefficient
 - d'autocorrélation, 92
 - de corrélation
 - multiple, 22
 - partielle, 23
 - de détermination, 19
 - régression multivariée, 22
 - de Theil, 75
- combinaison linéaire
 - de deux matrices, 5
 - de deux vecteurs, 2
- complémentaire d'un événement, 32
- corrélations partielles, 23
- correlation, 17
 - coefficient, 17
 - coefficient de détermination, 17
- covariance, 17
- dérivation d'une application
 - linéaire, 12
- dérivation d'une forme
 - linéaire, 12
 - quadratique, 13
- déterminant, 8
- dimension d'un sous-espace vectoriel, 4
- distance entre deux vecteurs, 4
- distribution
 - bivariée, 37
 - conditionnelle, 38
 - de probabilité, 34
 - de probabilité des estimateurs, 48
 - marginale, 37
 - normale multivariée, 39
- droite de régression, 18
- écart-type marginal, 17
- équations simultanées, 108
- erreur
 - de deuxième espèce, 41
 - de première espèce, 41
- espérance
 - d'une variable
 - binomiale, 35
 - de Poisson, 35
 - indicatrice, 34
 - d'une variable aléatoire continue, 36
- espace
 - euclidien, 4
 - norme, 4
 - vectoriel, 2
- estimateur
 - convergent, 48
 - des moindres carrés généralisés, 78
 - du maximum de vraisemblance, 47
 - efficace, 47
 - par les variables instrumentales, 109
 - projecteur oblique, 78
 - sans biais, 47
- estimation
 - avec des termes d'erreur, 100
 - par les moindres carrés (ordinaires), 44
- événements, 32
 - indépendants, 33
 - mutuellement exclusifs, 32
- expérience aléatoire, 32
- facteur d'inflation, 75
- fonction
 - de densité
 - conditionnelle, 37
 - d'une variable aléatoire continue, 36
 - marginale, 37
 - de répartition
 - d'une variable aléatoire continue, 36
 - jointe, 37
 - de score, 124
 - de vraisemblance, 47
- forme
 - linéaire, 9
 - quadratique, 9
- gradient, 12
- hétéroscédasticité

- détection, 80
- estimation avec, 85
- homoscédasticité, 43
- image d'une matrice, 10
- indépendance, 38
- inférence statistique, 32, 40
- intervalle de confiance, 41
- méthode
 - Backward, 76
 - de Klein, 75
 - des moindres carrés généralisés, 77
 - du maximum de vraisemblance, 79
 - Forward, 76
 - Stepwise, 76
- matric
 - rang, 7
- matrice, 5
 - carrée, 6
 - définie positive, 9
 - déterminant, 8
 - de plein rang, 44
 - de projection, 11
 - de variance-covariance, 22
 - des corrélations, 23
 - des variables explicatives, 57
 - diagonale, 6
 - idempotente, 10, 11, 78
 - trace, 12
 - identité, 7
 - image, 10
 - inverse généralisée, 73
 - de Moore-Penrose, 73
 - inversible, 7
 - irréversible, 7
 - noyau, 10
 - orthogonale, 9
 - produit d'une matrice et d'un vecteur, 6
 - propriétés, 8
 - pseudo-inverse, 73
 - semi-définie positive, 9
 - sous-espace, 10
 - symétrique, 6
 - trace, 7
 - transposition, 6
 - valeurs propres, 9
 - vecteurs propres, 9
- modélisation, 40, 92
- modèle
 - à choix discret, 119
 - de probabilité linéaire, 120
 - forme réduite, 110
 - linéaire général, 43
 - définition, 43
 - hypothèses, 43
 - logit, 120
 - multinormal, 123
 - probit, 120
- moindres carrés, 16
 - doubles, 109, 116
 - généralisés, 77
 - estimateur, 78
 - intérêt, 80
 - méthode, 77
 - indirects, 116
 - ordinaires
 - estimateur, 78
 - principe, 18, 21
- moyenne, 16
 - conditionnelle, 38
 - marginale, 17, 37
- multicolinéarité, 44, 73
- multiplication par un scalaire, 2
- norme, 4
- noyau d'une matrice, 10
- orthogonal d'un sous-espace vectoriel, 5
- paramètres marginaux, 17
- prévision ponctuelle d'une valeur, 63
- principe des moindres carrés, 18
 - régression multivariée, 21
- probabilité, 33
- processus autorégressif d'ordre un, 92
- produit
 - d'une matrice et d'un vecteur, 5
 - matriciel, 6
 - scalaire, 4
- projecteur oblique, 12
- projection, 4, 10
 - orthogonale, 10
 - dans l'image, 11
 - dans le noyau, 11
- puissance d'un test, 41
- régression, 18
 - données centrées, 25
 - droite de régression, 18
 - multivariée, 21
 - variance de régression, 19
- résidus, 19
 - graphique, 81
- règle bayésienne, 123
- rang d'une matrice, 7
- représentation
 - graphique de deux variables, 16
 - matricielle des données, 21
- série statistique bivariée, 16
- scalaire, 2
- somme des carrés
 - des résidus, 25
 - expliquée par la régression, 24

- inter-groupes, 60, 63
- intra-groupes, 60, 63
- totale des écarts à la moyenne, 24
- sous-espace vectoriel, 3
 - base, 3
 - base canonique de \mathbb{R} , 3
 - dimension, 4
 - orthogonal, 5
 - système générateur, 3
- statistique
 - d'Anderson, 124
 - exhaustive, 48
- système
 - complet d'événements, 33
 - générateur d'un sous-espace vectoriel, 3
- terme d'erreur, 43–45, 48, 49, 57, 67, 71, 77–79, 86, 87, 92, 100, 101, 106, 107
- test
 - d'hypothèses
 - composites, 41
 - simples, 41
 - d'un seul coefficient de régression, 51, 52
 - de Durbin-Watson, 98
 - de Farrar et Glauber, 75
 - de Fisher sur un coefficient de régression, 56
 - de Goldfeld-Quant, 84
 - de Wald sur les coefficients de régression, 53
 - de White, 83
 - global sur les coefficients de régression, 54
- théorème
 - de Bayès, 33
 - de diagonalisation, 9
 - de Gauss-Markov, 45
 - généralisé, 78
 - de Pythagore, 4
 - des probabilités totales, 33
 - des trois perpendiculaires, 12
- trace d'une matrice, 7
 - idempotente, 12
- transposition, 2
- valeurs
 - ajustées, 19
 - propres, 9
 - d'une matrice idempotente, 11
- variable, 32
 - aléatoire, 34
 - discrète, 34
 - indépendante, 38
 - bernoullienne, 34
 - binomiale, 35
 - de Fisher, 39
 - de Poisson, 35
 - de Student, 39
 - endogène, 110
 - exogène, 110
 - indicatrice, 34
 - instrumentale, 108
 - khi-carrée, 39
 - latente, 122
 - méthodes de choix, 76
 - normale, 37
 - multivariée, 39
 - uniforme, 36
- variance, 16
 - conditionnelle, 38
 - d'une variable
 - binomiale, 35
 - de Poisson, 35
 - indicatrice, 34
 - d'une variable aléatoire continue, 36
 - de régression, 19
 - régression multivariée, 22
 - marginale, 17, 20, 37
 - résiduelle, 20
 - régression multivariée, 22
- vecteur
 - colonne, 2
 - des résidus, 22
 - des valeurs ajustées, 21
 - ligne, 2
 - projection, 4
- vecteurs, 2
 - linéairement indépendants, 3
 - orthogonaux, 4
 - propres, 9