

# RÉSUMÉ DU COURS DE MODÈLES DE RÉGRESSION

Yves Tillé

10 janvier 2011



# Chapitre 1

## Régression bivariée

### 1.1 Série statistique bivariée

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.

#### 1.1.1 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 1.1** On mesure le poids  $Y$  et la taille  $X$  de 20 individus.

TABLE 1.1 – 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

#### 1.1.2 Analyse des variables

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

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

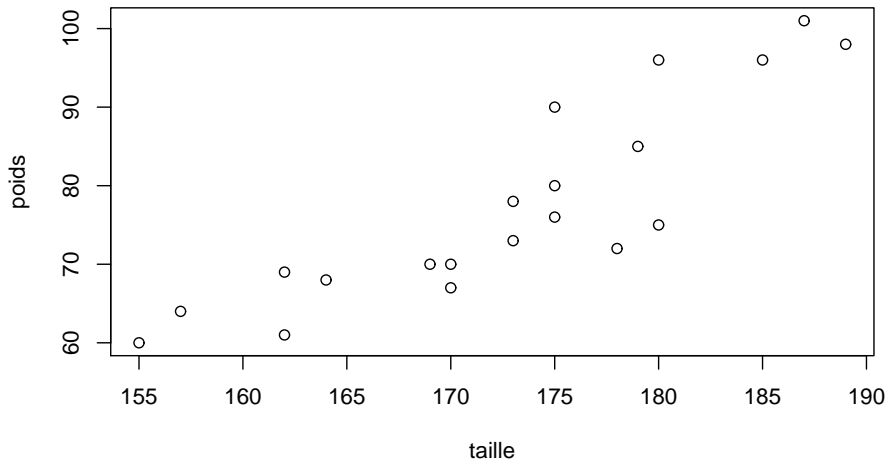


FIGURE 1.1 – Le nuage de points

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

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

### 1.1.3 Covariance

La *covariance* est définie

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

#### Remarque 1.1

- 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

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

### 1.1.4 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 corrélation peut être interprété géométriquement. Considérons les deux vecteurs centrés de  $\mathbb{R}^n$  :

$$\begin{aligned} \tilde{\mathbf{x}} &= (x_1 - \bar{x}, \dots, x_i - \bar{x}, \dots, x_n - \bar{x})' \\ \tilde{\mathbf{y}} &= (y_1 - \bar{y}, \dots, y_i - \bar{y}, \dots, y_n - \bar{y})'. \end{aligned}$$

Par définition, le cosinus de l'angle entre,  $\tilde{\mathbf{x}}$  et  $\tilde{\mathbf{y}}$  vaut

$$\cos(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \frac{\langle \tilde{\mathbf{x}}, \tilde{\mathbf{y}} \rangle}{\|\tilde{\mathbf{x}}\| \times \|\tilde{\mathbf{y}}\|},$$

où  $\langle \tilde{\mathbf{x}}, \tilde{\mathbf{y}} \rangle$  est le produit scalaire entre les vecteurs  $\tilde{\mathbf{x}}$  et  $\tilde{\mathbf{y}}$

$$\langle \tilde{\mathbf{x}}, \tilde{\mathbf{y}} \rangle = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

et  $\|\tilde{\mathbf{x}}\|$  (rep.  $\|\tilde{\mathbf{y}}\|$ ) est la norme de  $\tilde{\mathbf{x}}$  (resp.  $\tilde{\mathbf{y}}$ ) :

$$\|\tilde{\mathbf{x}}\| = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \text{ et } \|\tilde{\mathbf{y}}\| = \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Le coefficient de corrélation est donc égal au cosinus de l'angle entre les vecteurs  $\tilde{\mathbf{x}}$  et  $\tilde{\mathbf{y}}$ . Comme un cosinus est toujours compris dans  $[-1, 1]$ , on obtient que :

$$-1 \leq r_{xy} \leq 1. \quad (1.1)$$

**Remarque 1.2** Le coefficient de corrélation mesure la dépendance linéaire entre deux variables.

- Si le coefficient de corrélation est positif, les points sont alignés le long d'une droite croissante.
- Si le coefficient de corrélation est négatif, les points sont alignés le long d'une droite décroissante.
- Si le coefficient de corrélation est nul ou proche de zéro, il n'y a pas de dépendance linéaire. On peut cependant avoir une dépendance non-linéaire avec un coefficient de corrélation nul.

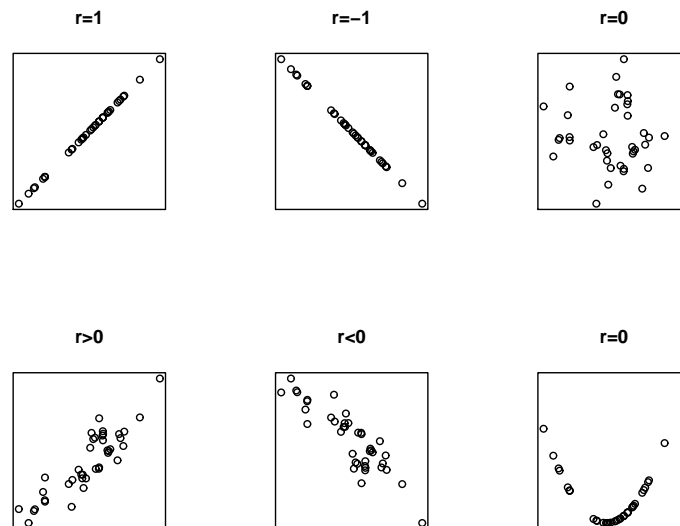


FIGURE 1.2 – Exemples de nuages de points et coefficients de corrélation

**Remarque 1.3** La présence d'une corrélation n'implique pas forcément une relation de causalité entre les deux variables.

Le *coefficient de détermination* (appelé aussi R-deux ou R-carré) est le carré du coefficient de corrélation.

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

De l'inégalité (1.1), on obtient directement que

$$0 \leq r_{xy}^2 \leq 1.$$

### 1.1.5 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 problème consiste à identifier une droite qui ajuste bien le nuage de points. Si les coefficients  $a$  et  $b$  étaient connus, on pourrait calculer les résidus de la régression définis par :

$$e_i = y_i - a - bx_i.$$

Le résidu  $e_i$  est l'erreur que l'on commet (voir Figure 1.3) en utilisant la droite de régression pour prédire  $y_i$  à partir de  $x_i$ . Les résidus peuvent être positifs ou négatifs.

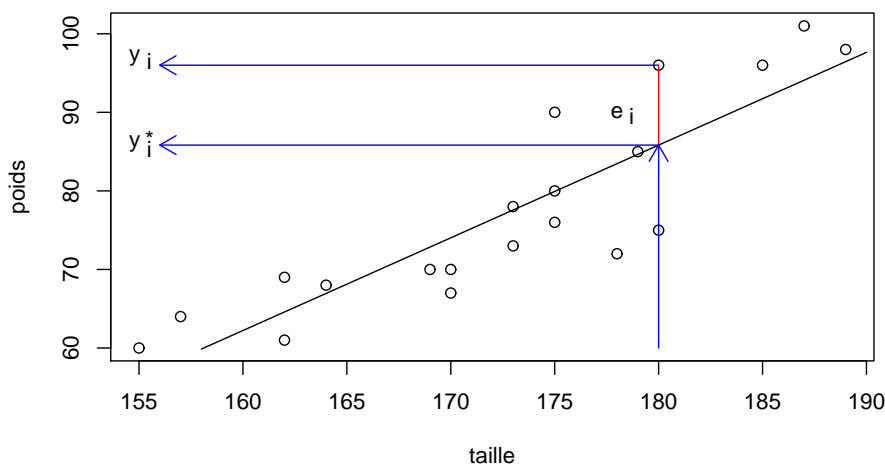


FIGURE 1.3 – Le nuage de points, le résidu

Pour déterminer la valeur des coefficients  $a$  et  $b$  on utilise le principe des *moindres carrés* qui consiste à chercher la droite qui minimise la somme des carrés des résidus :

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

**Théorème 1.1** Les coefficients  $a$  et  $b$  qui minimisent le critère des moindres carrés sont donnés par :

$$b = \frac{s_{xy}}{s_x^2} \quad \text{et} \quad a = \bar{y} - b\bar{x}.$$

**Démonstration** Le minimum  $M(a, b)$  en  $a, b$  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. En divisant les deux équations par  $-2n$ , on obtient :

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

ou encore

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n y_i - \frac{1}{n} \sum_{i=1}^n a - b \frac{1}{n} \sum_{i=1}^n x_i = 0 \\ \frac{1}{n} \sum_{i=1}^n y_i x_i - \frac{1}{n} \sum_{i=1}^n a x_i - \frac{1}{n} \sum_{i=1}^n b x_i^2 = 0, \end{cases}$$

ce qui s'écrit aussi

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

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

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

En remplaçant  $a$  par  $\bar{y} - b\bar{x}$  dans la seconde équation, on a

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

ce qui donne

$$s_{xy} - b s_x^2 = 0.$$

Donc

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

On a donc identifié les deux paramètres

$$\begin{cases} b = \frac{s_{xy}}{s_x^2} \text{ (la pente)} \\ a = \bar{y} - b\bar{x} = \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x} \text{ (la constante)}. \end{cases}$$

Pour vérifier qu'il s'agit bien d'un minimum, on doit montrer que la matrice hessienne des dérivées secondes est définie positive. Cette matrice vaut :

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 M(a,b)}{\partial a^2} & \frac{\partial^2 M(a,b)}{\partial a \partial b} \\ \frac{\partial^2 M(a,b)}{\partial a \partial b} & \frac{\partial^2 M(a,b)}{\partial b^2} \end{pmatrix}$$

On a

$$\frac{\partial^2 M(a,b)}{\partial a^2} = 2n, \quad \frac{\partial^2 M(a,b)}{\partial b^2} = 2 \sum_{i=1}^n x_i^2, \quad \text{et} \quad \frac{\partial^2 M(a,b)}{\partial a \partial b} = 2 \sum_{i=1}^n x_i = 2n\bar{x}.$$

La matrice hessienne vaut donc

$$\mathbf{H} = 2 \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}.$$

et peut s'écrire  $\mathbf{H} = 2\mathbf{X}'\mathbf{X}$ , où

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

Pour tout vecteur  $\mathbf{u} \in \mathbb{R}^2$ , les formes quadratiques  $\mathbf{u}'\mathbf{H}\mathbf{u}$  peuvent s'écrire  $2\mathbf{v}'\mathbf{v}$  en posant  $\mathbf{v} = \mathbf{X}\mathbf{u}$ . Comme  $\mathbf{v}'\mathbf{v}$  est toujours positif, la matrice  $\mathbf{H}$  est définie positive.  $\square$

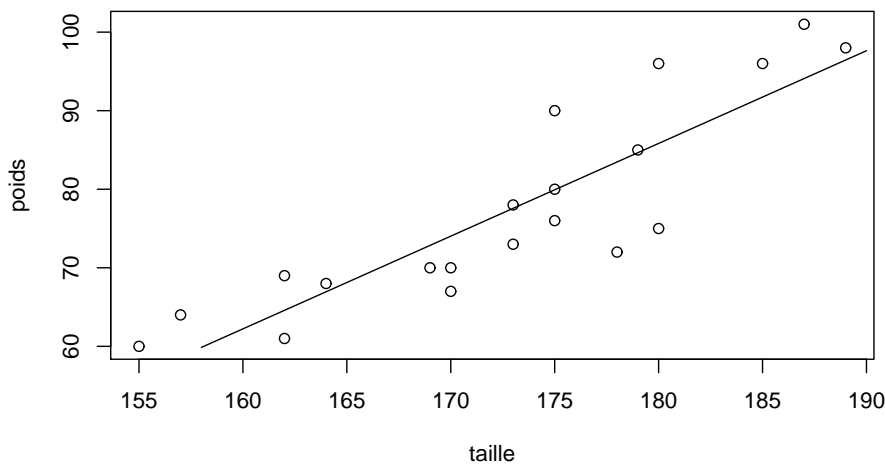
La droite de régression /indexdroite!de régression est donc

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

ce qui peut s'écrire aussi

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

FIGURE 1.4 – La droite de régression



**Remarque 1.4** La droite de régression de  $y$  en  $x$  n'est pas la même que la droite de régression de  $x$  en  $y$ .

### 1.1.6 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 1.5** La moyenne des valeurs ajustées est égale à la moyenne des valeurs observées  $\bar{y}$ . En effet,

$$\frac{1}{n} \sum_{i=1}^n y_i^* = \frac{1}{n} \sum_{i=1}^n (a + bx_i) = a + b \frac{1}{n} \sum_{i=1}^n x_i = a + b\bar{x}.$$

Or,  $\bar{y} = a + b\bar{x}$ , car le point  $(\bar{x}, \bar{y})$  appartient à la droite de régression.

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.



**Propriété 1.1** *La moyenne des résidus est nulle.*

**Démonstration**

$$\frac{1}{n} \sum_{i=1}^n e_i = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^*) = \bar{y} - \bar{y} = 0.$$

**Propriété 1.2**

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

**Démonstration**

$$\begin{aligned} \sum_{i=1}^n x_i e_i &= \sum_{i=1}^n x_i (y_i - y_i^*) \\ &= \sum_{i=1}^n x_i (y_i - a - b x_i) \\ &= \sum_{i=1}^n x_i (y_i - \bar{y} - b(x_i - \bar{x})) \quad \text{car } a = \bar{y} - b\bar{x} \\ &= \sum_{i=1}^n x_i (y_i - \bar{y}) - b \sum_{i=1}^n x_i (x_i - \bar{x}) \end{aligned} \tag{1.2}$$

Or

$$\begin{aligned} \sum_{i=1}^n x_i (x_i - \bar{x}) &= \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \\ &= n \left( \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x} \underbrace{\frac{1}{n} \sum_{i=1}^n x_i}_{\bar{x}} \right) \\ &= n \left( \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right) \\ &= n s_x^2. \end{aligned} \tag{1.3}$$

De plus,

$$\begin{aligned} \sum_{i=1}^n x_i (y_i - \bar{y}) &= \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i \\ &= n \left( \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{y} \underbrace{\frac{1}{n} \sum_{i=1}^n x_i}_{\bar{x}} \right) \\ &= n \left( \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \right) \\ &= n s_{xy}. \end{aligned} \tag{1.4}$$

En considérant (1.3) et (1.4), l'Expression (1.2) devient :

$$\sum_{i=1}^n x_i e_i = \sum_{i=1}^n x_i (y_i - \bar{y}) - b \sum_{i=1}^n x_i (x_i - \bar{x}) = n s_{xy} - b n s_x^2 = 0,$$

car  $b = s_{xy} / s_x^2$ . □

### Propriété 1.3

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

**Démonstration** Cette propriété découle des Propriétés (1.1) et (1.2). En effet,

$$\sum_{i=1}^n y_i^* e_i = \sum_{i=1}^n (a + bx_i) e_i = a \underbrace{\sum_{i=1}^n e_i}_{=0} + b \underbrace{\sum_{i=1}^n x_i e_i}_{=0} = 0.$$

□

La régression peut s'interpréter géométriquement. Considérons les deux vecteurs de  $\mathbb{R}^n$  :

$$\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)' \text{ et } \mathbf{y} = (y_1, \dots, y_i, \dots, y_n)'$$

La régression permet de déterminer les deux vecteurs des valeurs ajustées et des résidus :

$$\mathbf{y}^* = (y_1^*, \dots, y_i^*, \dots, y_n^*)' \text{ et } \mathbf{e} = (e_1, \dots, e_i, \dots, e_n)'$$

De plus considérons le vecteur de  $\mathbb{R}^n$  composé de uns :

$$\mathbf{1}_n = (1, \dots, 1, \dots, 1)'$$

La régression permet la décomposition du vecteur  $\mathbf{y}$

$$\mathbf{y} = \mathbf{y}^* + \mathbf{e},$$

où

$$\mathbf{y}^* = a\mathbf{1}_n + b\mathbf{x}.$$

On a donc les relations suivantes

- $\mathbf{y}^*$  et  $\mathbf{e}$  sont orthogonaux, i.e.  $\langle \mathbf{y}^*, \mathbf{e} \rangle = 0$ .
- $\mathbf{x}$  et  $\mathbf{e}$  sont orthogonaux, i.e.  $\langle \mathbf{x}, \mathbf{e} \rangle = 0$ .
- $\mathbf{1}_n$  et  $\mathbf{e}$  sont orthogonaux, i.e.  $\langle \mathbf{1}_n, \mathbf{e} \rangle = 0$ .

La régression décompose le vecteur  $\mathbf{y}$  en deux vecteurs orthogonaux  $\mathbf{y}^*$  et  $\mathbf{e}$ . Le vecteur  $\mathbf{e}$  est par ailleurs orthogonal au vecteur contenant la variable explicative et au vecteur  $\mathbf{1}_n$ .

### 1.1.7 Sommes de carrés et variances

**Définition 1.1** On appelle somme des carrés totale la quantité

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

La variance marginale peut alors être définie par

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

**Définition 1.2** On appelle somme des carrés de la régression la quantité

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

**Définition 1.3** La variance de régression est la variance des valeurs ajustées.

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

**Définition 1.4** On appelle somme des carrés des résidus (ou résiduelle) la quantité

$$SC_{RES} = \sum_{i=1}^n e_i^2.$$

**Définition 1.5** La variance résiduelle est la variance des résidus.

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

Note : Il n'est pas nécessaire de centrer les résidus sur leurs moyennes pour calculer la variance, car la moyenne des résidus est nulle.

**Théorème 1.2**

$$SC_{TOT} = SC_{REGR} + SC_{RES}.$$

**Démonstration**

$$\begin{aligned} SC_{TOT} &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - y_i^* + y_i^* - \bar{y})^2 = \sum_{i=1}^n (y_i - y_i^*)^2 + \sum_{i=1}^n (y_i^* - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - y_i^*)(y_i^* - \bar{y}) \\ &= SC_{RES} + SC_{REGR} + 2 \sum_{i=1}^n (y_i - y_i^*)(y_i^* - \bar{y}). \end{aligned}$$

Le troisième terme est nul. En effet,

$$\sum_{i=1}^n (y_i - y_i^*)(y_i^* - \bar{y}) = \sum_{i=1}^n (y_i - a - bx_i)(a + bx_i - \bar{y})$$

En remplaçant  $a$  par  $\bar{y} - b\bar{x}$ , on obtient

$$\begin{aligned} \sum_{i=1}^n (y_i - y_i^*)(y_i^* - \bar{y}) &= \sum_{i=1}^n [y_i - \bar{y} - b(x_i - \bar{x})] b(x_i - \bar{x}) = \sum_{i=1}^n [(y_i - \bar{y}) - b(x_i - \bar{x})] b(x_i - \bar{x}) \\ &= b \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) - b^2 \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = bn s_{xy} - b^2 n s_x^2 = \frac{s_{xy}}{s_x^2} n s_{xy} - \frac{s_{xy}^2}{s_x^4} n s_x^2 = 0. \end{aligned}$$

□

### 1.1.8 Décomposition de la variance

**Théorème 1.3** 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 la variance des résidus.

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

**Théorème 1.4** 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 1.5** La variance marginale est la somme de la variance de régression et de la variance résiduelle,

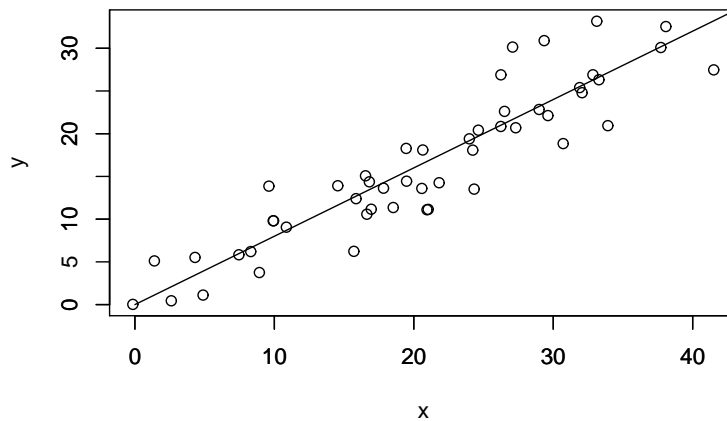
$$s_y^2 = s_{y^*}^2 + s_e^2.$$

La démonstration découle directement des deux théorèmes précédents.

## 1.2 La régression par l'origine (sans constante)

Dans certaines applications, on peut raisonnablement penser que la droite de régression passe par l'origine comme dans la Figure 1.5.

FIGURE 1.5 – Droite passant par l'origine



La droite de régression s'écrit alors simplement

$$y = bx.$$

Le critère à minimiser devient

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

En annulant la dérivée de  $M(b)$  par rapport à  $b$ , nous obtenons,

$$\frac{dM(b)}{db} = -2 \sum_{i=1}^n x_i(y_i - bx_i) = -2 \sum_{i=1}^n x_i y_i + 2b \sum_{i=1}^n x_i^2 = 0.$$

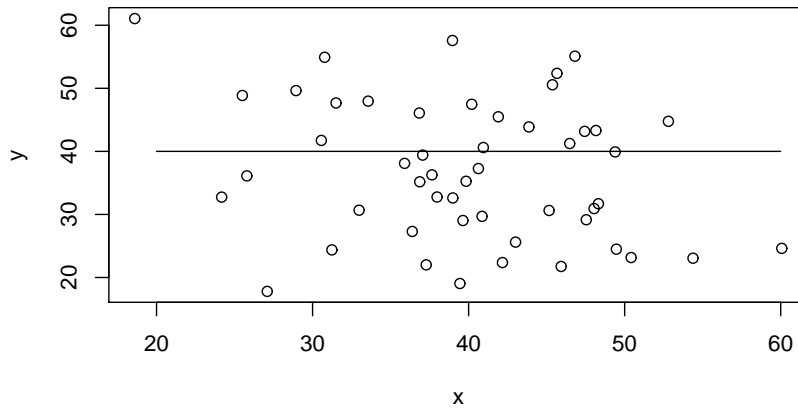
On obtient :

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

### 1.3 La régression avec une pente nulle

Un autre cas particulier est la régression avec une pente nulle. Dans ce cas, les points sont alignés le long d'une droite horizontale comme dans la Figure 1.6.

FIGURE 1.6 – Régression avec une pente nulle



La droite de régression s'écrit alors simplement

$$y = a.$$

Le critère à minimiser devient

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

En annulant la dérivée de  $M(a)$  par rapport à  $a$ , on obtient :

$$\frac{dM(a)}{da} = - \sum_{i=1}^n 2(y_i - a) = -2 \sum_{i=1}^n y_i + 2 \sum_{i=1}^n a = -2n\bar{y} + 2na = 0.$$

Donc,

$$a = \bar{y},$$

ce qui signifie que la variable  $x$  n'a aucune corrélation avec la variable  $y$  et ne permet donc pas de prédire cette variable.

# Chapitre 2

## Éléments d'algèbre linéaire

### 2.1 Espace vectoriel

#### 2.1.1 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 2.1** 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 2.2** Si  $\mathbf{a} = (3 \ 0)$ , la transposée de  $\mathbf{a}$  est

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

#### 2.1.2 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  $c_1, c_2 \in \mathbb{R}$ .

### 2.1.3 Vecteurs linéairement indépendants

**Définition 2.1** 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  $a_1 = a_2 = \dots = a_J = 0$ .

### 2.1.4 Sous-espace vectoriel

**Définition 2.2** Un sous-ensemble non-vidé  $V$  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  $a \in \mathbb{R}$ .

### 2.1.5 Système générateur d'un sous-espace vectoriel

**Définition 2.3** 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  $\mathbf{v} = a_1\mathbf{u}_1 + \dots + a_p\mathbf{u}_p$ .

### 2.1.6 Base d'un sous-espace vectoriel

**Définition 2.4** 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  $\mathbf{u}_1, \dots, \mathbf{u}_p$ .

### 2.1.7 Base canonique de $\mathbb{R}^n$

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

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

### 2.1.8 Dimension d'un sous-espace vectoriel

**Définition 2.5** 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  $V$ .

## 2.2 Espace euclidien

### 2.2.1 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 2.6** Un espace euclidien est un espace vectoriel muni d'un produit scalaire.

## 2.2.2 Norme

**Définition 2.7** 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.

## 2.2.3 Distance entre deux vecteurs

**Définition 2.8** 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 2.9** La projection d'un vecteur  $\mathbf{u}$  sur un vecteur  $\mathbf{v}$  est définie par

$$p_v(\mathbf{u}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v}}{\|\mathbf{v}\|^2}. \quad (2.1)$$

## 2.2.4 Vecteurs orthogonaux

**Définition 2.10** 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 2.1** (de Pythagore) Si  $\mathbf{u}$  et  $\mathbf{v}$  sont orthogonaux, alors

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

## 2.2.5 Orthogonal d'un sous-espace vectoriel

**Définition 2.11** Un vecteur  $\mathbf{u}$  est orthogonal à un sous-espace vectoriel  $V$  si et seulement s'il est orthogonal à tous les vecteurs de  $V$ , on note alors

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

**Définition 2.12** Les sous-espaces  $V$  et  $W$  sont dits orthogonaux, si tout vecteur de  $V$  est orthogonal à tout vecteur de  $W$ .

**Définition 2.13** L'ensemble de tous les vecteurs orthogonaux à  $V$  est appelé l'orthogonal de  $V$  et est noté  $V^\perp$ .

**Propriété 2.1**

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



## 2.3 Application linéaire et matrices

### 2.3.1 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  $a \in \mathbb{R}$

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

### 2.3.2 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.

### 2.3.3 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  $\mathbf{A}\mathbf{u}$  est donné par

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

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

### 2.3.4 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  $(I \times K)$ .

### 2.3.5 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 2.1** Soient  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  de dimensions respectives  $(I \times J), (J \times K)$  et  $(K \times L)$ , alors la transposée de  $\mathbf{ABC}$  vaut

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

### 2.3.6 Matrices carrées, symétriques et diagonales

**Définition 2.14** 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 2.15** Une matrice est dite symétrique si elle est égale à sa transposée.

Une matrice symétrique est donc toujours carrée.

**Définition 2.16** 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 2.17** 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  $3 \times 3$ .

### 2.3.7 Rang d'une matrice

**Définition 2.18** Le rang d'une matrice est le nombre maximum de lignes (ou de colonnes) linéairement indépendantes.

**Propriété 2.2** Le rang est toujours inférieur ou égal au minimum du nombre de lignes et du nombre de colonnes de la matrice.

**Définition 2.19** 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 2.3** Le rang d'un produit de matrices est inférieur ou égal au rang de chaque matrice.

### 2.3.8 Trace d'une matrice

**Définition 2.20** La trace d'une matrice carrée est la somme des éléments de sa diagonale.

- Propriété 2.4**
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  $\text{trace}(\mathbf{AB}) \neq \text{trace}(\mathbf{A})\text{trace}(\mathbf{B})$ .

### 2.3.9 Matrices inversibles

**Définition 2.21** 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 2.5** Si une matrice carrée est de plein rang, alors elle est inversible.

### 2.3.10 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  $\mathbf{F}^{-1}\mathbf{F}$ .

### 2.3.11 Déterminant

**Définition 2.22** 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 ligne  $i$  et la colonne  $j$  de la matrice  $\mathbf{A}$ .

**Exemple 2.3** Soit  $\mathbf{A}$  une matrice  $(2 \times 2)$ ,

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

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 2.4** Soit une matrice  $\mathbf{A}$  de dimension  $(3 \times 3)$ , 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 2.6

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  $J \times J$ ),

### 2.3.12 Quelques propriétés

**Propriété 2.7** 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.  $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$ .

### 2.3.13 Matrices orthogonales

**Définition 2.23** Une matrice  $\mathbf{\Gamma}$  est dite orthogonale si son inverse est égale à sa transposée :

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

### 2.3.14 Valeurs propres et vecteurs propres

**Définition 2.24** Soit  $\mathbf{A}$  une matrice  $J \times J$ .  $\lambda_i$  est une valeur propre de  $\mathbf{A}$  si  $\lambda_i$  est une solution de l'équation

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

### Propriété 2.8

- 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 2.25** Le vecteur  $\mathbf{u}_i \neq 0$  est un vecteur propre de  $\mathbf{A}$  associé à la valeur propre  $\lambda_i$  si

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

**Propriété 2.9** Si  $\mathbf{A}$  est une matrice  $J \times J$  réelle symétrique, il existe  $J$  vecteurs propres normés et orthogonaux.

**Théorème 2.2** (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  $\mathbf{\Gamma}$  dont les colonnes sont les  $J$  vecteurs propres de  $\mathbf{A}$ , et la matrice diagonale  $\mathbf{\Lambda}$  ayant sur sa diagonale principale les  $J$  valeurs propres. Alors

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

### 2.3.15 Formes et applications linéaires, formes quadratiques

**Définition 2.26** 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 2.27** Une matrice  $\mathbf{A}$  de dimension ( $I \times I$ ) est dite définie positive si

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

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

**Définition 2.28** 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 2.10** 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 2.11** 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 2.12** Une matrice définie positive est toujours inversible.

### 2.3.16 Image et noyau d'une matrice

**Définition 2.29** 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 2.30** 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 2.2** Le sous-espace  $\text{Im}(\mathbf{B})$  est l'orthogonal de  $\text{Ker}(\mathbf{B}')$ .

**Propriété 2.13** 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  $\mathbf{X}$ .

## 2.4 Projection et matrice idempotente

### 2.4.1 Projection

L'opération de projection se déduit du théorème suivant :

**Théorème 2.3** 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  $V^\perp$ .

### 2.4.2 Projection orthogonale

**Définition 2.31** 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 ( $\mathbf{u}^* \in V$ ).

### 2.4.3 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 2.3** 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 (2.9).

### 2.4.4 Matrice idempotente

**Définition 2.32** Une matrice  $\mathbf{P}$  est dite idempotente si  $\mathbf{P}\mathbf{P} = \mathbf{P}$ .

Une matrice de projection est idempotente.

**Remarque 2.4** 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 2.4** *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

$$\mathbf{P}\mathbf{u} = \lambda\mathbf{u}, \tag{2.2}$$

où  $\lambda$  est la valeur propre associée à  $\mathbf{u}$ . En multipliant (2.2) 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  $\lambda$  vaut 0 ou 1. □

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

**Remarque 2.5** 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  $\mathbf{P}_X$ .

## 2.4.5 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  $\mathbf{X}$ .

## 2.4.6 Théorème des trois perpendiculaires

**Théorème 2.5** *Soit  $V$  et  $W$  deux sous-espaces vectoriels tels que  $V \subset W$ , alors*

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

## 2.5 Dérivée par rapport à un vecteur

### 2.5.1 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 :

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

### 2.5.2 Dérivation d'une forme linéaire

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

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

### 2.5.3 Dérivation 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

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

Donc,

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

### 2.5.4 Dérivée d'une forme quadratique

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

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

Donc,

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



et

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

Si la matrice  $\mathbf{A}$  est symétrique, on a  $\mathbf{A} = \mathbf{A}'$  et donc

$$\frac{\partial \mathbf{x}' \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A} \mathbf{x}.$$

# Chapitre 3

## La régression multivariée

### 3.1 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.

### 3.2 Principe des moindres carrés

La régression de  $\mathbf{y}$  en  $\mathbf{X}$  consiste à chercher un vecteur de coefficients de régression :

$$\mathbf{b} = (b_1, \dots, b_j, \dots, b_p)$$

qui permet d'ajuster au mieux les  $y_i$  par les  $x_{ij}$  au sens où la somme des carrés des résidus est minimisée. La régression s'écrit :

$$y_i = b_1 + b_2 x_{i2} + \cdots + b_j x_{ij} + \cdots + b_p x_{ip} + e_i = \sum_{j=1}^p b_j x_{ij} + e_i, i = 1, \dots, n.$$

Cette équation peut s'écrire de manière matricielle :

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

La somme des carrés des résidus est le critère à minimiser

$$\begin{aligned} Q(b_1, \dots, b_p) &= \sum_{i=1}^n e_i^2 \\ &= \sum_{i=1}^n [y_i - (b_1 + b_2 x_{i2} + \dots + b_j x_{ij} + \dots + b_p x_{ip})]^2 \\ &= \sum_{i=1}^n \left( y_i - \sum_{j=1}^p b_j x_{ij} \right)^2. \end{aligned}$$

Il est cependant plus facile d'utiliser une écriture matricielle. Comme on peut écrire

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b} = \begin{pmatrix} y_1 - \sum_{j=1}^p b_j x_{1j} \\ \vdots \\ y_i - \sum_{j=1}^p b_j x_{ij} \\ \vdots \\ y_n - \sum_{j=1}^p b_j x_{nj} \end{pmatrix},$$

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

$$Q(b_1, \dots, b_p) = Q(\mathbf{b}) = \sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2.$$

Remarquons que  $Q(\mathbf{b})$  peut également s'écrire

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

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

**Théorème 3.1** *Le vecteur de coefficient de régression au sens des moindres carrés est donné par*

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

*sous réserve que la matrice  $\mathbf{X}'\mathbf{X}$  soit inversible.*

### Démonstration

Calculons le vecteur des dérivées partielles  $Q(\mathbf{b})$  par rapport à  $\mathbf{b}$ ,

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

En annulant le vecteur des dérivées partielles, on obtient un système de  $p$  équations à  $p$  inconnues

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

ce qui donne :

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

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

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

Pour montrer que l'on a bien obtenu un minimum, on calcule la matrice hessienne des dérivées secondes

$$\mathbf{H} = \frac{\partial^2 Q(\mathbf{b})}{\partial \mathbf{b} \partial \mathbf{b}'} = 2\mathbf{X}'\mathbf{X}.$$

Cette matrice est définie positive, on a donc bien un minimum. □

### 3.3 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  $\mathbf{X}$ ).

$$\mathbf{e} = \mathbf{P}_X^\perp \mathbf{y}, \tag{3.1}$$

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 3.1

- $\mathbf{y} = \mathbf{y}^* + \mathbf{e}$ ,
- $\mathbf{y}^*$  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  $\mathbf{e}'\mathbf{X} = \mathbf{0}$ .

### 3.4 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 variance 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 des résidus :

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

### 3.5 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.

### 3.6 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

$$\Sigma = \begin{pmatrix} s_2^2 & \cdots & s_{2j} & \cdots & s_{2p} \\ \vdots & & \vdots & & \vdots \\ s_{j2} & \cdots & s_j^2 & \cdots & s_{jp} \\ \vdots & & \vdots & & \vdots \\ s_{p2} & \cdots & s_{pj} & \cdots & s_p^2 \end{pmatrix}, \quad (3.2)$$

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  $\mathbf{X}$ .

Considérons la matrice des valeurs centrées pour les  $p-1$  dernières variables :

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

La matrice de covariance peut alors s'écrire comme un simple produit matriciel :

$$\Sigma = \frac{\mathbf{X}_c' \mathbf{X}_c}{n}.$$

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

$$\mathbf{R} = \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} = \mathbf{S}^{-1} \frac{\mathbf{X}' \mathbf{X}_c}{n} \mathbf{S}^{-1}.$$

## 3.7 Corrélations partielles

### 3.7.1 Cas général

Les corrélations partielles sont des outils importants mettant en jeu deux variables  $y$  et  $z$  et un groupe de variables explicatives contenues dans la matrice  $\mathbf{X}$ . L'objectif est de mesurer la corrélation entre les variables  $y$  et  $z$  auxquelles on a enlevé la partie explicable par les variables de  $\mathbf{X}$ . Pour ce faire, on calcule le vecteur de coefficients  $\mathbf{b}_{y|X}$  de la régression de  $y$  en  $\mathbf{X}$ , et le vecteur de coefficients  $\mathbf{b}_{z|X}$  de la régression de  $z$  en  $\mathbf{X}$ . On peut ensuite calculer :

- le vecteur des résidus de la régression de  $y$  en  $\mathbf{X}$

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

- le vecteur des résidus de la régression de  $z$  en  $\mathbf{X}$

$$\mathbf{e}_{z|X} = \mathbf{z} - \mathbf{X} \mathbf{b}_{z|X} = \mathbf{P}_X^\perp \mathbf{z},$$

où  $\mathbf{P}_X^\perp = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  est le projecteur sur l'orthogonal du sous-espace engendré par les colonnes de la matrice  $\mathbf{X}$ . Le coefficient de corrélation partielle est le coefficient de corrélation entre les deux vecteurs de résidus  $\mathbf{e}_{y|X}$  et  $\mathbf{e}_{z|X}$ .

Comme la première colonne de la matrice  $\mathbf{X}$  est composée de uns, la somme des résidus est nulle, la covariance des résidus s'écrit :

$$\text{cov}(\mathbf{e}_{y|X}, \mathbf{e}_{z|X}) = \frac{1}{n} \sum_{i=1}^n e_{iy|X} e_{iz|X} = \frac{1}{n} \mathbf{e}'_{y|X} \mathbf{e}_{z|X} = \frac{1}{n} \mathbf{y}' \mathbf{P}_X^\perp \mathbf{P}_X^\perp \mathbf{z} = \frac{1}{n} \mathbf{y}' \mathbf{P}_X^\perp \mathbf{z}$$

car  $\mathbf{P}_X^\perp$  est idempotente. Les variances de ces deux résidus s'écrivent

$$\text{var}(\mathbf{e}_{y|X}) = \frac{1}{n} \sum_{i=1}^n e_{iy|X}^2 = \frac{1}{n} \mathbf{e}'_{y|X} \mathbf{e}_{y|X} = \frac{1}{n} \mathbf{y}' \mathbf{P}_X^\perp \mathbf{P}_X^\perp \mathbf{y} = \frac{1}{n} \mathbf{y}' \mathbf{P}_X^\perp \mathbf{y},$$

et

$$\text{var}(\mathbf{e}_{z|X}) = \frac{1}{n} \sum_{i=1}^n e_{iz|X}^2 = \frac{1}{n} \mathbf{e}'_{z|X} \mathbf{e}_{z|X} = \frac{1}{n} \mathbf{z}' \mathbf{P}_X^\perp \mathbf{P}_X^\perp \mathbf{z} = \frac{1}{n} \mathbf{z}' \mathbf{P}_X^\perp \mathbf{z}.$$

Le coefficient de corrélation partielle s'écrit :

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

### 3.7.2 Cas où une seule variable $x$ est utilisée

Si une seule variable indépendante est utilisée, c'est-à-dire si la matrice  $\mathbf{X}$  est composée d'une ligne et d'une constante :

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

alors on obtient par le Théorème 1.4, page 11, que

$$\frac{1}{n} \sum_{i=1}^n e_{iy|X}^2 = (1 - r_{xy}^2) s_y^2$$

et

$$\frac{1}{n} \sum_{i=1}^n e_{iz|X}^2 = (1 - r_{xz}^2) s_z^2$$

où  $r_{xy}$  est le coefficient de corrélation entre  $x$  et  $y$  et  $r_{xz}$  est le coefficient de corrélation entre  $x$  et  $z$ .

De plus, en notant  $b_{y|x} = s_{xy}/s_x^2$  la pente de la droite de régression de  $y$  en  $x$  et  $b_{z|x} = s_{xz}/s_x^2$  la pente de la droite de régression de  $z$  en  $x$ , on a

$$\begin{aligned} & \text{cov}(\mathbf{e}_{y|X}, \mathbf{e}_{z|X}) \\ &= \frac{1}{n} \sum_{i=1}^n e_{iy|X} e_{iz|X} \\ &= \frac{1}{n} \sum_{i=1}^n [(y_i - \bar{y}) - b_{y|x}(x_i - \bar{x})] [(z_i - \bar{z}) - b_{z|x}(x_i - \bar{x})] \\ &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z}) - \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}) b_{z|x}(x_i - \bar{x}) \\ &\quad - \frac{1}{n} \sum_{i=1}^n b_{y|x}(x_i - \bar{x})(z_i - \bar{z}) + \frac{1}{n} \sum_{i=1}^n b_{y|x}(x_i - \bar{x}) b_{z|x}(x_i - \bar{x}) \\ &= s_{yz} - s_{xy} b_{z|x} - s_{xz} b_{y|x} + s_x^2 b_{y|x} b_{z|x} \\ &= s_y s_z (r_{yz} - r_{xy} r_{xz} - r_{xz} r_{xy} + r_{xy} r_{xz}) \\ &= s_y s_z (r_{yz} - r_{xy} r_{xz}). \end{aligned}$$

Le coefficient de corrélation partielle vaut donc :

$$r_{yz|x} = \frac{\text{cov}(\mathbf{e}_{y|X}, \mathbf{e}_{z|X})}{\sqrt{\text{var}(\mathbf{e}_{y|X}) \text{var}(\mathbf{e}_{z|X})}} = \frac{r_{yz} - r_{xy} r_{xz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{xz}^2)}}.$$

Dans ce cas, les corrélations partielles peuvent être calculées au moyen des corrélations simples.

### 3.8 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 3.2** *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 peut donc écrire :

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

□

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

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

### 3.9 Décomposition en sommes de carrés

**Théorème 3.3** *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

$$SC_{res} = \mathbf{e}'\mathbf{e} = \sum_{i=1}^n (y_i - y_i^*)^2 = \sum_{i=1}^n e_i^2. \quad (3.3)$$

**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

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

□



### 3.10 Régression avec les données centrées

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

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

Dans ce cas, la régression multiple s'écrit :

$$y_i = b_1 + x_{i2}b_2 + x_{i3}b_3 + \cdots + x_{ip}b_p + e_i. \quad (3.4)$$

On peut aussi travailler avec les données centrées. En sommant sur les  $i$  et en divisant par  $n$  l'Équation (3.4), on obtient :

$$\bar{y} = b_1 + \bar{x}_2b_2 + \bar{x}_3b_3 + \cdots + \bar{x}_pb_p, \quad (3.5)$$

et donc en soustrayant (3.5) à (3.4), on a finalement :

$$y_i - \bar{y} = (x_{i2} - \bar{x}_2)b_2 + (x_{i3} - \bar{x}_3)b_3 + \cdots + (x_{ip} - \bar{x}_p)b_p + e_i. \quad (3.6)$$

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 :

$$\mathbf{P}_c = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n} = \begin{pmatrix} 1 - 1/n & -1/n & -1/n & \cdots & -1/n \\ -1/n & 1 - 1/n & -1/n & \cdots & -1/n \\ -1/n & -1/n & 1 - 1/n & \cdots & -1/n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & -1/n & \cdots & 1 - 1/n \end{pmatrix}, \quad (3.7)$$

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

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

$$\tilde{\mathbf{b}} = (\mathbf{X}'_c \mathbf{X}_c)^{-1} \mathbf{X}'_c \mathbf{y}_c = \left( \frac{\mathbf{X}'_c \mathbf{X}_c}{n} \right)^{-1} \frac{\mathbf{X}'_c \mathbf{y}_c}{n}. \quad (3.8)$$

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  $\Sigma$  donnée en (3.2).

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

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 totale en une somme de deux termes :

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

où

– la somme des carrés totale

$$SC_{tot} = \mathbf{y}'_c \mathbf{y}_c = \sum_{i=1}^n (y_i - \bar{y})^2, \quad (3.9)$$

– la somme des carrés expliquée par la régression,

$$SC_{regr} = \tilde{\mathbf{b}}' \mathbf{X}'_c \mathbf{X}_c \tilde{\mathbf{b}}, \quad (3.10)$$

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

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

– la somme des carrés des résidus

$$SC_{res} = \mathbf{e}' \mathbf{e} = \sum_{i=1}^n e_i^2. \quad (3.11)$$

### 3.11 Retour au cas bivarié

### 3.12 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

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

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  $x$ .

### 3.13 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

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

Il reste à déduire  $b_1$  de l'équation

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

ce qui donne

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

# Chapitre 4

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

### 4.1 Probabilités

#### 4.1.1 Événement

Une expérience est dite aléatoire si on ne peut pas prédire *a priori* son résultat. On note  $\omega$  un résultat possible de cette expérience aléatoire. L'ensemble de tous les résultats possibles est noté  $\Omega$ . 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 comme “avoir deux fois pile” ou “avoir au moins une fois pile”. Formellement, un événement est un sous-ensemble de  $\Omega$ .

- L'événement “avoir deux fois pile” est le sous ensemble  $\{(P, P)\}$ .
- L'événement “avoir au moins une fois pile” est le sous ensemble  $\{(P, P), (F, P), (P, F)\}$ .

L'ensemble  $\Omega$  est appelé événement certain, et l'ensemble vide  $\emptyset$  est appelé événement impossible.

#### 4.1.2 Opérations sur les événements

Sur les événements, on peut appliquer les opérations habituelles de la théorie des ensembles.

##### L'union

L'événement  $A \cup B$  est réalisé dès que  $A$  ou  $B$  est réalisé. Dans un lancer de dé, si l'événement  $A$  est “obtenir un nombre pair” et l'événement  $B$  “obtenir un multiple de 3”, l'événement  $A \cup B$  est l'événement “obtenir un nombre pair OU un multiple de 3”, c'est-à-dire  $\{2, 3, 4, 6\}$ .

##### L'intersection

L'événement  $A \cap B$  est réalisé dès que  $A$  et  $B$  sont réalisés conjointement dans la même expérience. Dans un lancer de dé, si l'événement  $A$  est “obtenir un nombre pair” et l'événement  $B$  “obtenir un multiple de 3”, l'événement  $A \cap B$  est l'événement “obtenir un nombre pair ET multiple de 3”, c'est-à-dire  $\{6\}$ .

##### La différence

L'événement  $A \setminus B$  est réalisé quand  $A$  est réalisé et que  $B$  ne l'est pas.

## Le complémentaire

Le complémentaire de l'événement  $A$  est l'événement  $\Omega \setminus A$ . Le complémentaire est noté  $\bar{A}$ .

**Exemple 4.1** 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

$$A = \{2, 4, 6\} \text{ et } \bar{A} = \{1, 3, 5\}.$$

### 4.1.3 Relations entre les événements

#### Événements mutuellement exclusifs

Si  $A \cap B = \emptyset$  on dit que  $A$  et  $B$  sont mutuellement exclusifs, ce qui signifie que  $A$  et  $B$  ne peuvent pas se produire ensemble.

**Exemple 4.2** 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-vidée et consiste en l'événement "obtenir 2".

#### Inclusion

Si  $A$  est inclus dans  $B$ , on écrit  $A \subset B$ . On dit que  $A$  implique  $B$ .

**Exemple 4.3** Si on jette un dé, on considère les événements  $A$  "obtenir 2" et  $B$  "obtenir un nombre pair".

$$A = \{2\} \text{ et } B = \{2, 4, 6\}.$$

On dit que  $A$  implique  $B$ .

### 4.1.4 Ensemble des parties d'un ensemble et système complet

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

**Exemple 4.4** Si on jette une pièce de monnaie alors  $\Omega = \{P, F\}$ , et

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

**Définition 4.1** Les événements  $A_1, \dots, A_n$  forment un système complet d'événements, s'ils constituent une partition de  $\Omega$ , c'est-à-dire si

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

TABLE 4.1 – Système complet d'événements

$A_1$	.....	$A_i$	.....	$A_n$
-------	-------	-------	-------	-------

### 4.1.5 Axiomatique des Probabilités

**Définition 4.2** Une probabilité  $P(\cdot)$  est une application de  $\mathcal{A}$  dans  $[0, 1]$ , telle que :

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

$$\Pr(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3) + \dots + \Pr(A_n).$$

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

**Propriété 4.1**  $\Pr(\emptyset) = 0$ .

**Démonstration** Comme  $\emptyset$  est d'intersection vide avec  $\emptyset$ , on a que

$$\Pr(\emptyset \cup \emptyset) = \Pr(\emptyset) + \Pr(\emptyset).$$

Donc,

$$\Pr(\emptyset) = 2\Pr(\emptyset),$$

ce qui implique que  $\Pr(\emptyset) = 0$ . □

**Propriété 4.2**  $\Pr(\bar{A}) = 1 - \Pr(A)$ .

**Démonstration** On sait que

$$A \cup \bar{A} = \Omega \text{ et } A \cap \bar{A} = \emptyset.$$

Ainsi, on a que

$$\Pr(\Omega) = \Pr(A \cup \bar{A}) = \Pr(A) + \Pr(\bar{A}).$$

Mais, par la définition d'une probabilité,  $\Pr(\Omega) = 1$ . Donc,

$$\Pr(A) + \Pr(\bar{A}) = 1$$

On en déduit que  $\Pr(\bar{A}) = 1 - \Pr(A)$ . □

**Propriété 4.3**  $\Pr(A) \leq \Pr(B)$  si  $A \subset B$ .

**Démonstration** Comme  $A \subset B$ , on a

$$B = (B \cap \bar{A}) \cup A.$$

Mais on a que

$$(B \cap \bar{A}) \cap A = \emptyset.$$

Ainsi, on a

$$\Pr(B) = \Pr(B \cap \bar{A}) + \Pr(A).$$

Or une probabilité est à valeur dans  $[0,1]$ , donc  $\Pr(B \cap \bar{A}) \geq 0$ . On a alors

$$\Pr(B) \geq \Pr(A).$$

□

**Propriété 4.4**  $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$ .

**Démonstration** On a que

$$A \cup B = A \cup (B \cap \bar{A}),$$

et

$$A \cap (B \cap \bar{A}) = \emptyset.$$

Ainsi

$$\Pr(A \cup B) = \Pr(A) + \Pr(B \cap \bar{A}).$$

Il reste à montrer que

$$\Pr(B) = \Pr(B \cap \bar{A}) + \Pr(A \cap B)$$

Mais

$$B = (B \cap \bar{A}) \cup (B \cap A)$$

avec

$$(B \cap \bar{A}) \cap (B \cap A) = \emptyset$$

Donc

$$\Pr(B) = \Pr(B \cap \bar{A}) + \Pr(B \cap A)$$

□

**Propriété 4.5**  $\Pr\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \Pr(A_i)$

**Démonstration** Notons respectivement

$$B_1 = A_1, \quad B_2 = (A_2 \setminus A_1), \quad B_3 = (A_3 \setminus (A_1 \cup A_2)),$$

$$B_4 = (A_4 \setminus (A_1 \cup A_2 \cup A_3)), \quad \dots, \quad B_n = (A_n \setminus (A_1 \cup A_2 \cup A_3 \cup \dots \cup A_{n-1})).$$

Comme

$$\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i,$$

et que  $B_i \cap B_j = \emptyset$  pour tout  $j \neq i$ , alors

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

De plus, comme pour tout  $i$ ,  $B_i \subset A_i$ , on a que  $\Pr(B_i) \leq \Pr(A_i)$ , ce qui donne finalement

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

□

**Propriété 4.6** Si  $A_1, \dots, A_n$  forment un système complet d'événements, alors

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

**Démonstration** Si  $A_1, \dots, A_n$  forment un système complet d'événements, alors

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

Mais on a, pour tout  $i, j$  tels que  $i \neq j$

$$(B \cap A_i) \cap (B \cap A_j) = \emptyset.$$

Finalement, on a que

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

□



### 4.1.6 Probabilités conditionnelles et indépendance

**Définition 4.3** Soient deux événements  $A$  et  $B$ , si  $\Pr(B) > 0$ , alors

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

**Exemple 4.5** Si on jette un dé, et que l'on considère les deux événements suivants :

- $A$  l'évènement 'avoir un nombre pair' et
- $B$  l'évènement 'avoir un nombre supérieur ou égal à 4'.

On a donc

- $\Pr(A) = \Pr(\{2, 4, 6\}) = \frac{1}{2}$ ,
- $\Pr(B) = \Pr(\{4, 5, 6\}) = \frac{3}{6} = \frac{1}{2}$ ,
- $\Pr(A \cap B) = \Pr(\{4, 6\}) = \frac{2}{6} = \frac{1}{3}$ ,
- $\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/3}{1/2} = \frac{2}{3}$ .

**Définition 4.4** 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

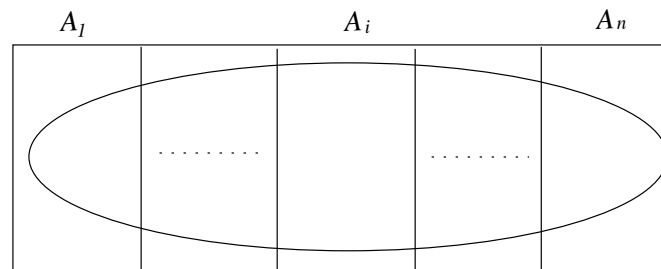
$$\Pr(A \cap B) = \Pr(A)\Pr(B).$$

### 4.1.7 Théorème des probabilités totales et théorème de Bayes

**Théorème 4.1** (des probabilités totales) Soit  $A_1, \dots, A_n$  un système complet d'événements, alors

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

TABLE 4.2 – Illustration du théorème des probabilités totales



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 4.2** (de Bayes) 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,

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

**Exemple 4.6** Supposons qu'une population d'adultes soit composée de 30% de fumeurs ( $A_1$ ) et de 70% de non-fumeurs ( $A_2$ ). Notons  $B$  l'événement "mourir d'un cancer du poumon". Supposons en outre que la probabilité de mourir d'un cancer du poumon est égale à  $\Pr(B|A_1) = 20\%$  si l'on est fumeur et de  $\Pr(B|A_2) = 1\%$  si l'on est non-fumeur. Le théorème de Bayes permet de calculer les probabilités a priori, c'est-à-dire la probabilité d'avoir été fumeur si on est mort d'un cancer du poumon. En effet, cette probabilité est notée  $\Pr(A_1|B)$  et peut être calculée par

$$\Pr(A_1|B) = \frac{\Pr(A_1)\Pr(B|A_1)}{\Pr(A_1)\Pr(B|A_1) + \Pr(A_2)\Pr(B|A_2)} = \frac{0.3 \times 0.2}{0.3 \times 0.2 + 0.7 \times 0.01} = \frac{0.06}{0.06 + 0.007} \approx 0.896.$$

La probabilité de ne pas avoir été non-fumeur si on est mort d'un cancer du poumon vaut quant à elle :

$$\Pr(A_2|B) = \frac{\Pr(A_2)\Pr(B|A_2)}{\Pr(A_1)\Pr(B|A_1) + \Pr(A_2)\Pr(B|A_2)} = \frac{0.7 \times 0.01}{0.3 \times 0.2 + 0.7 \times 0.01} = \frac{0.07}{0.06 + 0.007} \approx 0.104.$$

## 4.2 Analyse combinatoire

### 4.2.1 Introduction

L'analyse combinatoire est l'étude mathématique de la manière de ranger des objets. L'analyse combinatoire est un outil utilisé dans le calcul des probabilités.

### 4.2.2 Permutations (sans répétition)

Une permutation sans répétition est un classement ordonné de  $n$  objets distincts. Considérons par exemple l'ensemble  $\{1, 2, 3\}$ . Il existe 6 manières d'ordonner ces trois chiffres :

$$\{1, 2, 3\}, \{1, 3, 2\}, \{2, 1, 3\}, \{2, 3, 1\}, \{3, 1, 2\}, \{3, 2, 1\}.$$

Si on dispose de  $n$  objets, chacun des  $n$  objets peut être placé à la première place. Il reste ensuite  $n-1$  objets qui peuvent être placés à la deuxième place, puis  $n-2$  objets pour la troisième place, et ainsi de suite. Le nombre de permutations possibles de  $n$  objets distincts vaut donc

$$n \times (n-1) \times (n-2) \times \dots \times 2 \times 1 = n!.$$

La notation  $n!$  se lit factorielle de  $n$  (voir Tableau 4.3).

TABLE 4.3 – Factorielle des nombres de 1 à 10

$n$	0	1	2	3	4	5	6	7	8	9	10
$n!$	1	1	2	6	24	120	720	5040	40320	362880	3628800

### 4.2.3 Permutations avec répétition

On peut également se poser la question du nombre de manières de ranger des objets qui ne sont pas tous distincts. Supposons que nous ayons 2 boules rouges (notées  $R$ ) et 3 boules blanches (notées  $B$ ). Il existe 10 permutations possibles qui sont :

$$\{R, R, B, B, B\}, \{R, B, R, B, B\}, \{R, B, B, R, B\}, \{R, B, B, B, R\}, \{B, R, R, B, B\}, \\ \{B, R, B, R, B\}, \{B, R, B, B, R\}, \{B, B, R, R, B\}, \{B, B, R, B, R\}, \{B, B, B, R, R\}.$$

Si l'on dispose de  $n$  objets appartenant à deux groupes de tailles  $n_1$  et  $n_2$ , le nombre de permutations avec répétition est

$$\frac{n!}{n_1!n_2!}.$$

Par exemple si l'on a 3 boules blanches et 2 boules rouges, on obtient

$$\frac{n!}{n_1!n_2!} = \frac{5!}{2!3!} = \frac{120}{2 \times 6} = 10.$$

Si l'on dispose de  $n$  objets appartenant à  $p$  groupes de tailles  $n_1, n_2, \dots, n_p$ , le nombre de permutations avec répétition est

$$\frac{n!}{n_1!n_2! \times \dots \times n_p!}.$$

### 4.2.4 Arrangements (sans répétition)

Soit  $n$  objets distincts. On appelle un arrangement une manière de sélectionner  $k$  objets parmi les  $n$  et de les ranger dans des boîtes numérotées de 1 à  $k$ .

Dans la première boîte, on peut mettre chacun des  $n$  objets. Dans la seconde boîte, on peut mettre chacun des  $n - 1$  objets restants, dans la troisième boîte, on peut mettre chacun des  $n - 2$  objets restants et ainsi de suite. Le nombre d'arrangements possibles est donc égal à :

$$A_n^k = n \times (n - 1) \times (n - 2) \times \dots \times (n - k + 1) = \frac{n!}{(n - k)!}.$$

### 4.2.5 Combinaisons

Soit  $n$  objets distincts. On appelle une combinaison une manière de sélectionner  $k$  objets parmi les  $n$  sans tenir compte de leur ordre. Le nombre de combinaisons est le nombre de sous-ensembles de taille  $k$  dans un ensemble de taille  $n$ . Soit l'ensemble  $\{1, 2, 3, 4, 5\}$ . Il existe 10 sous-ensembles de taille 3 qui sont :

$$\{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{1, 3, 4\}, \{1, 3, 5\}, \{1, 4, 5\}, \{2, 3, 4\}, \{2, 3, 5\}, \{2, 4, 5\}, \{3, 4, 5\}.$$

De manière générale, quel est le nombre de combinaisons de  $k$  objets parmi  $n$ ? Commençons par calculer le nombre de manières différentes de sélectionner  $k$  objets parmi  $n$  en tenant compte de l'ordre : c'est le nombre d'arrangements sans répétition  $A_n^k$ . Comme il existe  $k!$  manières d'ordonner ces  $k$  éléments, si l'on ne veut pas tenir compte de l'ordre on divise  $A_n^k$  par  $k!$ . Le nombre de combinaisons de  $k$  objets parmi  $n$  vaut donc

$$\frac{A_n^k}{k!} = \frac{n!}{k!(n - k)!}.$$

Le nombre de combinaisons de  $k$  objets parmi  $n$  s'écrit parfois  $\binom{n}{k}$  et parfois  $C_n^k$  :

$$\binom{n}{k} = C_n^k = \frac{n!}{k!(n - k)!}.$$

Par exemple, si on cherche à déterminer le nombre de combinaisons de 3 objets parmi 5, on a

$$\binom{5}{3} = C_5^3 = \frac{5!}{3!(5 - 3)!} = \frac{120}{6 \times 2} = 10.$$

## 4.3 Variables aléatoires

### 4.3.1 Définition

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

**Définition 4.5** Une variable aléatoire  $X$  est une application de l'ensemble fondamental  $\Omega$  dans  $\mathbb{R}$ .

**Exemple 4.7** 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  $\Omega$  a une probabilité  $1/4$ . Une variable aléatoire va associer une valeur à chacun des éléments de  $\Omega$ . Considérons la variable aléatoire représentant le nombre de "Faces" obtenus :

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

C'est une variable aléatoire discrète dont la distribution de probabilités est présentée en Figure 4.1.

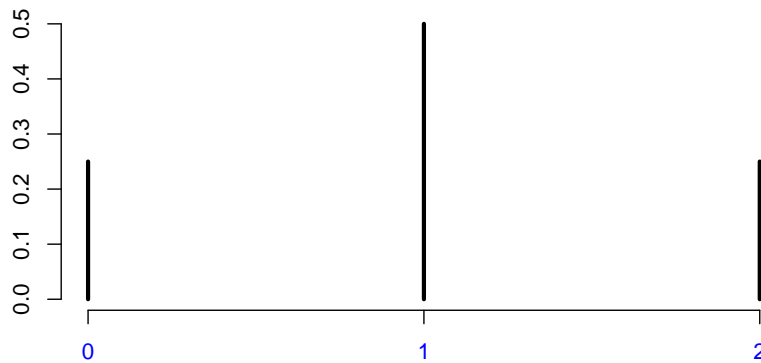


FIGURE 4.1 – Distribution de "faces" obtenus.

## 4.4 Variables aléatoires discrètes

### 4.4.1 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 définie de la manière suivante :

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

et sa variance

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

On peut aussi calculer les moments et tous les autres paramètres.

#### 4.4.2 Variable indicatrice ou bernoullienne

La variable indicatrice  $X$  de paramètre  $p \in [0, 1]$  a la distribution de probabilités 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 4.8** 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  $p = 18/(18+12) = 0.6$ .

#### 4.4.3 Variable binomiale

La variable aléatoire binomiale de paramètres  $n$  et  $p$  correspond à l'expérience suivante. On renouvelle  $n$  fois de manière indépendante une épreuve de Bernoulli de paramètre  $p$ , où  $p$  est la probabilité de succès pour une expérience élémentaire. Ensuite, on note  $X$  le nombre de succès obtenus. Le nombre de succès est une variable aléatoire prenant des valeurs entières de 0 à  $n$  et ayant une distribution binomiale.

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

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

où  $q = 1 - p$ , et

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

De manière synthétique, si  $X$  a une distribution binomiale, on note :

$$X \sim \mathcal{B}(n, p).$$

**Rappel** Cette variable est appelée binomiale car sa distribution de probabilités est un terme du développement du binôme de Newton  $(p + q)^n$ .

$$\begin{aligned} (p + q)^0 &= 1 \\ (p + q)^1 &= p + q = 1 \\ (p + q)^2 &= p^2 + 2pq + q^2 = 1 \\ (p + q)^3 &= p^3 + 3p^2q + 3pq^2 + q^3 = 1 \\ (p + q)^4 &= p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4 = 1 \\ &\vdots \\ (p + q)^n &= \sum_{x=0}^n \binom{n}{x} p^x q^{n-x} = 1. \end{aligned}$$

La somme de ces probabilités vaut 1. En effet

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

L'espérance se calcule de la manière suivante :

$$\begin{aligned}
 E(X) &= \sum_{x=0}^n x \Pr(X = x) \\
 &= \sum_{x=0}^n x \binom{n}{x} p^x q^{n-x} \\
 &= \sum_{x=1}^n x \binom{n}{x} p^x q^{n-x} \text{ (on peut enlever le terme } x = 0 \text{)} \\
 &= \sum_{x=1}^n n \binom{n-1}{x-1} p^x q^{n-x} \\
 &= np \sum_{x=1}^n \binom{n-1}{x-1} p^{x-1} q^{(n-1)-(x-1)} \\
 &= np \sum_{z=0}^{n-1} \binom{n-1}{z} p^z q^{(n-1)-z} \text{ (en posant } z = x - 1 \text{)} \\
 &= np(p + q)^{n-1} \\
 &= np.
 \end{aligned}$$

La variance est donnée (sans démonstration) par

$$\text{var}(X) = npq.$$

**Exemple 4.9** 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

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

La distribution de probabilités de la variable  $X$  est présentée dans la Figure 4.2.

**Exemple 4.10** Supposons que, dans une population d'électeurs, 60% des électeurs s'appêtent à voter pour le candidat  $A$  et 40% pour le candidat  $B$  et que l'on sélectionne un échantillon aléatoire de 10 électeurs avec remise dans cette population. Soit  $X$  le nombre de personnes s'appêtent à voter pour le candidat  $A$  dans l'échantillon. La variable  $X$  a une distribution binomiale de paramètres  $n = 10$  et  $p = 0.6$  et donc

$$\Pr(X = x) = \binom{10}{x} 0.6^x (0.4)^{10-x}, x = 0, 1, \dots, n-1, n.$$

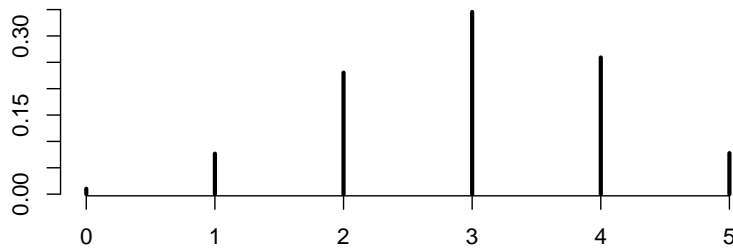


FIGURE 4.2 – Distribution d’une variable aléatoire binomiale avec  $n = 5$  et  $p = 0.6$ .

#### 4.4.4 Variable de Poisson

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

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

On note alors  $X \sim \mathcal{P}(\lambda)$ . La somme des probabilités est bien égale à 1, en effet

$$\sum_{x=0}^{\infty} \Pr(X = x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

L’espérance et la variance d’une loi de Poisson sont égales au paramètre  $\lambda$ . En effet

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \Pr(X = x) \\ &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= e^{-\lambda} \lambda \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} \text{ en posant } z = x - 1 \\ &= e^{-\lambda} \lambda e^{\lambda} \\ &= \lambda. \end{aligned}$$

En outre, il est possible de montrer que

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

La distribution de probabilités d’une variable de Poisson  $\mathcal{P}(\lambda = 1)$  est présentée dans la Figure 4.3.

#### En langage R

```
# # distributions de probabilités discrètes\index{distribution!de
probabilités discrètes}# # nombre de faces obtenues en lançant
deux pièces plot(0:2,dbinom(0:2, 2,0.5),type = "h", lwd=3,
ylim=c(0,0.5),xlab="",ylab="",xaxt = "n",frame = FALSE)
axis(1, 0:2, 0:2, col.axis = "blue")
# binomiale B(5,0.6)
plot(dbinom(0:5, 5,0.6),type = "h",
lwd=3,xlab="",ylab="",main="",frame=FALSE)
```

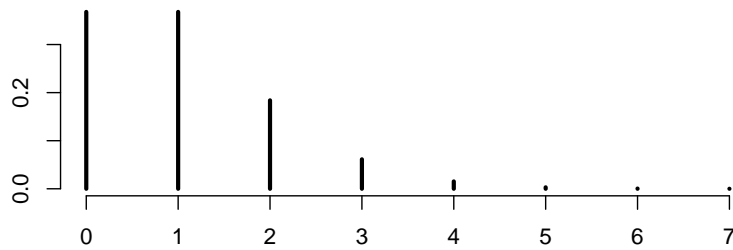


FIGURE 4.3 – Distribution d’une variable de Poisson avec  $\lambda = 1$ .

```
# Poisson P(1)
plot(dpois(0:7, 1), type = "h",
     lwd=3, xlab="", ylab="", main="", frame=FALSE)
```

## 4.5 Variable aléatoire continue

### 4.5.1 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 :

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

Dans la Figure 4.4, la probabilité  $\Pr[X \leq a]$  est l’aire sous la densité de  $-\infty$  à  $a$ .

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

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



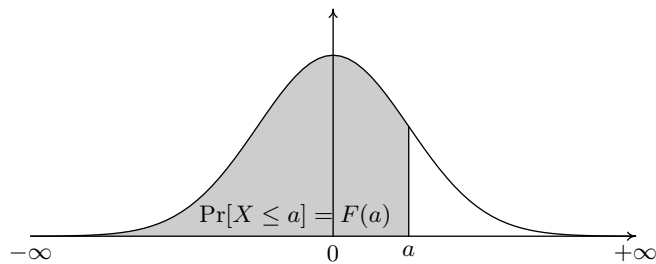


FIGURE 4.4 – Probabilité que la variable aléatoire soit inférieure à  $a$

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

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

#### 4.5.2 Variable uniforme

Une variable aléatoire  $X$  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 calculer l'espérance et la variance :

##### Propriété 4.7

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

##### Démonstration

$$\begin{aligned} \mu &= E(X) = \int_a^b xf(x)dx = \int_a^b x \frac{1}{b - a} dx = \frac{1}{b - a} \int_a^b x dx = \frac{1}{b - a} \left[ \frac{x^2}{2} \right]_a^b \\ &= \frac{1}{b - a} \left( \frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{1}{b - a} \frac{1}{2} (b + a)(b - a) = \frac{a + b}{2}. \end{aligned}$$

□

##### Propriété 4.8

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

### Démonstration

De manière générale, une variance peut toujours s'écrire comme un moment à l'origine d'ordre 2 moins le carré de la moyenne. En effet,

$$\begin{aligned}\sigma^2 &= \text{var}(X) = \int_a^b (x - \mu)^2 f(x) dx = \int_a^b (x^2 + \mu^2 - 2x\mu) f(x) dx \\ &= \int_a^b x^2 f(x) dx + \int_a^b \mu^2 f(x) dx - 2\mu \int_a^b x f(x) dx = \int_a^b x^2 f(x) dx + \mu^2 - 2\mu^2 = \int_a^b x^2 f(x) dx - \mu^2.\end{aligned}$$

On calcule ensuite un moment à l'origine d'ordre 2 :

$$\begin{aligned}\int_a^b x^2 f(x) dx &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left[ \frac{x^3}{3} \right]_a^b \\ &= \frac{1}{b-a} \left( \frac{b^3}{3} - \frac{a^3}{3} \right) = \frac{1}{b-a} \frac{1}{3} (b^2 + ab + a^2)(b-a) = \frac{b^2 + ab + a^2}{3}.\end{aligned}$$

On obtient enfin la variance par différence :

$$\begin{aligned}\sigma^2 &= \int_a^b x^2 f(x) dx - \mu^2 = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{4b^2 + 4ab + 4a^2}{12} - \frac{3a^2 + 6ab + 3b^2}{12} \\ &= \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}.\end{aligned}$$

□

Les logiciels génèrent en général des variables aléatoires uniformes dans  $[0,1]$ . Les Figures 4.5 et 4.6 représentent respectivement les fonctions de densité et de répartition d'une variable uniforme.

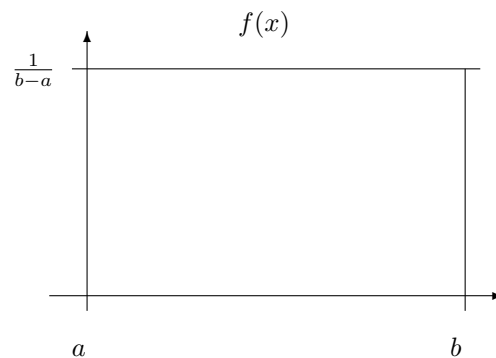


FIGURE 4.5 – Fonction de densité d'une variable uniforme

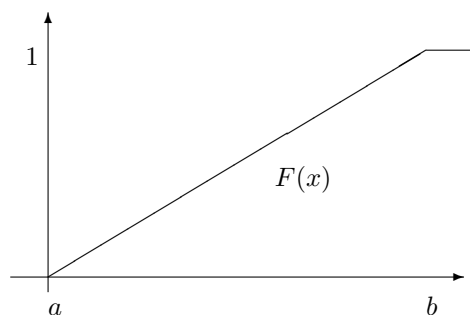


FIGURE 4.6 – Fonction de répartition d'une variable uniforme

### 4.5.3 Variable normale

Une variable aléatoire  $X$  est dite normale si sa densité vaut

$$f_{\mu, \sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2, \quad (4.1)$$

où  $\mu \in \mathbb{R}$  et  $\sigma \in \mathbb{R}^+$  sont les paramètres de la distribution. Le paramètre  $\mu$  est appelé la moyenne et le paramètre  $\sigma$  l'écart-type de la variable normale.

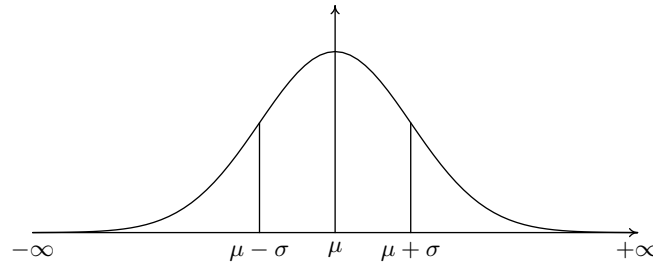


FIGURE 4.7 – Fonction de densité d'une variable normale

De manière synthétique, pour noter que  $X$  a une distribution normale de moyenne  $\mu$  et de variance  $\sigma^2$  on écrit :

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

On peut montrer (sans démonstration) que

$$\int_{-\infty}^{\infty} f_{\mu, \sigma^2}(x) dx = 1,$$

$$E(X) = \int_{-\infty}^{\infty} x f_{\mu, \sigma^2}(x) dx = \mu,$$

et

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

La fonction de répartition vaut

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

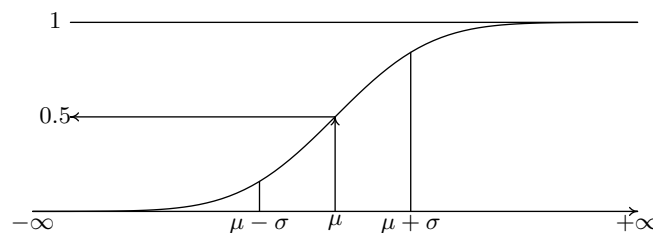


FIGURE 4.8 – Fonction de répartition d'une variable normale

### 4.5.4 Variable normale centrée réduite

La variable aléatoire normale centrée réduite est une variable normale, d'espérance nulle,  $\mu = 0$  et de variance  $\sigma^2 = 1$ . Sa fonction de densité vaut

$$f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{x^2}{2}.$$

et sa répartition vaut

$$\Phi(x) = F_{0,1}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du.$$

Du fait de la symétrie de la densité, on a la relation

$$\Phi(-x) = 1 - \Phi(x),$$

qui se comprend facilement en examinant la Figure 4.9.

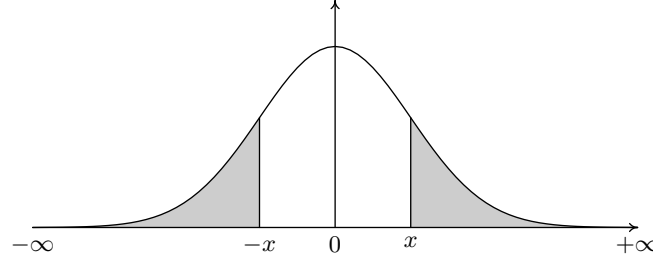


FIGURE 4.9 – Densité d’une normale centrée réduite, symétrie

De plus, le calcul de la répartition d’une variable normale de moyenne  $\mu$  et de variance  $\sigma^2$  peut toujours être ramené à une normale centrée réduite.

**Propriété 4.9**

$$F_{\mu,\sigma^2}(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

**Démonstration**

On a

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

En posant

$$z = \frac{u - \mu}{\sigma},$$

on obtient  $u = z\sigma + \mu$ , et donc  $du = \sigma dz$ . Donc,

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

□

Les tables de la variable normale ne sont données que pour la normale centrée réduite. Les tables ne donnent  $\Phi(x)$  que pour les valeurs positives de  $x$ , car les valeurs négatives peuvent être trouvées par la relation de symétrie.

### 4.5.5 Distribution exponentielle

Une variable aléatoire  $X$  a une distribution exponentielle si sa fonction de densité est donnée par :

$$f(x) = \begin{cases} \lambda \exp(-\lambda x), & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

Le paramètre  $\lambda$  est positif. Quand  $x > 0$ , sa fonction de répartition vaut :

$$F(x) = \int_0^x f(u) du = \int_0^x \lambda e^{-\lambda u} du = [-e^{-\lambda u}]_0^x = 1 - e^{-\lambda x}.$$

On peut alors calculer la moyenne :

**Propriété 4.10**  $E(X) = \frac{1}{\lambda}.$

### Démonstration

$$E(X) = \int_0^{\infty} x f(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \left[ -\frac{1+x\lambda}{\lambda} e^{-\lambda x} \right]_0^{\infty} = \left( 0 + \frac{1}{\lambda} \right) = \frac{1}{\lambda}.$$

□

Il est également possible de montrer que la variance vaut :

$$\text{var}(X) = \frac{1}{\lambda^2}.$$

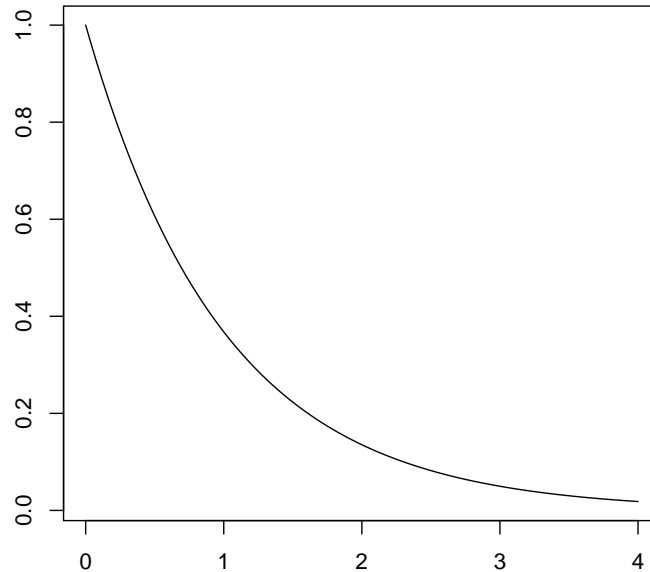


FIGURE 4.10 – Fonction de densité d’une variable exponentielle avec  $\lambda = 1$ .

## 4.6 Distribution bivariée, cas continu

### 4.6.1 Généralités

Deux variables aléatoires ont une distribution bivariée si elles ont une distribution jointe. Soient  $X$  et  $Y$  deux variables aléatoires continues, leur distribution de densité jointe  $f(x, y)$  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 espérances marginales, et les variances marginales :

$$\mu_x = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx, \text{ et } \mu_y = E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy,$$

$$\sigma_X^2 = \text{var}(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_X(x) dx, \text{ et } \sigma_Y^2 = \text{var}(Y) = \int_{-\infty}^{\infty} (y - \mu_y)^2 f_Y(y) dy.$$

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 espérances conditionnelles, et les variances conditionnelles :

$$\mu_x(y) = E(X|Y = y) = \int_{-\infty}^{\infty} x f(x|y) dx, \text{ et } \mu_y(x) = E(Y|X = x) = \int_{-\infty}^{\infty} y f(y|x) dy,$$

$$\sigma_X^2(y) = \text{var}(X|Y = y) = \int_{-\infty}^{\infty} \{x - \mu_x(y)\}^2 f(x|y) dx,$$

et

$$\sigma_Y^2(x) = \text{var}(Y|X = x) = \int_{-\infty}^{\infty} \{y - \mu_y(x)\}^2 f(y|x) dy.$$

Enfin, la covariance entre  $X$  et  $Y$  est définie par

$$\sigma_{xy} = \text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy.$$

Cette covariance peut également s'écrire

$$\sigma_{xy} = \text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy - \mu_x \mu_y.$$

En effet,

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \mu_y f(x, y) dx dy \\ & \quad - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \mu_x f(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu_x \mu_y f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy - \mu_x \mu_y - \mu_x \mu_y + \mu_x \mu_y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy - \mu_x \mu_y. \end{aligned}$$

#### 4.6.2 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 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

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

### 4.6.3 Propriétés des espérances et des variances

De manière générale, pour des variables aléatoires  $X$  et  $Y$ , et avec  $a$  et  $b$  constants, on a les résultats suivants.

#### Propriété 4.11

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

#### Démonstration

$$E(a + bX) = \int_{\mathbb{R}} (a + bx)f_X(x)dx = a \int_{\mathbb{R}} f_X(x)dx + b \int_{\mathbb{R}} xf_X(x)dx = a + bE(X).$$

□

#### Propriété 4.12

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

#### Démonstration

$$\begin{aligned} E(aY + bX) &= \int_{\mathbb{R}} \int_{\mathbb{R}} (ay + bx)f(x, y)dxdy = a \int_{\mathbb{R}} \int_{\mathbb{R}} yf(x, y)dxdy + b \int_{\mathbb{R}} \int_{\mathbb{R}} xf(x, y)dxdy \\ &= a \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x, y)dxdy + b \int_{\mathbb{R}} x \int_{\mathbb{R}} f(x, y)dydx = a \int_{\mathbb{R}} yf_Y(y)dy + b \int_{\mathbb{R}} xf_X(x)dx = aE(Y) + bE(X) \end{aligned}$$

□

#### Propriété 4.13

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

#### Démonstration

$$\begin{aligned} \text{var}(a + bX) &= \int_{\mathbb{R}} [a + bx - E(a + bX)]^2 f_X(x)dx = \int_{\mathbb{R}} [a + bx - (a + bE(X))]^2 f_X(x)dx \\ &= \int_{\mathbb{R}} [bx - bE(X)]^2 f_X(x)dx = b^2 \int_{\mathbb{R}} [x - E(X)]^2 f_X(x)dx = b^2\text{var}(X). \end{aligned}$$

□

#### Propriété 4.14

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

#### Démonstration

$$\begin{aligned} \text{var}(X + Y) &= \int_{\mathbb{R}} \int_{\mathbb{R}} [x + y - E(X + Y)]^2 f(x, y)dxdy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} [x - E(X) + y - E(Y)]^2 f(x, y)dxdy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} [x - E(X)]^2 + [y - E(Y)]^2 + 2[x - E(X)][y - E(Y)]f(x, y)dxdy \\ &= \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \end{aligned}$$

□

**Propriété 4.15** De plus, si  $X$  et  $Y$  sont indépendantes,  $E(XY) = E(X)E(Y)$ .

## Démonstration

$$E(XY) = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf(x)f(y)dxdy = \int_{\mathbb{R}} xf(x)dx \int_{\mathbb{R}} yf(y)dy = E(X)E(Y).$$

□

Si  $X$  et  $Y$  sont indépendantes, on a  $\text{cov}(X, Y) = E(X)E(Y) - E(X)E(Y) = 0$ , et donc

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

### 4.6.4 Somme de variables aléatoires

Soit  $X$  et  $Y$  deux variables aléatoires continues possédant une distribution jointe  $f(x, y)$ . On s'intéresse à la distribution de probabilité de la somme de ces deux variables aléatoires  $Z = X + Y$ . La fonction de répartition de  $Z$  s'écrit :

$$\begin{aligned} F_Z(z) &= \Pr(Z \leq z) = \Pr(X + Y \leq z) = \int_{-\infty}^{\infty} P(X + Y \leq z | Y = y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} P(X \leq z - y | Y = y) f_Y(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f(x|y) dx f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f(x, y) dxdy = \int_{-\infty}^{\infty} \int_{-\infty}^z f(x - y, y) dxdy = \int_{-\infty}^z \int_{-\infty}^{\infty} f(x - y, y) dy dx. \end{aligned}$$

La fonction de densité de  $Z = X + Y$  vaut donc

$$f_Z(z) = \int_{-\infty}^{\infty} f(z - y, y) dy.$$

### 4.6.5 Loi normale bivariée

Les variables  $X$  et  $Y$  suivent une loi normale bivariée si leur densité jointe est donnée par

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[ \frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right\}. \quad (4.2)$$

La fonction de densité dépend de 5 paramètres

- les deux moyennes marginales  $\mu_x \in \mathbb{R}$  et  $\mu_y \in \mathbb{R}$ ,
- les deux variances marginales  $\sigma_x^2 > 0$  et  $\sigma_y^2 > 0$ ,
- le coefficient de corrélation  $-1 < \rho < 1$ .

Un exemple de normale bivariée est présentée dans la Figure 4.11.

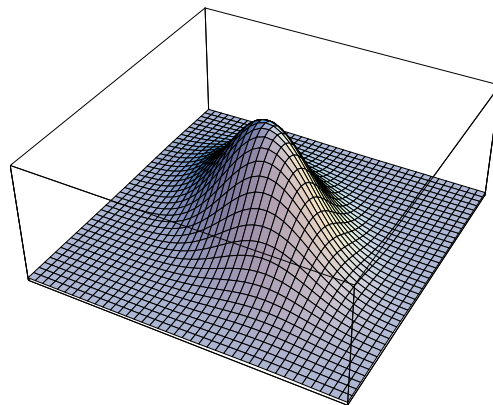


FIGURE 4.11 – Densité d'une normale bivariée

La Figure 4.12 montre le nuage de points de 1000 réalisations d'une normale bivariée avec les paramètres suivants :  $\mu_x = 8$ ,  $\mu_y = 20$ ,  $\sigma_x^2 = 9$ ,  $\sigma_y^2 = 25$ ,  $\rho = 0.6$ .

**En langage R**



```

a=8; b=3 ;c=12 ; d=4
X=a+ b*rnorm(2000)
Y=c+X+d*rnorm(2000)
plot(X,Y,type="p")

```

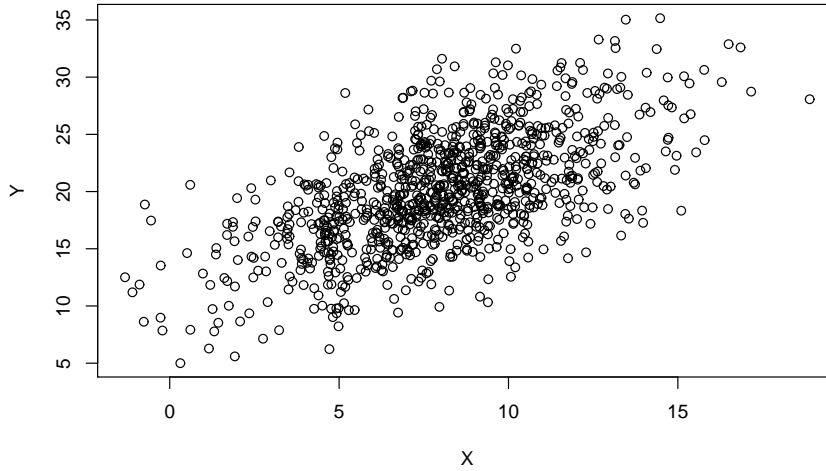


FIGURE 4.12 – Nuage de points de réalisations d’une normale bivariée

**Théorème 4.3** *Les deux distributions marginales d’une distribution normale bivariée ont une distribution normale donnée par :*

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{\sigma_x \sqrt{2\pi}} \exp -\frac{(x - \mu_x)^2}{2\sigma_x^2}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \frac{1}{\sigma_y \sqrt{2\pi}} \exp -\frac{(y - \mu_y)^2}{2\sigma_y^2}$$

**Démonstration** (pour  $f_X(x)$ )

On peut vérifier que la densité jointe peut également s’écrire :

$$f(x, y) = \left( \frac{1}{\sigma_x \sqrt{2\pi}} \exp -\frac{(x - \mu_x)^2}{2\sigma_x^2} \right) \frac{1}{\sigma_y(x) \sqrt{2\pi}} \exp \left\{ \frac{-1}{2} \left( \frac{y - \mu_y(x)}{\sigma_y(x)} \right)^2 \right\},$$

où

$$\mu_y(x) = \mu_y + \frac{\sigma_y \rho}{\sigma_x} (x - \mu_x) \text{ et } \sigma_y^2(x) = \sigma_y^2 (1 - \rho^2).$$

On a

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

$$= \left( \frac{1}{\sigma_x \sqrt{2\pi}} \exp -\frac{(x - \mu_x)^2}{2\sigma_x^2} \right) \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sigma_y(x) \sqrt{2\pi}} \exp \left\{ \frac{-1}{2} \left( \frac{y - \mu_y(x)}{\sigma_y(x)} \right)^2 \right\} dy}_{=1}.$$

□

Le Théorème 4.3 montre que les deux distributions marginales sont normales, que  $\mu_x$  et  $\mu_y$  sont les moyennes marginales, et que  $\sigma_x^2$  et  $\sigma_y^2$  sont les deux variance marginales de la distribution jointes. On peut également montrer à partir du Théorème 4.3 que le volume sous la courbe vaut bien 1. En effet

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

Attention, la réciproque du Théorème 4.3 n'est pas nécessairement vraie. Une distribution bivariée dont les deux distributions marginales sont normales, n'est pas nécessairement normale.

**Théorème 4.4** *Toutes les distributions conditionnelles d'une distribution normale bivariée ont une distribution normale donnée par :*

$$f(y|x) = \frac{1}{\sigma_y(x)\sqrt{2\pi}} \exp \left\{ \frac{-1}{2} \left( \frac{y - \mu_y(x)}{\sigma_y(x)} \right)^2 \right\}$$

où

$$\mu_y(x) = \mu_y + \frac{\sigma_y \rho}{\sigma_x} (x - \mu_x) \text{ et } \sigma_y^2(x) = \sigma_y^2(1 - \rho^2).$$

et

$$f(x|y) = \frac{1}{\sigma_x(y)\sqrt{2\pi}} \exp \left\{ \frac{-1}{2} \left( \frac{x - \mu_x(y)}{\sigma_x(y)} \right)^2 \right\}$$

où

$$\mu_x(y) = \mu_x + \frac{\sigma_x \rho}{\sigma_y} (y - \mu_y) \text{ et } \sigma_x^2(y) = \sigma_x^2(1 - \rho^2).$$

**Démonstration** (pour  $f(y|x)$ )

$$\begin{aligned} f(y|x) &= \frac{f(x, y)}{f_X(x)} \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[ \frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right\} \\ &= \frac{1}{\sigma_x\sqrt{2\pi}} \exp \left\{ -\frac{(x-\mu_x)^2}{2\sigma_x^2} \right\} \\ &= \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[ \frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] + \frac{(x-\mu_x)^2}{2\sigma_x^2} \right\} \\ &= \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[ \frac{\rho^2(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right\} \\ &= \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}} \exp \left\{ \frac{-1}{2\sqrt{1-\rho^2}} \left( \frac{y-\mu_y}{\sigma_y} - \frac{\rho(x-\mu_x)}{\sigma_x} \right)^2 \right\} \\ &= \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}} \exp \left\{ \frac{-1}{2\sqrt{1-\rho^2}} \left( \frac{y-\mu_y - \frac{\rho\sigma_y}{\sigma_x}(x-\mu_x)}{\sigma_y} \right)^2 \right\} \\ &= \frac{1}{\sigma_y(x)\sqrt{2\pi}} \exp \left\{ \frac{-1}{2} \left( \frac{y-\mu_y(x)}{\sigma_y(x)} \right)^2 \right\}. \end{aligned}$$

□

Le Théorème 4.4 montre que toutes les distributions conditionnelles sont également normales. La variance conditionnelle de  $Y$  pour une valeur fixée de  $x$  de la variable  $X$  vaut :

$$E(Y|X = x) = \mu_y(x) = \mu_y + \frac{\sigma_y \rho}{\sigma_x} (x - \mu_x).$$

De même, l'espérance conditionnelle de  $X$  pour une valeur fixée de  $y$  de la variable  $Y$  vaut :

$$E(X|Y = y) = \mu_x(y) = \mu_x + \frac{\sigma_x \rho}{\sigma_y} (y - \mu_y).$$

La variance conditionnelle de  $Y$  pour une valeur fixée de  $x$  de la variable  $X$  vaut :

$$\text{var}(Y|X = x) = \sigma_y^2(x) = \sigma_y^2(1 - \rho^2).$$

Cette variance conditionnelle ne dépend pas de  $x$ . La variance conditionnelle de  $X$  pour une valeur fixée de  $y$  de la variable  $Y$  vaut :

$$\text{var}(X|Y = y) = \sigma_x^2(y) = \sigma_x^2(1 - \rho^2),$$

et ne dépend pas de  $y$ . Cette variance conditionnelle ne dépend pas de  $y$ . Les variances conditionnelles sont donc homoscedastiques (même variance).

**Théorème 4.5**

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dy dx = \sigma_x \sigma_y \rho.$$

**Démonstration**

La covariance peut également s'écrire

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dy dx - \mu_x \mu_y.$$

On a :

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f(y|x) dy dx = \int_{-\infty}^{\infty} x f_X(x) \int_{-\infty}^{\infty} y f(y|x) dy dx \\ &= \int_{-\infty}^{\infty} x f_X(x) \left[ \mu_y + \frac{\sigma_y \rho}{\sigma_x} (x - \mu_x) \right] dx = \mu_y \int_{-\infty}^{\infty} x f_X(x) dx + \frac{\sigma_y \rho}{\sigma_x} \int_{-\infty}^{\infty} x f_X(x) (x - \mu_x) dx \\ &= \mu_y \mu_x + \frac{\sigma_y \rho}{\sigma_x} \sigma_x^2 = \mu_y \mu_x + \sigma_x \sigma_y \rho. \end{aligned}$$

Donc

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy = \sigma_x \sigma_y \rho. \quad \square$$

Le paramètre  $\rho$  est bien un coefficient de corrélation entre les variables  $X$  et  $Y$  car il peut s'écrire :

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\sigma_x \sigma_y \rho}{\sigma_x \sigma_y} = \rho.$$

**Théorème 4.6** Si les deux variables  $X$  et  $Y$  ont une distribution normale bivariée et que leur coefficient de corrélation est nul, alors  $X$  et  $Y$  sont indépendantes.

**Démonstration**

Si  $\rho = 0$ , alors de l'Expression 4.2, la distribution jointe vaut :

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[ \frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right] \right\} \\ &= \left( \frac{1}{\sqrt{2\pi}\sigma_x} \exp \left\{ -\frac{(x - \mu_x)^2}{2\sigma_x^2} \right\} \right) \left( \frac{1}{\sqrt{2\pi}\sigma_y} \exp \left\{ -\frac{(y - \mu_y)^2}{2\sigma_y^2} \right\} \right) \\ &= f_X(x) f_Y(y). \end{aligned}$$

Dans ce cas, la densité jointe peut s'écrire comme le produit des deux densités marginales. Les deux variables sont donc indépendantes.  $\square$

## 4.7 Vecteurs aléatoires

### 4.7.1 Généralités

Un vecteur aléatoire de variables aléatoires continues  $\mathbf{X} = (X_1, \dots, X_p)'$  est un vecteur colonne de variables aléatoires possédant une fonction de densité jointe notée  $f(x_1, \dots, x_p)$  telle que

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_p) dx_1 \dots dx_p = 1.$$

L'espérance mathématique d'un vecteur aléatoire est le vecteur des espérances de ses composantes :

$$\mathbf{E}(\mathbf{X}) = \boldsymbol{\mu} = \mathbf{E} \begin{pmatrix} X_1 \\ \vdots \\ X_j \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \mathbf{E}(X_1) \\ \vdots \\ \mathbf{E}(X_j) \\ \vdots \\ \mathbf{E}(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_j \\ \vdots \\ \mu_p \end{pmatrix}.$$

La variance (ou matrice variance-covariance) d'un vecteur aléatoire de  $\mathbb{R}^p$  est une matrice définie par

$$\text{var}(\mathbf{X}) = \text{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = \text{E}(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}\boldsymbol{\mu}' = \begin{pmatrix} \text{var}(X_1) & \cdots & \text{cov}(X_1, X_j) & \cdots & \text{cov}(X_1, X_p) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \text{cov}(X_1, X_j) & \cdots & \text{var}(X_j) & \cdots & \text{cov}(X_j, X_p) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \text{cov}(X_1, X_p) & \cdots & \text{cov}(X_j, X_p) & \cdots & \text{var}(X_p) \end{pmatrix}.$$

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

$$\text{E}(\mathbf{A}\mathbf{X}) = \text{E} \begin{pmatrix} \sum_{j=1}^p a_{1j} X_j \\ \vdots \\ \sum_{j=1}^p a_{ij} X_j \\ \vdots \\ \sum_{j=1}^p a_{qj} X_j \end{pmatrix} = \begin{pmatrix} \text{E} \left( \sum_{j=1}^p a_{1j} X_j \right) \\ \vdots \\ \text{E} \left( \sum_{j=1}^p a_{ij} X_j \right) \\ \vdots \\ \text{E} \left( \sum_{j=1}^p a_{qj} X_j \right) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^p a_{1j} \text{E}(X_j) \\ \vdots \\ \sum_{j=1}^p a_{ij} \text{E}(X_j) \\ \vdots \\ \sum_{j=1}^p a_{qj} \text{E}(X_j) \end{pmatrix} = \mathbf{A}\text{E}(\mathbf{X}) = \mathbf{A}\boldsymbol{\mu},$$

et

$$\begin{aligned} \text{var}(\mathbf{A}\mathbf{X}) &= \text{E} \{ [\mathbf{A}\mathbf{X} - \text{E}(\mathbf{A}\mathbf{X})][\mathbf{A}\mathbf{X} - \text{E}(\mathbf{A}\mathbf{X})]' \} \\ &= \text{E} \{ [\mathbf{A}\mathbf{X} - \text{E}(\mathbf{A}\mathbf{X})][\mathbf{X}'\mathbf{A}' - \text{E}(\mathbf{X}'\mathbf{A}')] \} \\ &= \text{E} \{ [\mathbf{A}\mathbf{X} - \mathbf{A}\text{E}(\mathbf{X})][\mathbf{X}'\mathbf{A}' - \text{E}(\mathbf{X}')'\mathbf{A}] \} \\ &= \mathbf{A}\text{E} \{ (\mathbf{X} - \text{E}(\mathbf{X}))(\mathbf{X}' - \text{E}(\mathbf{X}')) \} \mathbf{A}' \\ &= \mathbf{A}\text{var}(\mathbf{X})\mathbf{A}' \\ &= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'. \end{aligned}$$

#### 4.7.2 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

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (4.3)$$

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

**Remarque 4.1** Si  $p = 1$ , on retrouve l'Expression (4.1). Si  $p = 2$ , on retrouve l'Expression (4.2).

**Théorème 4.7** Si la matrice variance-covariance d'un vecteur multinormal est diagonale, alors les variables aléatoires de ce vecteur sont indépendantes.

**Démonstration** Si la matrice variance-covariance peut s'écrire  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ , toutes les composantes du vecteur  $\mathbf{X}$  sont non-corrélées. Dans ce cas,

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

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  $X_j$ . □

Dans le cas multinormal (et seulement dans ce cas), l'absence de corrélation implique donc l'indépendance des variables aléatoires.

Un résultat particulièrement important est le suivant :

**Propriété 4.16** *Toute combinaison linéaire d'un vecteur de variables aléatoires normales est normale (Cependant sa matrice variance-covariance n'est pas nécessairement de plein rang).*

(Sans démonstration)

Donc, si  $\mathbf{X}$  est un vecteur multinormal de moyenne  $\boldsymbol{\mu}$  et de matrice variance-covariance  $\Sigma$  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

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

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

**Propriété 4.17** *Toute projection d'un vecteur des variables aléatoires normales est normale.*

## 4.8 Autres variables aléatoires

### 4.8.1 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 la densité de cette variable aléatoire vaut

$$\frac{(1/2)^{p/2}}{\Gamma(p/2)} x^{p/2-1} e^{-x/2}$$

où

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt.$$

On peut également montrer que

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

et que

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

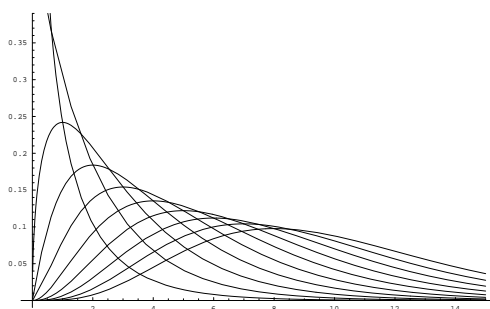


FIGURE 4.13 – Densité d’une variable de chi-carré avec  $p = 1, 2, \dots, 10$

### 4.8.2 Variable de Student

Soit une variable aléatoire  $X$  normale centrée réduite, et une variable aléatoire khi-carré  $\chi_p^2$  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.

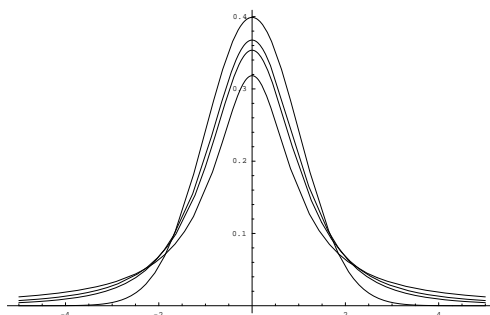


FIGURE 4.14 – Densités de variables de Student avec  $p = 1, 2$  et  $3$  et d’une variable normale

### 4.8.3 Variable de Fisher

Soient deux variables aléatoires khi-carrés indépendantes  $\chi_p^2, \chi_q^2$ , respectivement à  $p$  et  $q$  degrés de liberté, alors la variable aléatoire

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

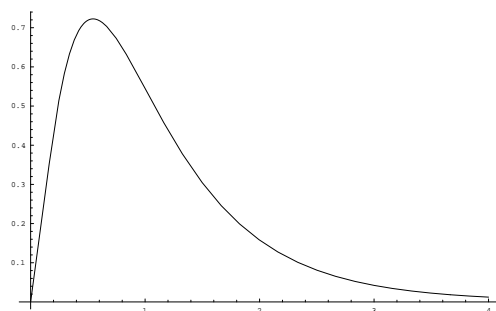


FIGURE 4.15 – Densité d’une variable de Fisher

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

**Remarque 4.2** 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.

## 4.9 Inférence statistique

### 4.9.1 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 randomisation est donc introduite à titre d’hypothèse.

### 4.9.2 Intervalle de confiance

Pour ne pas donner sèchement la valeur d’un estimateur  $\hat{\theta}$  d’un paramètre  $\theta$ , on préfère produire un intervalle  $[L^-, L^+]$  dans lequel pourrait se trouver le paramètre inconnu avec une certaine probabilité que l’on note  $1 - \alpha$  ( $\alpha$  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).

### 4.9.3 Tests d’hypothèses

#### Tests d’hypothèses simples

Le test d’hypothèses consiste à énoncer deux hypothèses sur un paramètre  $\theta$ , 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 fautive, cette erreur est appelée erreur de deuxième espèce.

La probabilité de commettre une erreur de première espèce est notée  $\alpha$ , et la probabilité de commettre une erreur de deuxième espèce est notée  $\beta$ . Dans la théorie des tests d’hypothèses, on fixe  $\alpha$  petit.

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

TABLE 4.4 – 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

TABLE 4.5 – 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  $\alpha$  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  $\theta$  est-il strictement plus grand qu'une certaine valeur  $\theta_0$ ?" 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 4.3** L'égalité doit toujours être dans l'hypothèse nulle, donc si la question est : " $\theta$  est-il strictement plus grand que  $\theta_0$ ?" 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  $\alpha$  particulier. Souvent la statistique de test est l'estimateur  $\hat{\theta}$  de  $\theta$ . 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  $T \in IA$



# Chapitre 5

## Le modèle de régression avec une variable explicative et une constante

### 5.1 Le modèle de régression

Considérons le modèle de régression

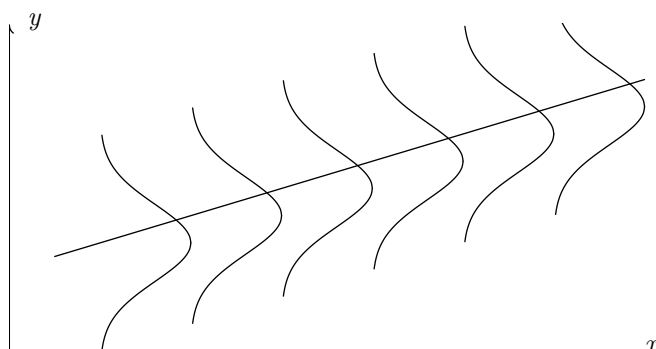
$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

où

1.  $x_i$  représente la valeur prise par la variable  $x$  sur l'individu  $i$ , les  $x_i$  sont supposés non-aléatoires,
2.  $\alpha$  est le coefficient de régression appelé 'constante' (en anglais *intercept*),
3.  $\beta$  est le coefficient de régression appelé 'pente',
4. les  $\varepsilon_i$  sont les termes d'erreurs du modèle et sont des variables aléatoires telles que
  - $E(\varepsilon_i) = 0$  pour tout  $i = 1, \dots, n$ ,
  - $E(\varepsilon_i \varepsilon_k) = 0$  pour tout  $i \neq k = 1, \dots, n$ ,
  - $E(\varepsilon_i^2) = \sigma_\varepsilon^2$  pour tout  $i = 1, \dots, n$ .

Dans un premier temps, nous ne faisons pas d'hypothèses sur la distribution de probabilité des  $\varepsilon_i$ . Les termes d'erreurs sont simplement homoscédastiques (même variance) et non corrélés. Cependant, dans beaucoup d'applications, on supposera que les  $\varepsilon_i$  ont une distribution normale. La Figure 5.1 donne une représentation géométrique du modèle linéaire. La droite représentée a comme équation  $y = \alpha + \beta x$ , et les termes d'erreurs ont une distribution normale autour de la droite de régression.

TABLE 5.1 – Représentation graphique du modèle de régression



Le modèle possède les propriétés suivantes :

1.  $E(y_i) = E(\alpha + \beta x_i + \varepsilon_i) = \alpha + \beta x_i + E(\varepsilon_i) = \alpha + \beta x_i$ .
2. Comme  $\alpha + \beta x_i$  n'est pas aléatoire,  $\text{var}(y_i) = \text{var}(\alpha + \beta x_i + \varepsilon_i) = \text{var}(\varepsilon_i) = \sigma_\varepsilon^2$ , pour tout  $i = 1, \dots, n$ .
3.  $\text{cov}(y_i, y_j) = \text{cov}(\alpha + \beta x_i + \varepsilon_i, \alpha + \beta x_j + \varepsilon_j) = \text{cov}(\varepsilon_i, \varepsilon_j) = 0$ .

Le modèle de régression peut donc être reformulé de manière plus synthétique comme suit :

- $E(y_i) = \alpha + \beta x_i$  pour tout  $i = 1, \dots, n$ ,
- $\text{var}(y_i) = \sigma_\varepsilon^2$  pour tout  $i = 1, \dots, n$ ,
- $\text{cov}(y_i, y_j) = 0$ .

## 5.2 Estimation des paramètres par les moindres carrés

Dans le Chapitre 1, nous avons vu que les paramètres  $\alpha$  et  $\beta$  peuvent être estimés par la méthode des moindres carrés. On cherche les estimateurs  $\hat{\alpha}$  et  $\hat{\beta}$  qui minimisent en  $\alpha$  et  $\beta$  le critère

$$M(\alpha, \beta) = \sum_{i=1}^n \varepsilon_i^2.$$

Comme  $\varepsilon_i = y_i - \alpha - \beta x_i$ ,

$$M(\alpha, \beta) = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2.$$

Les estimateurs obtenus en minimisant  $M(\alpha, \beta)$  en  $\alpha$  et  $\beta$  sont donnés par :

$$\hat{\beta} = \frac{s_{xy}}{s_x^2} \quad \text{et} \quad \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}.$$

Ces estimateurs sont des variables aléatoires car ils dépendent des  $y_i$  qui sont des variables aléatoires.

**Théorème 5.1** *Les estimateurs  $\hat{\beta}$  et  $\hat{\alpha}$  sont des estimateurs linéaires en  $y_i$ , au sens où ils peuvent s'écrire sous la forme :*

$$\hat{\beta} = \sum_{i=1}^n c_i y_i \quad \text{et} \quad \hat{\alpha} = \sum_{i=1}^n d_i y_i,$$

avec des coefficients  $c_i$  et  $d_i$  qui ne sont pas aléatoires.

### Démonstration

1. Si on prend

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

on a

$$\sum_{i=1}^n c_i y_i = \frac{1}{n s_x^2} \sum_{i=1}^n (x_i - \bar{x}) y_i = \frac{s_{xy}}{s_x^2} = \hat{\beta}.$$

En effet,

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

2. Si on prend

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

on a

$$\sum_{i=1}^n d_i y_i = \frac{1}{n} \sum_{i=1}^n y_i - \bar{x} \frac{1}{n s_x^2} \sum_{i=1}^n (x_i - \bar{x}) y_i = \bar{y} - \bar{x} \frac{s_{xy}}{s_x^2} = \bar{y} - \bar{x} \hat{\beta} = \hat{\alpha}.$$

**Théorème 5.2** *Les estimateurs  $\hat{\alpha}$  et  $\hat{\beta}$  sont sans biais pour les paramètres  $\alpha$  et  $\beta$ .*

### Démonstration

Comme

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{n s_x^2}, \tag{5.1}$$

on a

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

et comme

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

et que

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

on obtient

$$\mathbb{E}(\hat{\beta}) = \frac{\beta n s_x^2}{n s_x^2} = \beta.$$

De plus, comme

$$\mathbb{E}(\bar{y}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(y_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\alpha + \beta x_i + \varepsilon_i) = \frac{1}{n} \sum_{i=1}^n (\alpha + \beta x_i) = \alpha + \beta \bar{x},$$

on a

$$\mathbb{E}(\hat{\alpha}) = \mathbb{E}(\bar{y}) - \bar{x}\mathbb{E}(\hat{\beta}) = \alpha + \beta \bar{x} - \beta \bar{x} = \alpha.$$

□

### Théorème 5.3

$$\begin{aligned} \text{var}(\hat{\beta}) &= \frac{\sigma_\varepsilon^2}{n s_x^2}, \\ \text{var}(\hat{\alpha}) &= \frac{\sigma_\varepsilon^2}{n s_x^2} (s_x^2 + \bar{x}^2), \\ \text{cov}(\hat{\alpha}, \hat{\beta}) &= -\frac{\bar{x} \sigma_\varepsilon^2}{n s_x^2}. \end{aligned}$$

**Démonstration** Dans le Théorème 5.1, nous avons vu que  $\hat{\alpha}$  et  $\hat{\beta}$  peuvent s'écrire :

$$\hat{\beta} = \sum_{i=1}^n c_i y_i \quad \text{et} \quad \hat{\alpha} = \sum_{i=1}^n d_i y_i,$$

où

$$c_i = \frac{1}{n} \frac{(x_i - \bar{x})}{s_x^2}, \quad \text{et} \quad d_i = \frac{1}{n} - \bar{x} \frac{1}{n} \frac{(x_i - \bar{x})}{s_x^2}.$$

Comme les  $y_i$  sont non corrélés, on a directement :

$$\begin{aligned} \text{var}(\hat{\beta}) &= \sum_{i=1}^n c_i^2 \text{var}(y_i) = \sigma_\varepsilon^2 \sum_{i=1}^n c_i^2, \\ \text{var}(\hat{\alpha}) &= \sum_{i=1}^n d_i^2 \text{var}(y_i) = \sigma_\varepsilon^2 \sum_{i=1}^n d_i^2, \\ \text{cov}(\hat{\alpha}, \hat{\beta}) &= \sum_{i=1}^n c_i d_i \text{var}(y_i) = \sigma_\varepsilon^2 \sum_{i=1}^n c_i d_i. \end{aligned}$$

Il ne reste qu'à remplacer  $c_i$  et  $d_i$  par leurs valeurs, pour obtenir

$$\begin{aligned} \text{var}(\hat{\beta}) &= \sigma_\varepsilon^2 \sum_{i=1}^n \frac{\frac{1}{n^2} (x_i - \bar{x})^2}{s_x^4} = \frac{\sigma_\varepsilon^2}{n} \frac{s_x^2}{s_x^4} = \frac{\sigma_\varepsilon^2}{n s_x^2}, \\ \text{var}(\hat{\alpha}) &= \sigma_\varepsilon^2 \sum_{i=1}^n \left( \frac{1}{n} - \bar{x} \frac{1}{n} \frac{(x_i - \bar{x})}{s_x^2} \right)^2 = \sigma_\varepsilon^2 \sum_{i=1}^n \left( \frac{1}{n^2} - 2 \frac{1}{n} \bar{x} \frac{1}{n} \frac{(x_i - \bar{x})}{s_x^2} + \bar{x}^2 \frac{1}{n^2} \frac{(x_i - \bar{x})^2}{s_x^4} \right) \\ &= \sigma_\varepsilon^2 \left( \frac{1}{n} - 0 + \bar{x}^2 \frac{1}{n} \frac{s_x^2}{s_x^4} \right) = \frac{\sigma_\varepsilon^2}{n} \left( 1 + \frac{\bar{x}^2}{s_x^2} \right) = \frac{\sigma_\varepsilon^2}{n s_x^2} (s_x^2 + \bar{x}^2), \end{aligned}$$

et

$$\begin{aligned}\text{cov}(\widehat{\alpha}, \widehat{\beta}) &= \sigma_\varepsilon^2 \sum_{i=1}^n \left( \frac{1}{n} - \bar{x} \frac{\frac{1}{n}(x_i - \bar{x})}{s_x^2} \right) \frac{\frac{1}{n}(x_i - \bar{x})}{s_x^2} = \frac{\sigma_\varepsilon^2}{n} \sum_{i=1}^n \left( \frac{(x_i - \bar{x})}{s_x^2 n} - \bar{x} \frac{\frac{1}{n}(x_i - \bar{x})^2}{s_x^4} \right) \\ &= \frac{\sigma_\varepsilon^2}{n} \left( 0 - \bar{x} \frac{s_x^2}{s_x^4} \right) = -\frac{\bar{x} \sigma_\varepsilon^2}{n s_x^2}.\end{aligned}$$

□

### 5.3 Valeurs ajustées et résidus

Les valeurs ajustées se calculent simplement par :

$$y_i^* = \widehat{\alpha} + \widehat{\beta}x_i, \text{ pour tout } i = 1, \dots, n,$$

et les résidus s'obtiennent par la différence entre les valeurs observées et les valeurs ajustées :

$$e_i = y_i - y_i^* = y_i - \widehat{\alpha} - \widehat{\beta}x_i, \text{ pour tout } i = 1, \dots, n.$$

Il ne faut pas confondre les résidus  $e_i$  avec les termes d'erreurs du modèle  $\varepsilon_i$ . Pour rappel, les résidus ont les propriétés suivantes :

$$\sum_{i=1}^n e_i = 0, \quad \sum_{i=1}^n e_i x_i = 0, \quad \sum_{i=1}^n e_i y_i^* = 0.$$

Les termes d'erreurs du modèle ne sont généralement pas de moyenne nulle, cependant

$$E\left(\sum_{i=1}^n \varepsilon_i\right) = \sum_{i=1}^n E(\varepsilon_i) = 0, \quad \text{et} \quad \text{var}\left(\sum_{i=1}^n \varepsilon_i\right) = \sum_{i=1}^n \text{var}(\varepsilon_i) = n\sigma_\varepsilon^2.$$

Les valeurs ajustées et les résidus sont des variables aléatoires, puisqu'ils dépendent tous deux des  $y_i$ . Leurs espérances valent :

$$E(y_i^*) = E(\widehat{\alpha}) + E(\widehat{\beta})x_i = \alpha + \beta x_i, \text{ pour tout } i = 1, \dots, n,$$

donc  $E(y_i^*) = E(y_i)$ , ce qui implique directement que

$$E(e_i) = 0, \text{ pour tout } i = 1, \dots, n.$$

La variance des  $y_i^*$  vaut

$$\text{var}(y_i^*) = \text{var}(\widehat{\alpha} + \widehat{\beta}x_i) = \text{var}(\widehat{\alpha}) + x_i^2 \text{var}(\widehat{\beta}) + 2x_i \text{cov}(\widehat{\alpha}, \widehat{\beta}), \text{ pour tout } i = 1, \dots, n.$$

En utilisant les résultats du Théorème 5.3, on a

$$\text{var}(y_i^*) = \frac{\sigma_\varepsilon^2}{n s_x^2} [(s_x^2 + \bar{x}^2) + x_i^2 - 2x_i \bar{x}] = \frac{\sigma_\varepsilon^2}{n s_x^2} [s_x^2 + (x_i - \bar{x})^2].$$

Enfin, on peut calculer la covariance entre la valeur observée  $y_i$  et la valeur ajustée  $y_i^*$  :

$$\begin{aligned}\text{cov}(y_i, y_i^*) &= \text{cov}(y_i, \widehat{\alpha} + \widehat{\beta}x_i) = \text{cov}\left(y_i, \sum_{j=1}^n d_j y_j + x_i \sum_{j=1}^n c_j y_j\right) = \text{cov}(y_i, d_i y_i) + x_i \text{cov}(y_i, c_i y_i) \\ &= d_i \text{var}(y_i) + x_i c_i \text{var}(y_i) = \sigma_\varepsilon^2 (d_i + x_i c_i) = \sigma_\varepsilon^2 \left[ \frac{1}{n} - \bar{x} \frac{\frac{1}{n}(x_i - \bar{x})}{s_x^2} + x_i \frac{1}{n} \frac{(x_i - \bar{x})}{s_x^2} \right] \\ &= \sigma_\varepsilon^2 \left[ \frac{1}{n} + \frac{\frac{1}{n}(x_i - \bar{x})^2}{s_x^2} \right] = \frac{\sigma_\varepsilon^2}{n s_x^2} [s_x^2 + (x_i - \bar{x})^2] = \text{var}(y_i^*).\end{aligned}$$

Ce dernier résultat permet dès lors de calculer la variance des résidus. En effet,

$$\begin{aligned} E(e_i^2) &= \text{var}(e_i) = \text{var}(y_i - y_i^*) = \text{var}(y_i) + \text{var}(y_i^*) - 2\text{cov}(y_i, y_i^*) = \text{var}(y_i) - \text{var}(y_i^*) \\ &= \sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^2}{ns_x^2} \left[ s_x^2 + (x_i - \bar{x})^2 \right] = \sigma_\varepsilon^2 \left\{ 1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{ns_x^2} \right\}. \end{aligned} \quad (5.2)$$

L'Égalité (5.2) nous montre également que

$$\text{var}(y_i) = \text{var}(y_i^*) + \text{var}(e_i),$$

et donc que  $\text{cov}(y_i^*, e_i) = 0$ .

**Théorème 5.4**

$$E\left(\sum_{i=1}^n e_i^2\right) = (n-2)\sigma_\varepsilon^2$$

Ce résultat fondamental s'obtient simplement en sommant l'Égalité (5.2). Il en découle que

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

est un estimateur sans biais de  $\sigma_\varepsilon^2$ .

Enfin on peut calculer l'espérance et la variance de  $\bar{y}$  :

$$E(\bar{y}) = E\left(\frac{1}{n} \sum_{i=1}^n y_i\right) = \frac{1}{n} \sum_{i=1}^n E(y_i) = \frac{1}{n} \sum_{i=1}^n (\alpha + \beta x_i) = \alpha + \beta \bar{x},$$

et

$$\text{var}(\bar{y}) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n y_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(y_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{n}.$$

On peut donc construire le Tableau 5.2 qui récapitule les résultats obtenus.

TABLE 5.2 – Tableau récapitulatif des espérances et variances

	$y_i$	$\varepsilon_i$	$y_i^*$	$e_i$	$\hat{\alpha}$	$\hat{\beta}$	$\bar{y}$
Espérance	$\alpha + \beta x_i$	0	$\alpha + \beta x_i$	0	$\alpha$	$\beta$	$\alpha + \beta \bar{x}$
Variance	$\sigma_\varepsilon^2$	$\sigma_\varepsilon^2$	$\frac{\sigma_\varepsilon^2}{ns_x^2} [s_x^2 + (x_i - \bar{x})^2]$	$\sigma_\varepsilon^2 \left\{ 1 - \frac{1}{n} + \frac{(x_i - \bar{x})^2}{ns_x^2} \right\}$	$\frac{\sigma_\varepsilon^2}{ns_x^2} (s_x^2 + \bar{x}^2)$	$\frac{\sigma_\varepsilon^2}{ns_x^2}$	$\frac{\sigma_\varepsilon^2}{n}$
Estimateur de Variance	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_\varepsilon^2$	$\frac{\hat{\sigma}_\varepsilon^2}{ns_x^2} [s_x^2 + (x_i - \bar{x})^2]$	$\hat{\sigma}_\varepsilon^2 \left\{ 1 - \frac{1}{n} + \frac{(x_i - \bar{x})^2}{ns_x^2} \right\}$	$\frac{\hat{\sigma}_\varepsilon^2}{ns_x^2} (s_x^2 + \bar{x}^2)$	$\frac{\hat{\sigma}_\varepsilon^2}{ns_x^2}$	$\frac{\hat{\sigma}_\varepsilon^2}{n}$

## 5.4 Espérance des sommes de carrés

Nous avons vu, que la somme des carrés totale se décompose en une somme de deux termes : la somme des carrés de la régression et la somme des carrés des résidus :

$$SC_{TOT} = SC_{REGR} + SC_{RES},$$

où

$$SC_{TOT} = \sum_{i=1}^n (y_i - \bar{y})^2, \quad SC_{REGR} = \sum_{i=1}^n (y_i^* - \bar{y})^2, \quad \text{et} \quad SC_{RES} = \sum_{i=1}^n e_i^2.$$

Comme

$$SC_{TOT} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2,$$

et que

$$E(\bar{y}^2) = \text{var}(\bar{y}) + E^2(\bar{y}) = \frac{\sigma_\varepsilon^2}{n} + (\alpha + \beta\bar{x})^2,$$

on a

$$\begin{aligned} E(\text{SC}_{TOT}) &= \sum_{i=1}^n E(y_i^2) - nE(\bar{y}^2) = \sum_{i=1}^n [\text{var}(y_i) + E^2(y_i)] - nE(\bar{y}^2) = n\sigma_\varepsilon^2 + \sum_{i=1}^n (\alpha + \beta x_i)^2 - \sigma_\varepsilon^2 - n(\alpha + \beta\bar{x})^2 \\ &= (n-1)\sigma_\varepsilon^2 + n\alpha^2 + 2n\alpha\beta\bar{x} + \beta^2 \sum_{i=1}^n x_i^2 - n\alpha^2 - 2n\alpha\beta\bar{x} - n\beta^2\bar{x}^2 = (n-1)\sigma_\varepsilon^2 + \beta^2 ns_x^2. \end{aligned}$$

Comme

$$\text{SC}_{REGR} = \sum_{i=1}^n (y_i^* - \bar{y})^2 = \sum_{i=1}^n [\hat{\alpha} + \hat{\beta}x_i - (\hat{\alpha} + \hat{\beta}\bar{x})]^2 = \hat{\beta}^2 \sum_{i=1}^n (x_i - \bar{x})^2 = \hat{\beta}^2 ns_x^2,$$

on a

$$E(\text{SC}_{REGR}) = ns_x^2 E(\hat{\beta}^2) = ns_x^2 [\text{var}(\hat{\beta}) + E^2(\hat{\beta})] = ns_x^2 \left[ \frac{\sigma_\varepsilon^2}{ns_x^2} + \beta^2 \right] = \sigma_\varepsilon^2 + \beta^2 ns_x^2.$$

Enfin, le Théorème 5.4 donne l'espérance de la somme des carrés des résidus :

$$E(\text{SC}_{RES}) = (n-2)\sigma_\varepsilon^2,$$

ce qui permet construire le Tableau récapitulatif 5.3.

TABLE 5.3 – Tableau récapitulatif des espérances des sommes de carrés

	$\text{SC}_{TOT}$	$\text{SC}_{REGR}$	$\text{SC}_{RES}$
Définition	$\sum_{i=1}^n (y_i - \bar{y})^2$	$\sum_{i=1}^n (y_i^* - \bar{y})^2$	$\sum_{i=1}^n e_i^2$
Espérance	$(n-1)\sigma_\varepsilon^2 + \beta^2 ns_x^2$	$\sigma_\varepsilon^2 + \beta^2 ns_x^2$	$(n-2)\sigma_\varepsilon^2$

## 5.5 Le modèle linéaire avec normalité des erreurs

Si les termes d'erreurs  $\varepsilon_i$  ont une distribution normale, alors ils sont indépendants. En effet, si des variables aléatoires ont une distribution multinormale et qu'ils sont non-corrélés, alors ils sont indépendants. Avec l'hypothèse de normalité des termes d'erreurs, le modèle linéaire se reformule comme suit :

- $y_i = \alpha + \beta x_i + \varepsilon_i$ , pour tout  $i = 1, \dots, n$ ,
- les  $\varepsilon_i$  sont des variables aléatoires normales indépendantes,

ou encore :

- $y_i \sim N(\alpha + \beta x_i, \sigma_\varepsilon^2)$ , pour tout  $i = 1, \dots, n$ ,
- les  $y_i$  sont des variables aléatoires normales indépendantes.

En effet, toute combinaison linéaire de variables normales a une distribution normale, et  $y_i$  est une combinaison linéaire de  $\varepsilon_i$ .

Supposer la normalité des résidus, implique que les données sont régies par un modèle paramétrique, au sens où l'on peut déterminer la distribution de probabilité des données observées. Cette distribution est une fonction qui dépend des paramètres inconnus à estimer. Si les données sont régies par un modèle paramétrique, on pourra utiliser la méthode du maximum de vraisemblance pour estimer les paramètres.

La densité de  $y_i$  est donnée par

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp - \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_\varepsilon^2},$$

et du fait de l'indépendance des  $y_i$ , la densité jointe de  $y_1, \dots, y_n$  peut s'écrire comme le produit des densités marginales

$$f(y_1, \dots, y_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp - \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_\varepsilon^2} = \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2}} \exp - \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_\varepsilon^2}.$$

La fonction de vraisemblance est égale à la fonction de densité, mais est vue comme une fonction des paramètres en considérant que les observations  $y_1, \dots, y_n$  sont fixées :

$$L(\alpha, \beta, \sigma_\varepsilon^2 | y_1, \dots, y_n) = \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2}} \exp - \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_\varepsilon^2}.$$

Pour estimer les paramètres, on va chercher les valeurs des paramètres les plus vraisemblables, c'est-à-dire les valeurs qui maximisent la fonction de vraisemblance pour les observations  $y_i$  et  $x_i$  obtenues. Pour réaliser cette optimisation, on utilise en général le logarithme de la fonction de vraisemblance. En effet, maximiser une fonction ou son logarithme donnera le même résultat, car le logarithme est une fonction croissante. La maximisation du logarithme s'avèrera beaucoup plus simple, car le logarithme transformera les produits en sommes, ce qui simplifiera le calcul des dérivées. La fonction de log-vraisemblance vaut :

$$\begin{aligned} \ell(\alpha, \beta, \sigma_\varepsilon^2 | y_1, \dots, y_n) &= \log L(\alpha, \beta, \sigma_\varepsilon^2 | y_1, \dots, y_n) = \log \left[ \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2}} \exp - \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_\varepsilon^2} \right] \\ &= -\frac{n}{2} \log \sigma_\varepsilon^2 - \frac{n}{2} \log 2\pi - \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_\varepsilon^2}. \end{aligned}$$

**Théorème 5.5** Les estimateurs du maximum de vraisemblance des paramètres  $\alpha$ ,  $\beta$  et  $\sigma_\varepsilon^2$  sont donnés par

$$\hat{\alpha}_{MV} = \bar{y} - \frac{s_{xy}^2}{s_x^2} \bar{x}, \quad \hat{\beta}_{MV} = \frac{s_{xy}}{s_x^2}, \quad \text{et} \quad \hat{\sigma}_{\varepsilon MV}^2 = \frac{1}{n} \sum_{i=1}^n e_i^2.$$

### Démonstration

En annulant la dérivée de la log-vraisemblance par rapport à  $\alpha$ , on obtient :

$$\frac{\partial \ell(\alpha, \beta, \sigma_\varepsilon^2 | y_1, \dots, y_n)}{\partial \alpha} = \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i)}{\sigma_\varepsilon^2} = 0,$$

ce qui donne, en multipliant cette équation par  $\sigma_\varepsilon^2/n$

$$\bar{y} - \alpha - \beta \bar{x} = 0. \tag{5.3}$$

En annulant la dérivée de la log-vraisemblance par rapport à  $\beta$ , on obtient :

$$\frac{\partial \ell(\alpha, \beta, \sigma_\varepsilon^2 | y_1, \dots, y_n)}{\partial \beta} = \sum_{i=1}^n \frac{x_i (y_i - \alpha - \beta x_i)}{\sigma_\varepsilon^2} = 0,$$

ce qui donne, en multipliant cette équation par  $\sigma_\varepsilon^2/n$

$$\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \alpha - \beta \frac{1}{n} \sum_{i=1}^n x_i^2 = 0. \tag{5.4}$$

L'Équation (5.3) nous donne  $\alpha = \bar{y} - \beta \bar{x}$ . Si l'on remplace  $\alpha$  par cette valeur dans l'Équation (5.4), on obtient :

$$\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} - \beta \frac{1}{n} \sum_{i=1}^n x_i^2 + \beta \bar{x}^2 = s_{xy}^2 - \beta s_x^2 = 0,$$

ce qui donne l'estimateur du maximum de vraisemblance (MV) de  $\beta$  :

$$\widehat{\beta}_{MV} = \frac{s_{xy}^2}{s_x^2}.$$

En utilisant l'Équation (5.3), on obtient l'estimateur du maximum de vraisemblance (MV) de  $\alpha$  :

$$\widehat{\alpha}_{MV} = \bar{y} - \frac{s_{xy}^2}{s_x^2} \bar{x}.$$

Il reste à calculer l'estimateur du maximum de vraisemblance de  $\sigma_\varepsilon^2$ . On annule la dérivée de la log-vraisemblance par rapport à  $\sigma_\varepsilon^2$ , pour obtenir :

$$\frac{\partial \ell(\alpha, \beta, \sigma_\varepsilon^2 | y_1, \dots, y_n)}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \sum_{i=1}^n \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_\varepsilon^4}.$$

En multipliant cette équation par  $2\sigma_\varepsilon^4/n$ , on a

$$\sigma_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2.$$

Si l'on remplace dans cette équation  $\alpha$  et  $\beta$  par leurs estimateurs du maximum de vraisemblance, on a

$$\widehat{\sigma}_{\varepsilon MV}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \widehat{\alpha}_{MV} - \widehat{\beta}_{MV} x_i)^2 = \frac{1}{n} \sum_{i=1}^n e_i^2.$$

□

On constate que :

- Les estimateurs du maximum de vraisemblance de  $\alpha$  et  $\beta$  sont égaux aux estimateurs des moindres carrés.
- L'estimateur du maximum de vraisemblance de  $\sigma_\varepsilon^2$  est un estimateur biaisé (mais asymptotiquement sans biais).

## 5.6 Distribution de probabilités des estimateurs

Nous avons vu que les estimateurs  $\widehat{\alpha}$  et  $\widehat{\beta}$  peuvent s'écrire comme une combinaison linéaire des  $y_i$ . Si les  $y_i$  ont une distribution normale, alors une combinaison linéaire de variables normales a également une distribution normale. On en déduit que  $\widehat{\alpha}$  et  $\widehat{\beta}$  ont une distribution normale. On peut donc écrire :

$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma_\varepsilon^2}{n s_x^2}\right) \quad \text{et} \quad \widehat{\alpha} \sim N\left(\alpha, \frac{\sigma_\varepsilon^2}{n s_x^2} (s_x^2 + \bar{x}^2)\right),$$

et en centrant et réduisant ces deux variables aléatoires

$$\frac{\widehat{\beta} - \mathbb{E}(\widehat{\beta})}{\sqrt{\text{var}(\widehat{\beta})}} = \frac{\sqrt{n} s_x (\widehat{\beta} - \beta)}{\sigma_\varepsilon} \sim N(0, 1) \quad \text{et} \quad \frac{\widehat{\alpha} - \mathbb{E}(\widehat{\alpha})}{\sqrt{\text{var}(\widehat{\alpha})}} = \frac{\sqrt{n} s_x (\widehat{\alpha} - \alpha)}{\sigma_\varepsilon \sqrt{s_x^2 + \bar{x}^2}} \sim N(0, 1). \quad (5.5)$$

Il est possible de montrer (nous le ferons plus loin dans un cadre plus général) que la variable aléatoire

$$\frac{\sum_{i=1}^n e_i^2}{\sigma_\varepsilon^2} = \frac{(n-2)\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{n-2}^2, \quad (5.6)$$

où  $\widehat{\sigma}_\varepsilon^2$  est l'estimateur sans biais de  $\sigma_\varepsilon^2$

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}.$$

Il est en outre possible de montrer que la somme des carrés des résidus est, sous l'hypothèse de normalité, une statistique indépendante de  $\widehat{\alpha}$  et  $\widehat{\beta}$ , ce que nous démontrerons plus loin dans un cadre plus général.



Selon la définition donnée en Section 4.8.2, on peut construire une variable de Student en divisant une variable normale centrée réduite par la racine d'une variable khi-carré divisée par son nombre de degrés de liberté, cette variable khi-carré devant être indépendante de la variable normale, ce qui s'écrit

$$\frac{N(0, 1)}{\sqrt{\chi_{(n-2)}^2/(n-2)}} \sim t_{n-2},$$

où  $t_{n-2}$  est une variable de Student à  $n-2$  degrés de liberté.

À partir des résultats (5.5) et (5.6), on a donc

$$\frac{\frac{\sqrt{ns_x}(\hat{\beta} - \beta)}{\sigma_\varepsilon}}{\sqrt{\frac{(n-2)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}/(n-2)}} \sim t_{n-2} \quad \text{et} \quad \frac{\frac{\sqrt{ns_x}(\hat{\alpha} - \alpha)}{\sigma_\varepsilon \sqrt{s_x^2 + \bar{x}^2}}}{\sqrt{\frac{(n-2)\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}/(n-2)}} \sim t_{n-2},$$

ce qui s'écrit plus simplement par

$$\frac{\sqrt{ns_x}(\hat{\beta} - \beta)}{\hat{\sigma}_\varepsilon} \sim t_{n-2} \quad \text{et} \quad \frac{\sqrt{ns_x}(\hat{\alpha} - \alpha)}{\hat{\sigma}_\varepsilon \sqrt{s_x^2 + \bar{x}^2}} \sim t_{n-2}, \quad (5.7)$$

ou encore

$$\frac{\hat{\beta} - \beta}{\sqrt{\widehat{\text{var}}(\hat{\beta})}} \sim t_{n-2} \quad \text{et} \quad \frac{\hat{\alpha} - \alpha}{\sqrt{\widehat{\text{var}}(\hat{\alpha})}} \sim t_{n-2},$$

Ces deux résultats fondamentaux vont permettre de construire des tests d'hypothèses sur les paramètres  $\alpha$  et  $\beta$ .

## 5.7 Intervalles de confiance sur les coefficients de régression

À partir des résultats de la section précédente, on peut construire des intervalles de confiance sur  $\alpha$  et  $\beta$  :

$$\begin{aligned} \text{IC}(0.95) &= [\hat{\alpha} - t_{n-2,0.975} \sqrt{\widehat{\text{var}}(\hat{\alpha})}; \hat{\alpha} + t_{n-2,0.975} \sqrt{\widehat{\text{var}}(\hat{\alpha})}], \\ \text{IC}(0.95) &= [\hat{\beta} - t_{n-2,0.975} \sqrt{\widehat{\text{var}}(\hat{\beta})}; \hat{\beta} + t_{n-2,0.975} \sqrt{\widehat{\text{var}}(\hat{\beta})}]. \end{aligned}$$

## 5.8 Test d'hypothèses sur un coefficient de régression

Les résultats de la section précédente vont permettre de tester des hypothèses sur les coefficients de régression.

$$(i) \begin{cases} H_0 : \alpha = \alpha_0 \\ H_1 : \alpha \neq \alpha_0 \end{cases} \quad \text{et} \quad (ii) \begin{cases} H_0 : \beta = \beta_0 \\ H_1 : \beta \neq \beta_0. \end{cases}$$

Les logiciels statistiques testent en général la nullité des coefficients et prennent donc comme valeur des paramètres sous l'hypothèse  $\alpha_0 = 0$  et  $\beta_0 = 0$ , mais il n'est pas plus difficile de construire une règle de décision pour une valeur générale de  $\alpha_0$  et  $\beta_0$ .

Grâce au résultat 5.7, on a respectivement sous les hypothèses nulles (i) et (ii) :

$$t_\beta = \frac{\sqrt{ns_x}(\hat{\beta} - \beta_0)}{\hat{\sigma}_\varepsilon} \sim t_{n-2} \quad \text{et} \quad t_\alpha = \frac{\sqrt{ns_x}(\hat{\alpha} - \alpha_0)}{\hat{\sigma}_\varepsilon \sqrt{s_x^2 + \bar{x}^2}} \sim t_{n-2},$$

On rejette les hypothèses nulles, (i) et (ii), si

$$|t_\alpha| > t_{n-2;1-\alpha/2} \quad \text{et} \quad |t_\beta| > t_{n-2;1-\alpha/2}$$

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

Les logiciels calculent les  $p$ -valeurs qui sont respectivement :

$$p_\alpha = \Pr(t_{n-2} > |t_\alpha| \mid H_0 \text{ vraie}) \quad \text{et} \quad p_\beta = \Pr(t_{n-2} > |t_\beta| \mid H_0 \text{ vraie}).$$

Si cette valeur  $p$  est plus petite que l'erreur de première espèce  $\alpha$  alors on rejette l'hypothèse nulle.

## 5.9 Le tableau d'analyse de la variance

Dans le cadre de la régression multivariée, le tableau d'analyse de la variance sert à réaliser un test d'hypothèse sur la nullité de tous les coefficients de régression à l'exception de la constante. Pour la régression avec seulement une variable explicative et une constante, ce tableau sert à tester la nullité de la pente de la droite de régression. Le test s'écrit donc :

$$\begin{cases} H_0 & \beta = 0 \\ H_1 & \beta \neq 0 \end{cases}$$

Le tableau contient :

- Les sommes des carrés totale, de la régression et des résidus.
- Les degrés de liberté associés aux sommes de carrés.
- Les carrés moyens.
- La statistique de test.
- la  $p$ -valeur.

TABLE 5.4 – Tableau d'analyse de la variance

Source de variation	Sommes des carrés	Degrés de liberté	Carrés moyens	$F_{obs}$	$p$
Régression	$SC_{regr}$	1	$CM_{regr} = \frac{SC_{regr}}{1}$	$F_c = CM_{regr}/CM_{res}$	$\Pr(F_{1,n-2} > F_c \mid H_0 \text{ vraie})$
Résiduelle	$SC_{res}$	$n - 2$	$CM_{res} = \frac{SC_{res}}{n - 2}$		
Totale	$SC_{tot}$	$n - 1$	$CM_{tot} = \frac{SC_{tot}}{n - 1}$		

Les degrés de liberté sont le nombre de composantes aléatoires libres du vecteur dont on fait la somme des carrés. Autrement dit, le nombre de degrés de liberté est le nombre de composantes aléatoires minimales du vecteur que l'on doit connaître pour pouvoir déterminer tout le vecteur. Ce nombre de degrés de liberté vaut :

- $n - 1$  pour la  $SC_{TOT}$ . En effet, la  $SC_{TOT}$  utilise  $n$  observations centrées  $y_i - \bar{y}$ . Ces observations centrées sont soumises à une contrainte :

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

Si  $n - 1$  valeurs  $y_i - \bar{y}$  sont connues, on peut calculer la valeur inconnue, on a donc  $n - 1$  degrés de liberté.

- 1 pour la  $SC_{REGR}$ . En effet, cette somme de carrés peut s'écrire  $SC_{REGR} = \hat{\beta}^2 n s_x^2$  et ne dépend donc que d'une composante aléatoire  $\hat{\beta}$ .
- $n - 2$  pour la  $SC_{RES}$ . En effet, on utilise  $n$  résidus  $e_i$ , mais les résidus sont soumis à deux contraintes :

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

Si  $n - 2$  résidus  $e_i$  sont connus, on peut calculer les deux valeurs restantes au moyen des contraintes. On a donc  $n - 2$  degrés de liberté.

Les carrés moyens sont les sommes de carrés divisées par les degrés de liberté :

$$CM_{TOT} = \frac{SC_{TOT}}{n - 1}, \quad CM_{REGR} = \frac{SC_{REGR}}{1}, \quad \text{et } CM_{RES} = \frac{SC_{RES}}{n - 2}.$$

Le Tableau récapitulatif 5.3 nous permet directement de calculer les espérances des carrés moyens

$$E(CM_{TOT}) = \sigma_\varepsilon^2 + \beta^2 s_x^2 \frac{n}{n - 1}, \quad E(CM_{REGR}) = \sigma_\varepsilon^2 + \beta^2 n s_x^2, \quad \text{et } E(CM_{RES}) = \sigma_\varepsilon^2.$$

On constate que si  $\beta = 0$ , alors les trois carrés moyens sont des estimateurs sans biais de  $\sigma_\varepsilon^2$ . Si  $\beta$  est différent de 0, le carré moyen de la régression sera beaucoup plus grand que les deux autres.

Enfin, la statistique de test est définie par :

$$F_c = \frac{CM_{REGR}}{CM_{RES}}$$

Il est possible de montrer que, sous  $H_0$ ,  $F_c$  a une distribution de Fisher à 1 et  $n - 2$  degrés de liberté, ce qui permet de calculer la  $p$ -valeur qui est défini par

$$p = \Pr(F_{1,n-2} > F_c \mid H_0 \text{ vraie}),$$

où  $F_{1,n-2}$  est une variable aléatoire de Fisher à 1 et  $n - 2$  degrés de liberté. Si  $p$  est inférieur à la probabilité de commettre une erreur de première espèce  $\alpha$  (généralement 5%) on rejette  $H_0$ .

## 5.10 Exemple et interprétation

Si on utilise le langage R, il est assez facile de réaliser une régression linéaire et de tester les hypothèses décrites ci-dessus.

```
poids=c(60,61,64,67,68,69,70,70,72,73,75,76,78,80,85,90,96,96,98,101)
taille=c(155,162,157,170,164,162,169,170,178,173,180,175,173,175,179,175,180,185,189,187)
modele=lm(poids ~taille)
modele
summary(modele)
anova(modele)
res=resid(modele)
summary(res)
ajust=predict(modele)
```

Dans ce code, on définit d'abord les variables poids et taille. Ensuite, on calcule un objet appelé `modele` contenant la régression de la variable poids par la variable taille. On peut enfin calculer un certain nombre de statistique à partir de l'objet `modele` : un résumé des statistiques principales, le tableau d'analyse de la variance, une variable contenant les résidus et une variable avec les valeurs ajustées.

```
> poids=c(60,61,64,67,68,69,70,70,72,73,75,76,78,80,85,90,96,96,98,101)
> taille=c(155,162,157,170,164,162,169,170,178,173,180,175,173,175,179,175,180,185,189,187)
> modele=lm(poids ~taille)
> modele
```

```
Call:
lm(formula = poids ~ taille)
```

```
Coefficients:
(Intercept)      taille
   -126.614         1.180
```

```
> summary(modele)
```

```
Call:
lm(formula = poids ~ taille)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-11.4692  -3.9532   0.3912   4.3054  10.1703
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
```

```

(Intercept) -126.6142    26.3003   -4.814 0.000139 ***
taille      1.1802      0.1519    7.770 3.7e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.27 on 18 degrees of freedom
Multiple R-Squared:  0.7703,    Adjusted R-squared:  0.7576
F-statistic: 60.37 on 1 and 18 DF,  p-value: 3.703e-07

> anova(modele)
Analysis of Variance Table

Response: poids
      Df Sum Sq Mean Sq F value    Pr(>F)
taille  1 2373.35 2373.35  60.374 3.703e-07 ***
Residuals 18  707.60   39.31
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> res=resid(modele)
> summary(res)
      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
-1.147e+01 -3.953e+00  3.912e-01 -8.327e-17  4.305e+00  1.017e+01
> ajust=predict(modele)

```

Les premiers résultats donnent les coefficients du modèle. La constante (en anglais *intercept*)  $\hat{\alpha} = -126.614$  et la pente  $\hat{\beta} = 1.180$ . Le modèle estimé est donc

$$\text{poids}_i = -126.614 + 1.180 \times \text{taille}_i + e_i.$$

Ensuite la commande `summary(modele)` permet d'examiner les quartiles des résidus, ce qui permet de se faire une idée sur la symétrie des résidus. Enfin, le Tableau 5.5 des coefficients nous donne les écart-types des coefficients, les statistiques  $t$  et la  $p$ -valeur.

TABLE 5.5 – Tableau des coefficients de régression

variable	Coefficients	Écart-type estimé	Erreur $t$	valeur $\Pr(>  t    H_0 \text{ vraie})$
constante	$\hat{\alpha} = -126.6142$	$\sqrt{\widehat{\text{var}}(\hat{\alpha})} = 26.3003$	$t_\alpha = \frac{\hat{\alpha}}{\sqrt{\widehat{\text{var}}(\hat{\alpha})}} = -4.814$	$p = 0.000139$
taille	$\hat{\beta} = 1.1802$	$\sqrt{\widehat{\text{var}}(\hat{\beta})} = 0.1519$	$t_\beta = \frac{\hat{\beta}}{\sqrt{\widehat{\text{var}}(\hat{\beta})}} = 7.770$	$p = 3.7e - 07$

Il ressort de ce tableau que les  $p$ -valeurs pour les coefficients sont très largement inférieures à 5%. On rejette donc les hypothèses que ces coefficients sont nuls. Ils sont donc significativement différents de zéros. Ensuite, le logiciel R affiche l'écart-type estimé des résidus  $\hat{\sigma}_e$ , le nombre de degrés de liberté associé à la somme des carrés de la régression, le  $R^2 = 0.7703$ , et le  $R^2$  corrigé 0.7576, ce qui indique que 77% de la variance de la variable 'poids' peut être expliquée par la variable 'taille'.

Enfin, quelques statistiques très résumées du tableau d'analyse de la variance sont présentées. La statistique  $F_{obs} = 60.37$  a sous  $H_0$  une distribution de Fisher à 1 et  $n - 2 = 20 - 2 = 18$  degrés de liberté. La  $p$ -valeur vaut  $3.703e - 07$ . On rejette donc  $H_0$ . On remarque que la  $p$ -valeur est la même que pour le test sur la pente de la droite de régression, car il s'agit du même test quand on a seulement une variable explicative.

Pour obtenir le tableau d'analyse de la variance, on utilise la commande : `anova(modele)` qui donne le tableau avec la somme des carrés due à la régression et la somme des carrés des résidus.

Enfin, il est possible de calculer les résidus et les valeurs ajustées au moyen des fonctions `predict` et `resid`.

TABLE 5.6 – Tableau d’analyse de la variance

	DL	Som. Car.	Car. Moy.	$F_{obs}$	$\Pr(> F   H_0 \text{ vraie})$
taille	1	2373.35	2373.35	60.374	3.703e-07
Residuals	18	707.60	39.31		

# Chapitre 6

## Le modèle linéaire général

### 6.1 Le modèle

#### 6.1.1 Définition du modèle linéaire général

Le modèle linéaire est une généralisation de la régression au cas où  $p$  variables explicatives sont utilisées pour expliquer la variable dépendante  $xy$ . 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

- $y_i$  est la valeur prise par la variable dépendante sur l'individu  $i$ , les  $y_i$  sont des variables aléatoires,
- $x_{ij}$  représente la valeur prise par la  $j$ ème variable explicative sur l'individu  $i$ , les  $x_{ij}$  sont supposés non-aléatoires,
- $\beta_j$  est la  $j$ ème composante du coefficient de régression,
- les  $\varepsilon_i$  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  $i$ .

### 6.2 Régression avec les résidus normaux

#### 6.2.1 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  $\varepsilon_i$  sont d'espérances nulles.
- Les termes d'erreur  $\varepsilon_i$  sont non-corrélés.
- Tous les  $\varepsilon_i$  ont la même variance (homoscédasticité).

#### 6.2.2 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  $\varepsilon_i$ .

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

### 6.2.3 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  $\sigma_\varepsilon^2$  est un scalaire.

Cette formulation est équivalente à la précédente.

## 6.3 Estimation du modèle

### 6.3.1 Estimation par les moindres carrés (ordinaires)

L'objectif est d'estimer  $\boldsymbol{\beta}$  et  $\sigma_\varepsilon^2$ . 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 3.2) 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 6.1** 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 6.1** 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 6.2**  $\text{var}(\hat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}$ .

### Démonstration

Comme

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

on a

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

□

On appelle erreur-standard ou erreur-type de  $\widehat{\beta}_j$  l'écart-type de cet estimateur

$$\sigma(\widehat{\beta}_j) = \sqrt{\text{var}(\widehat{\beta}_j)} = \sigma_{\varepsilon} \sqrt{[(\mathbf{X}'\mathbf{X})^{-1}]_{jj}}$$

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

**Théorème 6.3** (de Gauss-Markov) *L'estimateur  $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  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,

$$\mathbf{B}\mathbf{X} = \mathbf{0}. \tag{6.1}$$

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 (6.1), on a finalement

$$\text{var}(\boldsymbol{\beta}^*) = \{\mathbf{B}\mathbf{B}' + (\mathbf{X}'\mathbf{X})^{-1}\} \sigma_{\varepsilon}^2. \tag{6.2}$$

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  $\sigma_{\varepsilon}^2$  pour estimer la variance de  $\widehat{\boldsymbol{\beta}}$ . Le vecteur des termes d'erreur  $\boldsymbol{\varepsilon}$  peut être estimé par :

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

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



**Lemme 6.1** 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  $\mathbb{E}(u_i u_j) = 0$ , quand  $j \neq i$ . Donc,

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

□

Grâce au Lemme 6.1, on peut calculer l'espérance de  $\mathbf{e}'\mathbf{e}$ .

**Théorème 6.4** Soit  $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ , alors

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

**Démonstration**

Nous avons vu en Section 3.1 que  $\mathbf{e}$  peut également s'écrire

$$\mathbf{e} = (\mathbf{I} - \mathbf{P}_X) \mathbf{y}, \quad (6.3)$$

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 6.1, on obtient

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

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

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

□

Le Théorème 6.4 nous permet de construire un estimateur sans biais pour  $\sigma_\varepsilon^2$  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  $(\mathbf{I} - \mathbf{P}_X)$ .

TABLE 6.1 – Tableau récapitulatif

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

### 6.3.2 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  $\varepsilon_i$  sont des variables aléatoires indépendantes ayant des distributions normales de moyennes nulles et de variance  $\sigma_\varepsilon^2$ .

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 (4.3), on a

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

On se trouve dans un problème paramétrique classique. Comme  $\mathbf{y}$  et  $\mathbf{X}$  sont observés, on va estimer les paramètres  $\boldsymbol{\beta}$  et  $\sigma_\varepsilon^2$ .

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 :

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

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 :

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

L'estimateur du maximum de vraisemblance de  $\sigma_\varepsilon^2$  est donné par

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

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

### 6.3.3 Propriétés des estimateurs du maximum de vraisemblance

Rappelons quelques propriétés des estimateurs :

- Un estimateur  $\hat{\theta}$  d'un paramètre  $\theta$  est sans biais, si  $E(\hat{\theta}) = \theta$  pour toute valeur de  $\theta$ .
- 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  $\varepsilon$  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  $\theta$  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  $\sigma_\varepsilon^2$  est en effet biaisé.

### 6.3.4 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

$$\hat{\beta} \sim \mathcal{N}(\beta, (\mathbf{X}'\mathbf{X})^{-1} \sigma_\varepsilon^2). \quad (6.4)$$

**Lemme 6.2** 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  $\mathbf{\Gamma}\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}\mathbf{\Gamma}'\mathbf{I})$ , et  $\mathbf{\Gamma}'\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}'\mathbf{I}\mathbf{\Gamma})$ . 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 6.5** 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  $\chi_p^2$  à  $p$  degrés de liberté.

#### Démonstration

La matrice  $\mathbf{P}$  admet une décomposition en valeurs propres et vecteurs propres. En vertu du Théorème 2.2, si  $\mathbf{\Lambda}$  représente la matrice diagonale ayant les valeurs propres  $\lambda_i$  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 6.2,  $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ . En vertu du Théorème 2.4, 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

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

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

**Corrolaire 6.1** Dans le modèle linéaire général avec des termes d'erreur normaux,

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

En effet,

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

donc

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

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 6.5.  $\square$

**Corrolaire 6.2** 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,

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

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

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

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

**Théorème 6.6** 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 6.3** Dans le modèle linéaire avec des termes d'erreur normaux,

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

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  $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon}$  or  $(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{P}_X^\perp = \mathbf{0}$ , ce qui implique directement le corollaire.

**Théorème 6.7** 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  $\mathbf{C}\mathbf{A} = \mathbf{0}$ .

### 6.3.5 Synthèse des résultats

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

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

# Chapitre 7

## Inférence dans le modèle linéaire

### 7.1 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. La variable normale est indépendante de la chi-carré, ce qui définit une variable de Student à  $n-p$  degrés de liberté. En simplifiant, on obtient que

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

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  $\alpha$ , donné par :

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

## 7.2 Test d'un seul coefficient de régression

### 7.2.1 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  $\sigma(\hat{\beta}_j)$  est l'erreur-type de  $\hat{\beta}_j$  :

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

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

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

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

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

on obtient

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

De plus,

$$\frac{\hat{\beta}_j - \beta_{j0}}{\sigma_\varepsilon \sqrt{[(\mathbf{X}'\mathbf{X})^{-1}]_{jj}}} = \frac{\hat{\beta}_j - \beta_{j0}}{\sigma(\hat{\beta}_j)} \sim \mathcal{N}(0, 1)$$

Sous  $H_0$ , la statistique

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

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

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

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

### 7.2.2 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  $\varepsilon_i$  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  $\sigma^2$  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 6.3,  $\hat{\mu}$  et  $\hat{\sigma}^2$  sont indépendants. De plus on a, par l'expression (6.4) :

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

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 :

$$\mathbf{P}_c = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n} = \begin{pmatrix} 1-1/n & -1/n & -1/n & \dots & -1/n \\ -1/n & 1-1/n & -1/n & \dots & -1/n \\ -1/n & -1/n & 1-1/n & \dots & -1/n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & -1/n & \dots & 1-1/n \end{pmatrix}. \quad (7.1)$$

Les variables aléatoires  $d$  et  $K$  sont indépendantes. De plus, par le Théorème 6.5,  $K \sim \chi_{n-1}^2$ . 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.

## 7.3 Tests de Wald sur les coefficients de régression

### 7.3.1 Test général d'une contrainte linéaire

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

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}, \quad (7.2)$$

contre l'hypothèse alternative

$$H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}, \quad (7.3)$$

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 7.1

- Le test  $H_0 : \beta_j = c$  s'obtient en prenant  $\mathbf{R} = (0 \cdots 0 \underbrace{1}_{j^{\text{ième}}} 0 \cdots 0)$  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 :

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \text{var}(\mathbf{R}\hat{\boldsymbol{\beta}})^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \frac{1}{\sigma_{\varepsilon}^2} \boldsymbol{\varepsilon}' \mathbf{W} \boldsymbol{\varepsilon}, \quad (7.4)$$

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 6.5, on obtient donc que

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

et donc

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \text{var}(\mathbf{R}\hat{\boldsymbol{\beta}})^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \frac{1}{\sigma_{\varepsilon}^2} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left\{ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right\}^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \sim \chi_q^2. \quad (7.5)$$

Si la forme quadratique (7.4) est grande, on soupçonne  $H_0$  d'être faux. Cependant, on ne peut réaliser directement un test  $\chi^2$  car l'expression (7.5) dépend de  $\sigma_{\varepsilon}^2$  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 6.7, 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

$$F_c = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left\{ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right\}^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \frac{1}{q}}{\mathbf{e}' \mathbf{e} \frac{1}{n-p}}. \quad (7.6)$$

Sous  $H_0$ , le numérateur et le dénominateur de  $F_c$  sont indépendants, et ont, à une constante près, une distribution  $\chi^2$ . La statistique de test  $F_c$  a donc une distribution de Fisher à  $q$  et  $n - p$  degrés de liberté. Donc, en notant  $\alpha$  l'erreur de première espèce, on rejette l'hypothèse 7.2, 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.



### 7.3.2 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 :

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

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 7.1**

$$\left\{ \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \right\}^{-1} = \tilde{\mathbf{X}}'\mathbf{P}_c\tilde{\mathbf{X}} = n\boldsymbol{\Sigma}, \quad (7.7)$$

où  $\mathbf{P}_c$  est l'opérateur qui centre les données déjà présenté dans l'Expression (3.7)

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

$\boldsymbol{\Sigma}$  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

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

où

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

et

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

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 (3.2). □

L'Expression (7.5) est alors la somme des carrés de la régression (voir Expression (3.10)) :

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

En considérant l'Expression (3.3), la statistique de test (7.6) devient :

$$F_c = \frac{\text{SC}_{regr}/(p-1)}{\text{SC}_{res}/(n-p)}, \quad (7.8)$$

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 7.1).

TABLE 7.1 – Tableau d'analyse de la variance

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

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.

### 7.3.3 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

$$\begin{aligned} - q &= 1, \\ - \mathbf{R} &= (0 \dots \underbrace{1}_{\text{unité } j} \dots 0), \end{aligned}$$

$$- \mathbf{r} = \beta_{j0}.$$

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 (7.6) 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 7.2.1. 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 4.8.3).

## 7.4 R-carré et R-carré ajusté

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

$$R^2 = \frac{\sum_{i=1}^n (y_i^* - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{SC_{Regr}}{SC_{Tot}} = 1 - \frac{SC_{Res}}{SC_{Tot}}.$$

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

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

## 7.5 Prévision ponctuelle d'une valeur

### 7.5.1 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 :

$$\hat{y}_j = (x_{j1} \cdots x_{jp}) \hat{\boldsymbol{\beta}}.$$

Le prédicteur peut également s'écrire

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

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

$$E(\hat{y}_j - y_j) = E\{\mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon} - \varepsilon_j\} = \mathbf{x}_j (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\boldsymbol{\varepsilon}) - E(\varepsilon_j) = 0.$$

Comme la valeur prédite se réfère à une nouvelle observation,

$$E(\varepsilon_j \boldsymbol{\varepsilon}) = \mathbf{0},$$

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  $\varepsilon_j$ .

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 :

$$\text{IC}(1-\alpha) = \left[ \hat{y}_j - t_{1-\alpha/2, n-p} \sqrt{\widehat{\text{var}}(\hat{y}_j - y_j)}; \hat{y}_j + t_{1-\alpha/2, n-p} \sqrt{\widehat{\text{var}}(\hat{y}_j - y_j)} \right].$$

## 7.5.2 Cas bivarié

Dans le cas où une seule variable explicative  $x$  et une constante sont utilisées, on a

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

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{ns_x^2} \begin{pmatrix} s_x^2 + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.$$

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

## 7.6 Un exemple

The savings data frame has 50 rows (countries) and 5 columns (Belsley, D., Kuh. E. and Welsch, R. (1980) "Regression Diagnostics" Wiley. ). The data is averaged over the period 1960-1970. The data are presented in Table 7.2.

- sr savings rate - personal saving divided by disposable income
- pop15 percent population under age of 15
- pop75 percent population over age of 75
- dpi per-capita disposable income in dollars

TABLE 7.2 – Data savings

Contry	sr	pop15	pop75	dpi	ddpi
Australia	11.43	29.35	2.87	2329.68	2.87
Austria	12.07	23.32	4.41	1507.99	3.93
Belgium	13.17	23.80	4.43	2108.47	3.82
Bolivia	5.75	41.89	1.67	189.13	0.22
Brazil	12.88	42.19	0.83	728.47	4.56
Canada	8.79	31.72	2.85	2982.88	2.43
Chile	0.60	39.74	1.34	662.86	2.67
China	11.90	44.75	0.67	289.52	6.51
Colombia	4.98	46.64	1.06	276.65	3.08
Costa Rica	10.78	47.64	1.14	471.24	2.80
Denmark	16.85	24.42	3.93	2496.53	3.99
Ecuador	3.59	46.31	1.19	287.77	2.19
Finland	11.24	27.84	2.37	1681.25	4.32
France	12.64	25.06	4.70	2213.82	4.52
Germany	12.55	23.31	3.35	2457.12	3.44
Greece	10.67	25.62	3.10	870.85	6.28
Guatamala	3.01	46.05	0.87	289.71	1.48
Honduras	7.70	47.32	0.58	232.44	3.19
Iceland	1.27	34.03	3.08	1900.10	1.12
India	9.00	41.31	0.96	88.94	1.54
Ireland	11.34	31.16	4.19	1139.95	2.99
Italy	14.28	24.52	3.48	1390.00	3.54
Japan	21.10	27.01	1.91	1257.28	8.21
Korea	3.98	41.74	0.91	207.68	5.81
Luxembourg	10.35	21.80	3.73	2449.39	1.57
Malta	15.48	32.54	2.47	601.05	8.12
Norway	10.25	25.95	3.67	2231.03	3.62
Netherlands	14.65	24.71	3.25	1740.70	7.66
New Zealand	10.67	32.61	3.17	1487.52	1.76
Nicaragua	7.30	45.04	1.21	325.54	2.48
Panama	4.44	43.56	1.20	568.56	3.61
Paraguay	2.02	41.18	1.05	220.56	1.03
Peru	12.70	44.19	1.28	400.06	0.67
Philippines	12.78	46.26	1.12	152.01	2.00
Portugal	12.49	28.96	2.85	579.51	7.48
South Africa	11.14	31.94	2.28	651.11	2.19
South Rhodesia	13.30	31.92	1.52	250.96	2.00
Spain	11.77	27.74	2.87	768.79	4.35
Sweden	6.86	21.44	4.54	3299.49	3.01
Switzerland	14.13	23.49	3.73	2630.96	2.70
Turkey	5.13	43.42	1.08	389.66	2.96
Tunisia	2.81	46.12	1.21	249.87	1.13
United Kingdom	7.81	23.27	4.46	1813.93	2.01
United States	7.56	29.81	3.43	4001.89	2.45
Venezuela	9.22	46.40	0.90	813.39	0.53
Zambia	18.56	45.25	0.56	138.33	5.14
Jamaica	7.72	41.12	1.73	380.47	10.23
Uruguay	9.24	28.13	2.72	766.54	1.88
Libya	8.89	43.69	2.07	123.58	16.71
Malaysia	4.71	47.20	0.66	242.69	5.08

– ddpi percent growth rate of dpi

```
#  
# On utilise le package faraway qui a été installé préalablement  
#  
library(faraway)  
#  
# On utilise les données savings  
#  
data(savings)  
attach(savings)  
#  
# Description des variable de "savings"  
#  
plot(savings)
```

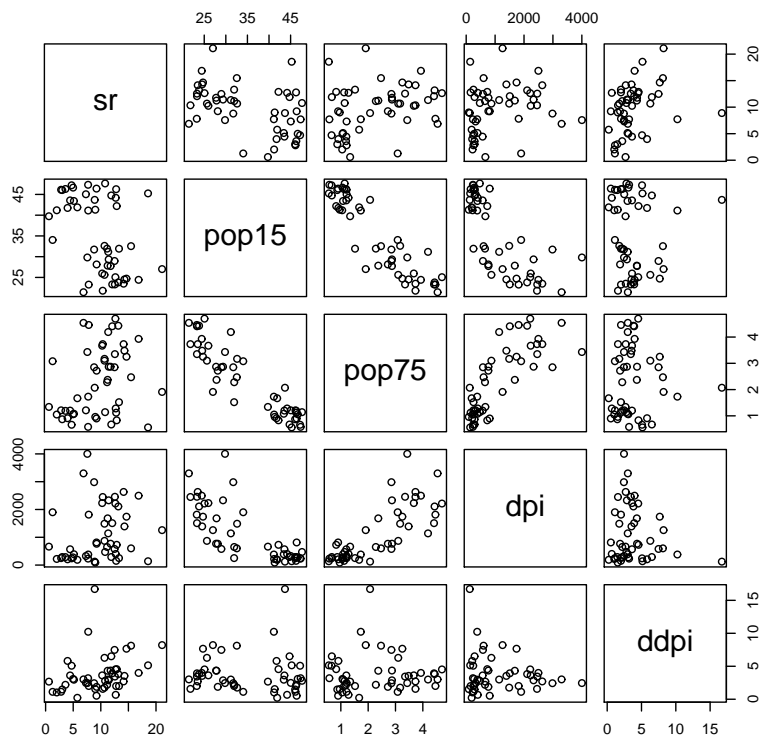


FIGURE 7.1 – Nuages de points de la base de données savings

```
> ll=lm(sr ~ pop15+pop75+dpi+ddpi)  
> summary(ll)
```

```
Call:  
lm(formula = sr ~ pop15 + pop75 + dpi + ddpi)
```

```
Residuals:  
    Min       1Q   Median       3Q      Max  
-8.2422 -2.6857 -0.2488  2.4280  9.7509
```

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 28.5660865  7.3545161   3.884 0.000334 ***
pop15       -0.4611931  0.1446422  -3.189 0.002603 **
pop75       -1.6914977  1.0835989  -1.561 0.125530
dpi         -0.0003369  0.0009311  -0.362 0.719173
ddpi         0.4096949  0.1961971   2.088 0.042471 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.803 on 45 degrees of freedom
Multiple R-Squared:  0.3385,    Adjusted R-squared:  0.2797
F-statistic: 5.756 on 4 and 45 DF,  p-value: 0.0007904

```

La statistique F vaut 5.756, et l'on rejette l'hypothèse que tous les coefficients de régression (à l'exception de la constante) sont nuls. On constate cependant que les variables `dpi` et `pop75` n'ont pas leurs coefficients de régression significativement différents de 0. On recommence la régression en enlevant la variable `dpi` la moins significative.

```

> ll=lm(sr ~ pop15+pop75+ddpi)
> summary(ll)

Call:
lm(formula = sr ~ pop15 + pop75 + ddpi)

Residuals:
    Min       1Q   Median       3Q      Max
-8.2539 -2.6159 -0.3913  2.3344  9.7070

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.1247     7.1838   3.915 0.000297 ***
pop15        -0.4518     0.1409  -3.206 0.002452 **
pop75        -1.8354     0.9984  -1.838 0.072473 .
ddpi          0.4278     0.1879   2.277 0.027478 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.767 on 46 degrees of freedom
Multiple R-Squared:  0.3365,    Adjusted R-squared:  0.2933
F-statistic: 7.778 on 3 and 46 DF,  p-value: 0.0002646

```

La variable `pop75` n'est pas significative. On recommence la régression en enlevant la variable `pop75`.

```

> ll=lm(sr ~ pop15+ddpi)
> summary(ll)

Call:
lm(formula = sr ~ pop15 + ddpi)

Residuals:
    Min       1Q   Median       3Q      Max
-7.58314 -2.86323  0.04535  2.22734 10.47530

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.59958     2.33439   6.682 2.48e-08 ***

```

```
pop15      -0.21638    0.06033   -3.586  0.000796 ***
ddpi       0.44283     0.19240    2.302  0.025837 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.861 on 47 degrees of freedom
Multiple R-Squared: 0.2878,    Adjusted R-squared: 0.2575
F-statistic: 9.496 on 2 and 47 DF,  p-value: 0.0003438
```



# Chapitre 8

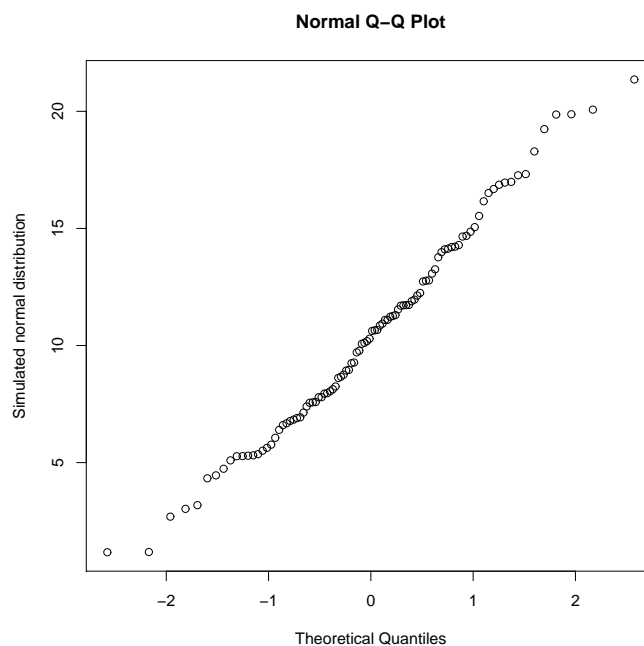
## Outils de détection en régression

### 8.1 Le qq-plot

Un qq-plot est un graphique statistique permettant de comparer la distribution de probabilité d'une distribution empirique issue d'un échantillon à la distribution théorique. Cette distribution théorique est en général une distribution normale.

Dans les Figures 8.1, 8.2 et 8.3, 100 variables aléatoires indépendantes respectivement normales, exponentielles, et de Student (avec 2 degrés de liberté) ont été générées. On compare ces trois échantillons avec les quantiles d'une variable aléatoire normale centrée réduite. Si les points sont alignés le long d'une droite, la distribution simulée est bien une distribution normale, ce qui n'est pas le cas pour les Figures 8.2 et 8.3.

FIGURE 8.1 – qq-plot de 100 variables normales simulées versus une distribution théorique normale



En langage R, on utilise le code suivant :

```
# Graphique 1
x=rnorm(100,10,5)
qqnorm(x,ylab = "Simulated normal distribution")
# Graphique 2
x=rexp(100,1)
```

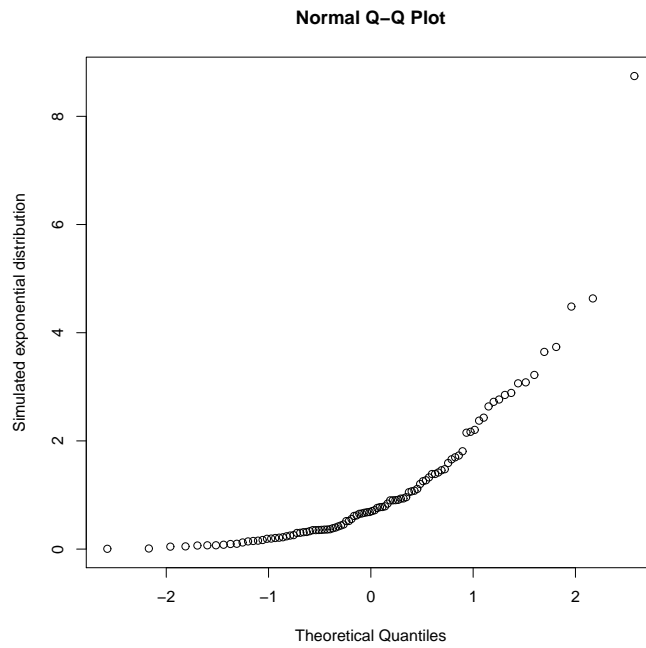
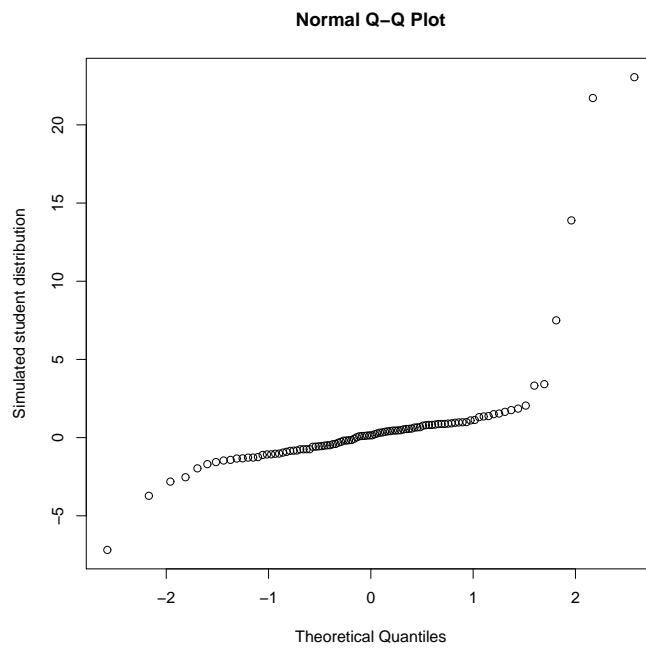


FIGURE 8.2 – qq-plot de 100 variables exponentielles simulées versus une distribution théorique normale

FIGURE 8.3 – qq-plot de 100 variables de Student avec 2 degrés de libertés simulées versus une distribution théorique normale



```
qqnorm(x,ylab = "Simulated exponential distribution")
# Graphique 3
x=rt(100,2)
qqnorm(x,ylab = "Simulated Student distribution")
```

En régression, on réalise en général un qq-plot des résidus de la régression. Ce graphique permet de se faire une idée de la normalité des résidus de la régression.

## 8.2 Points leviers

Les valeurs ajustées peuvent s'écrire comme la projection du vecteur  $\mathbf{y}$  sur le sous-espace engendré par les colonnes de la matrice  $\mathbf{X}$ , ce qui s'écrit :

$$\mathbf{y}^* = \mathbf{P}_X \mathbf{y}. \quad (8.1)$$

La matrice  $\mathbf{P}_X$  est une matrice idempotente et est parfois notée  $\mathbf{H}$  et appelée *hat-matrix* (matrice chapeau) car  $\mathbf{H}$  est un opérateur qui permet de passer de  $\mathbf{y}$  aux valeurs ajustées appelées aussi valeurs prédites  $\mathbf{y}^*$  de  $\mathbf{y}$ . Le vecteur  $\mathbf{y}^*$  est parfois noté  $\hat{\mathbf{y}}$ . L'opérateur  $\mathbf{H}$  est celui qui met un chapeau sur  $\mathbf{y}$  au sens où on peut ré-écrire l'Expression (8.1) sous la forme.

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}.$$

La variance de  $\mathbf{y}^*$  est

$$\text{var}(\mathbf{y}^*) = \text{var}(\mathbf{P}_X \mathbf{y}) = \mathbf{P}_X \text{var}(\mathbf{y}) \mathbf{P}_X = \mathbf{P}_X \mathbf{I} \sigma_\varepsilon^2 \mathbf{P}_X = \sigma_\varepsilon^2 \mathbf{P}_X.$$

En notant  $h_i$  la composante correspondant à la  $i$ ème ligne et la  $i$ ème colonne de  $\mathbf{P}_X$  i.e.  $h_i = [\mathbf{P}_X]_{ii}$ , on a

$$\text{var}(y_i^*) = h_i \sigma_\varepsilon^2.$$

La valeur  $h_i$  est appelée le levier de l'observation  $i$ . En effet, plus  $h_i$  est élevé, plus  $y_i$  contribuera à  $y_i^*$ . Nous avons montré que toutes les valeurs propres des matrices idempotente sont égales à 0 ou à 1. Donc, pour tous les vecteurs normés  $\mathbf{u}$ , on a  $0 \leq \mathbf{u}' \mathbf{P} \mathbf{u} \leq 1$ . En prenant un vecteur  $\mathbf{v}$  composé de 0 et d'un seul 1 à la  $i$ ème composante on obtient  $\mathbf{v}' \mathbf{P} \mathbf{v} = h_i$ , ce qui implique que

$$0 \leq h_i \leq 1, \text{ pour tout } i = 1, \dots, n.$$

De plus, comme  $\mathbf{P}_X$  est une matrice idempotente de rang  $p$ , la trace de  $\mathbf{P}_X$  vaut  $p$ . Donc

$$\frac{1}{n} \sum_{h=1}^n h_i = \frac{p}{n}.$$

Un individu ayant une valeur de  $h_i > p/n$  a donc un levier important. Plus précisément, on dit en général que  $i$  est un point levier si  $h_i > 2p/n$ .

## 8.3 Résidus standardisés et studentisés

Les résidus du modèle peuvent s'écrire comme la projection du vecteur  $\mathbf{y}$  sur l'orthogonal du sous-espace engendré par les colonnes de la matrice  $\mathbf{X}$ , ce qui s'écrit

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

La variance de  $\mathbf{e}$  est donc

$$\text{var}(\mathbf{e}) = \text{var}[(\mathbf{I} - \mathbf{P}_X) \mathbf{y}] = (\mathbf{I} - \mathbf{P}_X) \text{var}(\mathbf{y}) (\mathbf{I} - \mathbf{P}_X)' = (\mathbf{I} - \mathbf{P}_X) \mathbf{I} \sigma_\varepsilon^2 (\mathbf{I} - \mathbf{P}_X) = \sigma_\varepsilon^2 (\mathbf{I} - \mathbf{P}_X).$$

On peut donc écrire

$$\text{var}(e_i) = (1 - h_i) \sigma_\varepsilon^2.$$

Les résidus ne sont donc pas homoscedastiques (de même variance) alors que les termes d'erreurs du modèle sont par définition homoscedastiques. En effet, selon la définition du modèle linéaire, on a

$$\text{var}(\varepsilon_i) = \sigma_\varepsilon^2.$$

Un résidu standardisé est un résidu divisé par son écart-type estimé :

$$r_i = \frac{e_i}{\sqrt{\widehat{\text{var}}(e_i)}} = \frac{e_i}{\hat{\sigma}_\varepsilon \sqrt{1 - h_i}}.$$

La statistique  $r_i$  a cependant un problème :  $e_i$  n'est pas indépendant de  $\hat{\sigma}_\varepsilon$ . Il n'est donc pas possible de déduire la distribution de probabilité des  $r_i$ . On préférera calculer des résidus studentisés défini par

$$t_i = \frac{e_i}{\hat{\sigma}_{(i)\varepsilon} \sqrt{1 - h_i}},$$

où  $\hat{\sigma}_{(i)\varepsilon}^2$  est la variance estimée des erreurs en réalisant une régression sans l'individu  $i$ . En procédant ainsi, le numérateur et le dénominateur de  $t_i$  sont indépendants. Les  $t_i$  ont donc une distribution de Student à  $n - p - 1$  degrés de liberté.

## 8.4 Distance de Cook

La distance de Cook est un indice calculable pour chaque unité statistique  $i$  qui permet de mesurer l'influence d'une observation sur la régression. La distance de Cook mesure l'effet produit par la suppression de l'observation  $i$  sur les valeurs ajustées :

$$D_i = \sum_{j=1}^n \frac{[y_j^* - y_j^*(i)]^2}{p \hat{\sigma}_\varepsilon^2}$$

où

- $y_j^*$  est la valeur ajustée de la régression,
- $y_j^*(i)$  est la valeur ajustée de la régression quand l'observation  $i$  est supprimée.

On considère en général que la distance de Cook est importante si elle est supérieure à 1.

Il est possible montrer que la distance de Cook peut être réécrite comme une fonction des leviers et des résidus :

$$D_i = \frac{e_i^2}{p \cdot \hat{\sigma}_\varepsilon^2} \left[ \frac{h_i}{(1 - h_i)^2} \right].$$

Cette expression est évidemment plus simple à calculer, car on ne doit pas calculer les  $n$  régressions en enlevant chacune des unités. Elle permet également de représenter sur un même graphique les leviers, les résidus, et la distance de Cook.

## 8.5 Un premier exemple

Au moyen du langage R on a généré 7 réalisations de deux variables factices  $X$  et  $Y$ , avec

$$X_i \sim N(10, 8), \quad Y_i = \frac{X}{3} + \varepsilon_i, \quad \varepsilon_i \sim N(0, 1).$$

Une 8ème observation a été ajoutée  $(x_8, y_8) = (30, 30)$ . Les données générées sont présentées dans le Tableau 8.1. La 8ème observation est un point aberrant (ou une valeur extrême) et apparaît clairement sur dans la Figure 8.4 où sont représentés le nuage de points et la droite de régression. La droite de régression vaut

$$y_i = -4.4932 + 0.9483x_i + e_i.$$

Le point aberrant va jouer un rôle de levier sur la droite de régression au sens où il va attirer la droite.

TABLE 8.1 – Variables générées : la 8ième observation est un point aberrant

i	1	2	3	4	5	6	7	8
X	7.843589	10.045501	-0.8862475	10.34440	7.673868	8.160382	18.415378	30
Y	3.570563	1.384940	0.9908995	4.48798	1.301344	2.706809	6.477405	30

Dans le Tableau 3, on a calculé les leviers  $h_i$ , résidus  $e_i$ , résidus standardisés  $r_i$ , résidus studentisés  $t_i$ , et distances de Cook  $D_i$  pour les 8 observations. On constate que pour l'observation 8,  $h_8 = 0.7121463 > 2p/n = 2 \times 2/8 = 1/2$ . La 8ème observation est donc un point levier. Cette observation est caractérisée par un résidu et un résidu studentisé fortement positif. La distance de Cook est très élevée pour le 8ème

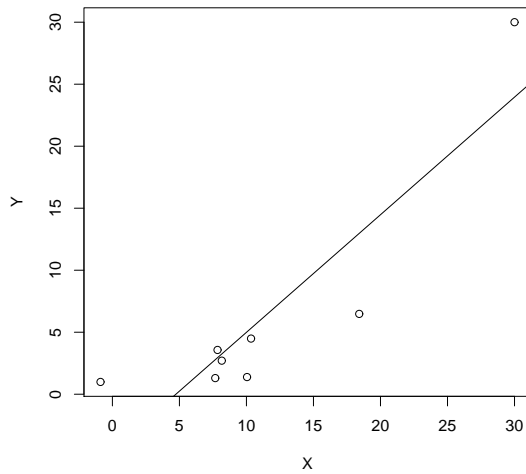


FIGURE 8.4 – Nuage de point et droite de régression : un point extrême joue un rôle de levier sur la droite de régression

observation ( $D_8 = 6.9452521$ ), ce qui signifie que la droite de régression sera très fortement modifiée si on enlève la 8ème observation.

Le langage R propose une série de graphiques pour diagnostiquer d'éventuels problèmes en régression linéaire. Les deux graphiques présentés en Figure 8.5 représentent respectivement les résidus standardisés en fonction des valeurs ajustées et un qq-plot des résidus. Dans le graphique représentant respectivement les résidus en fonction des valeurs ajustées, un lissage a été fait automatiquement pour pouvoir détecter une éventuelle tendance dans les résidus. Sans point aberrant, ce lissage devrait être proche d'une droite horizontale, ce qui n'est pas le cas dans cet exemple. Dans le qq-plot, les points ne sont pas non plus alignés le long d'une droite, les résidus n'ont donc pas une distribution normale.

Le premier graphique présenté en Figure 8.6 représente les racines des valeurs absolues des résidus standardisés  $\sqrt{|r_i|}$  en fonction des valeurs ajustées  $y_i^*$ . À nouveau, un lisseur permet d'évaluer la tendance. Ce graphique permet de voir si la dispersion des résidus dépend ou non des valeurs ajustées. Dans le cas présent, il y a une dépendance, qui découle de la présence d'un point aberrant. Le second graphique présenté en Figure 8.6 présente les résidus en fonction des leviers. Les points qui se trouvent au delà des traits tirés sont les points leviers, car leur distance de Cook dépasse respectivement 0.5 et 1. La 8ième observation est clairement aberrante. On en conclut que la régression devrait être réalisée sans cette observation.

TABLE 8.2 – Leviers  $h_i$ , résidus  $e_i$ , résidus standardisés  $r_i$ , résidus studentisés  $t_i$ , et distances de Cook  $D_i$  pour les 8 observations

i	1	2	3	4	5	6	7	8
$h_i$	0.1471869	0.1283639	0.3846443	0.1270841	0.1493245	0.1434598	0.2077901	0.7121463
$e_i$	0.6253298	-3.6484699	6.3245836	-0.8288918	-1.4829345	-0.5388546	-6.4935561	6.0427936
$r_i$	0.1424592	-0.8221510	1.6962008	-0.1866466	-0.3382583	-0.1224916	-1.5348676	2.3695196
$t_i$	0.1302674	-0.7967367	2.1462635	-0.1708811	-0.3117732	-0.1119591	-1.7978602	8.5349781
$D_i$	0.0017513	0.0497715	0.8992028	0.0025359	0.0100423	0.0012565	0.3089558	6.9452521

Tout ce traitement statistique a été réalisé au moyen de commandes R suivantes :

**En langage R**

```
X=rnorm(7,10,8); Y=rnorm(7,0,1)+X*1/3
X=c(X,30); Y=c(Y,30)
g=lm(Y~X)
```

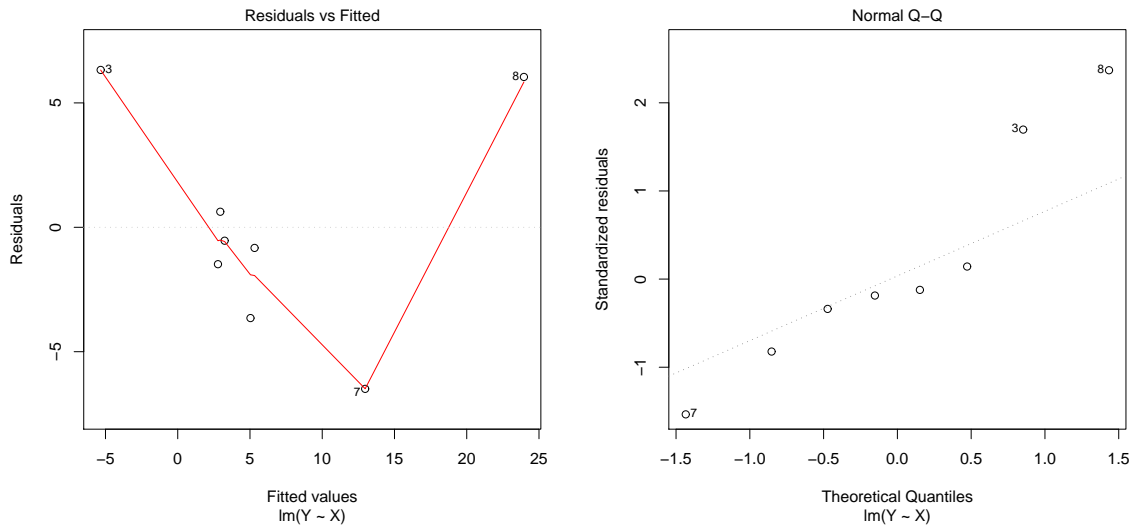


FIGURE 8.5 – Résidus standardisés en fonction des valeurs ajustées et qq-plot des résidus

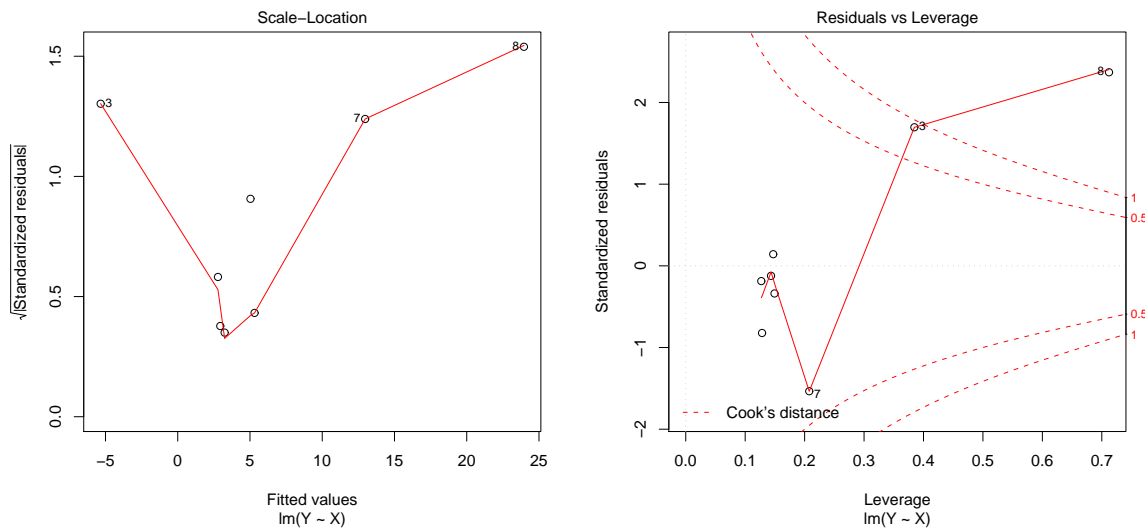


FIGURE 8.6 – Graphique des racines des valeurs absolues des résidus standardisés  $\sqrt{|r_i|}$  en fonction des valeurs ajustées et graphique des résidus en fonction des leviers

```

postscript("q0.ps", onefile = TRUE, width=6, height=6, horizontal=FALSE)
plot(X, Y); abline(g$coef); dev.off()
t(cbind(X, Y))
XX=cbind(rep(1, times=8), X)
h=diag(XX%*%solve(t(XX)%*%XX)%*%t(XX))
e=residuals(g); r=rstandard(g); t=rstudent(g)
sigma2=sum(e^2)/(8-2); D=e^2/(2*sigma2)*h/(1-h)^2
postscript("q.ps", onefile = TRUE, width=6, height=6, horizontal=FALSE)
plot(g); dev.off()

```

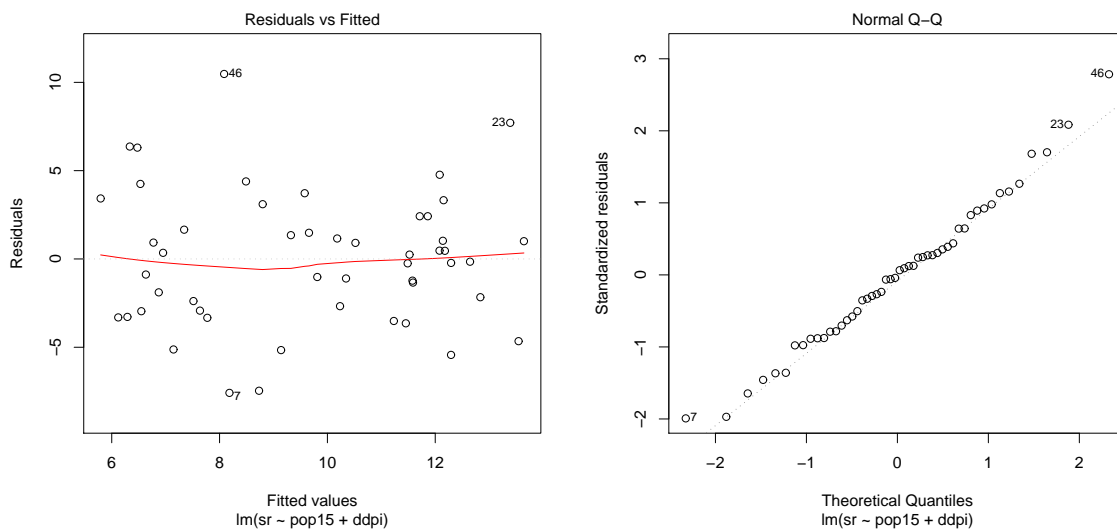
## 8.6 Un second exemple

On utilise les données `savings` de la base de données et on demande les graphiques de la régression qui sont présentés en Figures 8.7 et 8.8. Un examen approfondi des graphiques met en évidence la particularités des individus numéros 7, 23, 46 et 49. L'individu 49 n'est pas un point levier mais à un levier et une distance de Cook relativement importante.

En langage R

```
library(faraway); data(savings); attach(savings)
ll=lm(sr~pop15+ddpi)
summary(ll)
postscript("ll.ps",onfile = TRUE,width=6, height=6,horizontal=FALSE)
plot(ll);dev.off()
```

FIGURE 8.7 – Résidus standardisés en fonction des valeurs ajustées et qq-plot des résidus



La commande R `savings[c(7,23,46,49),]` permet d'identifier ces pays et d'analyser leurs particularité en retournant au Tableau 7.2. Ce pays sont respectivement le Chili, le Japon, la Zambie et la Lybie.

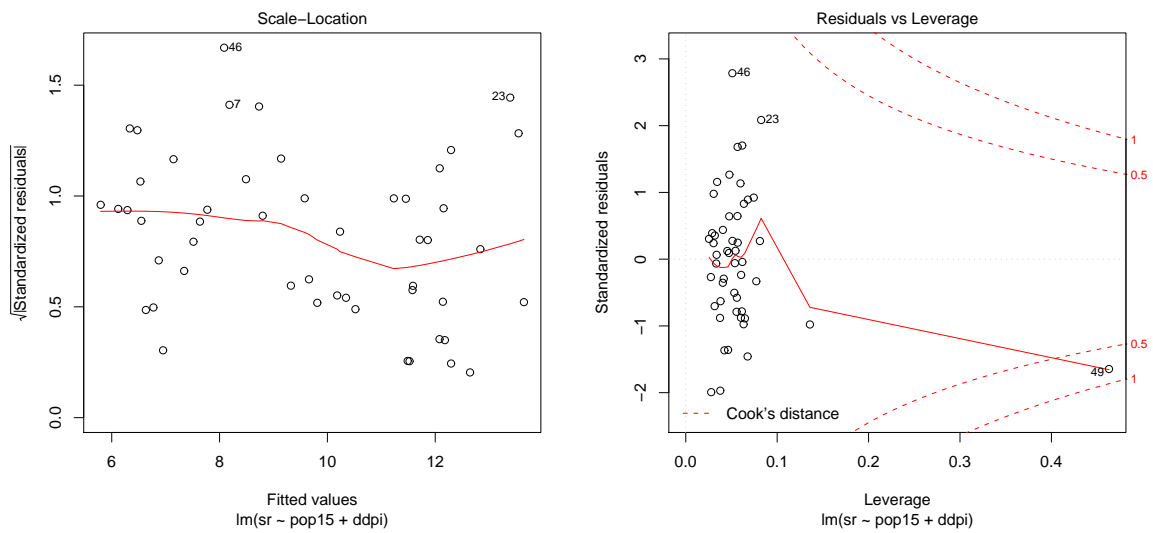


FIGURE 8.8 – Graphique des racines des valeurs absolues des résidus standardisés  $\sqrt{|r_i|}$  en fonction des valeurs ajustées et graphique des résidus en fonction des leviers



## Chapitre 9

# Analyse de la variance à un facteur

### 9.1 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 :

$$y_{ih} = \mu_h + \varepsilon_{ih}, \quad (9.1)$$

pour tout  $h = 1, \dots, H$ , et  $i = 1, \dots, n_h$ , où les  $\mu_h$  sont  $H$  constantes et les  $\varepsilon_i$  sont des termes d'erreur indépendants, identiquement distribués ayant une distribution normale de moyenne nulle et de variance  $\sigma_\varepsilon^2$ . Le modèle (9.1) 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

$$\begin{cases} H_0 & \mu_1 = \mu_2 = \dots = \mu_H \\ H_1 & \text{au moins un des } \mu_h \text{ est différent des autres.} \end{cases} \quad (9.2)$$

### 9.2 Méthode 1

La première méthode consiste à écrire le modèle (9.1) 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,

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \hline \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \hline 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \quad (9.3)$$

On peut dès lors écrire le modèle (9.1) 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  $\boldsymbol{\beta}$  est donné par

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_h \\ \vdots \\ \bar{y}_H \end{pmatrix},$$

où  $\bar{y}_h$  est la moyenne du groupe  $h$  et l'estimateur de  $\mu_h$  :

$$\widehat{\mu}_h = \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{ih},$$

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 (9.2), on va utiliser la méthode de Wald développée dans la section (7.3). Le test (9.2) est un cas particulier du test (7.2) 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 :

$$\mathbf{R}\boldsymbol{\beta} = \begin{pmatrix} \mu_1 - \frac{1}{n} \sum_{h=1}^H n_h \mu_h \\ \vdots \\ \mu_h - \frac{1}{n} \sum_{h=1}^H n_h \mu_h \\ \vdots \\ \mu_{H-1} - \frac{1}{n} \sum_{h=1}^H n_h \mu_h \end{pmatrix} = \begin{pmatrix} \mu_1 - \mu \\ \vdots \\ \mu_h - \mu \\ \vdots \\ \mu_{H-1} - \mu \end{pmatrix},$$

et, de la même manière,

$$\mathbf{R}\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \bar{y}_1 - \bar{y} \\ \vdots \\ \bar{y}_h - \bar{y} \\ \vdots \\ \bar{y}_{H-1} - \bar{y} \end{pmatrix},$$

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  $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$  équivaut, dans ce cas, à tester l'hypothèse nulle de (9.2). Pour calculer la statistique du test donné en (7.6), 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

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \{\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\}^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) = \sum_{h=1}^H n_h (\bar{y}_h - \bar{y})^2,$$

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 (7.6), 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 (7.6). Comme  $q = H-1$ , on a

$$F_c = \frac{SC_{INTER}/(H-1)}{SC_{INTRA}/(n-H)}. \quad (9.4)$$

On construit le Tableau 9.1 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.

TABLE 9.1 – 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	$SC_{TOT}$	$n - 1$	$CM_{TOT} = \frac{SC_{TOT}}{n-1}$	

### 9.3 Méthode 2

Une autre manière d’écrire le modèle (9.1) 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

$$y_{ih} = \mu + \alpha_h + \varepsilon_{ih}, \quad (9.5)$$

avec la contrainte que

$$\sum_{h=1}^H n_h \alpha_h = 0. \quad (9.6)$$

Le modèle (9.5) 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

$$\alpha_H = -\frac{1}{n_H} \sum_{h=1}^{H-1} n_h \alpha_h, \quad (9.7)$$

ce qui donne

$$\begin{cases} y_{ih} = \mu + \alpha_h + \varepsilon_{ih} & \text{si } 1 \leq h \leq H - 1 \\ y_{iH} = \mu - \frac{1}{n_H} \sum_{h=1}^{H-1} n_h \alpha_h + \varepsilon_{iH} & \text{sinon.} \end{cases} \quad (9.8)$$

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  $\alpha_h$  suffit, en vertu de l’Expression (9.7). Le modèle (9.5) s’écrit comme un modèle linéaire général

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

où

$$\boldsymbol{\beta}' = (\mu \ \alpha_1 \ \alpha_2 \ \dots \ \alpha_{H-1}),$$

et la matrice  $\mathbf{X}$  est de dimension  $n \times H$  et est donnée par

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \hline 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \hline \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \hline 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \\ \hline 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \\ 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \\ 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \end{pmatrix}. \quad (9.9)$$

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 (7.3.2). 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  $\beta$ .

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} \bar{y} \\ \bar{y}_1 - \bar{y} \\ \vdots \\ \bar{y}_{H-1} - \bar{y} \end{pmatrix}.$$

L'estimateur de  $\mu$  est donc  $\bar{y}$  et les estimateurs  $\alpha_h$  sont

$$\widehat{\alpha}_h = \bar{y}_h - \bar{y}, h = 1, \dots, H-1.$$

Les valeurs ajustées valent  $\mathbf{X}\widehat{\beta}$ , ce qui donne, si  $h \leq H-1$

$$y_{ih}^* = \widehat{\mu} + \widehat{\alpha}_h = \bar{y}_h,$$

et si  $h = H$ ,

$$y_{iH}^* = \widehat{\mu} - \sum_{h=1}^{H-1} \frac{\widehat{\alpha}_h n_h}{n_H} = \bar{y}_H.$$

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 (7.8) et vaut

$$F_c = \frac{SC_{INTER}/(H-1)}{SC_{INTRA}/(n-H)}$$

et est exactement la même que (9.4). En posant le modèle différemment, on estime d'autres paramètres, mais les résidus, les valeurs ajustées, et le test sont identiques.

## 9.4 Exemple d'analyse de la variance à un facteur

### 9.4.1 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 9.2.

### 9.4.2 Les résultats

Le traitement statistique nous donne les résultats présentés dans les Tableaux 9.3 et 9.4.

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.

TABLE 9.2 – 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

TABLE 9.3 – 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

TABLE 9.4 – Tableau d'analyse de la variance

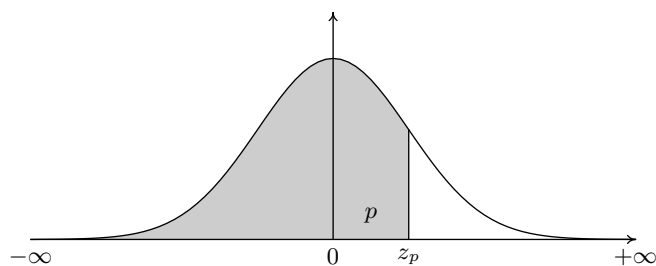
	sommes de carrés	degrés de liberté	carrés moyens	$F$	Sign.
Inter Groupes	10141.815	2	5070.907	1.176	0.317
Intra Groupes	219866.778	51	4311.113		
Total	230008.593	53			



# Chapitre 10

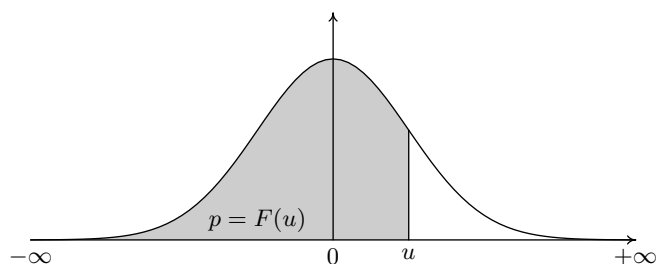
## Tables statistiques

TABLE 10.1 – Table des quantiles d'une variable normale centrée réduite



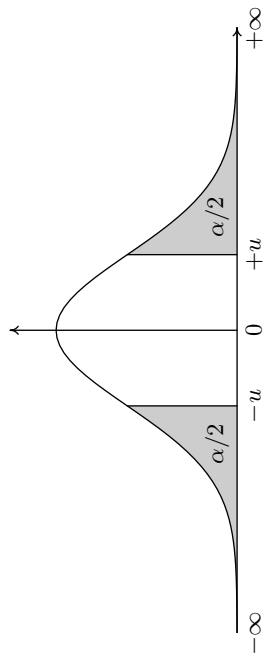
Ordre du quantile ( $p$ )	Quantile ( $z_p$ )	Ordre du quantile ( $p$ )	Quantile ( $z_p$ )
0.500	0.0000	0.975	1.9600
0.550	0.1257	0.976	1.9774
0.600	0.2533	0.977	1.9954
0.650	0.3853	0.978	2.0141
0.700	0.5244	0.979	2.0335
0.750	0.6745	0.990	2.3263
0.800	0.8416	0.991	2.3656
0.850	1.0364	0.992	2.4089
0.900	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

TABLE 10.2 – 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

TABLE 10.3 – Quantiles de la loi normale centrée réduite  
*( $u$  : valeur ayant la probabilité  $\alpha$  d'être dépassé en valeur absolue)*



$\alpha$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	$\infty$	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

TABLE 10.4 – Table des quantiles d’une variable  $\chi^2$  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

TABLE 10.5 – 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
$\infty$	1.645	1.960	2.327	2.576

TABLE 10.6 – 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	$\infty$
$n_2=1$	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
$\infty$	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

TABLE 10.7 – 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	$\infty$
$n_2=1$	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
$\infty$	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

TABLE 10.8 – 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.



TABLE 10.9 – Quantiles du coefficient de corrélation de Pearson d’une variable aléatoire normale bivariée sous l’hypothèse que  $\rho = 0$ . La taille de l’échantillon est notée  $n$

$n$	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

# Liste des tableaux

1.1	Taille et poids de 20 individus . . . . .	2
4.1	Système complet d'événements . . . . .	38
4.2	Illustration du théorème des probabilités totales . . . . .	40
4.3	Factorielle des nombres de 1 à 10 . . . . .	41
4.4	Erreur de première et seconde espèce . . . . .	63
4.5	Probabilité de commettre les erreurs . . . . .	63
5.1	Représentation graphique du modèle de régression . . . . .	64
5.2	Tableau récapitulatif des espérances et variances . . . . .	68
5.3	Tableau récapitulatif des espérances des sommes de carrés . . . . .	69
5.4	Tableau d'analyse de la variance . . . . .	73
5.5	Tableau des coefficients de régression . . . . .	75
5.6	Tableau d'analyse de la variance . . . . .	76
6.1	Tableau récapitulatif . . . . .	81
7.1	Tableau d'analyse de la variance . . . . .	89
7.2	Data savings . . . . .	92
8.1	Variables générées : la 8ième observation est un point aberrant . . . . .	99
8.2	Leviers $h_i$ , résidus $e_i$ , résidus standardisés $r_i$ , résidus studentisés $t_i$ , et distances de Cook $D_i$ pour les 8 observations . . . . .	100
9.1	Tableau d'analyse de la variance à un facteur . . . . .	108
9.2	Nombre de mots selon les groupes . . . . .	111
9.3	Moyennes selon les groupes . . . . .	111
9.4	Tableau d'analyse de la variance . . . . .	111
10.1	Table des quantiles d'une variable normale centrée réduite . . . . .	112
10.2	Fonction de répartition de la loi normale centrée réduite . . . . .	113
10.3	Quantiles de la loi normale centrée réduite . . . . .	114
10.4	Table des quantiles d'une variable $\chi^2$ à $n$ degrés de liberté . . . . .	115
10.5	Table des quantiles d'une variable de Student à $n$ degrés de liberté . . . . .	116
10.6	Table des quantiles d'ordre 0.95 d'une variable de Fisher à $n_1$ et $n_2$ degrés de liberté . . . . .	117
10.7	Table des quantiles d'ordre 0.99 d'une variable de Fisher à $n_1$ et $n_2$ degrés de liberté . . . . .	118
10.8	Valeur critique du test de Durbin-Watson au seuil de 5% . . . . .	119
10.9	Quantiles du coefficient de corrélation de Pearson d'une variable aléatoire normale bivariée sous l'hypothèse que $\rho = 0$ . La taille de l'échantillon est notée $n$ . . . . .	120

# Table des figures

1.1	Le nuage de points . . . . .	3
1.2	Exemples de nuages de points et coefficients de corrélation . . . . .	4
1.3	Le nuage de points, le résidu . . . . .	5
1.4	La droite de régression . . . . .	7
1.5	Droite passant par l'origine . . . . .	11
1.6	Régression avec une pente nulle . . . . .	12
4.1	Distribution de "faces" obtenus. . . . .	43
4.2	Distribution d'une variable aléatoire binomiale avec $n = 5$ et $p = 0.6$ . . . . .	46
4.3	Distribution d'une variable de Poisson avec $\lambda = 1$ . . . . .	47
4.4	Probabilité que la variable aléatoire soit inférieure à $a$ . . . . .	48
4.5	Fonction de densité d'une variable uniforme . . . . .	49
4.6	Fonction de répartition d'une variable uniforme . . . . .	49
4.7	Fonction de densité d'une variable normale . . . . .	50
4.8	Fonction de répartition d'une variable normale . . . . .	50
4.9	Densité d'une normale centrée réduite, symétrie . . . . .	51
4.10	Fonction de densité d'une variable exponentielle avec $\lambda = 1$ . . . . .	52
4.11	Densité d'une normale bivariée . . . . .	55
4.12	Nuage de points de réalisations d'une normale bivariée . . . . .	56
4.13	Densité d'une variable de chi-carré avec $p = 1, 2, \dots, 10$ . . . . .	61
4.14	Densités de variables de Student avec $p = 1, 2$ et $3$ et d'une variable normale . . . . .	61
4.15	Densité d'une variable de Fisher . . . . .	62
7.1	Nuages de points de la base de données savings . . . . .	93
8.1	qq-plot de 100 variables normales simulées versus une distribution théorique normale . . . . .	96
8.2	qq-plot de 100 variables exponentielles simulées versus une distribution théorique normale . . . . .	97
8.3	qq-plot de 100 variables de Student avec 2 degrés de libertés simulées versus une distribution théorique normale . . . . .	97
8.4	Nuage de point et droite de régression : un point extrême joue un rôle de levier sur la droite de régression . . . . .	100
8.5	Résidus standardisés en fonction des valeurs ajustées et qq-plot des résidus . . . . .	101
8.6	Graphique des racines des valeurs absolues des résidus standardisés $\sqrt{ r_i }$ en fonction des valeurs ajustées et graphique des résidus en fonction des leviers . . . . .	101
8.7	Résidus standardisés en fonction des valeurs ajustées et qq-plot des résidus . . . . .	102
8.8	Graphique des racines des valeurs absolues des résidus standardisés $\sqrt{ r_i }$ en fonction des valeurs ajustées et graphique des résidus en fonction des leviers . . . . .	103

# Table des matières

<b>1</b>	<b>Régression bivariée</b>	<b>2</b>
1.1	Série statistique bivariée . . . . .	2
1.1.1	Représentation graphique de deux variables . . . . .	2
1.1.2	Analyse des variables . . . . .	2
1.1.3	Covariance . . . . .	3
1.1.4	Corrélation . . . . .	3
1.1.5	Droite de régression . . . . .	5
1.1.6	Résidus et valeurs ajustées . . . . .	7
1.1.7	Sommes de carrés et variances . . . . .	9
1.1.8	Décomposition de la variance . . . . .	10
1.2	La régression par l'origine (sans constante) . . . . .	11
1.3	La régression avec une pente nulle . . . . .	12
<b>2</b>	<b>Éléments d'algèbre linéaire</b>	<b>13</b>
2.1	Espace vectoriel . . . . .	13
2.1.1	Vecteur . . . . .	13
2.1.2	Multiplication par un scalaire et addition . . . . .	13
2.1.3	Vecteurs linéairement indépendants . . . . .	14
2.1.4	Sous-espace vectoriel . . . . .	14
2.1.5	Système générateur d'un sous-espace vectoriel . . . . .	14
2.1.6	Base d'un sous-espace vectoriel . . . . .	14
2.1.7	Base canonique de $\mathbb{R}^n$ . . . . .	14
2.1.8	Dimension d'un sous-espace vectoriel . . . . .	14
2.2	Espace euclidien . . . . .	14
2.2.1	Produit scalaire . . . . .	14
2.2.2	Norme . . . . .	15
2.2.3	Distance entre deux vecteurs . . . . .	15
2.2.4	Vecteurs orthogonaux . . . . .	15
2.2.5	Orthogonal d'un sous-espace vectoriel . . . . .	15
2.3	Application linéaire et matrices . . . . .	16
2.3.1	Application linéaire . . . . .	16
2.3.2	Matrice . . . . .	16
2.3.3	Produit d'une matrice et d'un vecteur . . . . .	16
2.3.4	Produit matriciel . . . . .	16
2.3.5	Transposition . . . . .	17
2.3.6	Matrices carrées, symétriques et diagonales . . . . .	17
2.3.7	Rang d'une matrice . . . . .	17
2.3.8	Trace d'une matrice . . . . .	18
2.3.9	Matrices inversibles . . . . .	18
2.3.10	Inversion par parties . . . . .	18
2.3.11	Déterminant . . . . .	18
2.3.12	Quelques propriétés . . . . .	19
2.3.13	Matrices orthogonales . . . . .	19
2.3.14	Valeurs propres et vecteurs propres . . . . .	19
2.3.15	Formes et applications linéaires, formes quadratiques . . . . .	20

2.3.16	Image et noyau d'une matrice . . . . .	20
2.4	Projection et matrice idempotente . . . . .	21
2.4.1	Projection . . . . .	21
2.4.2	Projection orthogonale . . . . .	21
2.4.3	Projection orthogonale dans l'image et le noyau d'une matrice . . . . .	21
2.4.4	Matrice idempotente . . . . .	21
2.4.5	Projecteurs obliques . . . . .	22
2.4.6	Théorème des trois perpendiculaires . . . . .	22
2.5	Dérivée par rapport à un vecteur . . . . .	23
2.5.1	Gradient . . . . .	23
2.5.2	Dérivation d'une forme linéaire . . . . .	23
2.5.3	Dérivation d'une application linéaire . . . . .	23
2.5.4	Dérivée d'une forme quadratique . . . . .	23
<b>3</b>	<b>La régression multivariée</b> . . . . .	<b>25</b>
3.1	Représentation matricielle des données . . . . .	25
3.2	Principe des moindres carrés . . . . .	25
3.3	Valeurs ajustées et résidus . . . . .	27
3.4	Variance de régression et variance résiduelle . . . . .	27
3.5	Coefficient de détermination . . . . .	28
3.6	Matrice de variance-covariance et matrice de corrélation . . . . .	28
3.7	Corrélations partielles . . . . .	29
3.7.1	Cas général . . . . .	29
3.7.2	Cas où une seule variable $x$ est utilisée . . . . .	29
3.8	Condition pour que la somme des résidus soit nulle . . . . .	30
3.9	Décomposition en sommes de carrés . . . . .	31
3.10	Régression avec les données centrées . . . . .	32
3.11	Retour au cas bivarié . . . . .	34
3.12	Méthode 1 . . . . .	34
3.13	Méthode 2 . . . . .	35
<b>4</b>	<b>Rappel sur le calcul des probabilités, les variables aléatoires, et l'inférence statistique</b> . . . . .	<b>36</b>
4.1	Probabilités . . . . .	36
4.1.1	Événement . . . . .	36
4.1.2	Opérations sur les événements . . . . .	36
4.1.3	Relations entre les événements . . . . .	37
4.1.4	Ensemble des parties d'un ensemble et système complet . . . . .	37
4.1.5	Axiomatique des Probabilités . . . . .	38
4.1.6	Probabilités conditionnelles et indépendance . . . . .	40
4.1.7	Théorème des probabilités totales et théorème de Bayes . . . . .	40
4.2	Analyse combinatoire . . . . .	41
4.2.1	Introduction . . . . .	41
4.2.2	Permutations (sans répétition) . . . . .	41
4.2.3	Permutations avec répétition . . . . .	42
4.2.4	Arrangements (sans répétition) . . . . .	42
4.2.5	Combinaisons . . . . .	42
4.3	Variables aléatoires . . . . .	43
4.3.1	Définition . . . . .	43
4.4	Variables aléatoires discrètes . . . . .	43
4.4.1	Définition, espérance et variance . . . . .	43
4.4.2	Variable indicatrice ou bernoullienne . . . . .	44
4.4.3	Variable binomiale . . . . .	44
4.4.4	Variable de Poisson . . . . .	46
4.5	Variable aléatoire continue . . . . .	47
4.5.1	Définition, espérance et variance . . . . .	47
4.5.2	Variable uniforme . . . . .	48
4.5.3	Variable normale . . . . .	50

4.5.4	Variable normale centrée réduite . . . . .	50
4.5.5	Distribution exponentielle . . . . .	51
4.6	Distribution bivariable, cas continu . . . . .	52
4.6.1	Généralités . . . . .	52
4.6.2	Indépendance de deux variables aléatoires . . . . .	53
4.6.3	Propriétés des espérances et des variances . . . . .	54
4.6.4	Somme de variables aléatoires . . . . .	55
4.6.5	Loi normale bivariée . . . . .	55
4.7	Vecteurs aléatoires . . . . .	58
4.7.1	Généralités . . . . .	58
4.7.2	Variable normale multivariée . . . . .	59
4.8	Autres variables aléatoires . . . . .	60
4.8.1	Variable khi-carrée . . . . .	60
4.8.2	Variable de Student . . . . .	61
4.8.3	Variable de Fisher . . . . .	61
4.9	Inférence statistique . . . . .	62
4.9.1	Modélisation . . . . .	62
4.9.2	Intervalle de confiance . . . . .	62
4.9.3	Tests d'hypothèses . . . . .	62
<b>5</b>	<b>Le modèle de régression avec une variable explicative et une constante</b>	<b>64</b>
5.1	Le modèle de régression . . . . .	64
5.2	Estimation des paramètres par les moindres carrés . . . . .	65
5.3	Valeurs ajustées et résidus . . . . .	67
5.4	Espérance des sommes de carrés . . . . .	68
5.5	Le modèle linéaire avec normalité des erreurs . . . . .	69
5.6	Distribution de probabilités des estimateurs . . . . .	71
5.7	Intervalles de confiance sur les coefficients de régression . . . . .	72
5.8	Test d'hypothèses sur un coefficient de régression . . . . .	72
5.9	Le tableau d'analyse de la variance . . . . .	73
5.10	Exemple et interprétation . . . . .	74
<b>6</b>	<b>Le modèle linéaire général</b>	<b>77</b>
6.1	Le modèle . . . . .	77
6.1.1	Définition du modèle linéaire général . . . . .	77
6.2	Régression avec les résidus normaux . . . . .	77
6.2.1	Hypothèses du modèle linéaire général . . . . .	77
6.2.2	Données observées, et formulation matricielle . . . . .	77
6.2.3	Autre présentation du modèle linéaire général . . . . .	78
6.3	Estimation du modèle . . . . .	78
6.3.1	Estimation par les moindres carrés (ordinaires) . . . . .	78
6.3.2	Estimateurs du maximum de vraisemblance . . . . .	81
6.3.3	Propriétés des estimateurs du maximum de vraisemblance . . . . .	82
6.3.4	Distribution de probabilité des estimateurs . . . . .	82
6.3.5	Synthèse des résultats . . . . .	83
<b>7</b>	<b>Inférence dans le modèle linéaire</b>	<b>84</b>
7.1	Intervalle de confiance sur un coefficient de régression . . . . .	84
7.2	Test d'un seul coefficient de régression . . . . .	85
7.2.1	Construction du test . . . . .	85
7.2.2	Modèle linéaire avec uniquement une constante . . . . .	85
7.3	Tests de Wald sur les coefficients de régression . . . . .	86
7.3.1	Test général d'une contrainte linéaire . . . . .	86
7.3.2	Test global des coefficients de régression . . . . .	88
7.3.3	Test de Fisher sur un coefficient de régression . . . . .	89
7.4	R-carré et R-carré ajusté . . . . .	90
7.5	Prévision ponctuelle d'une valeur . . . . .	90

7.5.1	Cas général . . . . .	90
7.5.2	Cas bivarié . . . . .	91
7.6	Un exemple . . . . .	91
<b>8</b>	<b>Outils de détection en régression</b>	<b>96</b>
8.1	Le qq-plot . . . . .	96
8.2	Points leviers . . . . .	98
8.3	Résidus standardisés et studentisés . . . . .	98
8.4	Distance de Cook . . . . .	99
8.5	Un premier exemple . . . . .	99
8.6	Un second exemple . . . . .	102
<b>9</b>	<b>Analyse de la variance à un facteur</b>	<b>104</b>
9.1	Le problème . . . . .	104
9.2	Méthode 1 . . . . .	104
9.3	Méthode 2 . . . . .	108
9.4	Exemple d'analyse de la variance à un facteur . . . . .	110
9.4.1	Les données . . . . .	110
9.4.2	Les résultats . . . . .	110
<b>10</b>	<b>Tables statistiques</b>	<b>112</b>

# Index

- analyse, 2
  - combinatoire, 41
  - de la variance, 73, 75
  - de la variance à un facteur, 104
- application linéaire, 16, 20
- arrangement, 42
- axiomatique, 38
- base
  - canonique, 14
  - d'un sous-espace vectoriel, 14
- Bernoulli, 44
- bernoullienne, 44
- binôme de Newton, 44
- cas bivarié, 34
- coefficient
  - de corrélation, 4, 30
    - multiple, 28
    - partielle, 29, 30
  - de détermination, 4, 11, 28
  - de régression, 64, 72
- combinaison, 42
  - linéaire
    - de deux vecteurs, 13
- combinaison linéaire
  - de deux matrices, 16
- complémentaire, 37
- corrélation, 3
  - coefficient, 3
  - partielle, 29
  - simple, 30
- covariance, 3, 53, 58, 67
- dérivée
  - partielle, 5, 23, 81
- dérivation
  - d'une application linéaire, 23
  - d'une forme linéaire, 23
- dérivation d'une forme
  - quadratique, 23
- déterminant, 18
- degré de liberté, 72, 74, 75, 80, 84, 85
- densité
  - jointe, 70
  - marginale, 58, 70
- différence, 36
- dimension d'un sous-espace vectoriel, 14
- distance
  - de Cook, 99
  - entre deux vecteurs, 15
- distribution
  - binomiale, 44, 45
  - bivarée, 57
  - bivariée, 52
  - conditionnelle, 53
  - d'une variable de Poisson, 47
  - de Fisher, 74
  - de probabilité, 43, 45, 62, 81
  - de probabilité des estimateurs, 71, 82
  - de Student, 85
  - exponentielle, 51
  - marginale, 52, 56, 57
  - multinormale, 81
  - normale bivariée, 56, 58
  - normale multivariée, 59
- données centrées, 35
- droite de régression, 5
- écart-type, 79
  - de la variable de régression, 50
  - des coefficients, 75
  - marginal, 3
- ensemble
  - parties d'un ensemble, 37
  - système complet, 37
- erreur
  - de deuxième espèce, 62
  - de première espèce, 62
- erreur-type, 79
- espérance, 43, 54
  - conditionnelle, 57
  - d'une variable
    - binomiale, 45
    - indicatrice, 44
  - des sommes des carrés, 68
  - propriétés, 54
- espace
  - euclidien, 14
  - norme, 15
  - vectoriel, 13
- estimateur
  - convergent, 82
  - du maximum de vraisemblance, 70, 81, 82
  - efficace, 82
  - linéaire, 65, 79



- sans biais, 80, 82
- estimation
  - par les moindres carrés (ordinaires), 78
- événements, 36
  - indépendants, 40
  - mutuellement exclusifs, 37
- expérience aléatoire, 36, 43
- fonction, 47
  - de densité, 50, 70
    - conditionnelle, 53
    - d'une variable aléatoire continue, 47
    - d'une variable exponentielle, 52
    - d'une variable normale multivariée, 59
    - d'une variable uniforme, 49
  - marginale, 52
  - de log-vraisemblance, 70
  - de répartition jointe, 52
  - de vraisemblance, 70, 81
- forme
  - linéaire, 20
  - quadratique, 20, 82, 83
- formulation matricielle, 77
- gradient, 23
- homoscédasticité, 77
- homoscédastique, 57, 64
- image d'une matrice, 20
- indépendance, 53
- inférence, 84
- inférence statistique, 36, 62
- intersection, 36
- intervalle
  - d'acceptation, 63
  - de confiance, 62, 72
  - sur un coefficient de régression, 84
- inversion par parties, 18
- log-vraisemblance, 70
- loi
  - binomiale, 44
  - de Poisson, 46
  - normale bivariée, 55
- matrice, 16
  - carée symétrique, 19
  - carrée, 17, 19
  - définie positive, 20
  - déterminant, 18
  - de corrélation, 28
  - de plein rang, 78
  - de projection, 21
  - de variance-covariance, 28
  - des corrélations, 29
  - des variables explicatives, 104
  - diagonale, 17, 82
  - hessienne, 6
  - idempotente, 21, 22, 32, 80, 83
    - trace, 22
  - identité, 17, 19
  - image, 20
  - inversible, 18
  - noyau, 20
  - orthogonale, 19, 82
  - produit d'une matrice et d'un vecteur, 16
  - propriétés, 19
  - rang, 17
  - semi-définie positive, 20
  - sous-espace, 21
  - symétrique, 17, 83
  - trace, 18
  - transposition, 17
  - valeurs propres, 19
  - variance-covariance, 33
  - vecteurs propres, 19
- modélisation, 62
- modèle
  - de régression, 64
  - estimé, 75
  - linéaire, 69
    - avec une constante, 85
  - linéaire général, 77, 83
    - définition, 77
    - hypothèses, 77
    - paramétrique, 69
- moindres carrés, 5, 65, 81
  - principe, 26
- moyenne, 2
  - conditionnelle, 53
  - marginale, 3, 52, 55, 56
- multicolinéarité, 78
- multiplication par un scalaire, 13
- normale bivariée, 55
- normalité des erreurs, 69
- norme, 15
- noyau d'une matrice, 20
- orthogonal d'un sous-espace vectoriel, 15
- paramètres marginaux, 3
- penne, 6
  - nulle, 12
- permutation
  - avec répétition, 42
  - sans répétition, 41
- point levier, 98
- prévision ponctuelle d'une valeur, 90
- principe des moindres carrés, 25
- probabilité, 36, 38
  - conditionnelle et indépendance, 40
  - théorème des probabilités totales, 40
- produit

- d'une matrice et d'un vecteur, 16
- matriciel, 16
- scalaire, 4, 14
- projecteur oblique, 22
- projecteur orthogonal, 21, 22
- projection, 15, 21
  - orthogonale, 21
    - dans l'image, 21
    - dans le noyau, 21
- propriétés des espérances et des variances, 54
- puissance d'un test, 63
- qq-plot, 96
- rang d'une matrice, 17
- régression, 5
  - bivariée, 2
  - données centrées, 32
  - linéaire, 74
  - multiple, 32
  - multivariée, 25, 64, 73
- représentation
  - graphique de deux variables, 2
  - matricielle des données, 25
- résidus, 7, 67
  - normaux, 77
  - standardisés et studentisés, 98
- série statistique bivariée, 2
- scalaire, 13
- somme
  - des carrés, 31, 33
    - de la régression, 9
    - des résidus, 5, 10, 31, 75
    - expliquée par la régression, 31
    - inter-groupes, 107, 110
    - intra-groupes, 107, 110
    - totale, 9
      - totale des écarts à la moyenne, 31
  - des carrés des résidus, 71
  - des résidus, 30
  - des variables aléatoires, 55
- sous ensemble, 36
- sous-espace vectoriel, 14, 21
  - base, 14
  - base canonique de  $\mathbb{R}$ , 14
  - dimension, 14
  - orthogonal, 15
  - système générateur, 14
- statistique exhaustive, 82
- système
  - complet d'événements, 37, 39, 40
  - générateur d'un sous-espace vectoriel, 14
- terme d'erreur, 64, 67, 77–79, 82, 83, 104
- test
  - d'hypothèses, 72, 73
    - composites, 63
    - simples, 62
  - d'un seul coefficient de régression, 85
  - de Fisher sur un coefficient de régression, 89
  - de Wald sur les coefficients de régression, 86
  - global sur les coefficients de régression, 88
- théorème
  - de Bayes, 40, 41
  - de diagonalisation, 20
  - de Gauss-Markov, 79
  - de Pythagore, 15
  - des probabilités totales, 40, 41
  - des trois perpendiculaires, 22
- théorie des ensembles, 36
- trace d'une matrice, 18
  - idempotente, 22
- transposition, 13
- union, 36
- valeur
  - ajustée, 7, 67
  - propre, 19, 82
    - d'une matrice idempotente, 22
- valeur propre, 22
- variable
  - aléatoire, 48
  - aléatoire, 36, 43, 50–54, 59, 65, 67, 71
    - continue, 47, 52
    - de Fisher, 74
    - de Student, 84
    - discrète, 43
    - indépendante, 53
    - normale indépendante, 69
    - uniforme, 49
  - binomiale, 44
  - constante, 30
  - dépendante, 33
  - de Fisher, 61
  - de Poisson, 46
  - de Student, 61
  - de student, 72
  - espérance, 43
  - explicative, 31, 33, 34, 75
  - indépendante, 29
  - indicatrice, 44
  - khi-carrée, 60
  - normale, 50
    - centrée réduite, 50
    - multivariée, 59
  - uniforme, 48, 49
- variance, 2, 9, 10, 43–54, 59, 60, 67, 79
  - conditionnelle, 53, 57
  - d'une variable
    - binomiale, 45
    - indicatrice, 44
  - de régression, 9–11

- régression multivariée, 27
- des résidus, 68
- marginale, 3, 11, 52, 55, 56
- minimum, 82
- propriétés, 54
- résiduelle, 10
  - régression multivariée, 27
- vecteur, 13
  - aléatoire, 58, 59, 78
  - colonne, 13, 14, 32
  - des résidus, 27, 29, 31
  - des valeurs ajustées, 27, 34
  - ligne, 13, 14
  - linéairement indépendant, 14
  - multinormal, 60
  - multinormal non-corrélé, 83
  - orthogonal, 15
  - projection, 15
  - propre, 19, 82